Bruce K. Driver

Probability Tools with Examples

June 17, 2010 *File:prob.tex*

Contents

Part III Stochastic Processes I

Part IV (Weak) Convergence of Random Sums

Part

Homework Problems

Math 280A Homework Problems Fall 2009

Problems are from Resnick, S. A Probability Path, Birkhauser, 1999 or from the lecture notes. The problems from the lecture notes are hyperlinked to their location.

-3.1 Homework 1. Due Wednesday, September 30, 2009

- Read over Chapter 1.
- Hand in Exercises 1.1, 1.2, and 1.3.

-3.2 Homework 2. Due Wednesday, October 7, 2009

- Look at Resnick, p. 20-27: 9, 12, 17, 19, 27, 30, 36, and Exercise 3.9 from the lecture notes.
- Hand in Resnick, p. 20-27: 5, 18, 23, 40*, 41, and Exercise 4.1 from the lecture notes.

*Notes on Resnick's #40: (i) $\mathcal{B}((0,1])$ should be $\mathcal{B}([0,1))$ in the statement of this problem, (ii) k is an integer, (iii) $r \geq 2$.

-3.3 Homework 3. Due Wednesday, October 21, 2009

- Look at Lecture note Exercises; 4.7, 4.8, 4.9
- Hand in Resnick, p. 63–70; 7* and 13.
- Hand in Lecture note Exercises: 4.3, 4.4, 4.5, 4.6, 4.10 – 4.15.

***Hint**: For #7 you might label the coupons as $\{1, 2, \ldots, N\}$ and let A_i be the event that the collector does **not** have the i^{th} – coupon after buying n - boxes of cereal.

-3.4 Homework 4. Due Wednesday, October 28, 2009

- Look at Lecture note Exercises; 5.5, 5.10.
- Look at Resnick, p. 63–70; 5, 14, 16, 19
- Hand in Resnick, p. 63–70; 3, 6, 11
- Hand in Lecture note Exercises: 5.6 – 5.9.

-3.5 Homework 5. Due Wednesday, November 4, 2009

- Look at Resnick, p. 85–90: 3, 7, 8, 12, 17, 21
- **Hand in** from Resnick, p. 85–90: 4, 6*, 9, 15, 18**.
  *Note: In #6, the random variable X is understood to take values in the extended real numbers.
  ** I would write the left side in terms of an expectation.
- Look at Lecture note Exercise 6.3, 6.7.
- **Hand in** Lecture note Exercises: 6.4, 6.6, 6.10.

-3.6 Homework 6. Due Wednesday, November 18, 2009

- Look at Lecture note Exercise 7.4, 7.9, 7.12, 7.17, 7.18, and 7.27.
- **Hand in** Lecture note Exercises: 7.5, 7.7, 7.8, 7.11, 7.13, 7.14, 7.16
- Look at from Resnik, p. 155–166: 6, 13, 26, 37
- **Hand in** from Resnick,p. 155–166: 7, 38

-3.7 Homework 7. Due Wednesday, November 25, 2009

- Look at Lecture note Exercise 9.12 – 9.14.
- Look at from Resnick§ 5.10: #18, 19, 20, 22, 31.
- **Hand in** Lecture note Exercises: 8.1, 8.2, 8.3, 8.4, 8.5, 9.4, 9.5, 9.6, 9.7, and 9.9.
- **Hand in** from Resnick § 5.10: #9, 29.

  See next page!

-3.8 Homework 8. Due Monday, December 7, 2009 by 11:00AM (Put under my office door if I am not in.)

- Look at Lecture note Exercise 10.1, 10.2, 10.7, 10.8, 10.10.
- Look at from Resnick § 4.5: 3, 5, 6, 8, 19, 28, 29.
- Look at from Resnick § 5.10: #6, 7, 8, 11, 13, 16, 22, 34

- **Hand in** Lecture note Exercises: 9.8, 10.9.
- **Hand in** from Resnick § 4.5: 1, 9*, 11, 18, 25. *Exercise 10.10 may be useful here.
- **Hand in** from Resnick § 5.10: #14, 26.

Math 280B Homework Problems Winter 2010

-2.1 Homework 1. Due Wednesday, January 13, 2010

- **Hand in** Lecture note Exercise 11.1, 11.2, 11.3, 11.4, 11.5, 11.6.
- Look at from Resnick § 5.10: #39

-2.2 Homework 2. Due Wednesday, January 20, 2010

- Look at from §6.7: 3, 4, 14, 15(Hint: see Corollary 12.9 **or** use $|a-b| = 2(a-b)^{+} - (a-b)$), 16, 17, 19, 24, 27, 30
- Look at Lecture note Exercise 12.12
- **Hand in** from Resnick §6.7: 1a, d, 12, 13, 18 (Also assume $\mathbb{E}X_n = 0$)*, 33.
- **Hand in** lecture note exercises: 12.1, 12.3

* For Problem 18, please add the missing assumption that the random variables should have mean zero. (The assertion to prove is false without this assumption.) With this assumption, $\mathrm{Var}(X) = \mathbb{E}[X^2]$. Also note that $\mathrm{Cov}(X,Y) = 0$ is equivalent to $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$.

-2.3 Homework 3. Due Wednesday, January 27, 2010

- Look at from §6.7:
- Look at Lecture note Exercise 13.3, 13.5
- **Hand in** from Resnick §6.7: 5*, 7 (Hint: Observe that $X_n \overset{d}{=} \sigma_n N(0,1)$.) *For one possible proof of #5 it is useful to first show $\{X_n\}_{n=1}^{\infty}$ are U.I. first.
- **Hand in** lecture note exercises: 13.2, 13.4, 13.6

-2.4 Homework 4. Due Wednesday, February 3, 2010

- Look at Resnick Chapter 10: 11
- **Hand in** lecture note exercises: 10.3, 14.1, 14.2, 14.3, 14.4.
- **Hand in** from Resnick §10.17**:** $2^{\dagger}$, 5*, $7^{\dagger\dagger}$, 8**

†In part 2b, please explain what convention you are using when the denominator is 0.

*A Poisson process, $\{N(t)\}_{t\geq 0}$, with parameter λ satisfies (by definition): (i) N has *independent increments*, so that $N(s)$ and $N(t)-N(s)$ are independent; (ii) if $0 \leq u < v$ then $N(v)-N(u)$ has the Poisson distribution with parameter $\lambda(v-u)$.

†† For 7a and 7b it is illuminating to find a formula for $\mathbb{E}[g(X_1)|X_1+X_2]$.

****Hint:** use Exercise 10.3 to first show $\mathrm{Cov}(Y, f(Y)) \leq 0$.

-2.5 Homework 5. Due Wednesday, February 10, 2010

- Look at the following Exercises from the Lecture Notes: 12.13, 16.2
- Do the following Exercises from the Lecture Notes: 12.14, 12.15, 14.6, 14.8, 14.9

-2.6 Homework 6. Due Friday, February 19, 2010

- Look at the following Exercises from the Lecture Notes: 17.5, 17.15, 17.16
- Do the following Exercises from the Lecture Notes: 17.1, 17.2, 17.3, 17.4, 17.6, 17.8.

-2.7 Homework 7. Due Monday, March 1, 2010

- Do the following Exercises from the Lecture Notes: 17.7, 17.9, 17.10, 17.11, 17.12, 17.13.

-2.8 Homework 8. Due Monday, March 8, 2010

- **Hand in** the following Exercises from the Lecture Notes: 18.1, 18.2, 18.3, 18.5,
- Resnick Chapter 10: **Hand in** 14, 15, 16, 33.

-2.9 Homework 9. (Not) Due Monday, March 15, 2010

The following homework will **not** be collected but it would certainly be good if you did the problems. Solutions will appear during finals week.

- **Hand in** the following Exercises from the Lecture Notes: 18.7, 18.18.
- **Look at** the following Exercises from the Lecture Notes: 18.4, 18.6, 18.8.
- Resnick Chapter 10.17: **Hand in** 19. For this problem please define $X_{n+1}/X_n = Z_{n+1}$ where

$$Z_{n+1} = \begin{cases} X_{n+1}/X_n & \text{if} \quad X_n \neq 0 \\ 1 & \text{if} \quad X_n = 0 = X_{n+1} \\ \infty \cdot X_{n+1} & \text{if } X_n = 0 \text{ and } X_{n+1} \neq 0. \end{cases}$$

Math 280C Homework Problems Spring 2010

-1.1 Homework 1. Due Wednesday, April 7, 2010

- Look at from Resnick §10.17: 22
- Look at Lecture note Exercises: 18.13, 18.14, 18.15, 18.16, 18.17, 18.20, 18.22, 18.26, 18.27, 18.31.
- **Hand in** from Resnick §10.17: 18b, 23, 28 (hint: consider $\ln X_n$)
- **Hand in** lecture note exercises: , 18.19, 18.21, 18.23, 18.24, 18.25

-1.2 Homework 2. Due Wednesday, April 14, 2010

- Look at from Resnick §10: #25
- Look at Lecture note Exercises: 20.2
- **Hand in** from Resnick §7: #1, #2, #28*, #42 (#15 was assigned here in error previously – see Exercise 20.4 on next homework.)

  [*Correction to #28: In the second part of the problem, the condition "$\mathbb{E}[X_iX_j] \leq \rho(i-j)$ for $i > j$" should be "$\mathbb{E}[X_iX_j] \leq \rho(i-j)$ for $i \geq j$".]
- **Hand in** lecture note exercises: 19.2, 19.3, 19.4, 20.1.

-1.3 Homework 3. Due Wednesday April 21, 2010

- Look at from Resnick §7: #12. **Hint:** let $\{U_n : n = 0, 1, 2, \ldots\}$ be i.i.d. random variables uniformly distributed on (0,1) and take $X_0 = U_0$ and then define X_n inductively so that $X_{n+1} = X_n \cdot U_{n+1}$.
- **Hand in** from Resnick §7: #13, #16, #33, #36 (assume each X_n is integrable!)
- **Hand in** lecture note exercises: 20.3, 20.4

Hints and comments. For Resnick 7.36; It must be assumed that $\mathbb{E}[X_n] < \infty$ for each n, else there is a subtraction-of-infinities problem. In addition the conclusion reached in the second part of the problem can fail to be true if the expectations are infinite. Use the assumptions to bound $\mathbb{E}[X_n]$ in terms of $\mathbb{E}[X_n : X_n \leq x]$. Then use the two series theorem of Exercise 20.4.

-1.4 Homework 4. Due Wednesday April 28, 2010

- **Look at** lecture note exercises: 21.4, 21.7
- **Hand in** from Resnick §8.8: #4a-d, #13 (Assume $\sigma_n^2 = \mathrm{Var}(N_n) > 0$ for all n.), #20, #31
- **Hand in** lecture note exercises: 21.2, 21.3, 21.5, 21.6

-1.5 Homework 5. Due Wednesday May 5, 2010

- **Look at** from Resnick §8.8: #14, #36
- Resnick Chapter 8: **Hand in** 8.7* (assume the central limit theorem here), 8.17, 8.30**, 8.34***

  Comments and hints:

1. * In 8.7 you will need to use the central limit theorem along with the δ – method.
2. ** For 8.30, ignore the part of the question referring to the moment generating function. **Hint:** use problem 8.31 and the convergence of types theorem.
3. *** For 8.34 use an adaptation of the δ – method. Example 21.46 may be helpful here as well.

-1.6 Homework 6. Due Wednesday May 12, 2010

- **Hand in** lecture note exercises: 21.9, 21.10, 21.11, 21.12, 22.2, 22.3, 22.4
- Resnick Chapter 9: **Look at:** 9.5 (special case of 21.11), 9.6, 9.22, 9.33
- Resnick Chapter 9: **Hand in** 9.9 a-e., 9.10

-1.7 Homework 7. Due Wednesday May 19, 2010

- **Look at** Resnick Chapter 9: 9.1 (now obsolete), 9.8
- **Hand in** Resnick Chapter 9: #11 (Exercise 21.11 may be useful here), 28, 34 (assume $\sum_n \sigma_n^2 > 0$), 35 (hint: show $P[\xi_n \neq 0 \text{ i.o. }] = 0.$), 38 (Hint: make use Proposition 10.61.)

-1.8 Homework 8. Due Wednesday June 2, 2010

- **Look at** lecture note exercises: 24.2 – 24.6, 25.2, and 27.3 – 27.5.
- **Hand in** lecture note exercises: 25.1, 26.1, and 26.2.

0

Math 286 Homework Problems Spring 2008

Part I

Background Material

Limsups, Liminfs and Extended Limits

Notation 1.1 *The* ***extended real numbers*** *is the set* $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, *i.e. it is* $\mathbb{R}$ *with two new points called* ∞ *and* $-\infty$. *We use the following conventions,* $\pm\infty \cdot 0 = 0$, $\pm\infty \cdot a = \pm\infty$ *if* $a \in \mathbb{R}$ *with* $a > 0$, $\pm\infty \cdot a = \mp\infty$ *if* $a \in \mathbb{R}$ *with* $a < 0$, $\pm\infty + a = \pm\infty$ *for any* $a \in \mathbb{R}$, $\infty + \infty = \infty$ *and* $-\infty - \infty = -\infty$ *while* $\infty - \infty$ *is not defined. A sequence* $a_n \in \bar{\mathbb{R}}$ *is said to converge to* ∞ $(-\infty)$ *if for all* $M \in \mathbb{R}$ *there exists* $m \in \mathbb{N}$ *such that* $a_n \geq M$ $(a_n \leq M)$ *for all* $n \geq m$.

Lemma 1.2. *Suppose* $\{a_n\}_{n=1}^\infty$ *and* $\{b_n\}_{n=1}^\infty$ *are convergent sequences in* $\bar{\mathbb{R}}$, *then:*

1. *If* $a_n \leq b_n$ *for*[1] *a.a.* n, *then* $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.
2. *If* $c \in \mathbb{R}$, *then* $\lim_{n\to\infty} (ca_n) = c \lim_{n\to\infty} a_n$.
3. $\{a_n + b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \tag{1.1}$$
provided the right side is not of the form $\infty - \infty$.
4. $\{a_n b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n \tag{1.2}$$
provided the right hand side is not of the for $\pm\infty \cdot 0$ *of* $0 \cdot (\pm\infty)$.

Before going to the proof consider the simple example where $a_n = n$ and $b_n = -\alpha n$ with $\alpha > 0$. Then

$$\lim (a_n + b_n) = \begin{cases} \infty & \text{if } \alpha < 1 \\ 0 & \text{if } \alpha = 1 \\ -\infty & \text{if } \alpha > 1 \end{cases}$$

while

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{``}=\text{''} \infty - \infty.$$

This shows that the requirement that the right side of Eq. (1.1) is not of form $\infty - \infty$ is necessary in Lemma 1.2. Similarly by considering the examples $a_n = n$ and $b_n = n^{-\alpha}$ with $\alpha > 0$ shows the necessity for assuming right hand side of Eq. (1.2) is not of the form $\infty \cdot 0$.

Proof. The proofs of items 1. and 2. are left to the reader.

Proof of Eq. (1.1). Let $a := \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. Case 1., suppose $b = \infty$ in which case we must assume $a > -\infty$. In this case, for every $M > 0$, there exists N such that $b_n \geq M$ and $a_n \geq a - 1$ for all $n \geq N$ and this implies

$$a_n + b_n \geq M + a - 1 \text{ for all } n \geq N.$$

Since M is arbitrary it follows that $a_n + b_n \to \infty$ as $n \to \infty$. The cases where $b = -\infty$ or $a = \pm\infty$ are handled similarly. Case 2. If $a, b \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|a - a_n| \leq \varepsilon \text{ and } |b - b_n| \leq \varepsilon \text{ for all } n \geq N.$$

Therefore,

$$|a + b - (a_n + b_n)| = |a - a_n + b - b_n| \leq |a - a_n| + |b - b_n| \leq 2\varepsilon$$

for all $n \geq N$. Since $\varepsilon > 0$ is arbitrary, it follows that $\lim_{n\to\infty} (a_n + b_n) = a + b$.

Proof of Eq. (1.2). It will be left to the reader to prove the case where $\lim a_n$ and $\lim b_n$ exist in $\mathbb{R}$. I will only consider the case where $a = \lim_{n\to\infty} a_n \neq 0$ and $\lim_{n\to\infty} b_n = \infty$ here. Let us also suppose that $a > 0$ (the case $a < 0$ is handled similarly) and let $\alpha := \min\left(\frac{a}{2}, 1\right)$. Given any $M < \infty$, there exists $N \in \mathbb{N}$ such that $a_n \geq \alpha$ and $b_n \geq M$ for all $n \geq N$ and for this choice of N, $a_n b_n \geq M\alpha$ for all $n \geq N$. Since $\alpha > 0$ is fixed and M is arbitrary it follows that $\lim_{n\to\infty} (a_n b_n) = \infty$ as desired. ■

For any subset $\Lambda \subset \bar{\mathbb{R}}$, let $\sup \Lambda$ and $\inf \Lambda$ denote the least upper bound and greatest lower bound of Λ respectively. The convention being that $\sup \Lambda = \infty$ if $\infty \in \Lambda$ or Λ is not bounded from above and $\inf \Lambda = -\infty$ if $-\infty \in \Lambda$ or Λ is not bounded from below. We will also use the **conventions** that $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.

Notation 1.3 *Suppose that* $\{x_n\}_{n=1}^\infty \subset \bar{\mathbb{R}}$ *is a sequence of numbers. Then*

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k : k \geq n\} \text{ and} \tag{1.3}$$
$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k : k \geq n\}. \tag{1.4}$$

[1] Here we use "a.a. n" as an abbreviation for almost all n. So $a_n \leq b_n$ a.a. n iff there exists $N < \infty$ such that $a_n \leq b_n$ for all $n \geq N$.

We will also write $\underline{\lim}$ *for* $\liminf_{n\to\infty}$ *and* $\overline{\lim}$ *for* $\limsup\limits_{n\to\infty}$.

Remark 1.4. Notice that if $a_k := \inf\{x_k : k \geq n\}$ and $b_k := \sup\{x_k : k \geq n\}$, then $\{a_k\}$ is an increasing sequence while $\{b_k\}$ is a decreasing sequence. Therefore the limits in Eq. (1.3) and Eq. (1.4) always exist in $\bar{\mathbb{R}}$ and

$$\liminf_{n\to\infty} x_n = \sup_n \inf\{x_k : k \geq n\} \text{ and}$$
$$\limsup_{n\to\infty} x_n = \inf_n \sup\{x_k : k \geq n\}.$$

The following proposition contains some basic properties of liminfs and limsups.

Proposition 1.5. *Let* $\{a_n\}_{n=1}^\infty$ *and* $\{b_n\}_{n=1}^\infty$ *be two sequences of real numbers. Then*

1. $\liminf_{n\to\infty} a_n \leq \limsup\limits_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} a_n$ *exists in* $\bar{\mathbb{R}}$ *iff*
$$\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n \in \bar{\mathbb{R}}.$$
2. *There is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \limsup\limits_{n\to\infty} a_n$. *Similarly, there is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \liminf_{n\to\infty} a_n$.
3. $$\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{1.5}$$
whenever the right side of this equation is not of the form $\infty - \infty$.
4. *If* $a_n \geq 0$ *and* $b_n \geq 0$ *for all* $n \in \mathbb{N}$, *then*
$$\limsup_{n\to\infty}(a_n b_n) \leq \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{1.6}$$
provided the right hand side of (1.6) is not of the form $0 \cdot \infty$ *or* $\infty \cdot 0$.

Proof. 1. Since

$$\inf\{a_k : k \geq n\} \leq \sup\{a_k : k \geq n\} \; \forall\, n,$$

$$\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

Now suppose that $\liminf_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n = a \in \mathbb{R}$. Then for all $\varepsilon > 0$, there is an integer N such that

$$a - \varepsilon \leq \inf\{a_k : k \geq N\} \leq \sup\{a_k : k \geq N\} \leq a + \varepsilon,$$

i.e.

$$a - \varepsilon \leq a_k \leq a + \varepsilon \text{ for all } k \geq N.$$

Hence by the definition of the limit, $\lim_{k\to\infty} a_k = a$. If $\liminf_{n\to\infty} a_n = \infty$, then we know for all $M \in (0,\infty)$ there is an integer N such that

$$M \leq \inf\{a_k : k \geq N\}$$

and hence $\lim_{n\to\infty} a_n = \infty$. The case where $\limsup\limits_{n\to\infty} a_n = -\infty$ is handled similarly.

Conversely, suppose that $\lim_{n\to\infty} a_n = A \in \bar{\mathbb{R}}$ exists. If $A \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$ such that $|A - a_n| \leq \varepsilon$ for all $n \geq N(\varepsilon)$, i.e.

$$A - \varepsilon \leq a_n \leq A + \varepsilon \text{ for all } n \geq N(\varepsilon).$$

From this we learn that

$$A - \varepsilon \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n \leq A + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$A \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n \leq A,$$

i.e. that $A = \liminf_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$. If $A = \infty$, then for all $M > 0$ there exists $N = N(M)$ such that $a_n \geq M$ for all $n \geq N$. This show that $\liminf_{n\to\infty} a_n \geq M$ and since M is arbitrary it follows that

$$\infty \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

The proof for the case $A = -\infty$ is analogous to the $A = \infty$ case.

2. – 4. The remaining items are left as an exercise to the reader. It may be useful to keep the following simple example in mind. Let $a_n = (-1)^n$ and $b_n = -a_n = (-1)^{n+1}$. Then $a_n + b_n = 0$ so that

$$0 = \lim_{n\to\infty}(a_n + b_n) = \liminf_{n\to\infty}(a_n + b_n) = \limsup_{n\to\infty}(a_n + b_n)$$

while

$$\liminf_{n\to\infty} a_n = \liminf_{n\to\infty} b_n = -1 \text{ and}$$
$$\limsup_{n\to\infty} a_n = \limsup_{n\to\infty} b_n = 1.$$

Thus in this case we have

$$\limsup_{n\to\infty}(a_n+b_n)<\limsup_{n\to\infty}a_n+\limsup_{n\to\infty}b_n \text{ and}$$
$$\liminf_{n\to\infty}(a_n+b_n)>\liminf_{n\to\infty}a_n+\liminf_{n\to\infty}b_n.$$

■

We will refer to the following basic proposition as the monotone convergence theorem for sums (MCT for short).

Proposition 1.6 (MCT for sums). *Suppose that for each $n\in\mathbb{N}$, $\{f_n(i)\}_{i=1}^\infty$ is a sequence in $[0,\infty]$ such that $\uparrow \lim_{n\to\infty} f_n(i)=f(i)$ by which we mean $f_n(i)\uparrow f(i)$ as $n\to\infty$. Then*

$$\lim_{n\to\infty}\sum_{i=1}^\infty f_n(i)=\sum_{i=1}^\infty f(i),\ \ i.e.$$
$$\lim_{n\to\infty}\sum_{i=1}^\infty f_n(i)=\sum_{i=1}^\infty \lim_{n\to\infty} f_n(i).$$

We allow for the possibility that these expression may equal to $+\infty$.

Proof. Let $M:=\uparrow \lim_{n\to\infty}\sum_{i=1}^\infty f_n(i)$. As $f_n(i)\le f(i)$ for all n it follows that $\sum_{i=1}^\infty f_n(i)\le\sum_{i=1}^\infty f(i)$ for all n and therefore passing to the limit shows $M\le\sum_{i=1}^\infty f(i)$. If $N\in\mathbb{N}$ we have,

$$\sum_{i=1}^N f(i)=\sum_{i=1}^N\lim_{n\to\infty}f_n(i)=\lim_{n\to\infty}\sum_{i=1}^N f_n(i)\le\lim_{n\to\infty}\sum_{i=1}^\infty f_n(i)=M.$$

Letting $N\uparrow\infty$ in this equation then shows $\sum_{i=1}^\infty f(i)\le M$ which completes the proof. ■

Proposition 1.7 (Tonelli's theorem for sums). *If $\{a_{kn}\}_{k,n=1}^\infty\subset[0,\infty]$, then*

$$\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}=\sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn}.$$

Here we allow for one and hence both sides to be infinite.

Proof. First Proof. Let $S_N(k):=\sum_{n=1}^N a_{kn}$, then by the MCT (Proposition 1.6),

$$\lim_{N\to\infty}\sum_{k=1}^\infty S_N(k)=\sum_{k=1}^\infty\lim_{N\to\infty}S_N(k)=\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}.$$

On the other hand,

$$\sum_{k=1}^\infty S_N(k)=\sum_{k=1}^\infty\sum_{n=1}^N a_{kn}=\sum_{n=1}^N\sum_{k=1}^\infty a_{kn}$$

so that

$$\lim_{N\to\infty}\sum_{k=1}^\infty S_N(k)=\lim_{N\to\infty}\sum_{n=1}^N\sum_{k=1}^\infty a_{kn}=\sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn}.$$

Second Proof. Let

$$M:=\sup\left\{\sum_{k=1}^K\sum_{n=1}^N a_{kn}:K,N\in\mathbb{N}\right\}=\sup\left\{\sum_{n=1}^N\sum_{k=1}^K a_{kn}:K,N\in\mathbb{N}\right\}$$

and

$$L:=\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}.$$

Since

$$L=\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}=\lim_{K\to\infty}\sum_{k=1}^K\sum_{n=1}^\infty a_{kn}=\lim_{K\to\infty}\lim_{N\to\infty}\sum_{k=1}^K\sum_{n=1}^N a_{kn}$$

and $\sum_{k=1}^K\sum_{n=1}^N a_{kn}\le M$ for all K and N, it follows that $L\le M$. Conversely,

$$\sum_{k=1}^K\sum_{n=1}^N a_{kn}\le\sum_{k=1}^K\sum_{n=1}^\infty a_{kn}\le\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}=L$$

and therefore taking the supremum of the left side of this inequality over K and N shows that $M\le L$. Thus we have shown

$$\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}=M.$$

By symmetry (or by a similar argument), we also have that $\sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn}=M$ and hence the proof is complete. ■

You are asked to prove the next three results in the exercises.

Proposition 1.8 (Fubini for sums). *Suppose $\{a_{kn}\}_{k,n=1}^\infty\subset\mathbb{R}$ such that*

$$\sum_{k=1}^\infty\sum_{n=1}^\infty|a_{kn}|=\sum_{n=1}^\infty\sum_{k=1}^\infty|a_{kn}|<\infty.$$

Then

$$\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}=\sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn}.$$

Example 1.9 (Counter example). Let $\{S_{mn}\}_{m,n=1}^{\infty}$ be any sequence of complex numbers such that $\lim_{m\to\infty} S_{mn} = 1$ for all n and $\lim_{n\to\infty} S_{mn} = 0$ for all n. For example, take $S_{mn} = 1_{m\geq n} + \frac{1}{n} 1_{m<n}$. Then define $\{a_{ij}\}_{i,j=1}^{\infty}$ so that

$$S_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}.$$

Then

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \lim_{m\to\infty} \lim_{n\to\infty} S_{mn} = 0 \neq 1 = \lim_{n\to\infty} \lim_{m\to\infty} S_{mn} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

To find a_{ij}, set $S_{mn} = 0$ if $m = 0$ or $n = 0$, then

$$S_{mn} - S_{m-1,n} = \sum_{j=1}^{n} a_{mj}$$

and

$$\begin{aligned} a_{mn} &= S_{mn} - S_{m-1,n} - (S_{m,n-1} - S_{m-1,n-1}) \\ &= S_{mn} - S_{m-1,n} - S_{m,n-1} + S_{m-1,n-1}. \end{aligned}$$

Proposition 1.10 (Fatou's Lemma for sums). *Suppose that for each $n \in \mathbb{N}$, $\{h_n(i)\}_{i=1}^{\infty}$ is any sequence in $[0,\infty]$, then*

$$\sum_{i=1}^{\infty} \liminf_{n\to\infty} h_n(i) \leq \liminf_{n\to\infty} \sum_{i=1}^{\infty} h_n(i).$$

The next proposition is referred to as the dominated convergence theorem (DCT for short) for sums.

Proposition 1.11 (DCT for sums). *Suppose that for each $n \in \mathbb{N}$, $\{f_n(i)\}_{i=1}^{\infty} \subset \mathbb{R}$ is a sequence and $\{g_n(i)\}_{i=1}^{\infty}$ is a sequence in $[0,\infty)$ such that;*

1. *$\sum_{i=1}^{\infty} g_n(i) < \infty$ for all n,*
2. *$f(i) = \lim_{n\to\infty} f_n(i)$ and $g(i) := \lim_{n\to\infty} g_n(i)$ exists for each i,*
3. *$|f_n(i)| \leq g_n(i)$ for all i and n,*
4. *$\lim_{n\to\infty} \sum_{i=1}^{\infty} g_n(i) = \sum_{i=1}^{\infty} g(i) < \infty$.*

Then

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} f_n(i) = \sum_{i=1}^{\infty} \lim_{n\to\infty} f_n(i) = \sum_{i=1}^{\infty} f(i).$$

(Often this proposition is used in the special case where $g_n = g$ for all n.)

Exercise 1.1. Prove Proposition 1.8. **Hint:** Let $a_{kn}^{+} := \max(a_{kn}, 0)$ and $a_{kn}^{-} = \max(-a_{kn}, 0)$ and observe that; $a_{kn} = a_{kn}^{+} - a_{kn}^{-}$ and $|a_{kn}^{+}| + |a_{kn}^{-}| = |a_{kn}|$. Now apply Proposition 1.7 with a_{kn} replaced by a_{kn}^{+} and a_{kn}^{-}.

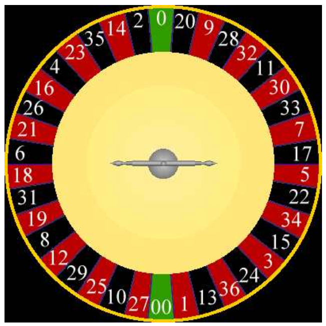
Exercise 1.2. Prove Proposition 1.10. **Hint:** apply the MCT by applying the monotone convergence theorem with $f_n(i) := \inf_{m\geq n} h_m(i)$.

Exercise 1.3. Prove Proposition 1.11. **Hint:** Apply Fatou's lemma twice. Once with $h_n(i) = g_n(i) + f_n(i)$ and once with $h_n(i) = g_n(i) - f_n(i)$.

2

Basic Probabilistic Notions

Definition 2.1. *A sample space Ω is a set which is to represents all possible outcomes of an "experiment."*



Example 2.2.
1. The sample space for flipping a coin one time could be taken to be, $\Omega = \{0, 1\}$.
2. The sample space for flipping a coin N -times could be taken to be, $\Omega = \{0, 1\}^N$ and for flipping an infinite number of times,
$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i \in \{0, 1\}\} = \{0, 1\}^{\mathbb{N}}.$$
3. If we have a roulette wheel with 38 entries, then we might take
$$\Omega = \{00, 0, 1, 2, \dots, 36\}$$
for one spin,
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^N$$
for N spins, and
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^{\mathbb{N}}$$
for an infinite number of spins.
4. If we throw darts at a board of radius R, we may take
$$\Omega = D_R := \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq R\right\}$$
for one throw,
$$\Omega = D_R^N$$
for N throws, and
$$\Omega = D_R^{\mathbb{N}}$$
for an infinite number of throws.
5. Suppose we release a perfume particle at location $x \in \mathbb{R}^3$ and follow its motion for all time, $0 \leq t < \infty$. In this case, we might take,
$$\Omega = \left\{\omega \in C\left([0, \infty), \mathbb{R}^3\right) : \omega(0) = x\right\}.$$

Definition 2.3. *An* ***event****, A, is a subset of Ω. Given $A \subset \Omega$ we also define the* ***indicator function of*** *A by*
$$1_A(\omega) := \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases}.$$

Example 2.4. Suppose that $\Omega = \{0, 1\}^{\mathbb{N}}$ is the sample space for flipping a coin an infinite number of times. Here $\omega_n = 1$ represents the fact that a head was thrown on the n^{th} – toss, while $\omega_n = 0$ represents a tail on the n^{th} – toss.

1. $A = \{\omega \in \Omega : \omega_3 = 1\}$ represents the event that the third toss was a head.
2. $A = \cup_{i=1}^{\infty} \{\omega \in \Omega : \omega_i = \omega_{i+1} = 1\}$ represents the event that (at least) two heads are tossed twice in a row at some time.
3. $A = \cap_{N=1}^{\infty} \cup_{n \geq N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where there are infinitely many heads tossed in the sequence.
4. $A = \cup_{N=1}^{\infty} \cap_{n \geq N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where heads occurs from some time onwards, i.e. $\omega \in A$ iff there exists, $N = N(\omega)$ such that $\omega_n = 1$ for all $n \geq N$.

Ideally we would like to assign a probability, $P(A)$, to all events $A \subset \Omega$. Given a physical experiment, we think of assigning this probability as follows. Run the experiment many times to get sample points, $\omega(n) \in \Omega$ for each $n \in \mathbb{N}$, then try to "define" $P(A)$ by

$$P(A) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} 1_A(\omega(k)) \tag{2.1}$$
$$= \lim_{N \to \infty} \frac{1}{N} \# \{1 \leq k \leq N : \omega(k) \in A\}. \tag{2.2}$$

That is we think of $P(A)$ as being the long term relative frequency that the event A occurred for the sequence of experiments, $\{\omega(k)\}_{k=1}^{\infty}$.

Similarly supposed that A and B are two events and we wish to know how likely the event A is given that we know that B has occurred. Thus we would like to compute:

$$P(A|B) = \lim_{N\to\infty} \frac{\#\{k : 1 \le k \le N \text{ and } \omega_k \in A\cap B\}}{\#\{k : 1 \le k \le N \text{ and } \omega_k \in B\}},$$

which represents the frequency that A occurs given that we know that B has occurred. This may be rewritten as

$$\begin{aligned} P(A|B) &= \lim_{N\to\infty} \frac{\frac{1}{N}\#\{k : 1 \le k \le N \text{ and } \omega_k \in A\cap B\}}{\frac{1}{N}\#\{k : 1 \le k \le N \text{ and } \omega_k \in B\}} \\ &= \frac{P(A\cap B)}{P(B)}. \end{aligned}$$

Definition 2.5. *If B is a non-null event, i.e. $P(B) > 0$, define the* ***conditional probability of*** *A* ***given*** *B by,*

$$P(A|B) := \frac{P(A\cap B)}{P(B)}.$$

There are of course a number of problems with this definition of P in Eq. (2.1) including the fact that it is not mathematical nor necessarily well defined. For example the limit may not exist. But ignoring these technicalities for the moment, let us point out three key properties that P should have.

1. $P(A) \in [0,1]$ for all $A \subset \Omega$.
2. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
3. **Additivity.** If A and B are disjoint event, i.e. $A\cap B = AB = \emptyset$, then $1_{A\cup B} = 1_A + 1_B$ so that

$$\begin{aligned} P(A\cup B) &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} 1_{A\cup B}(\omega(k)) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N}\left[1_A(\omega(k)) + 1_B(\omega(k))\right] \\ &= \lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=1}^{N} 1_A(\omega(k)) + \frac{1}{N}\sum_{k=1}^{N} 1_B(\omega(k))\right] \\ &= P(A) + P(B). \end{aligned}$$

4. **Countable Additivity.** If $\{A_j\}_{j=1}^{\infty}$ are pairwise disjoint events (i.e. $A_j \cap A_k = \emptyset$ for all $j \ne k$), then again, $1_{\cup_{j=1}^{\infty} A_j} = \sum_{j=1}^{\infty} 1_{A_j}$ and therefore we might hope that,

$$\begin{aligned} P\left(\cup_{j=1}^{\infty} A_j\right) &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} 1_{\cup_{j=1}^{\infty} A_j}(\omega(k)) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{\infty} 1_{A_j}(\omega(k)) \\ &= \lim_{N\to\infty} \sum_{j=1}^{\infty} \frac{1}{N}\sum_{k=1}^{N} 1_{A_j}(\omega(k)) \\ &\overset{?}{=} \sum_{j=1}^{\infty} \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} 1_{A_j}(\omega(k)) \text{ (by a leap of faith)} \\ &= \sum_{j=1}^{\infty} P(A_j). \end{aligned}$$

Example 2.6. Let us consider the tossing of a coin N times with a fair coin. In this case we would expect that every $\omega \in \Omega$ is equally likely, i.e. $P(\{\omega\}) = \frac{1}{2^N}$. Assuming this we are then forced to define

$$P(A) = \frac{1}{2^N}\#(A).$$

Observe that this probability has the following property. Suppose that $\sigma \in \{0,1\}^k$ is a given sequence, then

$$P(\{\omega : (\omega_1, \dots, \omega_k) = \sigma\}) = \frac{1}{2^N}\cdot 2^{N-k} = \frac{1}{2^k}.$$

That is if we ignore the flips after time k, the resulting probabilities are the same as if we only flipped the coin k times.

Example 2.7. The previous example suggests that if we flip a fair coin an infinite number of times, so that now $\Omega = \{0,1\}^{\mathbb{N}}$, then we should define

$$P(\{\omega \in \Omega : (\omega_1, \dots, \omega_k) = \sigma\}) = \frac{1}{2^k} \tag{2.3}$$

for any $k \ge 1$ and $\sigma \in \{0,1\}^k$. Assuming there exists a probability, $P : 2^{\Omega} \to [0,1]$ such that Eq. (2.3) holds, we would like to compute, for example, the probability of the event B where an infinite number of heads are tossed. To try to compute this, let

$$\begin{aligned} A_n &= \{\omega \in \Omega : \omega_n = 1\} = \{\text{heads at time } n\} \\ B_N &:= \cup_{n\ge N} A_n = \{\text{at least one heads at time } N \text{ or later}\} \end{aligned}$$

and

$$B = \cap_{N=1}^{\infty} B_N = \{A_n \text{ i.o.}\} = \cap_{N=1}^{\infty} \cup_{n\ge N} A_n.$$

Since

$$B_N^c = \cap_{n \geq N} A_n^c \subset \cap_{M \geq n \geq N} A_n^c = \{\omega \in \Omega : \omega_N = \omega_{N+1} = \cdots = \omega_M = 0\},$$

we see that

$$P(B_N^c) \leq \frac{1}{2^{M-N}} \to 0 \text{ as } M \to \infty.$$

Therefore, $P(B_N) = 1$ for all N. If we assume that P is continuous under taking decreasing limits we may conclude, using $B_N \downarrow B$, that

$$P(B) = \lim_{N \to \infty} P(B_N) = 1.$$

Without this continuity assumption we would not be able to compute $P(B)$.

The unfortunate fact is that we can not always assign a desired probability function, $P(A)$, for all $A \subset \Omega$. For example we have the following negative theorem.

Theorem 2.8 (No-Go Theorem). *Let $S = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle. Then there is no probability function, $P : 2^S \to [0,1]$ such that $P(S) = 1$, P is invariant under rotations, and P is continuous under taking decreasing limits.*

Proof. We are going to use the fact proved below in Proposition 5.3, that the continuity condition on P is equivalent to the $\sigma-$additivity of P. For $z \in S$ and $N \subset S$ let

$$zN := \{zn \in S : n \in N\}, \tag{2.4}$$

that is to say $e^{i\theta}N$ is the set N rotated counter clockwise by angle θ. By assumption, we are supposing that

$$P(zN) = P(N) \tag{2.5}$$

for all $z \in S$ and $N \subset S$.

Let

$$R := \{z = e^{i2\pi t} : t \in \mathbb{Q}\} = \{z = e^{i2\pi t} : t \in [0,1) \cap \mathbb{Q}\}$$

– a countable subgroup of S. As above R acts on S by rotations and divides S up into equivalence classes, where $z, w \in S$ are equivalent if $z = rw$ for some $r \in R$. Choose (using the axiom of choice) one representative point n from each of these equivalence classes and let $N \subset S$ be the set of these representative points. Then every point $z \in S$ may be uniquely written as $z = nr$ with $n \in N$ and $r \in R$. That is to say

$$S = \sum_{r \in R} (rN) \tag{2.6}$$

where $\sum_\alpha A_\alpha$ is used to denote the union of pair-wise disjoint sets $\{A_\alpha\}$. By Eqs. (2.5) and (2.6),

$$1 = P(S) = \sum_{r \in R} P(rN) = \sum_{r \in R} P(N). \tag{2.7}$$

We have thus arrived at a contradiction, since the right side of Eq. (2.7) is either equal to 0 or to ∞ depending on whether $P(N) = 0$ or $P(N) > 0$. ∎

To avoid this problem, we are going to have to relinquish the idea that P should necessarily be defined on all of 2^Ω. So we are going to only define P on particular subsets, $\mathcal{B} \subset 2^\Omega$. We will developed this below.

Part II

Formal Development

3

Preliminaries

3.1 Set Operations

Let $\mathbb{N}$ denote the positive integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ be the non-negative integers and $\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N})$ – the positive and negative integers including 0, $\mathbb{Q}$ the rational numbers, $\mathbb{R}$ the real numbers, and $\mathbb{C}$ the complex numbers. We will also use $\mathbb{F}$ to stand for either of the fields $\mathbb{R}$ or $\mathbb{C}$.

Notation 3.1 *Given two sets X and Y, let Y^X denote the collection of all functions $f : X \to Y$. If $X = \mathbb{N}$, we will say that $f \in Y^{\mathbb{N}}$ is a sequence with values in Y and often write f_n for $f(n)$ and express f as $\{f_n\}_{n=1}^\infty$. If $X = \{1, 2, \ldots, N\}$, we will write Y^N in place of $Y^{\{1,2,\ldots,N\}}$ and denote $f \in Y^N$ by $f = (f_1, f_2, \ldots, f_N)$ where $f_n = f(n)$.*

Notation 3.2 *More generally if $\{X_\alpha : \alpha \in A\}$ is a collection of non-empty sets, let $X_A = \prod_{\alpha \in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map defined by $\pi_\alpha(x) = x_\alpha$. If If $X_\alpha = X$ for some fixed space X, then we will write $\prod_{\alpha \in A} X_\alpha$ as X^A rather than X_A.*

Recall that an element $x \in X_A$ is a "**choice function**," i.e. an assignment $x_\alpha := x(\alpha) \in X_\alpha$ for each $\alpha \in A$. The **axiom of choice** states that $X_A \neq \emptyset$ provided that $X_\alpha \neq \emptyset$ for each $\alpha \in A$.

Notation 3.3 *Given a set X, let 2^X denote the **power set** of X – the collection of all subsets of X including the empty set.*

The reason for writing the power set of X as 2^X is that if we think of 2 meaning $\{0, 1\}$, then an element of $a \in 2^X = \{0, 1\}^X$ is completely determined by the set

$$A := \{x \in X : a(x) = 1\} \subset X.$$

In this way elements in $\{0, 1\}^X$ are in one to one correspondence with subsets of X.

For $A \in 2^X$ let

$$A^c := X \setminus A = \{x \in X : x \notin A\}$$

and more generally if $A, B \subset X$ let

$$B \setminus A := \{x \in B : x \notin A\} = B \cap A^c.$$

We also define the symmetric difference of A and B by

$$A \triangle B := (B \setminus A) \cup (A \setminus B).$$

As usual if $\{A_\alpha\}_{\alpha \in I}$ is an indexed collection of subsets of X we define the union and the intersection of this collection by

$$\cup_{\alpha \in I} A_\alpha := \{x \in X : \exists\, \alpha \in I \ni x \in A_\alpha\} \text{ and}$$
$$\cap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \,\forall\, \alpha \in I\}.$$

Notation 3.4 *We will also write $\sum_{\alpha \in I} A_\alpha$ for $\cup_{\alpha \in I} A_\alpha$ in the case that $\{A_\alpha\}_{\alpha \in I}$ are pairwise disjoint, i.e. $A_\alpha \cap A_\beta = \emptyset$ if $\alpha \neq \beta$.*

Notice that $\cup$ is closely related to $\exists$ and $\cap$ is closely related to $\forall$. For example let $\{A_n\}_{n=1}^\infty$ be a sequence of subsets from X and define

$$\inf_{k \geq n} A_n := \cap_{k \geq n} A_k,$$

$$\sup_{k \geq n} A_n := \cup_{k \geq n} A_k,$$

$$\limsup_{n \to \infty} A_n := \{A_n \text{ i.o.}\} := \{x \in X : \#\{n : x \in A_n\} = \infty\}$$

and

$$\liminf_{n \to \infty} A_n := \{A_n \text{ a.a.}\} := \{x \in X : x \in A_n \text{ for all } n \text{ sufficiently large}\}.$$

(One should read $\{A_n \text{ i.o.}\}$ as A_n infinitely often and $\{A_n \text{ a.a.}\}$ as A_n almost always.) Then $x \in \{A_n \text{ i.o.}\}$ iff

$$\forall N \in \mathbb{N}\ \exists\ n \geq N \ni x \in A_n$$

and this may be expressed as

$$\{A_n \text{ i.o.}\} = \cap_{N=1}^\infty \cup_{n \geq N} A_n.$$

Similarly, $x \in \{A_n \text{ a.a.}\}$ iff

$$\exists\ N \in \mathbb{N} \ni \forall\ n \geq N, \ \ x \in A_n$$

which may be written as

$$\{A_n \text{ a.a.}\} = \cup_{N=1}^\infty \cap_{n \geq N} A_n.$$

Definition 3.5. *Given a set $A \subset X$, let*

$$1_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$

be the ***indicator function*** *of A.*

Lemma 3.6. *We have:*

1. $(\cup_n A_n)^c = \cap_n A_n^c$,
2. $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ a.a.}\}$,
3. $\limsup_{n\to\infty} A_n = \{x \in X : \sum_{n=1}^\infty 1_{A_n}(x) = \infty\}$,
4. $\liminf_{n\to\infty} A_n = \left\{x \in X : \sum_{n=1}^\infty 1_{A_n^c}(x) < \infty\right\}$,
5. $\sup_{k\ge n} 1_{A_k}(x) = 1_{\cup_{k\ge n} A_k} = 1_{\sup_{k\ge n} A_k}$,
6. $\inf_{k\ge n} 1_{A_k}(x) = 1_{\cap_{k\ge n} A_k} = 1_{\inf_{k\ge n} A_k}$,
7. $1_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} 1_{A_n}$, and
8. $1_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} 1_{A_n}$.

Definition 3.7. *A set X is said to be* ***countable*** *if is empty or there is an injective function $f : X \to \mathbb{N}$, otherwise X is said to be* ***uncountable.***

Lemma 3.8 (Basic Properties of Countable Sets).

1. *If $A \subset X$ is a subset of a countable set X then A is countable.*
2. *Any infinite subset $\Lambda \subset \mathbb{N}$ is in one to one correspondence with $\mathbb{N}$.*
3. *A non-empty set X is countable iff there exists a surjective map, $g : \mathbb{N} \to X$.*
4. *If X and Y are countable then $X \times Y$ is countable.*
5. *Suppose for each $m \in \mathbb{N}$ that A_m is a countable subset of a set X, then $A = \cup_{m=1}^\infty A_m$ is countable. In short, the countable union of countable sets is still countable.*
6. *If X is an infinite set and Y is a set with at least two elements, then Y^X is uncountable. In particular 2^X is uncountable for any infinite set X.*

Proof. 1. If $f : X \to \mathbb{N}$ is an injective map then so is the restriction, $f|_A$, of f to the subset A. 2. Let $f(1) = \min \Lambda$ and define f inductively by

$$f(n+1) = \min(\Lambda \setminus \{f(1), \dots, f(n)\}).$$

Since Λ is infinite the process continues indefinitely. The function $f : \mathbb{N} \to \Lambda$ defined this way is a bijection.

3. If $g : \mathbb{N} \to X$ is a surjective map, let

$$f(x) = \min g^{-1}(\{x\}) = \min\{n \in \mathbb{N} : f(n) = x\}.$$

Then $f : X \to \mathbb{N}$ is injective which combined with item

2. (taking $\Lambda = f(X)$) shows X is countable. Conversely if $f : X \to \mathbb{N}$ is injective let $x_0 \in X$ be a fixed point and define $g : \mathbb{N} \to X$ by $g(n) = f^{-1}(n)$ for $n \in f(X)$ and $g(n) = x_0$ otherwise.

4. Let us first construct a bijection, h, from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$. To do this put the elements of $\mathbb{N} \times \mathbb{N}$ into an array of the form

$$\begin{pmatrix} (1,1) & (1,2) & (1,3) & \dots \\ (2,1) & (2,2) & (2,3) & \dots \\ (3,1) & (3,2) & (3,3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and then "count" these elements by counting the sets $\{(i,j) : i+j = k\}$ one at a time. For example let $h(1) = (1,1)$, $h(2) = (2,1)$, $h(3) = (1,2)$, $h(4) = (3,1)$, $h(5) = (2,2)$, $h(6) = (1,3)$ and so on. If $f : \mathbb{N} \to X$ and $g : \mathbb{N} \to Y$ are surjective functions, then the function $(f \times g) \circ h : \mathbb{N} \to X \times Y$ is surjective where $(f \times g)(m,n) := (f(m), g(n))$ for all $(m,n) \in \mathbb{N} \times \mathbb{N}$.

5. If $A = \emptyset$ then A is countable by definition so we may assume $A \neq \emptyset$. With out loss of generality we may assume $A_1 \neq \emptyset$ and by replacing A_m by A_1 if necessary we may also assume $A_m \neq \emptyset$ for all m. For each $m \in \mathbb{N}$ let $a_m : \mathbb{N} \to A_m$ be a surjective function and then define $f : \mathbb{N} \times \mathbb{N} \to \cup_{m=1}^\infty A_m$ by $f(m,n) := a_m(n)$. The function f is surjective and hence so is the composition, $f \circ h : \mathbb{N} \to \cup_{m=1}^\infty A_m$, where $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ is the bijection defined above.

6. Let us begin by showing $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$ is uncountable. For sake of contradiction suppose $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is a surjection and write $f(n)$ as $(f_1(n), f_2(n), f_3(n), \dots)$. Now define $a \in \{0,1\}^{\mathbb{N}}$ by $a_n := 1 - f_n(n)$. By construction $f_n(n) \neq a_n$ for all n and so $a \notin f(\mathbb{N})$. This contradicts the assumption that f is surjective and shows $2^{\mathbb{N}}$ is uncountable. For the general case, since $Y_0^X \subset Y^X$ for any subset $Y_0 \subset Y$, if Y_0^X is uncountable then so is Y^X. In this way we may assume Y_0 is a two point set which may as well be $Y_0 = \{0,1\}$. Moreover, since X is an infinite set we may find an injective map $x : \mathbb{N} \to X$ and use this to set up an injection, $i : 2^{\mathbb{N}} \to 2^X$ by setting $i(A) := \{x_n : n \in \mathbb{N}\} \subset X$ for all $A \subset \mathbb{N}$. If 2^X were countable we could find a surjective map $f : 2^X \to \mathbb{N}$ in which case $f \circ i : 2^{\mathbb{N}} \to \mathbb{N}$ would be surjective as well. However this is impossible since we have already seed that $2^{\mathbb{N}}$ is uncountable. ■

3.2 Exercises

Let $f : X \to Y$ be a function and $\{A_i\}_{i\in I}$ be an indexed family of subsets of Y, verify the following assertions.

Exercise 3.1. $(\cap_{i\in I} A_i)^c = \cup_{i\in I} A_i^c$.

Exercise 3.2. Suppose that $B \subset Y$, show that $B \setminus (\cup_{i\in I} A_i) = \cap_{i\in I}(B \setminus A_i)$.

Exercise 3.3. $f^{-1}(\cup_{i\in I} A_i) = \cup_{i\in I} f^{-1}(A_i)$.

Exercise 3.4. $f^{-1}(\cap_{i\in I} A_i) = \cap_{i\in I} f^{-1}(A_i)$.

Exercise 3.5. Find a counterexample which shows that $f(C \cap D) = f(C) \cap f(D)$ need not hold.

Example 3.9. Let $X = \{a, b, c\}$ and $Y = \{1, 2\}$ and define $f(a) = f(b) = 1$ and $f(c) = 2$. Then $\emptyset = f(\{a\} \cap \{b\}) \neq f(\{a\}) \cap f(\{b\}) = \{1\}$ and $\{1, 2\} = f(\{a\}^c) \neq f(\{a\})^c = \{2\}$.

3.3 Algebraic sub-structures of sets

Definition 3.10. *A collection of subsets $\mathcal{A}$ of a set X is a π – **system** or **multiplicative system** if $\mathcal{A}$ is closed under taking finite intersections.*

Definition 3.11. *A collection of subsets $\mathcal{A}$ of a set X is an **algebra (Field)** if*

1. *$\emptyset, X \in \mathcal{A}$*
2. *$A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$*
3. *$\mathcal{A}$ is closed under finite unions, i.e. if $A_1, \ldots, A_n \in \mathcal{A}$ then $A_1 \cup \cdots \cup A_n \in \mathcal{A}$. In view of conditions 1. and 2., 3. is equivalent to*

3′. *$\mathcal{A}$ is closed under finite intersections.*

Definition 3.12. *A collection of subsets $\mathcal{B}$ of X is a σ – **algebra** (or sometimes called a σ – **field**) if $\mathcal{B}$ is an algebra which also closed under countable unions, i.e. if $\{A_i\}_{i=1}^\infty \subset \mathcal{B}$, then $\cup_{i=1}^\infty A_i \in \mathcal{B}$. (Notice that since $\mathcal{B}$ is also closed under taking complements, $\mathcal{B}$ is also closed under taking countable intersections.)*

Example 3.13. Here are some examples of algebras.

1. $\mathcal{B} = 2^X$, then $\mathcal{B}$ is a σ – algebra.
2. $\mathcal{B} = \{\emptyset, X\}$ is a σ – algebra called the trivial σ – field.
3. Let $X = \{1, 2, 3\}$, then $\mathcal{A} = \{\emptyset, X, \{1\}, \{2, 3\}\}$ is an algebra while, $\mathcal{S} := \{\emptyset, X, \{2, 3\}\}$ is a not an algebra but is a π – system.

Proposition 3.14. *Let $\mathcal{E}$ be any collection of subsets of X. Then there exists a unique smallest algebra $\mathcal{A}(\mathcal{E})$ and σ – algebra $\sigma(\mathcal{E})$ which contains $\mathcal{E}$.*

Proof. Simply take

$$\mathcal{A}(\mathcal{E}) := \bigcap\{\mathcal{A} : \mathcal{A} \text{ is an algebra such that } \mathcal{E} \subset \mathcal{A}\}$$

and

$$\sigma(\mathcal{E}) := \bigcap\{\mathcal{M} : \mathcal{M} \text{ is a } \sigma - \text{algebra such that } \mathcal{E} \subset \mathcal{M}\}.$$

■

Example 3.15. Suppose $X = \{1, 2, 3\}$ and $\mathcal{E} = \{\emptyset, X, \{1, 2\}, \{1, 3\}\}$, see Figure 3.1. Then

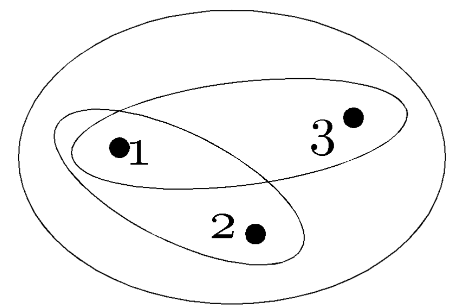


Fig. 3.1. A collection of subsets.

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = 2^X.$$

On the other hand if $\mathcal{E} = \{\{1, 2\}\}$, then $\mathcal{A}(\mathcal{E}) = \{\emptyset, X, \{1, 2\}, \{3\}\}$.

Exercise 3.6. Suppose that $\mathcal{E}_i \subset 2^X$ for $i = 1, 2$. Show that $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \mathcal{A}(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \mathcal{A}(\mathcal{E}_1)$. Similarly show, $\sigma(\mathcal{E}_1) = \sigma(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \sigma(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \sigma(\mathcal{E}_1)$. Give a simple example where $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ while $\mathcal{E}_1 \neq \mathcal{E}_2$.

In this course we will often be interested in the Borel σ – algebra on a topological space.

Definition 3.16 (Borel σ – field). *The **Borel σ – algebra**, $\mathcal{B} = \mathcal{B}_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$, on $\mathbb{R}$ is the smallest σ -field containing all of the open subsets of $\mathbb{R}$. More generally if (X, τ) is a topological space, the Borel σ – algebra on X is $\mathcal{B}_X := \sigma(\tau)$ – i.e. the smallest σ – algebra containing all open (closed) subsets of X.*

Exercise 3.7. Verify the Borel σ – algebra, $\mathcal{B}_{\mathbb{R}}$, is generated by any of the following collection of sets:

$$1.\ \{(a,\infty): a \in \mathbb{R}\},\ 2.\ \{(a,\infty): a \in \mathbb{Q}\} \text{ or } 3.\ \{[a,\infty): a \in \mathbb{Q}\}.$$

Hint: make use of Exercise 3.6.

We will postpone a more in depth study of σ – algebras until later. For now, let us concentrate on understanding the the simpler notion of an algebra.

Definition 3.17. *Let X be a set. We say that a family of sets $\mathcal{F} \subset 2^X$ is a **partition** of X if distinct members of $\mathcal{F}$ are disjoint and if X is the union of the sets in $\mathcal{F}$.*

Example 3.18. Let X be a set and $\mathcal{E} = \{A_1, \ldots, A_n\}$ where $A_1, \ldots, A_n$ is a partition of X. In this case

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset \{1,2,\ldots,n\}\}$$

where $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. Notice that

$$\#\left(\mathcal{A}(\mathcal{E})\right) = \#(2^{\{1,2,\ldots,n\}}) = 2^n.$$

Example 3.19. Suppose that X is a set and that $\mathcal{A} \subset 2^X$ is a finite algebra, i.e. $\#(\mathcal{A}) < \infty$. For each $x \in X$ let

$$A_x = \cap\{A \in \mathcal{A} : x \in A\} \in \mathcal{A},$$

wherein we have used $\mathcal{A}$ is finite to insure $A_x \in \mathcal{A}$. Hence A_x is the smallest set in $\mathcal{A}$ which contains x.

Now suppose that $y \in X$. If $x \in A_y$ then $A_x \subset A_y$ so that $A_x \cap A_y = A_x$. On the other hand, if $x \notin A_y$ then $x \in A_x \setminus A_y$ and therefore $A_x \subset A_x \setminus A_y$, i.e. $A_x \cap A_y = \emptyset$. Therefore we have shown, either $A_x \cap A_y = \emptyset$ or $A_x \cap A_y = A_x$. By reversing the roles of x and y it also follows that either $A_y \cap A_x = \emptyset$ or $A_y \cap A_x = A_y$. Therefore we may conclude, either $A_x = A_y$ or $A_x \cap A_y = \emptyset$ for all $x, y \in X$.

Let us now define $\{B_i\}_{i=1}^k$ to be an enumeration of $\{A_x\}_{x\in X}$. It is a straightforward to conclude that

$$\mathcal{A} = \{\cup_{i\in\Lambda} B_i : \Lambda \subset \{1,2,\ldots,k\}\}.$$

For example observe that for any $A \in \mathcal{A}$, we have $A = \cup_{x\in A} A_x = \cup_{i\in\Lambda} B_i$ where $\Lambda := \{i : B_i \subset A\}$.

Proposition 3.20. *Suppose that $\mathcal{B} \subset 2^X$ is a σ – algebra and $\mathcal{B}$ is at most a countable set. Then there exists a unique **finite** partition $\mathcal{F}$ of X such that $\mathcal{F} \subset \mathcal{B}$ and every element $B \in \mathcal{B}$ is of the form*

$$B = \cup\{A \in \mathcal{F} : A \subset B\}. \tag{3.1}$$

In particular $\mathcal{B}$ is actually a finite set and $\#(\mathcal{B}) = 2^n$ for some $n \in \mathbb{N}$.

Proof. We proceed as in Example 3.19. For each $x \in X$ let

$$A_x = \cap\{A \in \mathcal{B} : x \in A\} \in \mathcal{B},$$

wherein we have used $\mathcal{B}$ is a countable σ – algebra to insure $A_x \in \mathcal{B}$. Just as above either $A_x \cap A_y = \emptyset$ or $A_x = A_y$ and therefore $\mathcal{F} = \{A_x : x \in X\} \subset \mathcal{B}$ is a (necessarily countable) partition of X for which Eq. (3.1) holds for all $B \in \mathcal{B}$.

Enumerate the elements of $\mathcal{F}$ as $\mathcal{F} = \{P_n\}_{n=1}^N$ where $N \in \mathbb{N}$ or $N = \infty$. If $N = \infty$, then the correspondence

$$a \in \{0,1\}^{\mathbb{N}} \to A_a = \cup\{P_n : a_n = 1\} \in \mathcal{B}$$

is bijective and therefore, by Lemma 3.8, $\mathcal{B}$ is uncountable. Thus any countable σ – algebra is necessarily finite. This finishes the proof modulo the uniqueness assertion which is left as an exercise to the reader. ■

Example 3.21 (Countable/Co-countable σ – Field). Let $X = \mathbb{R}$ and $\mathcal{E} := \{\{x\} : x \in \mathbb{R}\}$. Then $\sigma(\mathcal{E})$ consists of those subsets, $A \subset \mathbb{R}$, such that A is countable or A^c is countable. Similarly, $\mathcal{A}(\mathcal{E})$ consists of those subsets, $A \subset \mathbb{R}$, such that A is finite or A^c is finite. More generally we have the following exercise.

Exercise 3.8. Let X be a set, I be an **infinite** index set, and $\mathcal{E} = \{A_i\}_{i\in I}$ be a partition of X. Prove the algebra, $\mathcal{A}(\mathcal{E})$, and that σ – algebra, $\sigma(\mathcal{E})$, generated by $\mathcal{E}$ are given by

$$\mathcal{A}(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \#(\Lambda) < \infty \text{ or } \#(\Lambda^c) < \infty\}$$

and

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \Lambda \text{ countable or } \Lambda^c \text{ countable}\}$$

respectively. Here we are using the convention that $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. In particular if I is countable, then

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I\}.$$

Proposition 3.22. *Let X be a set and $\mathcal{E} \subset 2^X$. Let $\mathcal{E}^c := \{A^c : A \in \mathcal{E}\}$ and $\mathcal{E}_c := \mathcal{E} \cup \{X, \emptyset\} \cup \mathcal{E}^c$ Then*

$$\mathcal{A}(\mathcal{E}) := \{\text{finite unions of finite intersections of elements from } \mathcal{E}_c\}. \tag{3.2}$$

Proof. Let $\mathcal{A}$ denote the right member of Eq. (3.2). From the definition of an algebra, it is clear that $\mathcal{E} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{E})$. Hence to finish that proof it suffices to show $\mathcal{A}$ is an algebra. The proof of these assertions are routine except for possibly showing that $\mathcal{A}$ is closed under complementation. To check $\mathcal{A}$ is closed under complementation, let $Z \in \mathcal{A}$ be expressed as

$$Z = \bigcup_{i=1}^{N} \bigcap_{j=1}^{K} A_{ij}$$

where $A_{ij} \in \mathcal{E}_c$. Therefore, writing $B_{ij} = A_{ij}^c \in \mathcal{E}_c$, we find that

$$Z^c = \bigcap_{i=1}^{N} \bigcup_{j=1}^{K} B_{ij} = \bigcup_{j_1,\dots,j_N=1}^{K} (B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}) \in \mathcal{A}$$

wherein we have used the fact that $B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}$ is a finite intersection of sets from $\mathcal{E}_c$. ∎

Remark 3.23. One might think that in general $\sigma(\mathcal{E})$ may be described as the countable unions of countable intersections of sets in $\mathcal{E}^c$. However this is in general **false,** since if

$$Z = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{\infty} A_{ij}$$

with $A_{ij} \in \mathcal{E}_c$, then

$$Z^c = \bigcup_{j_1=1, j_2=1, \dots j_N=1, \dots}^{\infty} \left(\bigcap_{\ell=1}^{\infty} A_{\ell, j_\ell}^c \right)$$

which is now an **uncountable** union. Thus the above description is not correct. In general it is complicated to explicitly describe $\sigma(\mathcal{E})$, see Proposition 1.23 on page 39 of Folland for details. Also see Proposition 3.20.

Exercise 3.9. Let τ be a topology on a set X and $\mathcal{A} = \mathcal{A}(\tau)$ be the algebra generated by τ. Show $\mathcal{A}$ is the collection of subsets of X which may be written as finite union of sets of the form $F \cap V$ where F is closed and V is open.

Solution to Exercise (3.9). In this case τ_c is the collection of sets which are either open or closed. Now if $V_i \subset_o X$ and $F_j \sqsubset X$ for each j, then $(\cap_{i=1}^n V_i) \cap \left(\cap_{j=1}^m F_j\right)$ is simply a set of the form $V \cap F$ where $V \subset_o X$ and $F \sqsubset X$. Therefore the result is an immediate consequence of Proposition 3.22.

Definition 3.24. *A set $\mathcal{S} \subset 2^X$ is said to be an **semialgebra or elementary class** provided that*

- $\emptyset \in \mathcal{S}$
- *$\mathcal{S}$ is closed under finite intersections*
- *if $E \in \mathcal{S}$, then E^c is a finite disjoint union of sets from $\mathcal{S}$. (In particular $X = \emptyset^c$ is a finite disjoint union of elements from $\mathcal{S}$.)*

Proposition 3.25. *Suppose $\mathcal{S} \subset 2^X$ is a semi-field, then $\mathcal{A} = \mathcal{A}(\mathcal{S})$ consists of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$.*

Proof. (Although it is possible to give a proof using Proposition 3.22, it is just as simple to give a direct proof.) Let $\mathcal{A}$ denote the collection of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$. Clearly $\mathcal{S} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{S})$ so it suffices to show $\mathcal{A}$ is an algebra since $\mathcal{A}(\mathcal{S})$ is the smallest algebra containing $\mathcal{S}$. By the properties of $\mathcal{S}$, we know that $\emptyset, X \in \mathcal{A}$. The following two steps now finish the proof.

1. ($\mathcal{A}$ is closed under finite intersections.) Suppose that $A_i = \sum_{F \in \Lambda_i} F \in \mathcal{A}$ where, for $i = 1, 2, \dots, n$, Λ_i is a finite collection of disjoint sets from $\mathcal{S}$. Then

$$\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \left(\sum_{F \in \Lambda_i} F \right) = \bigcup_{(F_1, \dots, F_n) \in \Lambda_1 \times \dots \times \Lambda_n} (F_1 \cap F_2 \cap \dots \cap F_n)$$

and this is a disjoint (you check) union of elements from $\mathcal{S}$. Therefore $\mathcal{A}$ is closed under finite intersections.

2. ($\mathcal{A}$ is closed under complementation.) If $A = \sum_{F \in \Lambda} F$ with Λ being a finite collection of disjoint sets from $\mathcal{S}$, then $A^c = \bigcap_{F \in \Lambda} F^c$. Since, by assumption, $F^c \in \mathcal{A}$ for all $F \in \Lambda \subset \mathcal{S}$ and $\mathcal{A}$ is closed under finite intersections by step 1., it follows that $A^c \in \mathcal{A}$. ∎

Example 3.26. Let $X = \mathbb{R}$, then

$$\begin{aligned} \mathcal{S} &:= \left\{(a, b] \cap \mathbb{R} : a, b \in \bar{\mathbb{R}}\right\} \\ &= \{(a, b] : a \in [-\infty, \infty) \text{ and } a < b < \infty\} \cup \{\emptyset, \mathbb{R}\} \end{aligned}$$

is a semi-field. The algebra, $\mathcal{A}(\mathcal{S})$, generated by $\mathcal{S}$ consists of finite disjoint unions of sets from $\mathcal{S}$. For example,

$$A = (0, \pi] \cup (2\pi, 7] \cup (11, \infty) \in \mathcal{A}(\mathcal{S}).$$

Exercise 3.10. Let $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be semi-fields. Show the collection

$$\mathcal{S} := \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$$

is also a semi-field.

Solution to Exercise (3.10). Clearly $\emptyset = \emptyset \times \emptyset \in \mathcal{E} = \mathcal{A} \times \mathcal{B}$. Let $A_i \in \mathcal{A}$ and $B_i \in \mathcal{B}$, then

$$\cap_{i=1}^{n}(A_i \times B_i) = (\cap_{i=1}^{n} A_i) \times (\cap_{i=1}^{n} B_i) \in \mathcal{A} \times \mathcal{B}$$

showing $\mathcal{E}$ is closed under finite intersections. For $A \times B \in \mathcal{E}$,

$$(A \times B)^c = (A^c \times B^c) \sum (A^c \times B) \sum (A \times B^c)$$

and by assumption $A^c = \sum_{i=1}^{n} A_i$ with $A_i \in \mathcal{A}$ and $B^c = \sum_{j=1}^{m} B_i$ with $B_j \in \mathcal{B}$. Therefore

$$A^c \times B^c = \left(\sum_{i=1}^{n} A_i\right) \times \left(\sum_{j=1}^{m} B_i\right) = \sum_{i=1,j=1}^{n,m} A_i \times B_i,$$
$$A^c \times B = \sum_{i=1}^{n} A_i \times B, \text{ and } A \times B^c = \sum_{j=1}^{m} A \times B_i$$

showing $(A \times B)^c$ may be written as finite disjoint union of elements from $\mathcal{S}$.

Finitely Additive Measures / Integration

Definition 4.1. *Suppose that $\mathcal{E} \subset 2^X$ is a collection of subsets of X and $\mu : \mathcal{E} \to [0, \infty]$ is a function. Then*

1. *μ is* ***additive or finitely additive on*** *$\mathcal{E}$ if*
$$\mu(E) = \sum_{i=1}^{n} \mu(E_i) \tag{4.1}$$
whenever $E = \sum_{i=1}^{n} E_i \in \mathcal{E}$ with $E_i \in \mathcal{E}$ for $i = 1, 2, \dots, n < \infty$.
2. *μ is σ –* ***additive (or countable additive) on*** *$\mathcal{E}$ if Eq. (4.1) holds even when $n = \infty$.*
3. *μ is* ***sub-additive (finitely sub-additive) on*** *$\mathcal{E}$ if*
$$\mu(E) \le \sum_{i=1}^{n} \mu(E_i)$$
whenever $E = \bigcup_{i=1}^{n} E_i \in \mathcal{E}$ with $n \in \mathbb{N} \cup \{\infty\}$ ($n \in \mathbb{N}$).
4. *μ is a* ***finitely additive measure*** *if $\mathcal{E} = \mathcal{A}$ is an algebra, $\mu(\emptyset) = 0$, and μ is finitely additive on $\mathcal{A}$.*
5. *μ is a* ***premeasure*** *if μ is a finitely additive measure which is σ – additive on $\mathcal{A}$.*
6. *μ is a* ***measure*** *if μ is a premeasure on a σ – algebra. Furthermore if $\mu(X) = 1$, we say μ is a* ***probability measure*** *on X.*

Proposition 4.2 (Basic properties of finitely additive measures). *Suppose μ is a finitely additive measure on an algebra, $\mathcal{A} \subset 2^X$, $A, B \in \mathcal{A}$ with $A \subset B$and $\{A_j\}_{j=1}^n \subset \mathcal{A}$, then :*

1. *(μ is* ***monotone****) $\mu(A) \le \mu(B)$ if $A \subset B$.*
2. *For $A, B \in \mathcal{A}$, the following* ***strong additivity formula*** *holds;*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{4.2}$$
3. *(μ is* ***finitely subbadditive****) $\mu(\cup_{j=1}^n A_j) \le \sum_{j=1}^n \mu(A_j)$.*
4. *μ is sub-additive on $\mathcal{A}$ iff*
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \text{ for } A = \sum_{i=1}^{\infty} A_i \tag{4.3}$$
where $A \in \mathcal{A}$ and $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ are pairwise disjoint sets.
5. *(μ is* ***countably superadditive****) If $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then*
$$\mu\left(\sum_{i=1}^{\infty} A_i\right) \ge \sum_{i=1}^{\infty} \mu(A_i). \tag{4.4}$$
(See Remark 4.9 for example where this inequality is strict.)
6. *A finitely additive measure, μ, is a premeasure iff μ is subadditive.*

Proof.

1. Since B is the disjoint union of A and $(B \setminus A)$ and $B \setminus A = B \cap A^c \in \mathcal{A}$ it follows that
$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A).$$
2. Since
$$A \cup B = [A \setminus (A \cap B)] \sum [B \setminus (A \cap B)] \sum A \cap B,$$
$$\begin{aligned}\mu(A \cup B) &= \mu(A \cup B \setminus (A \cap B)) + \mu(A \cap B) \\ &= \mu(A \setminus (A \cap B)) + \mu(B \setminus (A \cap B)) + \mu(A \cap B).\end{aligned}$$
Adding $\mu(A \cap B)$ to both sides of this equation proves Eq. (4.2).
3. Let $\widetilde{E}_j = E_j \setminus (E_1 \cup \dots \cup E_{j-1})$ so that the $\tilde{E}_j$'s are pair-wise disjoint and $E = \cup_{j=1}^n \widetilde{E}_j$. Since $\tilde{E}_j \subset E_j$ it follows from the monotonicity of μ that
$$\mu(E) = \sum_{j=1}^{n} \mu(\widetilde{E}_j) \le \sum_{j=1}^{n} \mu(E_j).$$
4. If $A = \bigcup_{i=1}^{\infty} B_i$ with $A \in \mathcal{A}$ and $B_i \in \mathcal{A}$, then $A = \sum_{i=1}^{\infty} A_i$ where $A_i := B_i \setminus (B_1 \cup \dots B_{i-1}) \in \mathcal{A}$ and $B_0 = \emptyset$. Therefore using the monotonicity of μ and Eq. (4.3)
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \le \sum_{i=1}^{\infty} \mu(B_i).$$
5. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then $\sum_{i=1}^{n} A_i \subset A$ for all n and so by the monotonicity and finite additivity of μ, $\sum_{i=1}^{n} \mu(A_i) \le \mu(A)$. Letting $n \to \infty$ in this equation shows μ is superadditive.
6. This is a combination of items 5. and 6.

■

4.1 Examples of Measures

Most σ − algebras and σ -additive measures are somewhat difficult to describe and define. However, there are a few special cases where we can describe explicitly what is going on.

Example 4.3. Suppose that Ω is a finite set, $\mathcal{B} := 2^{\Omega}$, and $p : \Omega \to [0,1]$ is a function such that

$$\sum_{\omega\in\Omega} p(\omega) = 1.$$

Then

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \subset \Omega$$

defines a measure on 2^{Ω}.

Example 4.4. Suppose that X is any set and $x \in X$ is a point. For $A \subset X$, let

$$\delta_x(A) = \begin{cases} 1 \text{ if } & x \in A \\ 0 \text{ if } & x \notin A. \end{cases}$$

Then $\mu = \delta_x$ is a measure on X called the Dirac delta measure at x.

Example 4.5. Suppose $\mathcal{B} \subset 2^X$ is a σ algebra, μ is a measure on $\mathcal{B}$, and $\lambda > 0$, then $\lambda \cdot \mu$ is also a measure on $\mathcal{B}$. Moreover, if J is an index set and $\{\mu_j\}_{j\in J}$ are all measures on $\mathcal{B}$, then $\mu = \sum_{j=1}^{\infty} \mu_j$, i.e.

$$\mu(A) := \sum_{j=1}^{\infty} \mu_j(A) \text{ for all } A \in \mathcal{B},$$

defines another measure on $\mathcal{B}$. To prove this we must show that μ is countably additive. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i \in \mathcal{B}$, then (using Tonelli for sums, Proposition 1.7),

$$\begin{aligned} \mu(A) &= \sum_{j=1}^{\infty} \mu_j(A) = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \mu_j(A_i) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu_j(A_i) = \sum_{i=1}^{\infty} \mu(A_i). \end{aligned}$$

Example 4.6. Suppose that X is a countable set and $\lambda : X \to [0,\infty]$ is a function. Let $X = \{x_n\}_{n=1}^{\infty}$ be an enumeration of X and then we may define a measure μ on 2^X by,

$$\mu = \mu_\lambda := \sum_{n=1}^{\infty} \lambda(x_n)\delta_{x_n}.$$

We will now show this measure is independent of our choice of enumeration of X by showing,

$$\mu(A) = \sum_{x\in A} \lambda(x) := \sup_{\Lambda\subset\subset A} \sum_{x\in\Lambda} \lambda(x) \ \forall\ A \subset X. \tag{4.5}$$

Here we are using the notation, $\Lambda \subset\subset A$ to indicate that Λ is a finite subset of A.

To verify Eq. (4.5), let $M := \sup_{\Lambda\subset\subset A} \sum_{x\in\Lambda} \lambda(x)$ and for each $N \in \mathbb{N}$ let

$$\Lambda_N := \{x_n : x_n \in A \text{ and } 1 \le n \le N\}.$$

Then by definition of μ,

$$\begin{aligned} \mu(A) &= \sum_{n=1}^{\infty} \lambda(x_n)\delta_{x_n}(A) = \lim_{N\to\infty} \sum_{n=1}^{N} \lambda(x_n) 1_{x_n\in A} \\ &= \lim_{N\to\infty} \sum_{x\in\Lambda_N} \lambda(x) \le M. \end{aligned}$$

On the other hand if $\Lambda \subset\subset A$, then

$$\sum_{x\in\Lambda} \lambda(x) = \sum_{n:\ x_n\in\Lambda} \lambda(x_n) = \mu(\Lambda) \le \mu(A)$$

from which it follows that $M \le \mu(A)$. This shows that μ is independent of how we enumerate X.

The above example has a natural extension to the case where X is uncountable and $\lambda : X \to [0,\infty]$ is any function. In this setting we simply may define $\mu : 2^X \to [0,\infty]$ using Eq. (4.5). We leave it to the reader to verify that this is indeed a measure on 2^X.

We will construct many more measure in Chapter 5 below. The starting point of these constructions will be the construction of finitely additive measures using the next proposition.

Proposition 4.7 (Construction of Finitely Additive Measures). *Suppose $\mathcal{S} \subset 2^X$ is a semi-algebra (see Definition 3.24) and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ is the algebra generated by $\mathcal{S}$. Then every additive function $\mu : \mathcal{S} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ extends uniquely to an additive measure (which we still denote by μ) on $\mathcal{A}$.*

Proof. Since (by Proposition 3.25) every element $A \in \mathcal{A}$ is of the form $A = \sum_i E_i$ for a finite collection of $E_i \in \mathcal{S}$, it is clear that if μ extends to a measure then the extension is unique and must be given by

$$\mu(A) = \sum_i \mu(E_i). \tag{4.6}$$

To prove existence, the main point is to show that $\mu(A)$ in Eq. (4.6) is well defined; i.e. if we also have $A = \sum_j F_j$ with $F_j \in \mathcal{S}$, then we must show

$$\sum_i \mu(E_i) = \sum_j \mu(F_j). \tag{4.7}$$

But $E_i = \sum_j (E_i \cap F_j)$ and the additivity of μ on $\mathcal{S}$ implies $\mu(E_i) = \sum_j \mu(E_i \cap F_j)$ and hence

$$\sum_i \mu(E_i) = \sum_i \sum_j \mu(E_i \cap F_j) = \sum_{i,j} \mu(E_i \cap F_j).$$

Similarly,

$$\sum_j \mu(F_j) = \sum_{i,j} \mu(E_i \cap F_j)$$

which combined with the previous equation shows that Eq. (4.7) holds. It is now easy to verify that μ extended to $\mathcal{A}$ as in Eq. (4.6) is an additive measure on $\mathcal{A}$. ■

Proposition 4.8. *Let $X = \mathbb{R}$, $\mathcal{S}$ be the semi-algebra,*

$$\mathcal{S} = \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}, \tag{4.8}$$

and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ be the algebra formed by taking finite disjoint unions of elements from $\mathcal{S}$, see Proposition 3.25. To each finitely additive probability measures $\mu : \mathcal{A} \to [0, \infty]$, there is a unique increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ and

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) \ \forall \ a \le b \text{ in } \bar{\mathbb{R}}. \tag{4.9}$$

Conversely, given an increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ there is a unique finitely additive measure $\mu = \mu_F$ on $\mathcal{A}$ such that the relation in Eq. (4.9) holds. (Eventually we will only be interested in the case where $F(-\infty) = \lim_{a \downarrow -\infty} F(a)$ and $F(\infty) = \lim_{b \uparrow \infty} F(b)$.)

Proof. Given a finitely additive probability measure μ, let

$$F(x) := \mu((-\infty, x] \cap \mathbb{R}) \text{ for all } x \in \bar{\mathbb{R}}.$$

Then $F(\infty) = 1$, $F(-\infty) = 0$ and for $b > a$,

$$F(b) - F(a) = \mu((-\infty, b] \cap \mathbb{R}) - \mu((-\infty, a]) = \mu((a, b] \cap \mathbb{R}).$$

Conversely, suppose $F : \bar{\mathbb{R}} \to [0, 1]$ as in the statement of the theorem is given. Define μ on $\mathcal{S}$ using the formula in Eq. (4.9). The argument will be completed by showing μ is additive on $\mathcal{S}$ and hence, by Proposition 4.7, has a unique extension to a finitely additive measure on $\mathcal{A}$. Suppose that

$$(a, b] = \sum_{i=1}^n (a_i, b_i].$$

By reordering $(a_i, b_i]$ if necessary, we may assume that

$$a = a_1 < b_1 = a_2 < b_2 = a_3 < \cdots < b_{n-1} = a_n < b_n = b.$$

Therefore, by the telescoping series argument,

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) = \sum_{i=1}^n [F(b_i) - F(a_i)] = \sum_{i=1}^n \mu((a_i, b_i] \cap \mathbb{R}).$$

■

Remark 4.9. Suppose that $F : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ is any non-decreasing function such that $F(\mathbb{R}) \subset \mathbb{R}$. Then the same methods used in the proof of Proposition 4.8 shows that there exists a unique finitely additive measure, $\mu = \mu_F$, on $\mathcal{A} = \mathcal{A}(\mathcal{S})$ such that Eq. (4.9) holds. If $F(\infty) > \lim_{b \uparrow \infty} F(b)$ and $A_i = (i, i+1]$ for $i \in \mathbb{N}$, then

$$\begin{aligned} \sum_{i=1}^\infty \mu_F(A_i) &= \sum_{i=1}^\infty (F(i+1) - F(i)) = \lim_{N \to \infty} \sum_{i=1}^N (F(i+1) - F(i)) \\ &= \lim_{N \to \infty} (F(N+1) - F(1)) < F(\infty) - F(1) = \mu_F(\cup_{i=1}^\infty A_i). \end{aligned}$$

This shows that strict inequality can hold in Eq. (4.4) and that μ_F is **not** a premeasure. Similarly one shows μ_F is **not** a premeasure if $F(-\infty) < \lim_{a \downarrow -\infty} F(a)$ or if F is **not** right continuous at some point $a \in \mathbb{R}$. Indeed, in the latter case consider

$$(a, a+1] = \sum_{n=1}^\infty (a + \frac{1}{n+1}, a + \frac{1}{n}].$$

Working as above we find,

$$\sum_{n=1}^\infty \mu_F\left((a + \frac{1}{n+1}, a + \frac{1}{n}]\right) = F(a+1) - F(a+)$$

while $\mu_F((a, a+1]) = F(a+1) - F(a)$. We will eventually show in Chapter 5 below that μ_F extends uniquely to a $\sigma-$additive measure on $\mathcal{B}_{\mathbb{R}}$ whenever F is increasing, right continuous, and $F(\pm\infty) = \lim_{x \to \pm\infty} F(x)$.

Before constructing σ – additive measures (see Chapter 5 below), we are going to pause to discuss a preliminary notion of integration and develop some of its properties. Hopefully this will help the reader to develop the necessary intuition before heading to the general theory. First we need to describe the functions we are allowed to integrate.

4.2 Simple Random Variables

Definition 4.10 (Simple random variables). *A function, $f : \Omega \to Y$ is said to be **simple** if $f(\Omega) \subset Y$ is a finite set. If $\mathcal{A} \subset 2^{\Omega}$ is an algebra, we say that a simple function $f : \Omega \to Y$ is **measurable** if $\{f = y\} := f^{-1}(\{y\}) \in \mathcal{A}$ for all $y \in Y$. A measurable simple function, $f : \Omega \to \mathbb{C}$, is called a **simple random variable** relative to $\mathcal{A}$.*

Notation 4.11 *Given an algebra, $\mathcal{A} \subset 2^{\Omega}$, let $\mathbb{S}(\mathcal{A})$ denote the collection of simple random variables from Ω to $\mathbb{C}$. For example if $A \in \mathcal{A}$, then $1_A \in \mathbb{S}(\mathcal{A})$ is a measurable simple function.*

Lemma 4.12. *Let $\mathcal{A} \subset 2^{\Omega}$ be an algebra, then;*

1. *$\mathbb{S}(\mathcal{A})$ is a sub-algebra of all functions from Ω to $\mathbb{C}$.*
2. *$f : \Omega \to \mathbb{C}$, is a $\mathcal{A}$ – simple random variable iff there exists $\alpha_i \in \mathbb{C}$ and $A_i \in \mathcal{A}$ for $1 \le i \le n$ for some $n \in \mathbb{N}$ such that*
$$f = \sum_{i=1}^{n} \alpha_i 1_{A_i}. \tag{4.10}$$
3. *For any function, $F : \mathbb{C} \to \mathbb{C}$, $F \circ f \in \mathbb{S}(\mathcal{A})$ for all $f \in \mathbb{S}(\mathcal{A})$. In particular, $|f| \in \mathbb{S}(\mathcal{A})$ if $f \in \mathbb{S}(\mathcal{A})$.*

Proof. 1. Let us observe that $1_\Omega = 1$ and $1_\emptyset = 0$ are in $\mathbb{S}(\mathcal{A})$. If $f, g \in \mathbb{S}(\mathcal{A})$ and $c \in \mathbb{C} \setminus \{0\}$, then

$$\{f + cg = \lambda\} = \bigcup_{a,b \in \mathbb{C}: a + cb = \lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{4.11}$$

and

$$\{f \cdot g = \lambda\} = \bigcup_{a,b \in \mathbb{C}: a \cdot b = \lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{4.12}$$

from which it follows that $f + cg$ and $f \cdot g$ are back in $\mathbb{S}(\mathcal{A})$.

2. Since $\mathbb{S}(\mathcal{A})$ is an algebra, every f of the form in Eq. (4.10) is in $\mathbb{S}(\mathcal{A})$. Conversely if $f \in \mathbb{S}(\mathcal{A})$ it follows by definition that $f = \sum_{\alpha \in f(\Omega)} \alpha 1_{\{f = \alpha\}}$ which is of the form in Eq. (4.10).

3. If $F : \mathbb{C} \to \mathbb{C}$, then

$$F \circ f = \sum_{\alpha \in f(\Omega)} F(\alpha) \cdot 1_{\{f = \alpha\}} \in \mathbb{S}(\mathcal{A}).$$

■

Exercise 4.1 ($\mathcal{A}$ – measurable simple functions). As in Example 3.19, let $\mathcal{A} \subset 2^X$ be a finite algebra and $\{B_1, \ldots, B_k\}$ be the partition of X associated to $\mathcal{A}$. Show that a function, $f : X \to \mathbb{C}$, is an $\mathcal{A}$ – simple function iff f is constant on B_i for each i. Thus any $\mathcal{A}$ – simple function is of the form,

$$f = \sum_{i=1}^{k} \alpha_i 1_{B_i} \tag{4.13}$$

for some $\alpha_i \in \mathbb{C}$.

Corollary 4.13. *Suppose that Λ is a finite set and $Z : X \to \Lambda$ is a function. Let*
$$\mathcal{A} := \mathcal{A}(Z) := Z^{-1}\left(2^{\Lambda}\right) := \left\{Z^{-1}(E) : E \subset \Lambda\right\}.$$
Then $\mathcal{A}$ is an algebra and $f : X \to \mathbb{C}$ is an $\mathcal{A}$ – simple function iff $f = F \circ Z$ for some function $F : \Lambda \to \mathbb{C}$.

Proof. For $\lambda \in \Lambda$, let

$$A_\lambda := \{Z = \lambda\} = \{x \in X : Z(x) = \lambda\}.$$

The $\{A_\lambda\}_{\lambda \in \Lambda}$ is the partition of X determined by $\mathcal{A}$. Therefore f is an $\mathcal{A}$ – simple function iff $f|_{A_\lambda}$ is constant for each $\lambda \in \Lambda$. Let us denote this constant value by $F(\lambda)$. As $Z = \lambda$ on A_λ, $F : \Lambda \to \mathbb{C}$ is a function such that $f = F \circ Z$.

Conversely if $F : \Lambda \to \mathbb{C}$ is a function and $f = F \circ Z$, then $f = F(\lambda)$ on A_λ, i.e. f is an $\mathcal{A}$ – simple function. ■

4.2.1 The algebraic structure of simple functions*

Definition 4.14. *A **simple function algebra**, $\mathbb{S}$, is a subalgebra[1] of the bounded complex functions on X such that $1 \in \mathbb{S}$ and each function in $\mathbb{S}$ is a simple function. If $\mathbb{S}$ is a simple function algebra, let*

$$\mathcal{A}(\mathbb{S}) := \{A \subset X : 1_A \in \mathbb{S}\}.$$

(It is easily checked that $\mathcal{A}(\mathbb{S})$ is a sub-algebra of 2^X.)

[1] To be more explicit we are assuming that $\mathbb{S}$ is a linear subspace of bounded functions which is closed under pointwise multiplication.

Lemma 4.15. *Suppose that $\mathbb{S}$ is a simple function algebra, $f \in \mathbb{S}$ and $\alpha \in f(X)$ – the range of f. Then $\{f = \alpha\} \in \mathcal{A}(\mathbb{S})$.*

Proof. Let $\{\lambda_i\}_{i=0}^n$ be an enumeration of $f(X)$ with $\lambda_0 = \alpha$. Then

$$g := \left[\prod_{i=1}^n (\alpha - \lambda_i)\right]^{-1} \prod_{i=1}^n (f - \lambda_i 1) \in \mathbb{S}.$$

Moreover, we see that $g = 0$ on $\cup_{i=1}^n \{f = \lambda_i\}$ while $g = 1$ on $\{f = \alpha\}$. So we have shown $g = 1_{\{f=\alpha\}} \in \mathbb{S}$ and therefore that $\{f = \alpha\} \in \mathcal{A}(\mathbb{S})$. ■

Exercise 4.2. Continuing the notation introduced above:

1. Show $\mathcal{A}(\mathbb{S})$ is an algebra of sets.
2. Show $\mathbb{S}(\mathcal{A})$ is a simple function algebra.
3. Show that the map

$$\mathcal{A} \in \left\{\text{Algebras} \subset 2^X\right\} \to \mathbb{S}(\mathcal{A}) \in \{\textbf{simple function algebras on } X\}$$

is bijective and the map, $\mathbb{S} \to \mathcal{A}(\mathbb{S})$, is the inverse map.

Solution to Exercise (4.2).

1. Since $0 = 1_\emptyset, 1 = 1_X \in \mathbb{S}$, it follows that $\emptyset$ and X are in $\mathcal{A}(\mathbb{S})$. If $A \in \mathcal{A}(\mathbb{S})$, then $1_{A^c} = 1 - 1_A \in \mathbb{S}$ and so $A^c \in \mathcal{A}(\mathbb{S})$. Finally, if $A, B \in \mathcal{A}(\mathbb{S})$ then $1_{A\cap B} = 1_A \cdot 1_B \in \mathbb{S}$ and thus $A \cap B \in \mathcal{A}(\mathbb{S})$.
2. If $f, g \in \mathbb{S}(\mathcal{A})$ and $c \in \mathbb{F}$, then

$$\{f + cg = \lambda\} = \bigcup_{a,b\in\mathbb{F}:a+cb=\lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A}$$

and

$$\{f \cdot g = \lambda\} = \bigcup_{a,b\in\mathbb{F}:a\cdot b=\lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A}$$

from which it follows that $f + cg$ and $f \cdot g$ are back in $\mathbb{S}(\mathcal{A})$.
3. If $f : \Omega \to \mathbb{C}$ is a simple function such that $1_{\{f=\lambda\}} \in \mathbb{S}$ for all $\lambda \in \mathbb{C}$, then $f = \sum_{\lambda\in\mathbb{C}} \lambda 1_{\{f=\lambda\}} \in \mathbb{S}$. Conversely, by Lemma 4.15, if $f \in \mathbb{S}$ then $1_{\{f=\lambda\}} \in \mathbb{S}$ for all $\lambda \in \mathbb{C}$. Therefore, a simple function, $f : X \to \mathbb{C}$ is in $\mathbb{S}$ iff $1_{\{f=\lambda\}} \in \mathbb{S}$ for all $\lambda \in \mathbb{C}$. With this preparation, we are now ready to complete the verification.
First off,

$$A \in \mathcal{A}(\mathbb{S}(\mathcal{A})) \iff 1_A \in \mathbb{S}(\mathcal{A}) \iff A \in \mathcal{A}$$

which shows that $\mathcal{A}(\mathbb{S}(\mathcal{A})) = \mathcal{A}$. Similarly,

$$\begin{aligned} f \in \mathbb{S}(\mathcal{A}(\mathbb{S})) &\iff \{f = \lambda\} \in \mathcal{A}(\mathbb{S}) \ \forall \lambda \in \mathbb{C} \\ &\iff 1_{\{f=\lambda\}} \in \mathbb{S} \ \forall \lambda \in \mathbb{C} \\ &\iff f \in \mathbb{S} \end{aligned}$$

which shows $\mathbb{S}(\mathcal{A}(\mathbb{S})) = \mathbb{S}$.

4.3 Simple Integration

Definition 4.16 (Simple Integral). *Suppose now that P is a finitely additive probability measure on an algebra $\mathcal{A} \subset 2^X$. For $f \in \mathbb{S}(\mathcal{A})$ the* ***integral or expectation,*** *$\mathbb{E}(f) = \mathbb{E}_P(f)$, is defined by*

$$\mathbb{E}_P(f) = \int_X f dP = \sum_{y\in\mathbb{C}} yP(f = y). \tag{4.14}$$

Example 4.17. Suppose that $A \in \mathcal{A}$, then

$$\mathbb{E}1_A = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A). \tag{4.15}$$

Remark 4.18. Let us recall that our intuitive notion of $P(A)$ was given as in Eq. (2.1) by

$$P(A) = \lim_{N\to\infty} \frac{1}{N} \sum 1_A(\omega(k))$$

where $\omega(k) \in \Omega$ was the result of the k^{th} "independent" experiment. If we use this interpretation back in Eq. (4.14) we arrive at,

$$\begin{aligned} \mathbb{E}(f) = \sum_{y\in\mathbb{C}} yP(f = y) &= \sum_{y\in\mathbb{C}} y \cdot \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{y\in\mathbb{C}} y \sum_{k=1}^N 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \sum_{y\in\mathbb{C}} f(\omega(k)) \cdot 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N f(\omega(k)). \end{aligned}$$

Thus informally, $\mathbb{E}f$ should represent the limiting average of the values of f over many "independent" experiments. We will come back to this later when we study the strong law of large numbers.

Proposition 4.19. *The expectation operator, $\mathbb{E} = \mathbb{E}_P : \mathbb{S}(\mathcal{A}) \to \mathbb{C}$, satisfies:*

1. *If $f \in \mathbb{S}(\mathcal{A})$ and $\lambda \in \mathbb{C}$, then*
$$\mathbb{E}(\lambda f) = \lambda \mathbb{E}(f). \tag{4.16}$$
2. *If $f, g \in \mathbb{S}(\mathcal{A})$, then*
$$\mathbb{E}(f+g) = \mathbb{E}(g) + \mathbb{E}(f). \tag{4.17}$$
Items 1. and 2. say that $\mathbb{E}(\cdot)$ is a linear functional on $\mathbb{S}(\mathcal{A})$.
3. *If $f = \sum_{j=1}^N \lambda_j 1_{A_j}$ for some $\lambda_j \in \mathbb{C}$ and some $A_j \in \mathbb{C}$, then*
$$\mathbb{E}(f) = \sum_{j=1}^N \lambda_j P(A_j). \tag{4.18}$$
4. *$\mathbb{E}$ is **positive**, i.e. $\mathbb{E}(f) \geq 0$ for all $0 \leq f \in \mathbb{S}(\mathcal{A})$. More generally, if $f, g \in \mathbb{S}(\mathcal{A})$ and $f \leq g$, then $\mathbb{E}(f) \leq \mathbb{E}(g)$.*
5. *For all $f \in \mathbb{S}(\mathcal{A})$,*
$$|\mathbb{E}f| \leq \mathbb{E}|f|. \tag{4.19}$$

Proof.

1. If $\lambda \neq 0$, then
$$\begin{aligned}\mathbb{E}(\lambda f) &= \sum_{y \in \mathbb{C}} y\ P(\lambda f = y) = \sum_{y \in \mathbb{C}} y\ P(f = y/\lambda) \\ &= \sum_{z \in \mathbb{C}} \lambda z\ P(f = z) = \lambda \mathbb{E}(f).\end{aligned}$$
The case $\lambda = 0$ is trivial.
2. Writing $\{f = a, g = b\}$ for $f^{-1}(\{a\}) \cap g^{-1}(\{b\})$, then
$$\begin{aligned}\mathbb{E}(f+g) &= \sum_{z \in \mathbb{C}} z\ P(f+g = z) \\ &= \sum_{z \in \mathbb{C}} z\ P\left(\sum_{a+b=z} \{f = a,\ g = b\}\right) \\ &= \sum_{z \in \mathbb{C}} z \sum_{a+b=z} P(\{f = a,\ g = b\}) \\ &= \sum_{z \in \mathbb{C}} \sum_{a+b=z} (a+b) P(\{f = a,\ g = b\}) \\ &= \sum_{a,b} (a+b) P(\{f = a,\ g = b\}).\end{aligned}$$
But
$$\begin{aligned}\sum_{a,b} aP(\{f = a,\ g = b\}) &= \sum_a a \sum_b P(\{f = a,\ g = b\}) \\ &= \sum_a aP(\cup_b \{f = a,\ g = b\}) \\ &= \sum_a aP(\{f = a\}) = \mathbb{E}f\end{aligned}$$
and similarly,
$$\sum_{a,b} bP(\{f = a,\ g = b\}) = \mathbb{E}g.$$
Equation (4.17) is now a consequence of the last three displayed equations.
3. If $f = \sum_{j=1}^N \lambda_j 1_{A_j}$, then
$$\mathbb{E}f = \mathbb{E}\left[\sum_{j=1}^N \lambda_j 1_{A_j}\right] = \sum_{j=1}^N \lambda_j \mathbb{E}1_{A_j} = \sum_{j=1}^N \lambda_j P(A_j).$$
4. If $f \geq 0$ then
$$\mathbb{E}(f) = \sum_{a \geq 0} aP(f = a) \geq 0$$
and if $f \leq g$, then $g - f \geq 0$ so that
$$\mathbb{E}(g) - \mathbb{E}(f) = \mathbb{E}(g - f) \geq 0.$$
5. By the triangle inequality,
$$|\mathbb{E}f| = \left|\sum_{\lambda \in \mathbb{C}} \lambda P(f = \lambda)\right| \leq \sum_{\lambda \in \mathbb{C}} |\lambda| P(f = \lambda) = \mathbb{E}|f|,$$
wherein the last equality we have used Eq. (4.18) and the fact that $|f| = \sum_{\lambda \in \mathbb{C}} |\lambda| 1_{f=\lambda}$.

■

Remark 4.20. If Ω is a finite set and $\mathcal{A} = 2^\Omega$, then
$$f(\cdot) = \sum_{\omega \in \Omega} f(\omega) 1_{\{\omega\}}$$
and hence
$$\mathbb{E}_P f = \sum_{\omega \in \Omega} f(\omega) P(\{\omega\}).$$

Remark 4.21. All of the results in Proposition 4.19 and Remark 4.20 remain valid when P is replaced by a finite measure, $\mu : \mathcal{A} \to [0,\infty)$, i.e. it is enough to assume $\mu(X) < \infty$.

Exercise 4.3. Let P is a finitely additive probability measure on an algebra $\mathcal{A} \subset 2^X$ and for $A, B \in \mathcal{A}$ let $\rho(A,B) := P(A \Delta B)$ where $A\Delta B = (A \setminus B) \cup (B \setminus A)$. Show;

1. $\rho(A,B) = \mathbb{E}|1_A - 1_B|$ and then use this (or not) to show
2. $\rho(A,C) \le \rho(A,B) + \rho(B,C)$ for all $A,B,C \in \mathcal{A}$.

Remark: it is now easy to see that $\rho : \mathcal{A} \times \mathcal{A} \to [0,1]$ satisfies the axioms of a metric except for the condition that $\rho(A,B) = 0$ does not imply that $A = B$ but only that $A = B$ modulo a set of probability zero.

Remark 4.22 (Chebyshev's Inequality). Suppose that $f \in \mathbb{S}(\mathcal{A})$, $\varepsilon > 0$, and $p > 0$, then

$$1_{|f|\ge\varepsilon} \le \frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon} \le \varepsilon^{-p} |f|^p$$

and therefore, see item 4. of Proposition 4.19,

$$P(\{|f| \ge \varepsilon\}) = \mathbb{E}\left[1_{|f|\ge\varepsilon}\right] \le \mathbb{E}\left[\frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon}\right] \le \varepsilon^{-p}\mathbb{E}|f|^p. \tag{4.20}$$

Observe that

$$|f|^p = \sum_{\lambda\in\mathbb{C}} |\lambda|^p 1_{\{f=\lambda\}}$$

is a simple random variable and $\{|f| \ge \varepsilon\} = \sum_{|\lambda|\ge\varepsilon} \{f = \lambda\} \in \mathcal{A}$ as well. Therefore, $\frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon}$ is still a simple random variable.

Lemma 4.23 (Inclusion Exclusion Formula). *If $A_n \in \mathcal{A}$ for $n = 1, 2, \dots, M$ such that $\mu\left(\cup_{n=1}^M A_n\right) < \infty$, then*

$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{k=1}^M (-1)^{k+1} \sum_{1\le n_1<n_2<\cdots<n_k\le M} \mu(A_{n_1} \cap \cdots \cap A_{n_k}). \tag{4.21}$$

Proof. This may be proved inductively from Eq. (4.2). We will give a different and perhaps more illuminating proof here. Let $A := \cup_{n=1}^M A_n$.

Since $A^c = \left(\cup_{n=1}^M A_n\right)^c = \cap_{n=1}^M A_n^c$, we have

$$\begin{aligned}
1 - 1_A = 1_{A^c} &= \prod_{n=1}^M 1_{A_n^c} = \prod_{n=1}^M (1 - 1_{A_n}) \\
&= 1 + \sum_{k=1}^M (-1)^k \sum_{1\le n_1<n_2<\cdots<n_k\le M} 1_{A_{n_1}} \cdots 1_{A_{n_k}} \\
&= 1 + \sum_{k=1}^M (-1)^k \sum_{1\le n_1<n_2<\cdots<n_k\le M} 1_{A_{n_1}\cap\cdots\cap A_{n_k}}
\end{aligned}$$

from which it follows that

$$1_{\cup_{n=1}^M A_n} = 1_A = \sum_{k=1}^M (-1)^{k+1} \sum_{1\le n_1<n_2<\cdots<n_k\le M} 1_{A_{n_1}\cap\cdots\cap A_{n_k}}. \tag{4.22}$$

Integrating this identity with respect to μ gives Eq. (4.21). ■

Remark 4.24. The following identity holds even when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$,

$$\begin{aligned}
\mu\left(\cup_{n=1}^M A_n\right) &+ \sum_{k=2 \;\&\; k \text{ even}}^M \sum_{1\le n_1<n_2<\cdots<n_k\le M} \mu(A_{n_1} \cap \cdots \cap A_{n_k}) \\
&= \sum_{k=1 \;\&\; k \text{ odd}}^M \sum_{1\le n_1<n_2<\cdots<n_k\le M} \mu(A_{n_1} \cap \cdots \cap A_{n_k}).
\end{aligned} \tag{4.23}$$

This can be proved by moving every term with a negative sign on the right side of Eq. (4.22) to the left side and then integrate the resulting identity. Alternatively, Eq. (4.23) follows directly from Eq. (4.21) if $\mu\left(\cup_{n=1}^M A_n\right) < \infty$ and when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$ one easily verifies that both sides of Eq. (4.23) are infinite.

To better understand Eq. (4.22), consider the case $M = 3$ where,

$$\begin{aligned}
1 - 1_A &= (1 - 1_{A_1})(1 - 1_{A_2})(1 - 1_{A_3}) \\
&= 1 - (1_{A_1} + 1_{A_2} + 1_{A_3}) \\
&\quad + 1_{A_1}1_{A_2} + 1_{A_1}1_{A_3} + 1_{A_2}1_{A_3} - 1_{A_1}1_{A_2}1_{A_3}
\end{aligned}$$

so that

$$1_{A_1\cup A_2\cup A_3} = 1_{A_1} + 1_{A_2} + 1_{A_3} - (1_{A_1\cap A_2} + 1_{A_1\cap A_3} + 1_{A_2\cap A_3}) + 1_{A_1\cap A_2\cap A_3}$$

Here is an alternate proof of Eq. (4.22). Let $\omega \in \Omega$ and by relabeling the sets $\{A_n\}$ if necessary, we may assume that $\omega \in A_1 \cap \cdots \cap A_m$ and $\omega \notin A_{m+1} \cup \cdots \cup A_M$ for some $0 \le m \le M$. (When $m = 0$, both sides of Eq. (4.22) are zero and so we will only consider the case where $1 \le m \le M$.) With this notation we have

$$\begin{aligned}
&\sum_{k=1}^{M} (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \dots < n_k \le M} 1_{A_{n_1} \cap \dots \cap A_{n_k}} (\omega) \\
&= \sum_{k=1}^{m} (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \dots < n_k \le m} 1_{A_{n_1} \cap \dots \cap A_{n_k}} (\omega) \\
&= \sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k} \\
&= 1 - \sum_{k=0}^{m} (-1)^{k} (1)^{n-k} \binom{m}{k} \\
&= 1 - (1-1)^m = 1.
\end{aligned}$$

This verifies Eq. (4.22) since $1_{\cup_{n=1}^{M} A_n} (\omega) = 1$.

Example 4.25 (Coincidences). Let Ω be the set of permutations (think of card shuffling), $\omega : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$, and define $P(A) := \frac{\#(A)}{n!}$ to be the uniform distribution (Haar measure) on Ω. We wish to compute the probability of the event, B, that a random permutation fixes some index i. To do this, let $A_i := \{\omega \in \Omega : \omega(i) = i\}$ and observe that $B = \cup_{i=1}^{n} A_i$. So by the Inclusion Exclusion Formula, we have

$$P(B) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < i_3 < \dots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k}).$$

Since

$$\begin{aligned}
P(A_{i_1} \cap \dots \cap A_{i_k}) &= P(\{\omega \in \Omega : \omega(i_1) = i_1, \dots, \omega(i_k) = i_k\}) \\
&= \frac{(n-k)!}{n!}
\end{aligned}$$

and

$$\#\{1 \le i_1 < i_2 < i_3 < \dots < i_k \le n\} = \binom{n}{k},$$

we find

$$P(B) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}. \tag{4.24}$$

For large n this gives,

$$P(B) = -\sum_{k=1}^{n} \frac{1}{k!} (-1)^k \cong 1 - \sum_{k=0}^{\infty} \frac{1}{k!} (-1)^k = 1 - e^{-1} \cong 0.632.$$

Example 4.26 (Expected number of coincidences). Continue the notation in Example 4.25. We now wish to compute the expected number of fixed points of a random permutation, ω, i.e. how many cards in the shuffled stack have not moved on average. To this end, let

$$X_i = 1_{A_i}$$

and observe that

$$N(\omega) = \sum_{i=1}^{n} X_i(\omega) = \sum_{i=1}^{n} 1_{\omega(i)=i} = \#\{i : \omega(i) = i\}.$$

denote the number of fixed points of ω. Hence we have

$$\mathbb{E}N = \sum_{i=1}^{n} \mathbb{E}X_i = \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} \frac{(n-1)!}{n!} = 1.$$

Let us check the above formulas when $n = 3$. In this case we have

ω	$N(\omega)$
1 2 3	3
1 3 2	1
2 1 3	1
2 3 1	0
3 1 2	0
3 2 1	1

and so

$$P(\exists \text{ a fixed point}) = \frac{4}{6} = \frac{2}{3} \cong 0.67 \cong 0.632$$

while

$$\sum_{k=1}^{3} (-1)^{k+1} \frac{1}{k!} = 1 - \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

and

$$\mathbb{E}N = \frac{1}{6}(3 + 1 + 1 + 0 + 0 + 1) = 1.$$

The next three problems generalize the results above. The following notation will be used throughout these exercises.

1. $(\Omega, \mathcal{A}, P)$ is a finitely additive probability space, so $P(\Omega) = 1$,
2. $A_i \in \mathcal{A}$ for $i = 1, 2, \dots, n$,
3. $N(\omega) := \sum_{i=1}^{n} 1_{A_i}(\omega) = \#\{i : \omega \in A_i\}$, and

4. $\{S_k\}_{k=1}^n$ are given by

$$S_k := \sum_{1\le i_1<\dots<i_k\le n} P\left(A_{i_1}\cap\dots\cap A_{i_k}\right) = \sum_{\Lambda\subset\{1,2,\dots,n\}\ni|\Lambda|=k} P\left(\cap_{i\in\Lambda}A_i\right).$$

Exercise 4.4. For $1\le k\le n$, show;

1. (as functions on Ω) that
$$\binom{N}{k} = \sum_{\Lambda\subset\{1,2,\dots,n\}\ni|\Lambda|=k} 1_{\cap_{i\in\Lambda}A_i}, \tag{4.25}$$
where by definition
$$\binom{m}{k} = \begin{cases} 0 & \text{if } \quad k>m \\ \frac{m!}{k!\cdot(m-k)!} & \text{if } 1\le k\le m \\ 1 & \text{if } \quad k=0\end{cases}. \tag{4.26}$$
2. Conclude from Eq. (4.25) that for all $z\in\mathbb{C}$,
$$(1+z)^N = 1+\sum_{k=1}^n z^k \sum_{1\le i_1<i_2<\dots<i_k\le n} 1_{A_{i_1}\cap\dots\cap A_{i_k}} \tag{4.27}$$
provided $(1+z)^0=1$ even when $z=-1$.
3. Conclude from Eq. (4.25) that $S_k=\mathbb{E}_P\binom{N}{k}$.

Exercise 4.5. Taking expectations of Eq. (4.27) implies,

$$\mathbb{E}\left[(1+z)^N\right] = 1+\sum_{k=1}^n S_k z^k. \tag{4.28}$$

Show that setting $z=-1$ in Eq. (4.28) gives another proof of the inclusion exclusion formula. **Hint:** use the definition of the expectation to write out $\mathbb{E}\left[(1+z)^N\right]$ explicitly.

Exercise 4.6. Let $1\le m\le n$. In this problem you are asked to compute the probability that there are exactly m – coincidences. Namely you should show,

$$\begin{aligned} P(N=m) &= \sum_{k=m}^n (-1)^{k-m}\binom{k}{m}S_k \\ &= \sum_{k=m}^n (-1)^{k-m}\binom{k}{m}\sum_{1\le i_1<\dots<i_k\le n} P\left(A_{i_1}\cap\dots\cap A_{i_k}\right)\end{aligned}$$

Hint: differentiate Eq. (4.28) m times with respect to z and then evaluate the result at $z=-1$. In order to do this you will find it useful to derive formulas for;

$$\frac{d^m}{dz^m}|_{z=-1}(1+z)^n \text{ and } \frac{d^m}{dz^m}|_{z=-1}z^k.$$

Example 4.27. Let us again go back to Example 4.26 where we computed,

$$S_k = \binom{n}{k}\frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Therefore it follows from Exercise 4.6 that

$$\begin{aligned} P(\exists \text{ exactly m fixed points}) = P(N=m) &= \sum_{k=m}^n (-1)^{k-m}\binom{k}{m}\frac{1}{k!} \\ &= \frac{1}{m!}\sum_{k=m}^n (-1)^{k-m}\frac{1}{(k-m)!}.\end{aligned}$$

So if n is much bigger than m we may conclude that

$$P(\exists \text{ exactly m fixed points}) \cong \frac{1}{m!}e^{-1}.$$

Let us check our results are consistent with Eq. (4.24);

$$\begin{aligned} P(\exists \text{ a fixed point}) &= \sum_{m=1}^n P(N=m) \\ &= \sum_{m=1}^n\sum_{k=m}^n (-1)^{k-m}\binom{k}{m}\frac{1}{k!} \\ &= \sum_{1\le m\le k\le n} (-1)^{k-m}\binom{k}{m}\frac{1}{k!} \\ &= \sum_{k=1}^n\sum_{m=1}^k (-1)^{k-m}\binom{k}{m}\frac{1}{k!} \\ &= \sum_{k=1}^n\left[\sum_{m=0}^k (-1)^{k-m}\binom{k}{m} - (-1)^k\right]\frac{1}{k!} \\ &= -\sum_{k=1}^n (-1)^k\frac{1}{k!}\end{aligned}$$

wherein we have used,

$$\sum_{m=0}^{k}(-1)^{k-m}\binom{k}{m}=(1-1)^k=0.$$

4.3.1 Appendix: Bonferroni Inequalities

In this appendix (see Feller Volume 1., p. 106-111 for more) we want to discuss what happens if we truncate the sums in the inclusion exclusion formula of Lemma 4.23. In order to do this we will need the following lemma whose combinatorial meaning was explained to me by Jeff Remmel.

Lemma 4.28. *Let $n\in\mathbb{N}_0$ and $0\le k\le n$, then*

$$\sum_{l=0}^{k}(-1)^l\binom{n}{l}=(-1)^k\binom{n-1}{k}1_{n>0}+1_{n=0}. \tag{4.29}$$

Proof. The case $n=0$ is trivial. We give two proofs for when $n\in\mathbb{N}$.

First proof. Just use induction on k. When $k=0$, Eq. (4.29) holds since $1=1$. The induction step is as follows,

$$\begin{aligned}\sum_{l=0}^{k+1}(-1)^l\binom{n}{l}&=(-1)^k\binom{n-1}{k}+\binom{n}{k+1}\\&=\frac{(-1)^{k+1}}{(k+1)!}\left[n(n-1)\dots(n-k)-(k+1)(n-1)\dots(n-k)\right]\\&=\frac{(-1)^{k+1}}{(k+1)!}\left[(n-1)\dots(n-k)(n-(k+1))\right]=(-1)^{k+1}\binom{n-1}{k+1}.\end{aligned}$$

Second proof. Let $X=\{1,2,\dots,n\}$ and observe that

$$\begin{aligned}m_k:=\sum_{l=0}^{k}(-1)^l\binom{n}{l}&=\sum_{l=0}^{k}(-1)^l\cdot\#\left(\Lambda\in 2^X:\#(\Lambda)=l\right)\\&=\sum_{\Lambda\in 2^X:\ \#(\Lambda)\le k}(-1)^{\#(\Lambda)}\end{aligned} \tag{4.30}$$

Define $T:2^X\to 2^X$ by

$$T(S)=\begin{cases}S\cup\{1\} & \text{if } 1\notin S\\ S\setminus\{1\} & \text{if } 1\in S\end{cases}.$$

Observe that T is a bijection of 2^X such that T takes even cardinality sets to odd cardinality sets and visa versa. Moreover, if we let

$$\Gamma_k:=\left\{\Lambda\in 2^X:\#(\Lambda)\le k\text{ and }1\in\Lambda\text{ if }\#(\Lambda)=k\right\},$$

then $T(\Gamma_k)=\Gamma_k$ for all $1\le k\le n$. Since

$$\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=\sum_{\Lambda\in\Gamma_k}(-1)^{\#(T(\Lambda))}=\sum_{\Lambda\in\Gamma_k}-(-1)^{\#(\Lambda)}$$

we see that $\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=0$. Using this observation with Eq. (4.30) implies

$$m_k=\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}+\sum_{\#(\Lambda)=k\ \&\ 1\notin\Lambda}(-1)^{\#(\Lambda)}=0+(-1)^k\binom{n-1}{k}.$$

■

Corollary 4.29 (Bonferroni Inequalitites). *Let $\mu:\mathcal{A}\to[0,\mu(X)]$ be a finitely additive finite measure on $\mathcal{A}\subset 2^X$, $A_n\in\mathcal{A}$ for $n=1,2,\dots,M$, $N:=\sum_{n=1}^{M}1_{A_n}$, and*

$$S_k:=\sum_{1\le i_1<\dots<i_k\le M}\mu\left(A_{i_1}\cap\dots\cap A_{i_k}\right)=\mathbb{E}_\mu\left[\binom{N}{k}\right].$$

Then for $1\le k\le M$,

$$\mu\left(\cup_{n=1}^{M}A_n\right)=\sum_{l=1}^{k}(-1)^{l+1}S_l+(-1)^k\,\mathbb{E}_\mu\left[\binom{N-1}{k}\right]. \tag{4.31}$$

This leads to the Bonferroni inequalities;

$$\mu\left(\cup_{n=1}^{M}A_n\right)\le\sum_{l=1}^{k}(-1)^{l+1}S_l\ \text{if } k \text{ is odd}$$

and

$$\mu\left(\cup_{n=1}^{M}A_n\right)\ge\sum_{l=1}^{k}(-1)^{l+1}S_l\ \text{if } k \text{ is even.}$$

Proof. By Lemma 4.28,

$$\sum_{l=0}^{k}(-1)^l\binom{N}{l}=(-1)^k\binom{N-1}{k}1_{N>0}+1_{N=0}.$$

Therefore integrating this equation with respect to μ gives,

$$\mu(X)+\sum_{l=1}^{k}(-1)^l S_l=\mu(N=0)+(-1)^k\,\mathbb{E}_\mu\binom{N-1}{k}$$

and therefore,

$$\mu\left(\cup_{n=1}^{M} A_n\right) = \mu\left(N>0\right) = \mu\left(X\right) - \mu\left(N=0\right)$$
$$= -\sum_{l=1}^{k}\left(-1\right)^{l} S_l + \left(-1\right)^{k} \mathbb{E}_\mu \binom{N-1}{k}.$$

The Bonferroni inequalities are a simple consequence of Eq. (4.31) and the fact that

$$\binom{N-1}{k} \geq 0 \implies \mathbb{E}_\mu \binom{N-1}{k} \geq 0.$$

■

4.3.2 Appendix: Riemann Stieljtes integral

In this subsection, let X be a set, $\mathcal{A} \subset 2^X$ be an algebra of sets, and $P := \mu : \mathcal{A} \to [0,\infty)$ be a finitely additive measure with $\mu(X) < \infty$. As above let

$$\mathbb{E}_\mu f := \int_X f d\mu := \sum_{\lambda\in\mathbb{C}} \lambda\mu(f=\lambda) \;\forall\; f \in \mathbb{S}(\mathcal{A}). \tag{4.32}$$

Notation 4.30 *For any function, $f : X \to \mathbb{C}$ let $\|f\|_u := \sup_{x\in X} |f(x)|$. Further, let $\bar{\mathbb{S}} := \overline{\mathbb{S}(\mathcal{A})}$ denote those functions, $f : X \to \mathbb{C}$ such that there exists $f_n \in \mathbb{S}(\mathcal{A})$ such that $\lim_{n\to\infty} \|f - f_n\|_u = 0$.*

Exercise 4.7. Prove the following statements.

1. For all $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}_\mu f| \leq \mu(X) \|f\|_u. \tag{4.33}$$
2. If $f \in \bar{\mathbb{S}}$ and $f_n \in \mathbb{S} := \mathbb{S}(\mathcal{A})$ such that $\lim_{n\to\infty} \|f - f_n\|_u = 0$, show $\lim_{n\to\infty} \mathbb{E}_\mu f_n$ exists. Also show that defining $\mathbb{E}_\mu f := \lim_{n\to\infty} \mathbb{E}_\mu f_n$ is well defined, i.e. you must show that $\lim_{n\to\infty} \mathbb{E}_\mu f_n = \lim_{n\to\infty} \mathbb{E}_\mu g_n$ if $g_n \in \mathbb{S}$ such that $\lim_{n\to\infty} \|f - g_n\|_u = 0$.
3. Show $\mathbb{E}_\mu : \bar{\mathbb{S}} \to \mathbb{C}$ is still linear and still satisfies Eq. (4.33).
4. Show $|f| \in \bar{\mathbb{S}}$ if $f \in \bar{\mathbb{S}}$ and that Eq. (4.19) is still valid, i.e. $|\mathbb{E}_\mu f| \leq \mathbb{E}_\mu |f|$ for all $f \in \bar{\mathbb{S}}$.

Let us now specialize the above results to the case where $X = [0,T]$ for some $T < \infty$. Let $\mathcal{S} := \{(a,b] : 0 \leq a \leq b \leq T\} \cup \{0\}$ which is easily seen to be a semi-algebra. The following proposition is fairly straightforward and will be left to the reader.

Proposition 4.31 (Riemann Stieljtes integral). *Let $F : [0,T] \to \mathbb{R}$ be an increasing function, then;*

1. *there exists a unique finitely additive measure, μ_F, on $\mathcal{A} := \mathcal{A}(\mathcal{S})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $0 \leq a \leq b \leq T$ and $\mu_F(\{0\}) = 0$. (In fact one could allow for $\mu_F(\{0\}) = \lambda$ for any $\lambda \geq 0$, but we would then have to write $\mu_{F,\lambda}$ rather than μ_F.)*
2. *Show $C([0,1],\mathbb{C}) \subset \overline{\mathbb{S}(\mathcal{A})}$. More precisely, suppose $\pi := \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is a partition of $[0,T]$ and $c = (c_1,\ldots,c_n) \in [0,T]^n$ with $t_{i-1} \leq c_i \leq t_i$ for each i. Then for $f \in C([0,1],\mathbb{C})$, let*
$$f_{\pi,c} := f(0)\, 1_{\{0\}} + \sum_{i=1}^{n} f(c_i)\, 1_{(t_{i-1},t_i]}. \tag{4.34}$$
Show that $\|f - f_{\pi,c}\|_u$ is small provided, $|\pi| := \max\{|t_i - t_{i-1}| : i = 1,2,\ldots,n\}$ is small.
3. *Using the above results, show*
$$\int_{[0,T]} f d\mu_F = \lim_{|\pi|\to 0} \sum_{i=1}^{n} f(c_i)\left(F(t_i) - F(t_{i-1})\right)$$
where the c_i may be chosen arbitrarily subject to the constraint that $t_{i-1} \leq c_i \leq t_i$.

It is customary to write $\int_0^T f dF$ for $\int_{[0,T]} f d\mu_F$. This integral satisfies the estimates,

$$\left|\int_{[0,T]} f d\mu_F\right| \leq \int_{[0,T]} |f|\, d\mu_F \leq \|f\|_u \left(F(T) - F(0)\right) \;\forall\; f \in \overline{\mathbb{S}(\mathcal{A})}.$$

When $F(t) = t$,

$$\int_0^T f dF = \int_0^T f(t)\, dt,$$

is the usual Riemann integral.

Exercise 4.8. Let $a \in (0,T)$, $\lambda > 0$, and

$$G(x) = \lambda \cdot 1_{x\geq a} = \begin{cases} \lambda \text{ if } x \geq a \\ 0 \text{ if } x < a \end{cases}.$$

1. Explicitly compute $\int_{[0,T]} f d\mu_G$ for all $f \in C([0,1],\mathbb{C})$.
2. If $F(x) = x + \lambda \cdot 1_{x\geq a}$ describe $\int_{[0,T]} f d\mu_F$ for all $f \in C([0,1],\mathbb{C})$. **Hint:** if $F(x) = G(x) + H(x)$ where G and H are two increasing functions on $[0,T]$, show
$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G + \int_{[0,T]} f d\mu_H.$$

Exercise 4.9. Suppose that $F,G:[0,T]\to\mathbb{R}$ are two increasing functions such that $F(0)=G(0)$, $F(T)=G(T)$, and $F(x)\neq G(x)$ for at most countably many points, $x\in(0,T)$. Show

$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G \text{ for all } f\in C([0,1],\mathbb{C}). \tag{4.35}$$

Note well, given $F(0)=G(0)$, $\mu_F=\mu_G$ on $\mathcal{A}$ iff $F=G$.

One of the points of the previous exercise is to show that Eq. (4.35) holds when $G(x):=F(x+)$ – the right continuous version of F. The exercise applies since and increasing function can have at most countably many jumps, see Remark 21.16. So if we only want to integrate continuous functions, we may always assume that $F:[0,T]\to\mathbb{R}$ is right continuous.

4.4 Simple Independence and the Weak Law of Large Numbers

To motivate the exercises in this section, let us imagine that we are following the outcomes of two "independent" experiments with values $\{\alpha_k\}_{k=1}^\infty\subset\Lambda_1$ and $\{\beta_k\}_{k=1}^\infty\subset\Lambda_2$ where Λ_1 and Λ_2 are two finite set of outcomes. Here we are using term independent in an intuitive form to mean that knowing the outcome of one of the experiments gives us no information about outcome of the other.

As an example of independent experiments, suppose that one experiment is the outcome of spinning a roulette wheel and the second is the outcome of rolling a dice. We expect these two experiments will be independent.

As an example of dependent experiments, suppose that dice roller now has two dice – one red and one black. The person rolling dice throws his black or red dice after the roulette ball has stopped and landed on either black or red respectively. If the black and the red dice are weighted differently, we expect that these two experiments are no longer independent.

Lemma 4.32 (Heuristic). *Suppose that $\{\alpha_k\}_{k=1}^\infty\subset\Lambda_1$ and $\{\beta_k\}_{k=1}^\infty\subset\Lambda_2$ are the outcomes of repeatedly running two experiments independent of each other and for $x\in\Lambda_1$ and $y\in\Lambda_2$,*

$$\begin{aligned} p(x,y) &:= \lim_{N\to\infty}\frac{1}{N}\#\{1\le k\le N:\alpha_k=x \text{ and } \beta_k=y\},\\ p_1(x) &:= \lim_{N\to\infty}\frac{1}{N}\#\{1\le k\le N:\alpha_k=x\}, \text{ and}\\ p_2(y) &:= \lim_{N\to\infty}\frac{1}{N}\#\{1\le k\le N:\beta_k=y\}. \end{aligned} \tag{4.36}$$

Then $p(x,y)=p_1(x)p_2(y)$. In particular this then implies for any $h:\Lambda_1\times\Lambda_2\to\mathbb{R}$ we have,

$$\mathbb{E}h=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N h(\alpha_k,\beta_k)=\sum_{(x,y)\in\Lambda_1\times\Lambda_2} h(x,y)\,p_1(x)\,p_2(y).$$

Proof. (Heuristic.) Let us imagine running the first experiment repeatedly with the results being recorded as, $\{\alpha_k^\ell\}_{k=1}^\infty$, where $\ell\in\mathbb{N}$ indicates the ℓ^{th} – run of the experiment. Then we have postulated that, independent of ℓ,

$$p(x,y):=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N 1_{\{\alpha_k^\ell=x \text{ and } \beta_k=y\}}=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N 1_{\{\alpha_k^\ell=x\}}\cdot 1_{\{\beta_k=y\}}$$

So for any $L\in\mathbb{N}$ we must also have,

$$\begin{aligned} p(x,y) &= \frac{1}{L}\sum_{\ell=1}^L p(x,y)=\frac{1}{L}\sum_{\ell=1}^L\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N 1_{\{\alpha_k^\ell=x\}}\cdot 1_{\{\beta_k=y\}}\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N\frac{1}{L}\sum_{\ell=1}^L 1_{\{\alpha_k^\ell=x\}}\cdot 1_{\{\beta_k=y\}}. \end{aligned}$$

Taking the limit of this equation as $L\to\infty$ and interchanging the order of the limits (this is faith based) implies,

$$p(x,y)=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N 1_{\{\beta_k=y\}}\cdot\lim_{L\to\infty}\frac{1}{L}\sum_{\ell=1}^L 1_{\{\alpha_k^\ell=x\}}. \tag{4.37}$$

Since for fixed k, $\{\alpha_k^\ell\}_{\ell=1}^\infty$ is just another run of the first experiment, by our postulate, we conclude that

$$\lim_{L\to\infty}\frac{1}{L}\sum_{\ell=1}^L 1_{\{\alpha_k^\ell=x\}}=p_1(x) \tag{4.38}$$

independent of the choice of k. Therefore combining Eqs. (4.36), (4.37), and (4.38) implies,

$$p(x,y)=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N 1_{\{\beta_k=y\}}\cdot p_1(x)=p_2(y)\,p_1(x).$$

■

To understand this "Lemma" in another but equivalent way, let $X_1:\Lambda_1\times\Lambda_2\to\Lambda_1$ and $X_2:\Lambda_1\times\Lambda_2\to\Lambda_2$ be the projection maps, $X_1(x,y)=x$ and

$X_2(x,y) = y$ respectively. Further suppose that $f : \Lambda_1 \to \mathbb{R}$ and $g : \Lambda_2 \to \mathbb{R}$ are functions, then using the heuristics Lemma 4.32 implies,

$$\begin{aligned}\mathbb{E}[f(X_1)g(X_2)] &= \sum_{(x,y)\in\Lambda_1\times\Lambda_2} f(x)g(y)p_1(x)p_2(y) \\ &= \sum_{x\in\Lambda_1} f(x)p_1(x)\cdot\sum_{y\in\Lambda_2} g(y)p_2(y) = \mathbb{E}f(X_1)\cdot\mathbb{E}g(X_2).\end{aligned}$$

Hopefully these heuristic computations will convince you that the mathematical notion of independence developed below is relevant. In what follows, we will use the obvious generalization of our "results" above to the setting of n – independent experiments. For notational simplicity we will now assume that $\Lambda_1 = \Lambda_2 = \dots = \Lambda_n = \Lambda$.

Let Λ be a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, and $X_i : \Omega \to \Lambda$ be defined by $X_i(\omega) = \omega_i$ for $\omega \in \Omega$ and $i = 1, 2, \dots, n$. We further suppose $p : \Omega \to [0,1]$ is a function such that

$$\sum_{\omega\in\Omega} p(\omega) = 1$$

and $P : 2^\Omega \to [0,1]$ is the probability measure defined by

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \in 2^\Omega. \tag{4.39}$$

Exercise 4.10 (Simple Independence 1.). Suppose $q_i : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda\in\Lambda} q_i(\lambda) = 1$ for $i = 1, 2, \dots, n$ and now define $p(\omega) = \prod_{i=1}^n q_i(\omega_i)$. Show for any functions, $f_i : \Lambda \to \mathbb{R}$ that

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)] = \prod_{i=1}^n \mathbb{E}_{Q_i} f_i$$

where Q_i is the measure on Λ defined by, $Q_i(\gamma) = \sum_{\lambda\in\gamma} q_i(\lambda)$ for all $\gamma \subset \Lambda$.

Exercise 4.11 (Simple Independence 2.). Prove the converse of the previous exercise. Namely, if

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)] \tag{4.40}$$

for any functions, $f_i : \Lambda \to \mathbb{R}$, then there exists functions $q_i : \Lambda \to [0,1]$ with $\sum_{\lambda\in\Lambda} q_i(\lambda) = 1$, such that $p(\omega) = \prod_{i=1}^n q_i(\omega_i)$.

Definition 4.33 (Independence). *We say simple random variables, $X_1, \dots, X_n$ with values in Λ on some probability space, $(\Omega, \mathcal{A}, P)$ are independent (more precisely P – independent) if Eq. (4.40) holds for all functions, $f_i : \Lambda \to \mathbb{R}$.*

Exercise 4.12 (Simple Independence 3.). Let $X_1, \dots, X_n : \Omega \to \Lambda$ and $P : 2^\Omega \to [0,1]$ be as described before Exercise 4.10. Show $X_1, \dots, X_n$ are independent iff

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1)\dots P(X_n \in A_n) \tag{4.41}$$

for all choices of $A_i \subset \Lambda$. Also explain why it is enough to restrict the A_i to single point subsets of Λ.

Exercise 4.13 (A Weak Law of Large Numbers). Suppose that $\Lambda \subset \mathbb{R}$ is a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, $p(\omega) = \prod_{i=1}^n q(\omega_i)$ where $q : \Lambda \to [0,1]$ such that $\sum_{\lambda\in\Lambda} q(\lambda) = 1$, and let $P : 2^\Omega \to [0,1]$ be the probability measure defined as in Eq. (4.39). Further let $X_i(\omega) = \omega_i$ for $i = 1, 2, \dots, n$, $\xi := \mathbb{E}X_i$, $\sigma^2 := \mathbb{E}(X_i - \xi)^2$, and

$$S_n = \frac{1}{n}(X_1 + \dots + X_n).$$

1. Show, $\xi = \sum_{\lambda\in\Lambda} \lambda\, q(\lambda)$ and
$$\sigma^2 = \sum_{\lambda\in\Lambda} (\lambda - \xi)^2 q(\lambda) = \sum_{\lambda\in\Lambda} \lambda^2 q(\lambda) - \xi^2. \tag{4.42}$$
2. Show, $\mathbb{E}S_n = \xi$.
3. Let $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Show
$$\mathbb{E}[(X_i - \xi)(X_j - \xi)] = \delta_{ij}\sigma^2.$$
4. Using $S_n - \xi$ may be expressed as, $\frac{1}{n}\sum_{i=1}^n (X_i - \xi)$, show
$$\mathbb{E}(S_n - \xi)^2 = \frac{1}{n}\sigma^2. \tag{4.43}$$
5. Conclude using Eq. (4.43) and Remark 4.22 that
$$P(|S_n - \xi| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2. \tag{4.44}$$

So for large n, S_n is concentrated near $\xi = \mathbb{E}X_i$ with probability approaching 1 for n large. This is a version of the weak law of large numbers.

Definition 4.34 (Covariance). *Let $(\Omega, \mathcal{B}, P)$ is a finitely additive probability. The* ***covariance,*** $\operatorname{Cov}(X,Y)$, *of $X, Y \in \mathbb{S}(\mathcal{B})$ is defined by*

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X - \xi_X)(Y - \xi_Y)] = \mathbb{E}[XY] - \mathbb{E}X\cdot\mathbb{E}Y$$

where $\xi_X := \mathbb{E}X$ and $\xi_Y := \mathbb{E}Y$. The variance of X,

$$\mathrm{Var}(X) := \mathrm{Cov}(X, X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$$

We say that X *and* Y *are* ***uncorrelated*** *if* $\mathrm{Cov}(X,Y) = 0$*, i.e.* $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$*. More generally we say* $\{X_k\}_{k=1}^n \subset \mathbb{S}(\mathcal{B})$ *are uncorrelated iff* $\mathrm{Cov}(X_i, X_j) = 0$ *for all* $i \neq j$*.*

Remark 4.35. 1. Observe that X and Y are independent iff $f(X)$ and $g(Y)$ are uncorrelated for all functions, f and g on the range of X and Y respectively. In particular if X and Y are independent then $\mathrm{Cov}(X,Y) = 0$.

2. If you look at your proof of the weak law of large numbers in Exercise 4.13 you will see that it suffices to assume that $\{X_i\}_{i=1}^n$ are uncorrelated rather than the stronger condition of being independent.

Exercise 4.14 (Bernoulli Random Variables). Let $\Lambda = \{0,1\}$, $X : \Lambda \to \mathbb{R}$ be defined by $X(0) = 0$ and $X(1) = 1$, $x \in [0,1]$, and define $Q = x\delta_1 + (1-x)\delta_0$, i.e. $Q(\{0\}) = 1 - x$ and $Q(\{1\}) = x$. Verify,

$$\xi(x) := \mathbb{E}_Q X = x \text{ and}$$
$$\sigma^2(x) := \mathbb{E}_Q (X - x)^2 = (1-x)x \leq 1/4.$$

Theorem 4.36 (Weierstrass Approximation Theorem via Bernstein's Polynomials.). *Suppose that* $f \in C([0,1], \mathbb{C})$ *and*

$$p_n(x) := \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Then

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |f(x) - p_n(x)| = 0.$$

Proof. Let $x \in [0,1]$, $\Lambda = \{0,1\}$, $q(0) = 1 - x$, $q(1) = x$, $\Omega = \Lambda^n$, and

$$P_x(\{\omega\}) = q(\omega_1)\dots q(\omega_n) = x^{\sum_{i=1}^n \omega_i} \cdot (1-x)^{1-\sum_{i=1}^n \omega_i}.$$

As above, let $S_n = \frac{1}{n}(X_1 + \cdots + X_n)$, where $X_i(\omega) = \omega_i$ and observe that

$$P_x\left(S_n = \frac{k}{n}\right) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Therefore, writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$, we have

$$\mathbb{E}_x[f(S_n)] = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} = p_n(x).$$

Hence we find

$$\begin{aligned}
|p_n(x) - f(x)| &= |\mathbb{E}_x f(S_n) - f(x)| = |\mathbb{E}_x[f(S_n) - f(x)]| \\
&\leq \mathbb{E}_x |f(S_n) - f(x)| \\
&= \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| \geq \varepsilon] \\
&\quad + \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| < \varepsilon] \\
&\leq 2M \cdot P_x(|S_n - x| \geq \varepsilon) + \delta(\varepsilon)
\end{aligned}$$

where

$$M := \max_{y\in[0,1]} |f(y)| \text{ and}$$
$$\delta(\varepsilon) := \sup\{|f(y) - f(x)| : x, y \in [0,1] \text{ and } |y - x| \leq \varepsilon\}$$

is the modulus of continuity of f. Now by the above exercises,

$$P_x(|S_n - x| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2} \quad \text{(see Figure 4.1)} \tag{4.45}$$

and hence we may conclude that

$$\max_{x\in[0,1]} |p_n(x) - f(x)| \leq \frac{M}{2n\varepsilon^2} + \delta(\varepsilon)$$

and therefore, that

$$\limsup_{n\to\infty} \max_{x\in[0,1]} |p_n(x) - f(x)| \leq \delta(\varepsilon).$$

This completes the proof, since by uniform continuity of f, $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$. ■

4.4.1 Complex Weierstrass Approximation Theorem

The main goal of this subsection is to prove Theorem 4.42 which states that any continuous 2π – periodic function on $\mathbb{R}$ may be well approximated by trigonometric polynomials. The main ingredient is the following two dimensional generalization of Theorem 4.36. All of the results in this section have natural generalization to higher dimensions as well , see Theorem 4.50.

Theorem 4.37 (Weierstrass Approximation Theorem). *Suppose that* $K = [0,1]^2$, $f \in C(K, \mathbb{C})$*, and*

$$p_n(x,y) := \sum_{k,l=0}^n f\left(\frac{k}{n}, \frac{l}{n}\right)\binom{n}{k}\binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}. \tag{4.46}$$

Then $p_n \to f$ *uniformly on* K*.*

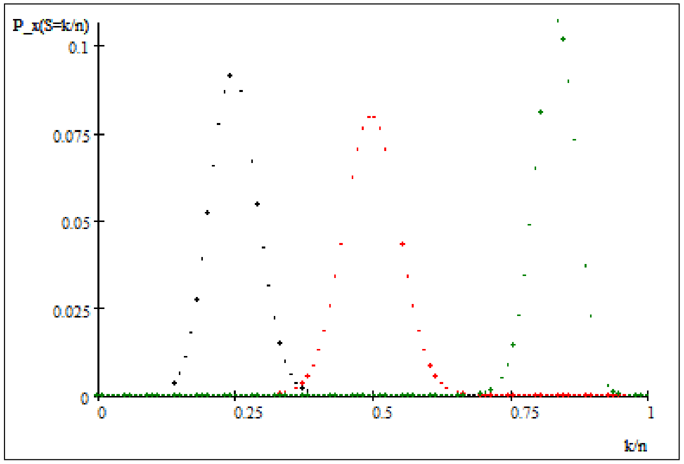


Fig. 4.1. Plots of $P_x(S_n = k/n)$ versus k/n for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

Proof. We are going to follow the argument given in the proof of Theorem 4.36. By considering the real and imaginary parts of f separately, it suffices to assume $f \in C([0,1]^2, \mathbb{R})$. For $(x,y) \in K$ and $n \in \mathbb{N}$ we may choose a collection of independent Bernoulli simple random variables $\{X_i, Y_i\}_{i=1}^n$ such that $P(X_i = 1) = x$ and $P(Y_i = 1) = y$ for all $1 \le i \le n$. Then letting $S_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $T_n := \frac{1}{n}\sum_{i=1}^n Y_i$, we have

$$\mathbb{E}[f(S_n, T_n)] = \sum_{k,l=0}^{n} f\left(\frac{k}{n}, \frac{l}{n}\right) P(n \cdot S_n = k, n \cdot T_n = l) = p_n(x,y)$$

where $p_n(x,y)$ is the polynomial given in Eq. (4.46) wherein the assumed independence is needed to show,

$$P(n \cdot S_n = k, n \cdot T_n = l) = \binom{n}{k}\binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}.$$

Thus if $M = \sup\{|f(x,y)| : (x,y) \in K\}$, $\varepsilon > 0$,

$$\delta_\varepsilon = \sup\{|f(x',y') - f(x,y)| : (x,y), (x',y') \in K \text{ and } \|x',y' - (x,y)\| \le \varepsilon\},$$

and

$$A := \{\|(S_n, T_n) - (x,y)\| > \varepsilon\},$$

we have,

$$\begin{aligned}
|f(x,y) - p_n(x,y)| &= |\mathbb{E}(f(x,y) - f((S_n, T_n)))| \\
&\le \mathbb{E}|f(x,y) - f((S_n, T_n))| \\
&= \mathbb{E}[|f(x,y) - f(S_n, T_n)| : A] \\
&\quad + \mathbb{E}[|f(x,y) - f(S_n, T_n)| : A^c] \\
&\le 2M \cdot P(A) + \delta_\varepsilon \cdot P(A^c) \\
&\le 2M \cdot P(A) + \delta_\varepsilon.
\end{aligned} \tag{4.47}$$

To estimate $P(A)$, observe that if

$$\|(S_n, T_n) - (x,y)\|^2 = (S_n - x)^2 + (T_n - y)^2 > \varepsilon^2,$$

then either,

$$(S_n - x)^2 > \varepsilon^2/2 \text{ or } (T_n - y)^2 > \varepsilon^2/2$$

and therefore by sub-additivity and Eq. (4.45) we know

$$\begin{aligned}
P(A) &\le P\left(|S_n - x| > \varepsilon/\sqrt{2}\right) + P\left(|T_n - y| > \varepsilon/\sqrt{2}\right) \\
&\le \frac{1}{2n\varepsilon^2} + \frac{1}{2n\varepsilon^2} = \frac{1}{n\varepsilon^2}.
\end{aligned} \tag{4.48}$$

Using this estimate in Eq. (4.47) gives,

$$|f(x,y) - p_n(x,y)| \le 2M \cdot \frac{1}{n\varepsilon^2} + \delta_\varepsilon$$

and as right is independent of $(x,y) \in K$ we may conclude,

$$\limsup_{n\to\infty} \sup_{(x,y)\in K} |f(x,y) - p_n(x,y)| \le \delta_\varepsilon$$

which completes the proof since $\delta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$ because f is uniformly continuous on K. ∎

Remark 4.38. We can easily improve our estimate on $P(A)$ in Eq. (4.48) by a factor of two as follows. As in the proof of Theorem 4.36,

$$\begin{aligned}
\mathbb{E}\left[\|(S_n, T_n) - (x,y)\|^2\right] &= \mathbb{E}\left[(S_n - x)^2 + (T_n - y)^2\right] \\
&= \mathrm{Var}(S_n) + \mathrm{Var}(T_n) \\
&= \frac{1}{n} x(1-x) + y(1-y) \le \frac{1}{2n}.
\end{aligned}$$

Therefore by Chebyshev's inequality,

$$P(A) = P(\|(S_n, T_n) - (x,y)\| > \varepsilon) \le \frac{1}{\varepsilon^2}\mathbb{E}\|(S_n, T_n) - (x,y)\|^2 \le \frac{1}{2n\varepsilon^2}.$$

Corollary 4.39. *Suppose that $K = [a,b] \times [c,d]$ is any compact rectangle in $\mathbb{R}^2$. Then every function, $f \in C(K,\mathbb{C})$, may be uniformly approximated by polynomial functions in $(x,y) \in \mathbb{R}^2$.*

Proof. Let $F(x,y) := f(a + x(b-a), c + y(d-c))$ – a continuous function of $(x,y) \in [0,1]^2$. Given $\varepsilon > 0$, we may use Theorem Theorem 4.37 to find a polynomial, $p(x,y)$, such that $\sup_{(x,y)\in[0,1]^2} |F(x,y) - p(x,y)| \le \varepsilon$. Letting $\xi = a + x(b-a)$ and $\eta := c + y(d-c)$, it now follows that

$$\sup_{(\xi.\eta)\in K} \left| f(\xi,\eta) - p\left(\frac{\xi - a}{b-a}, \frac{\eta - c}{d-c}\right)\right| \le \varepsilon$$

which completes the proof since $p\left(\frac{\xi-a}{b-a}, \frac{\eta-c}{d-c}\right)$ is a polynomial in (ξ,η). ∎

Here is a version of the complex Weierstrass approximation theorem.

Theorem 4.40 (Complex Weierstrass Approximation Theorem). *Suppose that $K \subset \mathbb{C}$ is a compact rectangle. Then there exists polynomials in $(z = x+iy, \bar z = x - iy)$, $p_n(z,\bar z)$ for $z \in \mathbb{C}$, such that $\sup_{z\in K} |q_n(z,\bar z) - f(z)| \to 0$ as $n \to \infty$ for every $f \in C(K,\mathbb{C})$.*

Proof. The mapping $(x,y) \in \mathbb{R}\times\mathbb{R} \to z = x + iy \in \mathbb{C}$ is an isomorphism of vector spaces. Letting $\bar z = x - iy$ as usual, we have $x = \frac{z+\bar z}{2}$ and $y = \frac{z - \bar z}{2i}$. Therefore under this identification any polynomial $p(x,y)$ on $\mathbb{R}\times\mathbb{R}$ may be written as a polynomial q in $(z,\bar z)$, namely

$$q(z,\bar z) = p\left(\frac{z+\bar z}{2}, \frac{z-\bar z}{2i}\right).$$

Conversely a polynomial q in $(z,\bar z)$ may be thought of as a polynomial p in (x,y), namely $p(x,y) = q(x+iy, x-iy)$. Hence the result now follows from Theorem 4.37. ∎

Example 4.41. Let $K = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathcal{A}$ be the set of polynomials in $(z,\bar z)$ restricted to S^1. Then $\mathcal{A}$ is dense in $C(S^1)$. To prove this first observe if $f \in C(S^1)$ then $F(z) = |z| f\left(\frac{z}{|z|}\right)$ for $z \ne 0$ and $F(0) = 0$ defines $F \in C(\mathbb{C})$ such that $F|_{S^1} = f$. By applying Theorem 4.40 to F restricted to a compact rectangle containing S^1 we may find $q_n(z,\bar z)$ converging uniformly to F on K and hence on S^1. Since $\bar z$ on S^1, we have shown polynomials in z and z^{-1} are dense in $C(S^1)$.

Theorem 4.42 (Density of Trigonometric Polynomials). *Any 2π – periodic continuous function, $f : \mathbb{R} \to \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form*

$$p(x) = \sum_{\lambda\in\Lambda} a_\lambda e^{i\lambda\cdot x}$$

where Λ is a finite subset of $\mathbb{Z}$ and $a_\lambda \in \mathbb{C}$ for all $\lambda \in \Lambda$.

Proof. For $z \in S^1$, define $F(z) := f(\theta)$ where $\theta \in \mathbb{R}$ is chosen so that $z = e^{i\theta}$. Since f is 2π – periodic, F is well defined since if θ solves $e^{i\theta} = z$ then all other solutions are of the form $\{\theta + 2\pi n : n \in \mathbb{Z}\}$. Since the map $\theta \to e^{i\theta}$ is a local homeomorphism, i.e. for any $J = (a,b)$ with $b - a < 2\pi$, the map $\theta \in J \xrightarrow{\phi} \tilde J := \left\{e^{i\theta} : \theta \in J\right\} \subset S^1$ is a homeomorphism, it follows that $F(z) = f \circ \phi^{-1}(z)$ for $z \in \tilde J$. This shows F is continuous when restricted to $\tilde J$. Since such sets cover S^1, it follows that F is continuous.

By Example 4.41, the polynomials in z and $\bar z = z^{-1}$ are dense in $C(S^1)$. Hence for any $\varepsilon > 0$ there exists

$$p(z,\bar z) = \sum_{0\le m,n\le N} a_{m,n} z^m \bar z^n$$

such that $|F(z) - p(z,\bar z)| \le \varepsilon$ for all $z \in S^1$. Taking $z = e^{i\theta}$ then implies

$$\sup_\theta \left| f(\theta) - p\left(e^{i\theta}, e^{-i\theta}\right)\right| \le \varepsilon$$

where

$$p\left(e^{i\theta}, e^{-i\theta}\right) = \sum_{0\le m,n\le N} a_{m,n} e^{i(m-n)\theta}$$

is the desired trigonometry polynomial. ∎

4.4.2 Product Measures and Fubini's Theorem

In the last part of this section we will extend some of the above ideas to more general "finitely additive measure spaces." A **finitely additive measure space** is a triple, $(X,\mathcal{A},\mu)$, where X is a set, $\mathcal{A} \subset 2^X$ is an algebra, and $\mu : \mathcal{A} \to [0,\infty]$ is a finitely additive measure. Let $(Y,\mathcal{B},\nu)$ be another finitely additive measure space.

Definition 4.43. *Let $\mathcal{A}\odot\mathcal{B}$ be the smallest sub-algebra of $2^{X\times Y}$ containing all sets of the form $\mathcal{S} := \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$. As we have seen in Exercise 3.10, $\mathcal{S}$ is a semi-algebra and therefore $\mathcal{A}\odot\mathcal{B}$ consists of subsets, $C \subset X \times Y$, which may be written as;*

$$C = \sum_{i=1}^n A_i \times B_i \text{ with } A_i \times B_i \in \mathcal{S}. \tag{4.49}$$

Theorem 4.44 (Product Measure and Fubini's Theorem). *Assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ for simplicity. Then there is a unique finitely additive measure, $\mu \odot \nu$, on $\mathcal{A} \odot \mathcal{B}$ such that $\mu \odot \nu(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Moreover if $f \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$ then;*

1. *$y \to f(x,y)$ is in $\mathbb{S}(\mathcal{B})$ for all $x \in X$ and $x \to f(x,y)$ is in $\mathbb{S}(\mathcal{A})$ for all $y \in Y$.*
2. *$x \to \int_Y f(x,y)\,d\nu(y)$ is in $\mathbb{S}(\mathcal{A})$ and $y \to \int_X f(x,y)\,d\mu(x)$ is in $\mathbb{S}(\mathcal{B})$.*
3. *we have,*

$$\begin{aligned}\int_X \left[\int_Y f(x,y)\,d\nu(y)\right] d\mu(x) \\ &= \int_{X\times Y} f(x,y)\,d(\mu\odot\nu)(x,y) \\ &= \int_Y \left[\int_X f(x,y)\,d\mu(x)\right] d\nu(y).\end{aligned}$$

We will refer to $\mu \odot \nu$ as the ***product measure*** *of μ and ν.*

Proof. According to Eq. (4.49),

$$1_C(x,y) = \sum_{i=1}^n 1_{A_i\times B_i}(x,y) = \sum_{i=1}^n 1_{A_i}(x)\,1_{B_i}(y)$$

from which it follows that $1_C(x,\cdot) \in \mathbb{S}(\mathcal{B})$ for each $x \in X$ and

$$\int_Y 1_C(x,y)\,d\nu(y) = \sum_{i=1}^n 1_{A_i}(x)\,\nu(B_i).$$

It now follows from this equation that $x \to \int_Y 1_C(x,y)\,d\nu(y) \in \mathbb{S}(\mathcal{A})$ and that

$$\int_X \left[\int_Y 1_C(x,y)\,d\nu(y)\right] d\mu(x) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i).$$

Similarly one shows that

$$\int_Y \left[\int_X 1_C(x,y)\,d\mu(x)\right] d\nu(y) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i).$$

In particular this shows that we may define

$$(\mu\odot\nu)(C) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i)$$

and with this definition we have,

$$\begin{aligned}\int_X \left[\int_Y 1_C(x,y)\,d\nu(y)\right] d\mu(x) \\ &= (\mu\odot\nu)(C) \\ &= \int_Y \left[\int_X 1_C(x,y)\,d\mu(x)\right] d\nu(y).\end{aligned}$$

From either of these representations it is easily seen that $\mu \odot \nu$ is a finitely additive measure on $\mathcal{A} \odot \mathcal{B}$ with the desired properties. Moreover, we have already verified the Theorem in the special case where $f = 1_C$ with $C \in \mathcal{A} \odot \mathcal{B}$. Since the general element, $f \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$, is a linear combination of such functions, it is easy to verify using the linearity of the integral and the fact that $\mathbb{S}(\mathcal{A})$ and $\mathbb{S}(\mathcal{B})$ are vector spaces that the theorem is true in general. ■

Example 4.45. Suppose that $f \in \mathbb{S}(\mathcal{A})$ and $g \in \mathbb{S}(\mathcal{B})$. Let $f \otimes g(x,y) := f(x)g(y)$. Since we have,

$$\begin{aligned}f\otimes g(x,y) &= \left(\sum_a a1_{f=a}(x)\right)\left(\sum_b b1_{g=b}(y)\right) \\ &= \sum_{a,b} ab1_{\{f=a\}\times\{g=b\}}(x,y)\end{aligned}$$

it follows that $f \otimes g \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$. Moreover, using Fubini's Theorem 4.44 it follows that

$$\int_{X\times Y} f\otimes g\ d(\mu\odot\nu) = \left[\int_X f\ d\mu\right]\left[\int_Y g\ d\nu\right].$$

4.5 Simple Conditional Expectation

In this section, $\mathcal{B}$ is a sub-algebra of 2^Ω, $P : \mathcal{B} \to [0,1]$ is a finitely additive probability measure, and $\mathcal{A} \subset \mathcal{B}$ is a finite sub-algebra. As in Example 3.19, for each $\omega \in \Omega$, let $A_\omega := \cap\{A \in \mathcal{A} : \omega \in A\}$ and recall that either $A_\omega = A_{\omega'}$ or $A_\omega \cap A_{\omega'} = \emptyset$ for all $\omega, \omega' \in \Omega$. In particular there is a partition, $\{B_1, \dots, B_n\}$, of Ω such that $A_\omega \in \{B_1, \dots, B_n\}$ for all $\omega \in \Omega$.

Definition 4.46 (Conditional expectation). *Let $X : \Omega \to \mathbb{R}$ be a $\mathcal{B}$ – simple random variable, i.e. $X \in \mathbb{S}(\mathcal{B})$, and*

$$\bar{X}(\omega) := \frac{1}{P(A_\omega)}\mathbb{E}[1_{A_\omega}X] \ \text{ for all } \omega\in\Omega, \tag{4.50}$$

where by convention, $\bar{X}(\omega) = 0$ *if* $P(A_\omega) = 0$. *We will denote* $\bar{X}$ *by* $\mathbb{E}[X|\mathcal{A}]$ *for* $\mathbb{E}_{\mathcal{A}} X$ *and call it the conditional expectation of* X *given* $\mathcal{A}$. *Alternatively we may write* $\bar{X}$ *as*

$$\bar{X} = \sum_{i=1}^{n} \frac{\mathbb{E}[1_{B_i} X]}{P(B_i)} 1_{B_i}, \tag{4.51}$$

again with the convention that $\mathbb{E}[1_{B_i} X] / P(B_i) = 0$ *if* $P(B_i) = 0$.

It should be noted, from Exercise 4.1, that $\bar{X} = \mathbb{E}_{\mathcal{A}} X \in \mathbb{S}(\mathcal{A})$. Heuristically, if $(\omega(1), \omega(2), \omega(3), \dots)$ is the sequence of outcomes of "independently" running our "experiment" repeatedly, then

$$\begin{aligned} \bar{X}|_{B_i} &= \frac{\mathbb{E}[1_{B_i} X]}{P(B_i)} \text{ "="} \frac{\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_{B_i}(\omega(n)) X(\omega(n))}{\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_{B_i}(\omega(n))} \\ &= \lim_{N\to\infty} \frac{\sum_{n=1}^{N} 1_{B_i}(\omega(n)) X(\omega(n))}{\sum_{n=1}^{N} 1_{B_i}(\omega(n))}. \end{aligned}$$

So to compute $\bar{X}|_{B_i}$ "empirically," we remove all experimental outcomes from the list, $(\omega(1), \omega(2), \omega(3), \dots) \in \Omega^{\mathbb{N}}$, which are not in B_i to form a new list, $(\bar{\omega}(1), \bar{\omega}(2), \bar{\omega}(3), \dots) \in B_i^{\mathbb{N}}$. We then compute $\bar{X}|_{B_i}$ using the empirical formula for the expectation of X relative to the "bar" list, i.e.

$$\bar{X}|_{B_i} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X(\bar{\omega}(n)).$$

Exercise 4.15 (Simple conditional expectation). Let $X \in \mathbb{S}(\mathcal{B})$ and, for simplicity, assume all functions are real valued. Prove the following assertions;

1. **(Orthogonal Projection Property 1.)** If $Z \in \mathbb{S}(\mathcal{A})$, then
$$\mathbb{E}[XZ] = \mathbb{E}[\bar{X}Z] = \mathbb{E}[\mathbb{E}_{\mathcal{A}} X \cdot Z] \tag{4.52}$$
and
$$(\mathbb{E}_{\mathcal{A}} Z)(\omega) = \begin{cases} Z(\omega) & \text{if } P(A_\omega) > 0 \\ 0 & \text{if } P(A_\omega) = 0 \end{cases}. \tag{4.53}$$
In particular, $\mathbb{E}_{\mathcal{A}}[\mathbb{E}_{\mathcal{A}} Z] = \mathbb{E}_{\mathcal{A}} Z$.
This basically says that $\mathbb{E}_{\mathcal{A}}$ is orthogonal projection from $\mathbb{S}(\mathcal{B})$ onto $\mathbb{S}(\mathcal{A})$ relative to the inner product
$$(f, g) = \mathbb{E}[fg] \text{ for all } f, g \in \mathbb{S}(\mathcal{B}).$$
2. **(Orthogonal Projection Property 2.)** If $Y \in \mathbb{S}(\mathcal{A})$ satisfies, $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all $Z \in \mathbb{S}(\mathcal{A})$, then $Y(\omega) = \bar{X}(\omega)$ whenever $P(A_\omega) > 0$. In particular, $P(Y \neq \bar{X}) = 0$. **Hint:** use item 1. to compute $\mathbb{E}\left[(\bar{X} - Y)^2\right]$.
3. **(Best Approximation Property.)** For any $Y \in \mathbb{S}(\mathcal{A})$,
$$\mathbb{E}\left[(X - \bar{X})^2\right] \le \mathbb{E}\left[(X - Y)^2\right] \tag{4.54}$$
with equality iff $\bar{X} = Y$ almost surely (a.s. for short), where $\bar{X} = Y$ a.s. iff $P(\bar{X} \neq Y) = 0$. In words, $\bar{X} = \mathbb{E}_{\mathcal{A}} X$ is the best ("L^2") approximation to X by an $\mathcal{A}$ – measurable random variable.
4. **(Contraction Property.)** $\mathbb{E}|\bar{X}| \le \mathbb{E}|X|$. (It is typically **not** true that $|\bar{X}(\omega)| \le |X(\omega)|$ for all ω.)
5. (**Pull Out Property.**) If $Z \in \mathbb{S}(\mathcal{A})$, then
$$\mathbb{E}_{\mathcal{A}}[ZX] = Z \mathbb{E}_{\mathcal{A}} X.$$

Example 4.47 (Heuristics of independence and conditional expectations). Let us suppose that we have an experiment consisting of spinning a spinner with values in $\Lambda_1 = \{1, 2, \dots, 10\}$ and rolling a die with values in $\Lambda_2 = \{1, 2, 3, 4, 5, 6\}$. So the outcome of an experiment is represented by a point, $\omega = (x, y) \in \Omega = \Lambda_1 \times \Lambda_2$. Let $X(x, y) = x$, $Y(x, y) = y$, $\mathcal{B} = 2^{\Omega}$, and

$$\mathcal{A} = \mathcal{A}(X) = X^{-1}\left(2^{\Lambda_1}\right) = \left\{X^{-1}(A) : A \subset \Lambda_1\right\} \subset \mathcal{B},$$

so that $\mathcal{A}$ is the smallest algebra of subsets of Ω such that $\{X = x\} \in \mathcal{A}$ for all $x \in \Lambda_1$. Notice that the partition associated to $\mathcal{A}$ is precisely

$$\{\{X = 1\}, \{X = 2\}, \dots, \{X = 10\}\}.$$

Let us now suppose that the spins of the spinner are "empirically independent" of the throws of the dice. As usual let us run the experiment repeatedly to produce a sequence of results, $\omega_n = (x_n, y_n)$ for all $n \in \mathbb{N}$. If $g : \Lambda_2 \to \mathbb{R}$ is a function, we have (heuristically) that

$$\begin{aligned} \mathbb{E}_{\mathcal{A}}[g(Y)](x, y) &= \lim_{N\to\infty} \frac{\sum_{n=1}^{N} g(Y(\omega(n))) 1_{X(\omega(n))=x}}{\sum_{n=1}^{N} 1_{X(\omega(n))=x}} \\ &= \lim_{N\to\infty} \frac{\sum_{n=1}^{N} g(y_n) 1_{x_n=x}}{\sum_{n=1}^{N} 1_{x_n=x}}. \end{aligned}$$

As the $\{y_n\}$ sequence of results are independent of the $\{x_n\}$ sequence, we should expect by the usual mantra[2] that

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} g(y_n) 1_{x_n=x}}{\sum_{n=1}^{N} 1_{x_n=x}} = \lim_{N\to\infty} \frac{1}{M(N)} \sum_{n=1}^{M(N)} g(\bar{y}_n) = \mathbb{E}[g(Y)],$$

[2] That is it should not matter which sequence of independent experiments are used to compute the time averages.

where $M(N) = \sum_{n=1}^{N} 1_{x_n=x}$ and $(\bar{y}_1, \bar{y}_2, \dots) = \{y_l : 1_{x_l=x}\}$. (We are also assuming here that $P(X=x) > 0$ so that we expect, $M(N) \sim P(X=x) N$ for N large, in particular $M(N) \to \infty$.) Thus under the assumption that X and Y are describing "independent" experiments we have heuristically deduced that $\mathbb{E}_{\mathcal{A}}[g(Y)] : \Omega \to \mathbb{R}$ is the constant function;

$$\mathbb{E}_{\mathcal{A}}[g(Y)](x,y) = \mathbb{E}[g(Y)] \text{ for all } (x,y) \in \Omega. \tag{4.55}$$

Let us further observe that if $f : \Lambda_1 \to \mathbb{R}$ is any other function, then $f(X)$ is an $\mathcal{A}$ – simple function and therefore by Eq. (4.55) and Exercise 4.15

$$\mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)] = \mathbb{E}[f(X) \cdot \mathbb{E}[g(Y)]] = \mathbb{E}[f(X) \cdot \mathbb{E}_{\mathcal{A}}[g(Y)]] = \mathbb{E}[f(X) \cdot g(Y)].$$

This observation along with Exercise 4.12 gives another "proof" of Lemma 4.32.

Lemma 4.48 (Conditional Expectation and Independence). *Let $\Omega = \Lambda_1 \times \Lambda_2$, X, Y, $\mathcal{B} = 2^{\Omega}$, and $\mathcal{A} = X^{-1}(2^{\Lambda_1})$, be as in Example 4.47 above. Assume that $P : \mathcal{B} \to [0,1]$ is a probability measure. If X and Y are P – independent, then Eq. (4.55) holds.*

Proof. From the definitions of conditional expectation and of independence we have,

$$\mathbb{E}_{\mathcal{A}}[g(Y)](x,y) = \frac{\mathbb{E}[1_{X=x} \cdot g(Y)]}{P(X=x)} = \frac{\mathbb{E}[1_{X=x}] \cdot \mathbb{E}[g(Y)]}{P(X=x)} = \mathbb{E}[g(Y)].$$

■

The following theorem summarizes much of what we (i.e. you) have shown regarding the underlying notion of independence of a pair of simple functions.

Theorem 4.49 (Independence result summary). *Let $(\Omega, \mathcal{B}, P)$ be a finitely additive probability space, Λ be a finite set, and $X, Y : \Omega \to \Lambda$ be two $\mathcal{B}$ – measurable simple functions, i.e. $\{X = x\} \in \mathcal{B}$ and $\{Y = y\} \in \mathcal{B}$ for all $x, y \in \Lambda$. Further let $\mathcal{A} = \mathcal{A}(X) := \mathcal{A}(\{X = x\} : x \in \Lambda)$. Then the following are equivalent;*

1. *$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ for all $x \in \Lambda$ and $y \in \Lambda$,*
2. *$\mathbb{E}[f(X) g(Y)] = \mathbb{E}[f(X)] \mathbb{E}[g(Y)]$ for all functions, $f : \Lambda \to \mathbb{R}$ and $g : \Lambda \to \mathbb{R}$,*
3. *$\mathbb{E}_{\mathcal{A}(X)}[g(Y)] = \mathbb{E}[g(Y)]$ for all $g : \Lambda \to \mathbb{R}$, and*
4. *$\mathbb{E}_{\mathcal{A}(Y)}[f(X)] = \mathbb{E}[f(X)]$ for all $f : \Lambda \to \mathbb{R}$.*

We say that X and Y are P – independent if any one (and hence all) of the above conditions holds.

4.6 Appendix: A Multi-dimensional Weirstrass Approximation Theorem

The following theorem is the multi-dimensional generalization of Theorem 4.36.

Theorem 4.50 (Weierstrass Approximation Theorem). *Suppose that $K = [a_1, b_1] \times \dots [a_d, b_d]$ with $-\infty < a_i < b_i < \infty$ is a compact rectangle in $\mathbb{R}^d$. Then for every $f \in C(K, \mathbb{C})$, there exists polynomials p_n on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on K.*

Proof. By a simple scaling and translation of the arguments of f we may assume without loss of generality that $K = [0,1]^d$. By considering the real and imaginary parts of f separately, it suffices to assume $f \in C([\mathbf{0}, \mathbf{1}], \mathbb{R})$.

Given $x \in K$, let $\{X_n = (X_n^1, \dots, X_n^d)\}_{n=1}^{\infty}$ be i.i.d. random vectors with values in $\mathbb{R}^d$ such that

$$P(X_n = \eta) = \prod_{i=1}^{d} (1 - x_i)^{1-\eta_i} x_i^{\eta_i}$$

for all $\eta = (\eta_1, \dots, \eta_d) \in \{0,1\}^d$. Since each X_n^j is a Bernoulli random variable with $P(X_n^j = 1) = x_j$, we know that

$$\mathbb{E}X_n = x \text{ and } \operatorname{Var}(X_n^j) = x_j - x_j^2 = x_j(1 - x_j).$$

As usual let $S_n = S_n := X_1 + \dots + X_n \in \mathbb{R}^d$, then

$$\mathbb{E}\left[\frac{S_n}{n}\right] = x \text{ and}$$

$$\begin{aligned}
\mathbb{E}\left[\left\|\frac{S_n}{n} - x\right\|^2\right] &= \sum_{j=1}^{d} \mathbb{E}\left(\frac{S_n^j}{n} - x_j\right)^2 = \sum_{j=1}^{d} \operatorname{Var}\left(\frac{S_n^j}{n} - x_j\right) \\
&= \sum_{j=1}^{d} \operatorname{Var}\left(\frac{S_n^j}{n}\right) = \frac{1}{n^2} \cdot \sum_{j=1}^{d} \sum_{k=1}^{n} \operatorname{Var}\left(X_k^j\right) \\
&= \frac{1}{n} \sum_{j=1}^{d} x_j (1 - x_j) \le \frac{d}{4n}.
\end{aligned}$$

This shows $S_n/n \to x$ in $L^2(P)$ and hence by Chebyshev's inequality, $S_n/n \xrightarrow{P} x$ in and by a continuity theorem, $f\left(\frac{S_n}{n}\right) \xrightarrow{P} f(x)$ as $n \to \infty$. This along with the dominated convergence theorem shows

$$p_n(x) := \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \to f(x) \text{ as } n \to \infty, \tag{4.56}$$

where

$$p_n(x) = \sum_{\eta:\{1,2,\dots,n\}\to\{0,1\}^d} f\left(\frac{\eta(1)+\dots+\eta(n)}{n}\right) P(X_1=\eta(1),\dots,X_n=\eta(n))$$
$$= \sum_{\eta:\{1,2,\dots,n\}\to\{0,1\}^d} f\left(\frac{\eta(1)+\dots+\eta(n)}{n}\right) \prod_{k=1}^{n}\prod_{i=1}^{d}(1-x_i)^{1-\eta_i(k)} x_i^{\eta_i(k)}$$

is a polynomial of degree nd. In fact more is true.

Suppose $\varepsilon>0$ is given, $M=\sup\{|f(x)| : x\in K\}$, and

$$\delta_\varepsilon = \sup\{|f(y)-f(x)| : x,y\in K \text{ and } \|y-x\|\le\varepsilon\}.$$

By uniform continuity of f on K, $\lim_{\varepsilon\downarrow 0}\delta_\varepsilon = 0$. Therefore,

$$\begin{aligned}|f(x)-p_n(x)| &= \left|\mathbb{E}\left(f(x)-f\left(\frac{S_n}{n}\right)\right)\right| \le \mathbb{E}\left|f(x)-f\left(\frac{S_n}{n}\right)\right| \\ &\le \mathbb{E}\left[\left|f(x)-f\left(\frac{S_n}{n}\right)\right| : \|S_n-x\|>\varepsilon\right] \\ &\quad + \mathbb{E}\left[\left|f(x)-f\left(\frac{S_n}{n}\right)\right| : \|S_n-x\|\le\varepsilon\right] \\ &\le 2MP(\|S_n-x\|>\varepsilon)+\delta_\varepsilon. \end{aligned} \tag{4.57}$$

By Chebyshev's inequality,

$$P(\|S_n-x\|>\varepsilon) \le \frac{1}{\varepsilon^2}\mathbb{E}\|S_n-x\|^2 = \frac{d}{4n\varepsilon^2},$$

and therefore, Eq. (4.57) yields the estimate

$$\sup_{x\in K}|f(x)-p_n(x)| \le \frac{2dM}{n\varepsilon^2}+\delta_\varepsilon$$

and hence

$$\limsup_{n\to\infty}\sup_{x\in K}|f(x)-p_n(x)| \le \delta_\varepsilon \to 0 \text{ as } \varepsilon\downarrow 0.$$

■

Here is a version of the complex Weirstrass approximation theorem.

Theorem 4.51 (Complex Weierstrass Approximation Theorem). *Suppose that $K\subset\mathbb{C}^d\cong\mathbb{R}^d\times\mathbb{R}^d$ is a compact rectangle. Then there exists polynomials in $(z=x+iy, \bar z = x-iy)$, $p_n(z,\bar z)$ for $z\in\mathbb{C}^d$, such that $\sup_{z\in K}|q_n(z,\bar z)-f(z)|\to 0$ as $n\to\infty$ for every $f\in C(K,\mathbb{C})$.*

Proof. The mapping $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d\to z=x+iy\in\mathbb{C}^d$ is an isomorphism of vector spaces. Letting $\bar z = x-iy$ as usual, we have $x=\frac{z+\bar z}{2}$ and $y=\frac{z-\bar z}{2i}$. Therefore under this identification any polynomial $p(x,y)$ on $\mathbb{R}^d\times\mathbb{R}^d$ may be written as a polynomial q in $(z,\bar z)$, namely

$$q(z,\bar z) = p\left(\frac{z+\bar z}{2},\frac{z-\bar z}{2i}\right).$$

Conversely a polynomial q in $(z,\bar z)$ may be thought of as a polynomial p in (x,y), namely $p(x,y)=q(x+iy,x-iy)$. Hence the result now follows from Theorem 4.50. ■

Example 4.52. Let $K=S^1=\{z\in\mathbb{C}:|z|=1\}$ and $\mathcal{A}$ be the set of polynomials in $(z,\bar z)$ restricted to S^1. Then $\mathcal{A}$ is dense in $C(S^1)$. To prove this first observe if $f\in C(S^1)$ then $F(z)=|z|\,f(\frac{z}{|z|})$ for $z\ne 0$ and $F(0)=0$ defines $F\in C(\mathbb{C})$ such that $F|_{S^1}=f$. By applying Theorem 4.51 to F restricted to a compact rectangle containing S^1 we may find $q_n(z,\bar z)$ converging uniformly to F on K and hence on S^1. Since $\bar z = z^{-1}$ on S^1, we have shown polynomials in z and z^{-1} are dense in $C(S^1)$. This example generalizes in an obvious way to $K=(S^1)^d\subset\mathbb{C}^d$.

Exercise 4.16. Use Example 4.52 to show that any 2π – periodic continuous function, $g:\mathbb{R}^d\to\mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form

$$p(x)=\sum_{\lambda\in\Lambda}a_\lambda e^{i\lambda\cdot x}$$

where Λ is a finite subset of $\mathbb{Z}^d$ and $a_\lambda\in\mathbb{C}$ for all $\lambda\in\Lambda$. **Hint:** start by showing there exists a unique continuous function, $f:(S^1)^d\to\mathbb{C}$ such that $f(e^{ix_1},\dots,e^{ix_d})=F(x)$ for all $x=(x_1,\dots,x_d)\in\mathbb{R}^d$.

Solution to Exercise (4.16). I will write out the solution when $d=1$. For $z\in S^1$, define $F(z):=f(e^{i\theta})$ where $\theta\in\mathbb{R}$ is chosen so that $z=e^{i\theta}$. Since f is 2π – periodic, F is well defined since if θ solves $e^{i\theta}=z$ then all other solutions are of the form $\{\theta+2\pi n : n\in\mathbb{Z}\}$. Since the map $\theta\to e^{i\theta}$ is a local homeomorphism, i.e. for any $J=(a,b)$ with $b-a<2\pi$, the map $\theta\in J\xrightarrow{\phi}\tilde J := \{e^{i\theta}:\theta\in J\}\subset S^1$ is a homeomorphism, it follows that $F(z)=f\circ\phi^{-1}(z)$ for $z\in\tilde J$. This shows F is continuous when restricted to $\tilde J$. Since such sets cover S^1, it follows that F is continuous. It now follows from Example 4.52 that polynomials in z and z^{-1} are dense in $C(S^1)$. Hence for any $\varepsilon>0$ there exists

$$p(z,\bar z)=\sum a_{m,n}z^m\bar z^n=\sum a_{m,n}z^m z^{-n}=\sum a_{m,n}z^{m-n}$$

such that $|F(z)-p(z,\bar z)|\le\varepsilon$ for all z. Taking $z=e^{i\theta}$ then implies there exists $b_n\in\mathbb{C}$ and $N\in\mathbb{N}$ such that

$$p_{\varepsilon}(\theta) := \sum_{n=-N}^{N} b_n e^{in\theta} \tag{4.58}$$

satisfies

$$\sup_{\theta} \left|\bar{f}(\theta) - p(\theta)\right| \le \varepsilon.$$

Exercise 4.17. Suppose $f \in C(\mathbb{R}, \mathbb{C})$ is a 2π – periodic function (i.e. $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$) and

$$\int_0^{2\pi} f(x) e^{inx} dx = 0 \text{ for all } n \in \mathbb{Z},$$

show again that $f \equiv 0$. **Hint**: Use Exercise 4.16.

Solution to Exercise (4.17). By assumption, $\int_0^{2\pi} f(\theta) e^{in\theta} d\theta = 0$ for all n and so by the linearity of the Riemann integral,

$$0 = \int_0^{2\pi} f(\theta) p_{\varepsilon}(\theta) d\theta. \tag{4.59}$$

Choose trigonometric polynomials, p_{ε}, as in Eq. (4.58) such that $p_{\varepsilon}(\theta) \to \bar{f}(\theta)$ uniformly in θ as $\varepsilon \downarrow 0$. Passing to the limit in Eq. (4.59) implies

$$0 = \lim_{\varepsilon \downarrow 0} \int_0^{2\pi} f(\theta) p_{\varepsilon}(\theta) d\theta = \int_0^{2\pi} f(\theta) \bar{f}(\theta) d\theta = \int_0^{2\pi} |f(\theta)|^2 d\theta.$$

From this it follows that $f \equiv 0$, for if $|f(\theta_0)| > 0$ for some θ_0 then $|f(\theta)| \ge \varepsilon > 0$ for θ in a neighborhood of θ_0 by continuity of f. It would then follow that $\int_0^{2\pi} |f(\theta)|^2 d\theta > 0$.

Countably Additive Measures

Let $\mathcal{A} \subset 2^{\Omega}$ be an algebra and $\mu : \mathcal{A} \to [0, \infty]$ be a finitely additive measure. Recall that μ is a **premeasure** on $\mathcal{A}$ if μ is σ – additive on $\mathcal{A}$. If μ is a premeasure on $\mathcal{A}$ and $\mathcal{A}$ is a σ – algebra (Definition 3.12), we say that μ is a **measure** on $(\Omega, \mathcal{A})$ and that $(\Omega, \mathcal{A})$ is a **measurable space.**

Definition 5.1. *Let $(\Omega, \mathcal{B})$ be a measurable space. We say that $P : \mathcal{B} \to [0,1]$ is a **probability measure on** $(\Omega, \mathcal{B})$ if P is a measure on $\mathcal{B}$ such that $P(\Omega) = 1$. In this case we say that $(\Omega, \mathcal{B}, P)$ a probability space.*

5.1 Overview

The goal of this chapter is develop methods for proving the existence of probability measures with desirable properties. The main results of this chapter may are summarized in the following theorem.

Theorem 5.2. *A finitely additive probability measure P on an algebra, $\mathcal{A} \subset 2^{\Omega}$, extends to σ – additive measure on $\sigma(\mathcal{A})$ iff P is a premeasure on $\mathcal{A}$. If the extension exists it is unique.*

Proof. The uniqueness assertion is proved Proposition 5.15 below. The existence assertion of the theorem in the content of Theorem 5.27. ■

In order to use this theorem it is necessary to determine when a finitely additive probability measure in is in fact a premeasure. The following Proposition is sometimes useful in this regard.

Proposition 5.3 (Equivalent premeasure conditions). *Suppose that P is a finitely additive probability measure on an algebra, $\mathcal{A} \subset 2^{\Omega}$. Then the following are equivalent:*

1. *P is a premeasure on $\mathcal{A}$, i.e. P is σ – additive on $\mathcal{A}$.*
2. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, $P(A_n) \uparrow P(A)$.*
3. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, $P(A_n) \downarrow P(A)$.*
4. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow \Omega$, $P(A_n) \uparrow 1$.*
5. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow \emptyset$, $P(A_n) \downarrow 0$.*

Proof. We will start by showing $1 \iff 2 \iff 3$.

$1. \implies 2.$ Suppose $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Let $A_n' := A_n \setminus A_{n-1}$ with $A_0 := \emptyset$. Then $\{A_n'\}_{n=1}^{\infty}$ are disjoint, $A_n = \cup_{k=1}^n A_k'$ and $A = \cup_{k=1}^{\infty} A_k'$. Therefore,

$$P(A) = \sum_{k=1}^{\infty} P(A_k') = \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k') = \lim_{n\to\infty} P\left(\cup_{k=1}^n A_k'\right) = \lim_{n\to\infty} P(A_n).$$

$2. \implies 1.$ If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ are disjoint and $A := \cup_{n=1}^{\infty} A_n \in \mathcal{A}$, then $\cup_{n=1}^N A_n \uparrow A$. Therefore,

$$P(A) = \lim_{N\to\infty} P\left(\cup_{n=1}^N A_n\right) = \lim_{N\to\infty} \sum_{n=1}^{N} P(A_n) = \sum_{n=1}^{\infty} P(A_n).$$

$2. \implies 3.$ If $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, then $A_n^c \uparrow A^c$ and therefore,

$$\lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n^c) = P(A^c) = 1 - P(A).$$

$3. \implies 2.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A_n^c \downarrow A^c$ and therefore we again have,

$$\lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n^c) = P(A^c) = 1 - P(A).$$

The same proof used for $2. \iff 3.$ shows $4. \iff 5$ and it is clear that $3. \implies 5.$ To finish the proof we will show $5. \implies 2.$

$5. \implies 2.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A \setminus A_n \downarrow \emptyset$ and therefore

$$\lim_{n\to\infty} [P(A) - P(A_n)] = \lim_{n\to\infty} P(A \setminus A_n) = 0.$$

■

Remark 5.4. Observe that the equivalence of items 1. and 2. in the above proposition hold without the restriction that $P(\Omega) = 1$ and in fact $P(\Omega) = \infty$ may be allowed for this equivalence.

Lemma 5.5. *If $\mu : \mathcal{A} \to [0, \infty]$ is a premeasure, then μ is countably sub-additive on $\mathcal{A}$.*

Proof. Suppose that $A_n \in \mathcal{A}$ with $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$. Let $A_1^{'} := A_1$ and for $n \geq 2$, let $A_n' := A_n \setminus (A_1 \cup \dots A_{n-1}) \in \mathcal{A}$. Then $\cup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} A_n'$ and therefore by the countable additivity and monotonicity of μ we have,

$$\mu\left(\cup_{n=1}^{\infty} A_n\right) = \mu\left(\sum_{n=1}^{\infty} A_n'\right) = \sum_{n=1}^{\infty} \mu\left(A_n'\right) \leq \sum_{n=1}^{\infty} \mu\left(A_n\right).$$

■

Let us now specialize to the case where $\Omega = \mathbb{R}$ and $\mathcal{A} = \mathcal{A}\left(\{(a,b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\}\right)$. In this case we will describe probability measures, P, on $\mathcal{B}_{\mathbb{R}}$ by their "cumulative distribution functions."

Definition 5.6. *Given a probability measure, P on $\mathcal{B}_{\mathbb{R}}$, the **cumulative distribution function** (CDF) of P is defined as the function, $F = F_P : \mathbb{R} \to [0,1]$ given as*

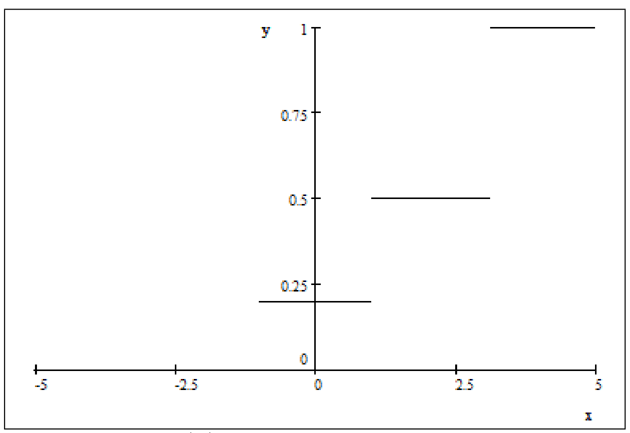
$$F(x) := P((-\infty, x]). \tag{5.1}$$

Example 5.7. Suppose that

$$P = p\delta_{-1} + q\delta_1 + r\delta_\pi$$

with $p, q, r > 0$ and $p + q + r = 1$. In this case,

$$F(x) = \begin{cases} 0 & \text{for } \quad x < -1 \\ p & \text{for } -1 \leq x < 1 \\ p+q & \text{for } \ 1 \leq x < \pi \\ 1 & \text{for } \pi \leq x < \infty \end{cases}.$$



A plot of $F(x)$ with $p = .2$, $q = .3$, and $r = .5$.

Lemma 5.8. *If $F = F_P : \mathbb{R} \to [0,1]$ is a distribution function for a probability measure, P, on $\mathcal{B}_{\mathbb{R}}$, then:*

1. *F is non-decreasing,*
2. *F is right continuous,*
3. *$F(-\infty) := \lim_{x \to -\infty} F(x) = 0$, and $F(\infty) := \lim_{x \to \infty} F(x) = 1$.*

Proof. The monotonicity of P shows that $F(x)$ in Eq. (5.1) is non-decreasing. For $b \in \mathbb{R}$ let $A_n = (-\infty, b_n]$ with $b_n \downarrow b$ as $n \to \infty$. The continuity of P implies

$$F(b_n) = P((-\infty, b_n]) \downarrow \mu((-\infty, b]) = F(b).$$

Since $\{b_n\}_{n=1}^{\infty}$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $F(b+) := \lim_{y \downarrow b} F(y) = F(b)$. This show that F is right continuous. Similar arguments show that $F(\infty) = 1$ and $F(-\infty) = 0$. ■

It turns out that Lemma 5.8 has the following important converse.

Theorem 5.9. *To each function $F : \mathbb{R} \to [0,1]$ satisfying properties 1. – 3.. in Lemma 5.8, there exists a unique probability measure, P_F, on $\mathcal{B}_{\mathbb{R}}$ such that*

$$P_F((a,b]) = F(b) - F(a) \ \text{ for all } \ -\infty < a \leq b < \infty.$$

Proof. The uniqueness assertion is proved in Corollary 5.17 below or see Exercises 5.2 and 5.11 below. The existence portion of the theorem is a special case of Theorem 5.33 below. ■

Example 5.10 (Uniform Distribution). The function,

$$F(x) := \begin{cases} 0 \text{ for } & x \leq 0 \\ x \text{ for } & 0 \leq x < 1 \\ 1 \text{ for } & 1 \leq x < \infty \end{cases},$$

is the distribution function for a measure, m on $\mathcal{B}_{\mathbb{R}}$ which is concentrated on $(0,1]$. The measure, m is called the **uniform distribution** or **Lebesgue measure** on $(0,1]$.

With this summary in hand, let us now start the formal development. We begin with uniqueness statement in Theorem 5.2.

5.2 $\pi - \lambda$ Theorem

Recall that a collection, $\mathcal{P} \subset 2^{\Omega}$, is a π – **class** or π – system if it is closed under finite intersections. We also need the notion of a λ –system.

Definition 5.11 (λ – system). *A collection of sets, $\mathcal{L} \subset 2^{\Omega}$, is λ – **class** or λ – **system** if*

a. $\Omega \in \mathcal{L}$

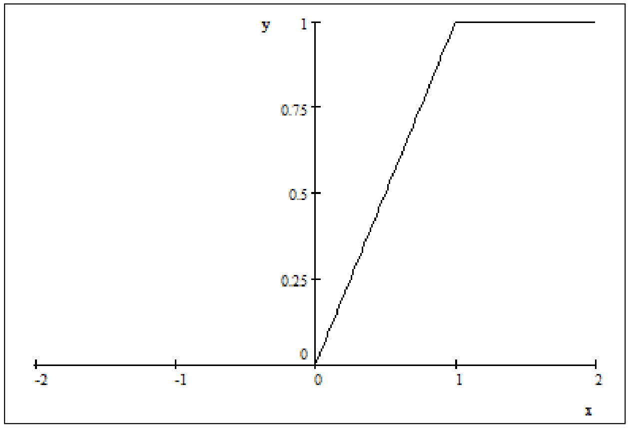


Fig. 5.1. The cumulative distribution function for the uniform distribution.

b. *If $A, B \in \mathcal{L}$ and $A \subset B$, then $B \setminus A \in \mathcal{L}$. (Closed under proper differences.)*
c. *If $A_n \in \mathcal{L}$ and $A_n \uparrow A$, then $A \in \mathcal{L}$. (Closed under countable increasing unions.)*

Remark 5.12. If $\mathcal{L}$ is a collection of subsets of Ω which is both a λ – class and a π – system then $\mathcal{L}$ is a σ – algebra. Indeed, since $A^c = \Omega \setminus A$, we see that any λ - system is closed under complementation. If $\mathcal{L}$ is also a π – system, it is closed under intersections and therefore $\mathcal{L}$ is an algebra. Since $\mathcal{L}$ is also closed under increasing unions, $\mathcal{L}$ is a σ – algebra.

Lemma 5.13 (Alternate Axioms for a λ – System*). *Suppose that $\mathcal{L} \subset 2^\Omega$ is a collection of subsets Ω. Then $\mathcal{L}$ is a λ – class iff λ satisfies the following postulates:*

1. *$\Omega \in \mathcal{L}$*
2. *$A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$. (Closed under complementation.)*
3. *If $\{A_n\}_{n=1}^\infty \subset \mathcal{L}$ are disjoint, then $\sum_{n=1}^\infty A_n \in \mathcal{L}$. (Closed under disjoint unions.)*

Proof. Suppose that $\mathcal{L}$ satisfies a. – c. above. Clearly then postulates 1. and 2. hold. Suppose that $A, B \in \mathcal{L}$ such that $A \cap B = \emptyset$, then $A \subset B^c$ and

$$A^c \cap B^c = B^c \setminus A \in \mathcal{L}.$$

Taking complements of this result shows $A \cup B \in \mathcal{L}$ as well. So by induction, $B_m := \sum_{n=1}^m A_n \in \mathcal{L}$. Since $B_m \uparrow \sum_{n=1}^\infty A_n$ it follows from postulate c. that $\sum_{n=1}^\infty A_n \in \mathcal{L}$.

Now suppose that $\mathcal{L}$ satisfies postulates 1. – 3. above. Notice that $\emptyset \in \mathcal{L}$ and by postulate 3., $\mathcal{L}$ is closed under finite disjoint unions. Therefore if $A, B \in \mathcal{L}$ with $A \subset B$, then $B^c \in \mathcal{L}$ and $A \cap B^c = \emptyset$ allows us to conclude that $A \cup B^c \in \mathcal{L}$. Taking complements of this result shows $B \setminus A = A^c \cap B \in \mathcal{L}$ as well, i.e. postulate b. holds. If $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $B_n := A_n \setminus A_{n-1} \in \mathcal{L}$ for all n, where by convention $A_0 = \emptyset$. Hence it follows by postulate 3 that $\cup_{n=1}^\infty A_n = \sum_{n=1}^\infty B_n \in \mathcal{L}$. ■

Theorem 5.14 (Dynkin's $\pi - \lambda$ Theorem). *If $\mathcal{L}$ is a λ class which contains a contains a π – class, $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. We start by proving the following assertion; for any element $C \in \mathcal{L}$, the collection of sets,

$$\mathcal{L}^C := \{D \in \mathcal{L} : C \cap D \in \mathcal{L}\},$$

is a λ – system. To prove this claim, observe that: a. $\Omega \in \mathcal{L}^C$, b. if $A \subset B$ with $A, B \in \mathcal{L}^C$, then $A \cap C$, $B \cap C \in \mathcal{L}$ with $A \cap C \subset B \cap C$ and therefore,

$$(B \setminus A) \cap C = [B \cap C] \setminus A = [B \cap C] \setminus [A \cap C] \in \mathcal{L}.$$

This shows that $\mathcal{L}^C$ is closed under proper differences. c. If $A_n \in \mathcal{L}^C$ with $A_n \uparrow A$, then $A_n \cap C \in \mathcal{L}$ and $A_n \cap C \uparrow A \cap C \in \mathcal{L}$, i.e. $A \in \mathcal{L}^C$. Hence we have verified $\mathcal{L}^C$ is still a λ – system.

For the rest of the proof, we may assume without loss of generality that $\mathcal{L}$ is the smallest λ – class containing $\mathcal{P}$ – if not just replace $\mathcal{L}$ by the intersection of all λ – classes containing $\mathcal{P}$. Then for $C \in \mathcal{P}$ we know that $\mathcal{L}^C \subset \mathcal{L}$ is a λ - class containing $\mathcal{P}$ and hence $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{P}$ was arbitrary, we have shown, $C \cap D \in \mathcal{L}$ for all $C \in \mathcal{P}$ and $D \in \mathcal{L}$. We may now conclude that if $C \in \mathcal{L}$, then $\mathcal{P} \subset \mathcal{L}^C \subset \mathcal{L}$ and hence again $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{L}$ is arbitrary, we have shown $C \cap D \in \mathcal{L}$ for all $C, D \in \mathcal{L}$, i.e. $\mathcal{L}$ is a π – system. So by Remark 5.12, $\mathcal{L}$ is a σ algebra. Since $\sigma(\mathcal{P})$ is the smallest σ – algebra containing $\mathcal{P}$ it follows that $\sigma(\mathcal{P}) \subset \mathcal{L}$. ■

As an immediate corollary, we have the following uniqueness result.

Proposition 5.15. *Suppose that $\mathcal{P} \subset 2^\Omega$ is a π – system. If P and Q are two probability*[1] *measures on $\sigma(\mathcal{P})$ such that $P = Q$ on $\mathcal{P}$, then $P = Q$ on $\sigma(\mathcal{P})$.*

Proof. Let $\mathcal{L} := \{A \in \sigma(\mathcal{P}) : P(A) = Q(A)\}$. One easily shows $\mathcal{L}$ is a λ – class which contains $\mathcal{P}$ by assumption. Indeed, $\Omega \in \mathcal{P} \subset \mathcal{L}$, if $A, B \in \mathcal{L}$ with $A \subset B$, then

$$P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$$

[1] More generally, P and Q could be two measures such that $P(\Omega) = Q(\Omega) < \infty$.

so that $B \setminus A \in \mathcal{L}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} Q(A_n) = Q(A)$ which shows $A \in \mathcal{L}$. Therefore $\sigma(\mathcal{P}) \subset \mathcal{L} = \sigma(\mathcal{P})$ and the proof is complete. ■

Example 5.16. Let $\Omega := \{a,b,c,d\}$ and let μ and ν be the probability measure on 2^Ω determined by, $\mu(\{x\}) = \frac{1}{4}$ for all $x \in \Omega$ and $\nu(\{a\}) = \nu(\{d\}) = \frac{1}{8}$ and $\nu(\{b\}) = \nu(\{c\}) = 3/8$. In this example,

$$\mathcal{L} := \left\{A \in 2^\Omega : P(A) = Q(A)\right\}$$

is λ – system which is not an algebra. Indeed, $A = \{a,b\}$ and $B = \{a,c\}$ are in $\mathcal{L}$ but $A \cap B \notin \mathcal{L}$.

Exercise 5.1. Suppose that μ and ν are two measures (not assumed to be finite) on a measure space, $(\Omega, \mathcal{B})$ such that $\mu = \nu$ on a π – system, $\mathcal{P}$. Further assume $\mathcal{B} = \sigma(\mathcal{P})$ and there exists $\Omega_n \in \mathcal{P}$ such that; i) $\mu(\Omega_n) = \nu(\Omega_n) < \infty$ for all n and ii) $\Omega_n \uparrow \Omega$ as $n \uparrow \infty$. Show $\mu = \nu$ on $\mathcal{B}$.

Hint: Consider the measures, $\mu_n(A) := \mu(A \cap \Omega_n)$ and $\nu_n(A) = \nu(A \cap \Omega_n)$.

Solution to Exercise (5.1). Let $\mu_n(A) := \mu(A \cap \Omega_n)$ and $\nu_n(A) = \nu(A \cap \Omega_n)$ for all $A \in \mathcal{B}$. Then μ_n and ν_n are finite measure such $\mu_n(\Omega) = \nu_n(\Omega)$ and $\mu_n = \nu_n$ on $\mathcal{P}$. Therefore by Proposition 5.15, $\mu_n = \nu_n$ on $\mathcal{B}$. So by the continuity properties of μ and ν, it follows that

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \mu_n(A) = \lim_{n\to\infty} \nu_n(A) = \lim_{n\to\infty} \nu(A \cap \Omega_n) = \nu(A)$$

for all $A \in \mathcal{B}$.

Corollary 5.17. *A probability measure, P, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is uniquely determined by its cumulative distribution function,*

$$F(x) := P((-\infty, x]).$$

Proof. This follows from Proposition 5.15 wherein we use the fact that $\mathcal{P} := \{(-\infty, x] : x \in \mathbb{R}\}$ is a π – system such that $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{P})$. ■

Remark 5.18. Corollary 5.17 generalizes to $\mathbb{R}^n$. Namely a probability measure, P, on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is uniquely determined by its CDF,

$$F(x) := P((-\infty, x]) \text{ for all } x \in \mathbb{R}^n$$

where now

$$(-\infty, x] := (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n].$$

5.2.1 A Density Result*

Exercise 5.2 (Density of $\mathcal{A}$ in $\sigma(\mathcal{A})$). Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} := \sigma(\mathcal{A})$, and P is a probability measure on $\mathcal{B}$. Let $\rho(A,B) := P(A \Delta B)$. The goal of this exercise is to use the $\pi - \lambda$ theorem to show that $\mathcal{A}$ is dense in $\mathcal{B}$ relative to the "metric," ρ. More precisely you are to show using the following outline that for every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that that $P(A \triangle B) < \varepsilon$.

1. Recall from Exercise 4.3 that $\rho(a, B) = P(A \triangle B) = \mathbb{E}|1_A - 1_B|$.
2. Observe; if $B = \cup B_i$ and $A = \cup_i A_i$, then
$$\begin{aligned} B \setminus A &= \cup_i [B_i \setminus A] \subset \cup_i (B_i \setminus A_i) \subset \cup_i A_i \triangle B_i \text{ and} \\ A \setminus B &= \cup_i [A_i \setminus B] \subset \cup_i (A_i \setminus B_i) \subset \cup_i A_i \triangle B_i \end{aligned}$$
so that
$$A \triangle B \subset \cup_i (A_i \triangle B_i).$$
3. We also have
$$\begin{aligned} (B_2 \setminus B_1) \setminus (A_2 \setminus A_1) &= B_2 \cap B_1^c \cap (A_2 \setminus A_1)^c \\ &= B_2 \cap B_1^c \cap (A_2 \cap A_1^c)^c \\ &= B_2 \cap B_1^c \cap (A_2^c \cup A_1) \\ &= [B_2 \cap B_1^c \cap A_2^c] \cup [B_2 \cap B_1^c \cap A_1] \\ &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1) \end{aligned}$$
and similarly,
$$(A_2 \setminus A_1) \setminus (B_2 \setminus B_1) \subset (A_2 \setminus B_2) \cup (B_1 \setminus A_1)$$
so that
$$\begin{aligned} (A_2 \setminus A_1) \triangle (B_2 \setminus B_1) &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1) \cup (A_2 \setminus B_2) \cup (B_1 \setminus A_1) \\ &= (A_1 \triangle B_1) \cup (A_2 \triangle B_2). \end{aligned}$$
4. Observe that $A_n \in \mathcal{B}$ and $A_n \uparrow A$, then
$$\begin{aligned} P(B \triangle A_n) &= P(B \setminus A_n) + P(A_n \setminus B) \\ &\to P(B \setminus A) + P(A \setminus B) = P(A \triangle B). \end{aligned}$$
5. Let $\mathcal{L}$ be the collection of sets $B \in \mathcal{B}$ for which the assertion of the theorem holds. Show $\mathcal{L}$ is a λ – system which contains $\mathcal{A}$.

Solution to Exercise (5.2). Since $\mathcal{L}$ contains the π – system, $\mathcal{A}$ it suffices by the $\pi - \lambda$ theorem to show $\mathcal{L}$ is a λ – system. Clearly, $\Omega \in \mathcal{L}$ since $\Omega \in \mathcal{A} \subset \mathcal{L}$. If $B_1 \subset B_2$ with $B_i \in \mathcal{L}$ and $\varepsilon > 0$, there exists $A_i \in \mathcal{A}$ such that $P(B_i \triangle A_i) = \mathbb{E}_P |1_{A_i} - 1_{B_i}| < \varepsilon/2$ and therefore,

$$\begin{aligned} P((B_2 \setminus B_1) \triangle (A_2 \setminus A_1)) &\le P((A_1 \triangle B_1) \cup (A_2 \triangle B_2)) \\ &\le P((A_1 \triangle B_1)) + P((A_2 \triangle B_2)) < \varepsilon. \end{aligned}$$

Also if $B_n \uparrow B$ with $B_n \in \mathcal{L}$, there exists $A_n \in \mathcal{A}$ such that $P(B_n \triangle A_n) < \varepsilon 2^{-n}$ and therefore,

$$P([\cup_n B_n] \triangle [\cup_n A_n]) \le \sum_{n=1}^{\infty} P(B_n \triangle A_n) < \varepsilon.$$

Moreover, if we let $B := \cup_n B_n$ and $A^N := \cup_{n=1}^N A_n$, then

$$P\left(B \triangle A^N\right) = P\left(B \setminus A^N\right) + P\left(A^N \setminus B\right) \to P(B \setminus A) + P(A \setminus B) = P(B \triangle A)$$

where $A := \cup_n A_n$. Hence it follows for N large enough that $P\left(B \triangle A^N\right) < \varepsilon$. Since $\varepsilon > 0$ was arbitrary we have shown $B \in \mathcal{L}$ as desired.

5.3 Construction of Measures

Definition 5.19. *Given a collection of subsets, $\mathcal{E}$, of Ω, let $\mathcal{E}_\sigma$ denote the collection of subsets of Ω which are finite or countable unions of sets from $\mathcal{E}$. Similarly let $\mathcal{E}_\delta$ denote the collection of subsets of Ω which are finite or countable intersections of sets from $\mathcal{E}$. We also write $\mathcal{E}_{\sigma\delta} = (\mathcal{E}_\sigma)_\delta$ and $\mathcal{E}_{\delta\sigma} = (\mathcal{E}_\delta)_\sigma$, etc.*

Lemma 5.20. *Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra. Then:*

1. *$\mathcal{A}_\sigma$ is closed under taking countable unions and finite intersections.*
2. *$\mathcal{A}_\delta$ is closed under taking countable intersections and finite unions.*
3. *$\{A^c : A \in \mathcal{A}_\sigma\} = \mathcal{A}_\delta$ and $\{A^c : A \in \mathcal{A}_\delta\} = \mathcal{A}_\sigma$.*

Proof. By construction $\mathcal{A}_\sigma$ is closed under countable unions. Moreover if $A = \cup_{i=1}^\infty A_i$ and $B = \cup_{j=1}^\infty B_j$ with $A_i, B_j \in \mathcal{A}$, then

$$A \cap B = \cup_{i,j=1}^\infty A_i \cap B_j \in \mathcal{A}_\sigma,$$

which shows that $\mathcal{A}_\sigma$ is also closed under finite intersections. Item 3. is straight forward and item 2. follows from items 1. and 3. ■

Remark 5.21. Let us recall from Proposition 5.3 and Remark 5.4 that a finitely additive measure $\mu : \mathcal{A} \to [0, \infty]$ is a premeasure on $\mathcal{A}$ iff $\mu(A_n) \uparrow \mu(A)$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Furthermore if $\mu(\Omega) < \infty$, then μ is a premeasure on $\mathcal{A}$ iff $\mu(A_n) \downarrow 0$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \downarrow \emptyset$.

Proposition 5.22. *Given a premeasure, $\mu : \mathcal{A} \to [0, \infty]$, we extend μ to $\mathcal{A}_\sigma$ by defining*

$$\mu(B) := \sup\{\mu(A) : \mathcal{A} \ni A \subset B\}. \tag{5.2}$$

This function $\mu : \mathcal{A}_\sigma \to [0, \infty]$ then satisfies;

1. ***(Monotonicity)** If $A, B \in \mathcal{A}_\sigma$ with $A \subset B$ then $\mu(A) \le \mu(B)$.*
2. ***(Continuity)** If $A_n \in \mathcal{A}$ and $A_n \uparrow A \in \mathcal{A}_\sigma$, then $\mu(A_n) \uparrow \mu(A)$ as $n \to \infty$.*
3. ***(Strong Additivity)** If $A, B \in \mathcal{A}_\sigma$, then*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{5.3}$$
4. ***(Sub-Additivity on $\mathcal{A}_\sigma$)** The function μ is sub-additive on $\mathcal{A}_\sigma$, i.e. if $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$, then*
$$\mu(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{5.4}$$
5. *(σ - Additivity **on** $\mathcal{A}_\sigma$) The function μ is countably additive on $\mathcal{A}_\sigma$.*

Proof. 1. and 2. Monotonicity follows directly from Eq. (5.2) which then implies $\mu(A_n) \le \mu(B)$ for all n. Therefore $M := \lim_{n\to\infty} \mu(A_n) \le \mu(B)$. To prove the reverse inequality, let $\mathcal{A} \ni A \subset B$. Then by the continuity of μ on $\mathcal{A}$ and the fact that $A_n \cap A \uparrow A$ we have $\mu(A_n \cap A) \uparrow \mu(A)$. As $\mu(A_n) \ge \mu(A_n \cap A)$ for all n it then follows that $M := \lim_{n\to\infty} \mu(A_n) \ge \mu(A)$. As $A \in \mathcal{A}$ with $A \subset B$ was arbitrary we may conclude,

$$\mu(B) = \sup\{\mu(A) : \mathcal{A} \ni A \subset B\} \le M.$$

3. Suppose that $A, B \in \mathcal{A}_\sigma$ and $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ are sequences in $\mathcal{A}$ such that $A_n \uparrow A$ and $B_n \uparrow B$ as $n \to \infty$. Then passing to the limit as $n \to \infty$ in the identity,

$$\mu(A_n \cup B_n) + \mu(A_n \cap B_n) = \mu(A_n) + \mu(B_n)$$

proves Eq. (5.3). In particular, it follows that μ is finitely additive on $\mathcal{A}_\sigma$.

4 and 5. Let $\{A_n\}_{n=1}^\infty$ be any sequence in $\mathcal{A}_\sigma$ and choose $\{A_{n,i}\}_{i=1}^\infty \subset \mathcal{A}$ such that $A_{n,i} \uparrow A_n$ as $i \to \infty$. Then we have,

$$\mu\left(\cup_{n=1}^N A_{n,N}\right) \le \sum_{n=1}^N \mu(A_{n,N}) \le \sum_{n=1}^N \mu(A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{5.5}$$

Since $\mathcal{A} \ni \cup_{n=1}^N A_{n,N} \uparrow \cup_{n=1}^\infty A_n \in \mathcal{A}_\sigma$, we may let $N \to \infty$ in Eq. (5.5) to conclude Eq. (5.4) holds. If we further assume that $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$ are pairwise disjoint, by the finite additivity and monotonicity of μ on $\mathcal{A}_\sigma$, we have

$$\sum_{n=1}^\infty \mu(A_n) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n) = \lim_{N\to\infty} \mu\left(\cup_{n=1}^N A_n\right) \le \mu\left(\cup_{n=1}^\infty A_n\right).$$

This inequality along with Eq. (5.4) shows that μ is σ – additive on $\mathcal{A}_\sigma$. ■

Suppose μ is a **finite** premeasure on an algebra, $\mathcal{A} \subset 2^\Omega$, and $A \in \mathcal{A}_\delta \cap \mathcal{A}_\sigma$. Since $A, A^c \in \mathcal{A}_\sigma$ and $\Omega = A \cup A^c$, it follows that $\mu(\Omega) = \mu(A) + \mu(A^c)$. From this observation we may extend μ to a function on $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ by defining

$$\mu(A) := \mu(\Omega) - \mu(A^c) \text{ for all } A \in \mathcal{A}_\delta. \tag{5.6}$$

Lemma 5.23. *Suppose μ is a finite premeasure on an algebra, $\mathcal{A} \subset 2^\Omega$, and μ has been extended to $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ as described in Proposition 5.22 and Eq. (5.6) above.*

1. *If $A \in \mathcal{A}_\delta$ then $\mu(A) = \inf\{\mu(B) : A \subset B \in \mathcal{A}\}$.*
2. *If $A \in \mathcal{A}_\delta$ and $A_n \in \mathcal{A}$ such that $A_n \downarrow A$, then $\mu(A) = \downarrow \lim_{n\to\infty} \mu(A_n)$.*
3. *μ is strongly additive when restricted to $\mathcal{A}_\delta$.*
4. *If $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset C$, then $\mu(C \setminus A) = \mu(C) - \mu(A)$.*

Proof.

1. Since $\mu(B) = \mu(\Omega) - \mu(B^c)$ and $A \subset B$ iff $B^c \subset A^c$, it follows that
$$\begin{aligned}\inf\{\mu(B) : A \subset B \in \mathcal{A}\} &= \inf\{\mu(\Omega) - \mu(B^c) : \mathcal{A} \ni B^c \subset A^c\} \\ &= \mu(\Omega) - \sup\{\mu(B) : \mathcal{A} \ni B \subset A^c\} \\ &= \mu(\Omega) - \mu(A^c) = \mu(A).\end{aligned}$$
2. Similarly, since $A_n^c \uparrow A^c \in \mathcal{A}_\sigma$, by the definition of $\mu(A)$ and Proposition 5.22 it follows that
$$\begin{aligned}\mu(A) = \mu(\Omega) - \mu(A^c) &= \mu(\Omega) - \uparrow \lim_{n\to\infty} \mu(A_n^c) \\ &= \downarrow \lim_{n\to\infty} [\mu(\Omega) - \mu(A_n^c)] = \downarrow \lim_{n\to\infty} \mu(A_n).\end{aligned}$$
3. Suppose $A, B \in \mathcal{A}_\delta$ and $A_n, B_n \in \mathcal{A}$ such that $A_n \downarrow A$ and $B_n \downarrow B$, then $A_n \cup B_n \downarrow A \cup B$ and $A_n \cap B_n \downarrow A \cap B$ and therefore,
$$\begin{aligned}\mu(A \cup B) + \mu(A \cap B) &= \lim_{n\to\infty} [\mu(A_n \cup B_n) + \mu(A_n \cap B_n)] \\ &= \lim_{n\to\infty} [\mu(A_n) + \mu(B_n)] = \mu(A) + \mu(B).\end{aligned}$$
All we really need is the finite additivity of μ which can be proved as follows. Suppose that $A, B \in \mathcal{A}_\delta$ are disjoint, then $A \cap B = \emptyset$ implies $A^c \cup B^c = \Omega$. So by the strong additivity of μ on $\mathcal{A}_\sigma$ it follows that
$$\mu(\Omega) + \mu(A^c \cap B^c) = \mu(A^c) + \mu(B^c)$$
from which it follows that
$$\begin{aligned}\mu(A \cup B) &= \mu(\Omega) - \mu(A^c \cap B^c) \\ &= \mu(\Omega) - [\mu(A^c) + \mu(B^c) - \mu(\Omega)] \\ &= \mu(A) + \mu(B).\end{aligned}$$
4. Since $A^c, C \in \mathcal{A}_\sigma$ we may use the strong additivity of μ on $\mathcal{A}_\sigma$ to conclude,
$$\mu(A^c \cup C) + \mu(A^c \cap C) = \mu(A^c) + \mu(C).$$
Because $\Omega = A^c \cup C$, and $\mu(A^c) = \mu(\Omega) - \mu(A)$, the above equation may be written as
$$\mu(\Omega) + \mu(C \setminus A) = \mu(\Omega) - \mu(A) + \mu(C)$$
which finishes the proof.

■

Notation 5.24 (Inner and outer measures) *Let $\mu : \mathcal{A} \to [0, \infty)$ be a finite premeasure extended to $\mathcal{A}_\sigma \cup \mathcal{A}_\delta$ as above. The for **any** $B \subset \Omega$ let*

$$\begin{aligned}\mu_*(B) &:= \sup\{\mu(A) : \mathcal{A}_\delta \ni A \subset B\} \text{ and} \\ \mu^*(B) &:= \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\}.\end{aligned}$$

*We refer to $\mu_*(B)$ and $\mu^*(B)$ as the **inner and outer** content of B respectively.*

If $B \subset \Omega$ has the same inner and outer content it is reasonable to define the measure of B as this common value. As we will see in Theorem 5.27 below, this extension becomes a σ – additive measure on a σ – algebra of subsets of Ω.

Definition 5.25 (Measurable Sets). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^\Omega$. We say that $B \subset \Omega$ is **measurable** if $\mu_*(B) = \mu^*(B)$. We will denote the collection of measurable subsets of Ω by $\mathcal{B} = \mathcal{B}(\mu)$ and define $\bar{\mu} : \mathcal{B} \to [0, \mu(\Omega)]$ by*

$$\bar{\mu}(B) := \mu_*(B) = \mu^*(B) \text{ for all } B \in \mathcal{B}. \tag{5.7}$$

Remark 5.26. Observe that $\mu_*(B) = \mu^*(B)$ iff for all $\varepsilon > 0$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu(C) - \mu(A) < \varepsilon,$$

wherein we have used Lemma 5.23 for the first equality. Moreover we will use below that if $B \in \mathcal{B}$ and $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$, then

$$\mu(A) \le \mu_*(B) = \bar{\mu}(B) = \mu^*(B) \le \mu(C). \tag{5.8}$$

Theorem 5.27 (Finite Premeasure Extension Theorem). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^\Omega$ and $\bar{\mu} : \mathcal{B} := \mathcal{B}(\mu) \to [0, \mu(\Omega)]$ be as in Definition 5.25. Then $\mathcal{B}$ is a σ – algebra on Ω which contains $\mathcal{A}$ and $\bar{\mu}$ is a σ – additive measure on $\mathcal{B}$. Moreover, $\bar{\mu}$ is the unique measure on $\mathcal{B}$ such that $\bar{\mu}|_{\mathcal{A}} = \mu$.*

Proof. It is clear that $\mathcal{A} \subset \mathcal{B}$ and that $\mathcal{B}$ is closed under complementation. Now suppose that $B_i \in \mathcal{B}$ for $i = 1, 2$ and $\varepsilon > 0$ is given. We may then choose $A_i \subset B_i \subset C_i$ such that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon$ for $i = 1, 2$. Then with $A = A_1 \cup A_2$, $B = B_1 \cup B_2$ and $C = C_1 \cup C_2$, we have $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$. Since

$$C \setminus A = (C_1 \setminus A) \cup (C_2 \setminus A) \subset (C_1 \setminus A_1) \cup (C_2 \setminus A_2),$$

it follows from the sub-additivity of μ that with

$$\mu(C \setminus A) \le \mu(C_1 \setminus A_1) + \mu(C_2 \setminus A_2) < 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have shown that $B \in \mathcal{B}$. Hence we now know that $\mathcal{B}$ is an algebra.

Because $\mathcal{B}$ is an algebra, to verify that $\mathcal{B}$ is a σ – algebra it suffices to show that $B = \sum_{n=1}^\infty B_n \in \mathcal{B}$ whenever $\{B_n\}_{n=1}^\infty$ is a disjoint sequence in $\mathcal{B}$. To prove $B \in \mathcal{B}$, let $\varepsilon > 0$ be given and choose $A_i \subset B_i \subset C_i$ such that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon 2^{-i}$ for all i. Since the $\{A_i\}_{i=1}^\infty$ are pairwise disjoint we may use Lemma 5.23 to show,

$$\begin{aligned}\sum_{i=1}^n \mu(C_i) &= \sum_{i=1}^n (\mu(A_i) + \mu(C_i \setminus A_i)) \\ &= \mu(\cup_{i=1}^n A_i) + \sum_{i=1}^n \mu(C_i \setminus A_i) \le \mu(\Omega) + \sum_{i=1}^n \varepsilon 2^{-i}.\end{aligned}$$

Passing to the limit, $n \to \infty$, in this equation then shows

$$\sum_{i=1}^\infty \mu(C_i) \le \mu(\Omega) + \varepsilon < \infty. \tag{5.9}$$

Let $B = \cup_{i=1}^\infty B_i$, $C := \cup_{i=1}^\infty C_i \in \mathcal{A}_\sigma$ and for $n \in \mathbb{N}$ let $A^n := \sum_{i=1}^n A_i \in \mathcal{A}_\delta$. Then $\mathcal{A}_\delta \ni A^n \subset B \subset C \in \mathcal{A}_\sigma$, $C \setminus A^n \in \mathcal{A}_\sigma$ and

$$C \setminus A^n = \cup_{i=1}^\infty (C_i \setminus A^n) \subset [\cup_{i=1}^n (C_i \setminus A_i)] \cup [\cup_{i=n+1}^\infty C_i] \in \mathcal{A}_\sigma.$$

Therefore, using the sub-additivity of μ on $\mathcal{A}_\sigma$ and the estimate in Eq. (5.9),

$$\begin{aligned}\mu(C \setminus A^n) &\le \sum_{i=1}^n \mu(C_i \setminus A_i) + \sum_{i=n+1}^\infty \mu(C_i) \\ &\le \varepsilon + \sum_{i=n+1}^\infty \mu(C_i) \to \varepsilon \text{ as } n \to \infty.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows that $B \in \mathcal{B}$ and that

$$\sum_{i=1}^n \mu(A_i) = \mu(A^n) \le \bar{\mu}(B) \le \mu(C) \le \sum_{i=1}^\infty \mu(C_i).$$

Letting $n \to \infty$ in this equation then shows,

$$\sum_{i=1}^\infty \mu(A_i) \le \bar{\mu}(B) \le \sum_{i=1}^\infty \mu(C_i). \tag{5.10}$$

On the other hand, since $A_i \subset B_i \subset C_i$, it follows (see Eq. (5.8) that

$$\sum_{i=1}^\infty \mu(A_i) \le \sum_{i=1}^\infty \bar{\mu}(B_i) \le \sum_{i=1}^\infty \mu(C_i). \tag{5.11}$$

As

$$\sum_{i=1}^\infty \mu(C_i) - \sum_{i=1}^\infty \mu(A_i) = \sum_{i=1}^\infty \mu(C_i \setminus A_i) \le \sum_{i=1}^\infty \varepsilon 2^{-i} = \varepsilon,$$

we may conclude from Eqs. (5.10) and (5.11) that

$$\left| \bar{\mu}(B) - \sum_{i=1}^\infty \bar{\mu}(B_i) \right| \le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have shown $\bar{\mu}(B) = \sum_{i=1}^\infty \bar{\mu}(B_i)$. This completes the proof that $\mathcal{B}$ is a σ - algebra and that $\bar{\mu}$ is a measure on $\mathcal{B}$.

Since we really had no choice as to how to extend μ, it is to be expected that the extension is unique. You are asked to supply the details in Exercise 5.3 below. ■

Exercise 5.3. Let μ, $\bar{\mu}$, $\mathcal{A}$, and $\mathcal{B} := \mathcal{B}(\mu)$ be as in Theorem 5.27. Further suppose that $\mathcal{B}_0 \subset 2^{\Omega}$ is a σ – algebra such that $\mathcal{A} \subset \mathcal{B}_0 \subset \mathcal{B}$ and $\nu : \mathcal{B}_0 \to [0, \mu(\Omega)]$ is a σ – additive measure on $\mathcal{B}_0$ such that $\nu = \mu$ on $\mathcal{A}$. Show that $\nu = \bar{\mu}$ on $\mathcal{B}_0$ as well. (When $\mathcal{B}_0 = \sigma(\mathcal{A})$ this exercise is of course a consequence of Proposition 5.15. It is not necessary to use this information to complete the exercise.)

Corollary 5.28. *Suppose that $\mathcal{A} \subset 2^{\Omega}$ is an algebra and $\mu : \mathcal{B}_0 := \sigma(\mathcal{A}) \to [0, \mu(\Omega)]$ is a σ – additive measure. Then for every $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$;*

1. *there exists $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$ and $\varepsilon > 0$ such that $\mu(C \setminus A) < \varepsilon$ and*
2. *there exists $A \in \mathcal{A}$ such that $\mu(A \Delta B) < \varepsilon$.*

Exercise 5.4. Prove corollary 5.28 by considering $\bar{\nu}$ where $\nu := \mu|_{\mathcal{A}}$. **Hint:** you may find Exercise 4.3 useful here.

Theorem 5.29. *Suppose that μ is a σ – finite premeasure on an algebra $\mathcal{A}$. Then*

$$\bar{\mu}(B) := \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\} \;\; \forall\, B \in \sigma(\mathcal{A}) \tag{5.12}$$

defines a measure on $\sigma(\mathcal{A})$ and this measure is the unique extension of μ on $\mathcal{A}$ to a measure on $\sigma(\mathcal{A})$. Recall that

$$\mu(C) = \sup\{\mu(A) : \mathcal{A} \ni A \subset C\}.$$

Proof. Let $\{\Omega_n\}_{n=1}^\infty \subset \mathcal{A}$ be chosen so that $\mu(\Omega_n) < \infty$ for all n and $\Omega_n \uparrow \Omega$ as $n \to \infty$ and let

$$\mu_n(A) := \mu_n(A \cap \Omega_n) \text{ for all } A \in \mathcal{A}.$$

Each μ_n is a premeasure (as is easily verified) on $\mathcal{A}$ and hence by Theorem 5.27 each μ_n has an extension, $\bar{\mu}_n$, to a measure on $\sigma(\mathcal{A})$. Since the measure $\bar{\mu}_n$ are increasing, $\bar{\mu} := \lim_{n\to\infty} \bar{\mu}_n$ is a measure which extends μ.

The proof will be completed by verifying that Eq. (5.12) holds. Let $B \in \sigma(\mathcal{A})$, $B_m = \Omega_m \cap B$ and $\varepsilon > 0$ be given. By Theorem 5.27, there exists $C_m \in \mathcal{A}_\sigma$ such that $B_m \subset C_m \subset \Omega_m$ and $\bar{\mu}(C_m \setminus B_m) = \bar{\mu}_m(C_m \setminus B_m) < \varepsilon 2^{-n}$. Then $C := \cup_{m=1}^\infty C_m \in \mathcal{A}_\sigma$ and

$$\bar{\mu}(C \setminus B) \le \bar{\mu}\left(\bigcup_{m=1}^\infty (C_m \setminus B)\right) \le \sum_{m=1}^\infty \bar{\mu}(C_m \setminus B) \le \sum_{m=1}^\infty \bar{\mu}(C_m \setminus B_m) < \varepsilon.$$

Thus

$$\bar{\mu}(B) \le \bar{\mu}(C) = \bar{\mu}(B) + \bar{\mu}(C \setminus B) \le \bar{\mu}(B) + \varepsilon$$

which, since $\varepsilon > 0$ is arbitrary, shows $\bar{\mu}$ satisfies Eq. (5.12). The uniqueness of the extension $\bar{\mu}$ is proved in Exercise 5.11. ∎

The following slight reformulation of Theorem 5.29 can be useful.

Corollary 5.30. *Let $\mathcal{A}$ be an algebra of sets, $\{\Omega_m\}_{m=1}^\infty \subset \mathcal{A}$ is a given sequence of sets such that $\Omega_m \uparrow \Omega$ as $m \to \infty$. Let*

$$\mathcal{A}_f := \{A \in \mathcal{A} : A \subset \Omega_m \text{ for some } m \in \mathbb{N}\}.$$

Notice that $\mathcal{A}_f$ is a ring, i.e. closed under differences, intersections and unions and contains the empty set. Further suppose that $\mu : \mathcal{A}_f \to [0, \infty)$ is an additive set function such that $\mu(A_n) \downarrow 0$ for any sequence, $\{A_n\} \subset \mathcal{A}_f$ such that $A_n \downarrow \emptyset$ as $n \to \infty$. Then μ extends uniquely to a σ – finite measure on $\mathcal{A}$.

Proof. Existence. By assumption, $\mu_m := \mu|_{\mathcal{A}_{\Omega_m}} : \mathcal{A}_{\Omega_m} \to [0, \infty)$ is a premeasure on $(\Omega_m, \mathcal{A}_{\Omega_m})$ and hence by Theorem 5.29 extends to a measure μ'_m on $(\Omega_m, \sigma(\mathcal{A}_{\Omega_m}) = \mathcal{B}_{\Omega_m})$. Let $\bar{\mu}_m(B) := \mu'_m(B \cap \Omega_m)$ for all $B \in \mathcal{B}$. Then $\{\bar{\mu}_m\}_{m=1}^\infty$ is an increasing sequence of measure on $(\Omega, \mathcal{B})$ and hence $\bar{\mu} := \lim_{m\to\infty} \bar{\mu}_m$ defines a measure on $(\Omega, \mathcal{B})$ such that $\bar{\mu}|_{\mathcal{A}_f} = \mu$.

Uniqueness. If μ_1 and μ_2 are two such extensions, then $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{A}$ and therefore by Proposition 5.15 or Exercise 5.11 we know that $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{B}$. We may now let $m \to \infty$ to see that in fact $\mu_1(B) = \mu_2(B)$ for all $B \in \mathcal{B}$, i.e. $\mu_1 = \mu_2$. ∎

5.4 Radon Measures on $\mathbb{R}$

We say that a measure, μ, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is a **Radon measure** if $\mu([a, b]) < \infty$ for all $-\infty < a < b < \infty$. In this section we will give a characterization of all Radon measures on $\mathbb{R}$. We first need the following general result characterizing premeasures on an algebra generated by a semi-algebra.

Proposition 5.31. *Suppose that $\mathcal{S} \subset 2^{\Omega}$ is a semi-algebra, $\mathcal{A} = \mathcal{A}(\mathcal{S})$ and $\mu : \mathcal{A} \to [0, \infty]$ is a finitely additive measure. Then μ is a premeasure on $\mathcal{A}$ iff μ is countably sub-additive on $\mathcal{S}$.*

Proof. Clearly if μ is a premeasure on $\mathcal{A}$ then μ is σ - additive and hence sub-additive on $\mathcal{S}$. Because of Proposition 4.2, to prove the converse it suffices to show that the sub-additivity of μ on $\mathcal{S}$ implies the sub-additivity of μ on $\mathcal{A}$.

So suppose $A = \sum_{n=1}^\infty A_n \in \mathcal{A}$ with each $A_n \in \mathcal{A}$. By Proposition 3.25 we may write $A = \sum_{j=1}^k E_j$ and $A_n = \sum_{i=1}^{N_n} E_{n,i}$ with $E_j, E_{n,i} \in \mathcal{S}$. Intersecting the identity, $A = \sum_{n=1}^\infty A_n$, with E_j implies

$$E_j = A \cap E_j = \sum_{n=1}^\infty A_n \cap E_j = \sum_{n=1}^\infty \sum_{i=1}^{N_n} E_{n,i} \cap E_j.$$

By the assumed sub-additivity of μ on $\mathcal{S}$,

$$\mu(E_j) \le \sum_{n=1}^{\infty}\sum_{i=1}^{N_n} \mu\left(E_{n,i} \cap E_j\right).$$

Summing this equation on j and using the finite additivity of μ shows

$$\begin{aligned}\mu(A) = \sum_{j=1}^{k} \mu(E_j) &\le \sum_{j=1}^{k}\sum_{n=1}^{\infty}\sum_{i=1}^{N_n} \mu\left(E_{n,i} \cap E_j\right) \\ &= \sum_{n=1}^{\infty}\sum_{i=1}^{N_n}\sum_{j=1}^{k} \mu\left(E_{n,i} \cap E_j\right) = \sum_{n=1}^{\infty}\sum_{i=1}^{N_n} \mu\left(E_{n,i}\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right).\end{aligned}$$

■

Suppose now that μ is a Radon measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that

$$\mu\left((a, b]\right) = F\left(b\right) - F\left(a\right) \text{ for all } -\infty < a \le b < \infty. \tag{5.13}$$

For example if $\mu\left(\mathbb{R}\right) < \infty$ we can take $F\left(x\right) = \mu\left((-\infty, x]\right)$ while if $\mu\left(\mathbb{R}\right) = \infty$ we might take

$$F\left(x\right) = \begin{cases} \mu\left((0, x]\right) & \text{if } x \ge 0 \\ -\mu\left((x, 0]\right) & \text{if } x \le 0 \end{cases}.$$

The function F is uniquely determined modulo translation by a constant.

Lemma 5.32. *If μ is a Radon measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that $\mu\left((a, b]\right) = F\left(b\right) - F\left(a\right)$, then F is increasing and right continuous.*

Proof. The function F is increasing by the monotonicity of μ. To see that F is right continuous, let $b \in \mathbb{R}$ and choose $a \in (-\infty, b)$ and any sequence $\{b_n\}_{n=1}^{\infty} \subset (b, \infty)$ such that $b_n \downarrow b$ as $n \to \infty$. Since $\mu\left((a, b_1]\right) < \infty$ and $(a, b_n] \downarrow (a, b]$ as $n \to \infty$, it follows that

$$F(b_n) - F(a) = \mu((a, b_n]) \downarrow \mu((a, b]) = F(b) - F(a).$$

Since $\{b_n\}_{n=1}^{\infty}$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $\lim_{y \downarrow b} F(y) = F(b)$. ■

The key result of this section is the converse to this lemma.

Theorem 5.33. *Suppose $F : \mathbb{R} \to \mathbb{R}$ is a right continuous increasing function. Then there exists a unique Radon measure, $\mu = \mu_F$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that Eq. (5.13) holds.*

Proof. Let $\mathcal{S} := \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}$, and $\mathcal{A} = \mathcal{A}\left(\mathcal{S}\right)$ consists of those sets, $A \subset \mathbb{R}$ which may be written as finite disjoint unions of sets from $\mathcal{S}$ as in Example 3.26. Recall that $\mathcal{B}_{\mathbb{R}} = \sigma\left(\mathcal{A}\right) = \sigma\left(\mathcal{S}\right)$. Further define $F\left(\pm\infty\right) := \lim_{x \to \pm\infty} F\left(x\right)$ and let $\mu = \mu_F$ be the finitely additive measure on $(\mathbb{R}, \mathcal{A})$ described in Proposition 4.8 and Remark 4.9. To finish the proof it suffices by Theorem 5.29 to show that μ is a premeasure on $\mathcal{A} = \mathcal{A}\left(\mathcal{S}\right)$ where $\mathcal{S} := \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}$. So in light of Proposition 5.31, to finish the proof it suffices to show μ is sub-additive on $\mathcal{S}$, i.e. we must show

$$\mu(J) \le \sum_{n=1}^{\infty} \mu(J_n). \tag{5.14}$$

where $J = \sum_{n=1}^{\infty} J_n$ with $J = (a, b] \cap \mathbb{R}$ and $J_n = (a_n, b_n] \cap \mathbb{R}$. Recall from Proposition 4.2 that the finite additivity of μ implies

$$\sum_{n=1}^{\infty} \mu(J_n) \le \mu\left(J\right). \tag{5.15}$$

We begin with the special case where $-\infty < a < b < \infty$. Our proof will be by "continuous induction." The strategy is to show $a \in \Lambda$ where

$$\Lambda := \left\{\alpha \in [a, b] : \mu(J \cap (\alpha, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\alpha, b])\right\}. \tag{5.16}$$

As $b \in J$, there exists an k such that $b \in J_k$ and hence $(a_k, b_k] = (a_k, b]$ for this k. It now easily follows that $J_k \subset \Lambda$ so that Λ is not empty. To finish the proof we are going to show $\bar{a} := \inf \Lambda \in \Lambda$ and that $\bar{a} = a$.

- If $\bar{a} \notin \Lambda$, there would exist $\alpha_m \in \Lambda$ such that $\alpha_m \downarrow \bar{a}$, i.e.

$$\mu(J \cap (\alpha_m, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\alpha_m, b]). \tag{5.17}$$

  Since $\mu(J_n \cap (\alpha_m, b]) \le \mu(J_n)$ and $\sum_{n=1}^{\infty} \mu\left(J_n\right) \le \mu\left(J\right) < \infty$ by Eq. (5.15), we may use the right continuity of F and the dominated convergence theorem for sums in order to pass to the limit as $m \to \infty$ in Eq. (5.17) to learn,

$$\mu(J \cap (\bar{a}, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\bar{a}, b]).$$

  This shows $\bar{a} \in \Lambda$ which is a contradiction to the original assumption that $\bar{a} \notin \Lambda$.
- If $\bar{a} > a$, then $\bar{a} \in J_l = (a_l, b_l]$ for some l. Letting $\alpha = a_l < \bar{a}$, we have,

$$\begin{aligned}
\mu(J\cap(\alpha,b]) &= \mu(J\cap(\alpha,\bar{a}]) + \mu(J\cap(\bar{a},b]) \\
&\le \mu(J_l\cap(\alpha,\bar{a}]) + \sum_{n=1}^{\infty}\mu(J_n\cap(\bar{a},b]) \\
&= \mu(J_l\cap(\alpha,\bar{a}]) + \mu\left(J_l\cap(\bar{a},b]\right) + \sum_{n\neq l}\mu(J_n\cap(\bar{a},b]) \\
&= \mu(J_l\cap(\alpha,b]) + \sum_{n\neq l}\mu(J_n\cap(\bar{a},b]) \\
&\le \sum_{n=1}^{\infty}\mu(J_n\cap(\alpha,b]).
\end{aligned}$$

This shows $\alpha\in\Lambda$ and $\alpha<\bar{a}$ which violates the definition of $\bar{a}$. Thus we must conclude that $\bar{a}=a$.

The hard work is now done but we still have to check the cases where $a=-\infty$ or $b=\infty$. For example, suppose that $b=\infty$ so that

$$J=(a,\infty)=\sum_{n=1}^{\infty}J_n$$

with $J_n=(a_n,b_n]\cap\mathbb{R}$. Then

$$I_M:=(a,M]=J\cap I_M=\sum_{n=1}^{\infty}J_n\cap I_M$$

and so by what we have already proved,

$$F(M)-F(a)=\mu(I_M)\le\sum_{n=1}^{\infty}\mu(J_n\cap I_M)\le\sum_{n=1}^{\infty}\mu(J_n).$$

Now let $M\to\infty$ in this last inequality to find that

$$\mu((a,\infty))=F(\infty)-F(a)\le\sum_{n=1}^{\infty}\mu(J_n).$$

The other cases where $a=-\infty$ and $b\in\mathbb{R}$ and $a=-\infty$ and $b=\infty$ are handled similarly. ■

5.4.1 Lebesgue Measure

If $F(x)=x$ for all $x\in\mathbb{R}$, we denote μ_F by m and call m Lebesgue measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$.

Theorem 5.34. *Lebesgue measure m is invariant under translations, i.e. for $B\in\mathcal{B}_{\mathbb{R}}$ and $x\in\mathbb{R}$,*

$$m(x+B)=m(B). \tag{5.18}$$

Lebesgue measure, m, is the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that $m((0,1])=1$ and Eq. (5.18) holds for $B\in\mathcal{B}_{\mathbb{R}}$ and $x\in\mathbb{R}$. Moreover, m has the scaling property

$$m(\lambda B)=|\lambda|\,m(B) \tag{5.19}$$

where $\lambda\in\mathbb{R}$, $B\in\mathcal{B}_{\mathbb{R}}$ and $\lambda B:=\{\lambda x : x\in B\}$.

Proof. Let $m_x(B):=m(x+B)$, then one easily shows that m_x is a measure on $\mathcal{B}_{\mathbb{R}}$ such that $m_x((a,b])=b-a$ for all $a<b$. Therefore, $m_x=m$ by the uniqueness assertion in Exercise 5.11. For the converse, suppose that m is translation invariant and $m((0,1])=1$. Given $n\in\mathbb{N}$, we have

$$(0,1]=\cup_{k=1}^{n}(\frac{k-1}{n},\frac{k}{n}]=\cup_{k=1}^{n}\left(\frac{k-1}{n}+(0,\frac{1}{n}]\right).$$

Therefore,

$$\begin{aligned}
1=m((0,1]) &= \sum_{k=1}^{n}m\left(\frac{k-1}{n}+(0,\frac{1}{n}]\right) \\
&= \sum_{k=1}^{n}m((0,\frac{1}{n}])=n\cdot m((0,\frac{1}{n}]).
\end{aligned}$$

That is to say

$$m((0,\frac{1}{n}])=1/n.$$

Similarly, $m((0,\frac{l}{n}])=l/n$ for all $l,n\in\mathbb{N}$ and therefore by the translation invariance of m,

$$m((a,b])=b-a \text{ for all } a,b\in\mathbb{Q} \text{ with } a<b.$$

Finally for $a,b\in\mathbb{R}$ such that $a<b$, choose $a_n,b_n\in\mathbb{Q}$ such that $b_n\downarrow b$ and $a_n\uparrow a$, then $(a_n,b_n]\downarrow(a,b]$ and thus

$$m((a,b])=\lim_{n\to\infty}m((a_n,b_n])=\lim_{n\to\infty}(b_n-a_n)=b-a,$$

i.e. m is Lebesgue measure. To prove Eq. (5.19) we may assume that $\lambda\neq 0$ since this case is trivial to prove. Now let $m_\lambda(B):=|\lambda|^{-1}m(\lambda B)$. It is easily checked that m_λ is again a measure on $\mathcal{B}_{\mathbb{R}}$ which satisfies

$$m_\lambda((a,b])=\lambda^{-1}m\left((\lambda a,\lambda b]\right)=\lambda^{-1}(\lambda b-\lambda a)=b-a$$

if $\lambda>0$ and

$$m_\lambda((a,b])=|\lambda|^{-1}m\left([\lambda b,\lambda a)\right)=-|\lambda|^{-1}(\lambda b-\lambda a)=b-a$$

if $\lambda<0$. Hence $m_\lambda=m$. ■

5.5 A Discrete Kolmogorov's Extension Theorem

For this section, let S be a finite or countable set (we refer to S as **state space**), $\Omega := S^\infty := S^{\mathbb{N}}$ (think of $\mathbb{N}$ as time and Ω as **path space**)

$$\mathcal{A}_n := \{B \times \Omega : B \subset S^n\} \text{ for all } n \in \mathbb{N},$$

$\mathcal{A} := \cup_{n=1}^\infty \mathcal{A}_n$, and $\mathcal{B} := \sigma(\mathcal{A})$. We call the elements, $A \subset \Omega$, the **cylinder subsets of** Ω. Notice that $A \subset \Omega$ is a cylinder set iff there exists $n \in \mathbb{N}$ and $B \subset S^n$ such that

$$A = B \times \Omega := \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}.$$

Also observe that we may write A as $A = B' \times \Omega$ where $B' = B \times S^k \subset S^{n+k}$ for any $k \geq 0$.

Exercise 5.5. Show;

1. $\mathcal{A}_n$ is a σ – algebra for each $n \in \mathbb{N}$,
2. $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ for all n, and
3. $\mathcal{A} \subset 2^\Omega$ is an algebra of subsets of Ω. (In fact, you might show that $\mathcal{A} = \cup_{n=1}^\infty \mathcal{A}_n$ is an algebra whenever $\{\mathcal{A}_n\}_{n=1}^\infty$ is an increasing sequence of algebras.)

Lemma 5.35 (Baby Tychonov Theorem). *Suppose $\{C_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of* ***non-empty*** *cylinder sets. Further assume there exists $N_n \in \mathbb{N}$ and $B_n \subset\subset S^{N_n}$ such that $C_n = B_n \times \Omega$. (This last assumption is vacuous when S is a finite set. Recall that we write $\Lambda \subset\subset A$ to indicate that Λ is a finite subset of A.) Then $\cap_{n=1}^\infty C_n \neq \emptyset$.*

Proof. Since $C_{n+1} \subset C_n$, if $N_n > N_{n+1}$, we would have $B_{n+1} \times S^{N_{n+1} - N_n} \subset B_n$. If S is an infinite set this would imply B_n is an infinite set and hence we must have $N_{n+1} \geq N_n$ for all n when $\#(S) = \infty$. On the other hand, if S is a finite set, we can always replace B_{n+1} by $B_{n+1} \times S^k$ for some appropriate k and arrange it so that $N_{n+1} \geq N_n$ for all n. So from now we assume that $N_{n+1} \geq N_n$.

Case 1. $\lim_{n\to\infty} N_n < \infty$ in which case there exists some $N \in \mathbb{N}$ such that $N_n = N$ for all large n. Thus for large N, $C_n = B_n \times \Omega$ with $B_n \subset\subset S^N$ and $B_{n+1} \subset B_n$ and hence $\#(B_n) \downarrow$ as $n \to \infty$. By assumption, $\lim_{n\to\infty} \#(B_n) \neq 0$ and therefore $\#(B_n) = k > 0$ for all n large. It then follows that there exists $n_0 \in \mathbb{N}$ such that $B_n = B_{n_0}$ for all $n \geq n_0$. Therefore $\cap_{n=1}^\infty C_n = B_{n_0} \times \Omega \neq \emptyset$.

Case 2. $\lim_{n\to\infty} N_n = \infty$. By assumption, there exists $\omega(n) = (\omega_1(n), \omega_2(n), \dots) \in \Omega$ such that $\omega(n) \in C_n$ for all n. Moreover, since $\omega(n) \in C_n \subset C_k$ for all $k \leq n$, it follows that

$$(\omega_1(n), \omega_2(n), \dots, \omega_{N_k}(n)) \in B_k \text{ for all } n \geq k \tag{5.20}$$

and as B_k is a finite set $\{\omega_i(n)\}_{n=1}^\infty$ must be a finite set for all $1 \leq i \leq N_k$. As $N_k \to \infty$ as $k \to \infty$ it follows that $\{\omega_i(n)\}_{n=1}^\infty$ is a finite set for all $i \in \mathbb{N}$. Using this observation, we may find, $s_1 \in S$ and an infinite subset, $\Gamma_1 \subset \mathbb{N}$ such that $\omega_1(n) = s_1$ for all $n \in \Gamma_1$. Similarly, there exists $s_2 \in S$ and an infinite set, $\Gamma_2 \subset \Gamma_1$, such that $\omega_2(n) = s_2$ for all $n \in \Gamma_2$. Continuing this procedure inductively, there exists (for all $j \in \mathbb{N}$) infinite subsets, $\Gamma_j \subset \mathbb{N}$ and points $s_j \in S$ such that $\Gamma_1 \supset \Gamma_2 \supset \Gamma_3 \supset \dots$ and $\omega_j(n) = s_j$ for all $n \in \Gamma_j$.

We are now going to complete the proof by showing $s := (s_1, s_2, \dots) \in \cap_{n=1}^\infty C_n$. By the construction above, for all $N \in \mathbb{N}$ we have

$$(\omega_1(n), \dots, \omega_N(n)) = (s_1, \dots, s_N) \text{ for all } n \in \Gamma_N.$$

Taking $N = N_k$ and $n \in \Gamma_{N_k}$ with $n \geq k$, we learn from Eq. (5.20) that

$$(s_1, \dots, s_{N_k}) = (\omega_1(n), \dots, \omega_{N_k}(n)) \in B_k.$$

But this is equivalent to showing $s \in C_k$. Since $k \in \mathbb{N}$ was arbitrary it follows that $s \in \cap_{n=1}^\infty C_n$. ■

Let $\bar{S} := S$ is S is a finite set and $\bar{S} = S \cup \{\infty\}$ if S is an infinite set. Here, ∞, is simply another point not in S which we call infinity Let $\{x_n\}_{n=1}^\infty \subset \bar{S}$ be a sequence, then we way $\lim_{n\to\infty} x_n = \infty$ if for every $A \subset\subset S$, $x_n \notin A$ for almost all n and we say that $\lim_{n\to\infty} x_n = s \in S$ if $x_n = s$ for almost all n. For example this is the usual notion of convergence for $S = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$ and $\bar{S} = S \cup \{0\} \subset [0,1]$, where 0 is playing the role of infinity here. Observe that either $\lim_{n\to\infty} x_n = \infty$ or there exists a finite subset $F \subset S$ such that $x_n \in F$ infinitely often. Moreover, there must be some point, $s \in F$ such that $x_n = s$ infinitely often. Thus if we let $\{n_1 < n_2 < \dots\} \subset \mathbb{N}$ be chosen such that $x_{n_k} = s$ for all k, then $\lim_{k\to\infty} x_{n_k} = s$. Thus we have shown that every sequence in $\bar{S}$ has a convergent subsequence.

Lemma 5.36 (Baby Tychonov Theorem I.). *Let $\bar{\Omega} := \bar{S}^{\mathbb{N}}$ and $\{\omega(n)\}_{n=1}^\infty$ be a sequence in $\bar{\Omega}$. Then there is a subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$ such that $\lim_{k\to\infty} \omega(n_k)$ exists in $\bar{\Omega}$ by which we mean, $\lim_{k\to\infty} \omega_i(n_k)$ exists in $\bar{S}$ for all $i \in \mathbb{N}$.*

Proof. This follows by the usual cantor's diagonalization argument. Indeed, let $\{n_k^1\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ be chosen so that $\lim_{k\to\infty} \omega_1(n_k^1) = s_1 \in \bar{S}$ exists. Then choose $\{n_k^2\}_{k=1}^\infty \subset \{n_k^1\}_{k=1}^\infty$ so that $\lim_{k\to\infty} \omega_2(n_k^2) = s_2 \in \bar{S}$ exists. Continue on this way to inductively choose

$$\{n_k^1\}_{k=1}^\infty \supset \{n_k^2\}_{k=1}^\infty \supset \dots \supset \{n_k^l\}_{k=1}^\infty \supset \dots$$

such that $\lim_{k\to\infty} \omega_l(n_k^l) = s_l \in \bar{S}$. The subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$, may now be defined by, $n_k = n_k^k$. ■

Corollary 5.37 (Baby Tychonov Theorem II.). *Suppose that $\{F_n\}_{n=1}^\infty \subset \bar{\Omega}$ is decreasing sequence of non-empty sets which are closed under taking sequential limits, then $\cap_{n=1}^\infty F_n \neq \emptyset$.*

Proof. Since $F_n \neq \emptyset$ there exists $\omega(n) \in F_n$ for all n. Using Lemma 5.36, there exists $\{n_k\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ such that $\omega := \lim_{k\to\infty} \omega(n_k)$ exits in $\bar{\Omega}$. Since $\omega(n_k) \in F_n$ for all $k \geq n$, it follows that $\omega \in F_n$ for all n, i.e. $\omega \in \cap_{n=1}^\infty F_n$ and hence $\cap_{n=1}^\infty F_n \neq \emptyset$. ∎

Example 5.38. Suppose that $1 \leq N_1 < N_2 < N_3 < \dots$, $F_n = K_n \times \Omega$ with $K_n \subset\subset S^{N_n}$ such that $\{F_n\}_{n=1}^\infty \subset \Omega$ is a decreasing sequence of non-empty sets. Then $\cap_{n=1}^\infty F_n \neq \emptyset$. To prove this, let $\bar{F}_n := K_n \times \bar{\Omega}$ in which case $\bar{F}_n$ are non – empty sets closed under taking limits. Therefore by Corollary 5.37, $\cap_n \bar{F}_n \neq \emptyset$. This completes the proof since it is easy to check that $\cap_{n=1}^\infty F_n = \cap_n \bar{F}_n \neq \emptyset$.

Corollary 5.39. *If S is a finite set and $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of non-empty cylinder sets, then $\cap_{n=1}^\infty A_n \neq \emptyset$.*

Proof. This follows directly from Example 5.38 since necessarily, $A_n = K_n \times \Omega$, for some $K_n \subset\subset S^{N_n}$. ∎

Theorem 5.40 (Kolmogorov's Extension Theorem I.). *Let us continue the notation above with the further assumption that S is a finite set. Then **every** finitely additive probability measure, $P : \mathcal{A} \to [0,1]$, has a unique extension to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

Proof. From Theorem 5.27, it suffices to show $\lim_{n\to\infty} P(A_n) = 0$ whenever $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ with $A_n \downarrow \emptyset$. However, by Lemma 5.35 with $C_n = A_n$, $A_n \in \mathcal{A}$ and $A_n \downarrow \emptyset$, we must have that $A_n = \emptyset$ for a.a. n and in particular $P(A_n) = 0$ for a.a. n. This certainly implies $\lim_{n\to\infty} P(A_n) = 0$. ∎

For the next three exercises, suppose that S is a finite set and continue the notation from above. Further suppose that $P : \sigma(\mathcal{A}) \to [0,1]$ is a probability measure and for $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, let

$$p_n(s_1, \dots, s_n) := P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}). \tag{5.21}$$

Exercise 5.6 (Consistency Conditions). If p_n is defined as above, show:

1. $\sum_{s\in S} p_1(s) = 1$ and
2. for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$,
$$p_n(s_1, \dots, s_n) = \sum_{s\in S} p_{n+1}(s_1, \dots, s_n, s).$$

Exercise 5.7 (Converse to 5.6). Suppose for each $n \in \mathbb{N}$ we are given functions, $p_n : S^n \to [0,1]$ such that the consistency conditions in Exercise 5.6 hold. Then there exists a unique probability measure, P on $\sigma(\mathcal{A})$ such that Eq. (5.21) holds for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$.

Example 5.41 (Existence of iid simple R.V.s). Suppose now that $q : S \to [0,1]$ is a function such that $\sum_{s\in S} q(s) = 1$. Then there exists a unique probability measure P on $\sigma(\mathcal{A})$ such that, for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, we have

$$P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}) = q(s_1) \dots q(s_n).$$

This is a special case of Exercise 5.7 with $p_n(s_1, \dots, s_n) := q(s_1) \dots q(s_n)$.

Theorem 5.42 (Kolmogorov's Extension Theorem II). *Suppose now that S is countably infinite set and $P : \mathcal{A} \to [0,1]$ is a finitely additive measure such that $P|_{\mathcal{A}_n}$ is a σ – additive measure for each $n \in \mathbb{N}$. Then P extends uniquely to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

Proof. From Theorem 5.27 it suffice to show; if $\{A_m\}_{m=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$, then $\cap_{m=1}^\infty A_m \neq \emptyset$. You are asked to verify this property of P in the next couple of exercises. ∎

For the next couple of exercises the hypothesis of Theorem 5.42 are to be assumed.

Exercise 5.8. Show for each $n \in \mathbb{N}$, $A \in \mathcal{A}_n$, and $\varepsilon > 0$ are given. Show there exists $F \in \mathcal{A}_n$ such that $F \subset A$, $F = K \times \Omega$ with $K \subset\subset S^n$, and $P(A \setminus F) < \varepsilon$.

Exercise 5.9. Let $\{A_m\}_{n=1}^\infty \subset \mathcal{A}$ be a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$. Using Exercise 5.8, choose $F_m = K_m \times \Omega \subset A_m$ with $K_m \subset\subset S^{N_n}$ and $P(A_m \setminus F_m) \leq \varepsilon/2^{m+1}$. Further define $C_m := F_1 \cap \dots \cap F_m$ for each m. Show;

1. Show $A_m \setminus C_m \subset (A_1 \setminus F_1) \cup (A_2 \setminus F_2) \cup \dots \cup (A_m \setminus F_m)$ and use this to conclude that $P(A_m \setminus C_m) \leq \varepsilon/2$.
2. Conclude C_m is not empty for m.
3. Use Lemma 5.35 to conclude that $\emptyset \neq \cap_{m=1}^\infty C_m \subset \cap_{m=1}^\infty A_m$.

Exercise 5.10. Convince yourself that the results of Exercise 5.6 and 5.7 are valid when S is a countable set. (See Example 4.6.)

In summary, the main result of this section states, to any sequence of functions, $p_n : S^n \to [0,1]$, such that $\sum_{\lambda\in S^n} p_n(\lambda) = 1$ and $\sum_{s\in S} p_{n+1}(\lambda, s) = p_n(\lambda)$ for all n and $\lambda \in S^n$, there exists a unique probability measure, P, on $\mathcal{B} := \sigma(\mathcal{A})$ such that

$$P(B \times \Omega) = \sum_{\lambda\in B} p_n(\lambda) \;\; \forall\; B \subset S^n \text{ and } n \in \mathbb{N}.$$

Example 5.43 (Markov Chain Probabilities). Let S be a finite or at most countable state space and $p : S \times S \to [0,1]$ be a **Markov kernel,** i.e.

$$\sum_{y\in S} p(x,y) = 1 \text{ for all } x \in S. \tag{5.22}$$

Also let $\pi : S \to [0,1]$ be a probability function, i.e. $\sum_{x\in S} \pi(x) = 1$. We now take

$$\Omega := S^{\mathbb{N}_0} = \{\omega = (s_0, s_1, \dots) : s_j \in S\}$$

and let $X_n : \Omega \to S$ be given by

$$X_n(s_0, s_1, \dots) = s_n \text{ for all } n \in \mathbb{N}_0.$$

Then there exists a unique probability measure, P_π, on $\sigma(\mathcal{A})$ such that

$$P_\pi(X_0 = x_0, \dots, X_n = x_n) = \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

for all $n \in \mathbb{N}_0$ and $x_0, x_1, \dots, x_n \in S$. To see such a measure exists, we need only verify that

$$p_n(x_0, \dots, x_n) := \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

verifies the hypothesis of Exercise 5.6 taking into account a shift of the n – index.

5.6 Appendix: Regularity and Uniqueness Results*

The goal of this appendix it to approximating measurable sets from inside and outside by classes of sets which are relatively easy to understand. Our first few results are already contained in Carathoédory's existence of measures proof. Nevertheless, we state these results again and give another somewhat independent proof.

Theorem 5.44 (Finite Regularity Result). *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} = \sigma(\mathcal{A})$ and $\mu : \mathcal{B} \to [0,\infty)$ is a finite measure, i.e. $\mu(\Omega) < \infty$. Then for every $\varepsilon > 0$ and $B \in \mathcal{B}$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$.*

Proof. Let $\mathcal{B}_0$ denote the collection of $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there here exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$. It is now clear that $\mathcal{A} \subset \mathcal{B}_0$ and that $\mathcal{B}_0$ is closed under complementation. Now suppose that $B_i \in \mathcal{B}_0$ for $i = 1, 2, \dots$ and $\varepsilon > 0$ is given. By assumption there exists $A_i \in \mathcal{A}_\delta$ and $C_i \in \mathcal{A}_\sigma$ such that $A_i \subset B_i \subset C_i$ and $\mu(C_i \setminus A_i) < 2^{-i}\varepsilon$.

Let $A := \cup_{i=1}^\infty A_i$, $A^N := \cup_{i=1}^N A_i \in \mathcal{A}_\delta$, $B := \cup_{i=1}^\infty B_i$, and $C := \cup_{i=1}^\infty C_i \in \mathcal{A}_\sigma$. Then $A^N \subset A \subset B \subset C$ and

$$C \setminus A = [\cup_{i=1}^\infty C_i] \setminus A = \cup_{i=1}^\infty [C_i \setminus A] \subset \cup_{i=1}^\infty [C_i \setminus A_i].$$

Therefore,

$$\mu(C \setminus A) = \mu(\cup_{i=1}^\infty [C_i \setminus A]) \le \sum_{i=1}^\infty \mu(C_i \setminus A) \le \sum_{i=1}^\infty \mu(C_i \setminus A_i) < \varepsilon.$$

Since $C \setminus A^N \downarrow C \setminus A$, it also follows that $\mu\left(C \setminus A^N\right) < \varepsilon$ for sufficiently large N and this shows $B = \cup_{i=1}^\infty B_i \in \mathcal{B}_0$. Hence $\mathcal{B}_0$ is a sub-σ-algebra of $\mathcal{B} = \sigma(\mathcal{A})$ which contains $\mathcal{A}$ which shows $\mathcal{B}_0 = \mathcal{B}$. ■

Many theorems in the sequel will require some control on the size of a measure μ. The relevant notion for our purposes (and most purposes) is that of a σ – finite measure defined next.

Definition 5.45. *Suppose Ω is a set, $\mathcal{E} \subset \mathcal{B} \subset 2^\Omega$ and $\mu : \mathcal{B} \to [0,\infty]$ is a function. The function μ is σ – finite on $\mathcal{E}$ if there exists $E_n \in \mathcal{E}$ such that $\mu(E_n) < \infty$ and $\Omega = \cup_{n=1}^\infty E_n$. If $\mathcal{B}$ is a σ – algebra and μ is a measure on $\mathcal{B}$ which is σ – finite on $\mathcal{B}$ we will say $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space.*

The reader should check that if μ is a finitely additive measure on an algebra, $\mathcal{B}$, then μ is σ – finite on $\mathcal{B}$ iff there exists $\Omega_n \in \mathcal{B}$ such that $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n) < \infty$.

Corollary 5.46 (σ – Finite Regularity Result). *Theorem 5.44 continues to hold under the weaker assumption that $\mu : \mathcal{B} \to [0,\infty]$ is a measure which is σ – finite on $\mathcal{A}$.*

Proof. Let $\Omega_n \in \mathcal{A}$ such that $\cup_{n=1}^\infty \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all n.Since $A \in \mathcal{B} \to \mu_n(A) := \mu(\Omega_n \cap A)$ is a finite measure on $A \in \mathcal{B}$ for each n, by Theorem 5.44, for every $B \in \mathcal{B}$ there exists $C_n \in \mathcal{A}_\sigma$ such that $B \subset C_n$ and $\mu(\Omega_n \cap [C_n \setminus B]) = \mu_n(C_n \setminus B) < 2^{-n}\varepsilon$. Now let $C := \cup_{n=1}^\infty [\Omega_n \cap C_n] \in \mathcal{A}_\sigma$ and observe that $B \subset C$ and

$$\begin{aligned}\mu(C \setminus B) &= \mu(\cup_{n=1}^\infty ([\Omega_n \cap C_n] \setminus B)) \\ &\le \sum_{n=1}^\infty \mu([\Omega_n \cap C_n] \setminus B) = \sum_{n=1}^\infty \mu(\Omega_n \cap [C_n \setminus B]) < \varepsilon.\end{aligned}$$

Applying this result to B^c shows there exists $D \in \mathcal{A}_\sigma$ such that $B^c \subset D$ and

$$\mu(B \setminus D^c) = \mu(D \setminus B^c) < \varepsilon.$$

So if we let $A := D^c \in \mathcal{A}_\delta$, then $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu([B \setminus A] \cup [(C \setminus B) \setminus A]) \le \mu(B \setminus A) + \mu(C \setminus B) < 2\varepsilon$$

and the result is proved. ■

Exercise 5.11. Suppose $\mathcal{A} \subset 2^{\Omega}$ is an algebra and μ and ν are two measures on $\mathcal{B} = \sigma(\mathcal{A})$.

a. Suppose that μ and ν are finite measures such that $\mu = \nu$ on $\mathcal{A}$. Show $\mu = \nu$.
b. Generalize the previous assertion to the case where you only assume that μ and ν are σ – finite on $\mathcal{A}$.

Corollary 5.47. *Suppose $\mathcal{A} \subset 2^{\Omega}$ is an algebra and $\mu : \mathcal{B} = \sigma(\mathcal{A}) \to [0, \infty]$ is a measure which is σ – finite on $\mathcal{A}$. Then for all $B \in \mathcal{B}$, there exists $A \in \mathcal{A}_{\delta\sigma}$ and $C \in \mathcal{A}_{\sigma\delta}$ such that $A \subset B \subset C$ and $\mu(C \setminus A) = 0$.*

Proof. By Theorem 5.44, given $B \in \mathcal{B}$, we may choose $A_n \in \mathcal{A}_\delta$ and $C_n \in \mathcal{A}_\sigma$ such that $A_n \subset B \subset C_n$ and $\mu(C_n \setminus B) \leq 1/n$ and $\mu(B \setminus A_n) \leq 1/n$. By replacing A_N by $\cup_{n=1}^N A_n$ and C_N by $\cap_{n=1}^N C_n$, we may assume that $A_n \uparrow$ and $C_n \downarrow$ as n increases. Let $A = \cup A_n \in \mathcal{A}_{\delta\sigma}$ and $C = \cap C_n \in \mathcal{A}_{\sigma\delta}$, then $A \subset B \subset C$ and

$$\begin{aligned}\mu(C \setminus A) &= \mu(C \setminus B) + \mu(B \setminus A) \leq \mu(C_n \setminus B) + \mu(B \setminus A_n) \\ &\leq 2/n \to 0 \text{ as } n \to \infty.\end{aligned}$$

■

Exercise 5.12. Let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n} = \sigma(\{\text{open subsets of } \mathbb{R}^n\})$ be the Borel σ – algebra on $\mathbb{R}^n$ and μ be a probability measure on $\mathcal{B}$. Further, let $\mathcal{B}_0$ denote those sets $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there exists $F \subset B \subset V$ such that F is closed, V is open, and $\mu(V \setminus F) < \varepsilon$. Show:

1. $\mathcal{B}_0$ contains all closed subsets of $\mathcal{B}$. **Hint:** given a closed subset, $F \subset \mathbb{R}^n$ and $k \in \mathbb{N}$, let $V_k := \cup_{x \in F} B(x, 1/k)$, where $B(x, \delta) := \{y \in \mathbb{R}^n : |y - x| < \delta\}$. Show, $V_k \downarrow F$ as $k \to \infty$.
2. Show $\mathcal{B}_0$ is a σ – algebra and use this along with the first part of this exercise to conclude $\mathcal{B} = \mathcal{B}_0$. **Hint:** follow closely the method used in the first step of the proof of Theorem 5.44.
3. Show for every $\varepsilon > 0$ and $B \in \mathcal{B}$, there exist a compact subset, $K \subset \mathbb{R}^n$, such that $K \subset B$ and $\mu(B \setminus K) < \varepsilon$. **Hint:** take $K := F \cap \{x \in \mathbb{R}^n : |x| \leq n\}$ for some sufficiently large n.

5.7 Appendix: Completions of Measure Spaces*

Definition 5.48. *A set $E \subset \Omega$ is a **null** set if $E \in \mathcal{B}$ and $\mu(E) = 0$. If P is some "property" which is either true or false for each $x \in \Omega$, we will use the terminology P a.e. (to be read P almost everywhere) to mean*

$$E := \{x \in \Omega : P \text{ is false for } x\}$$

is a null set. For example if f and g are two measurable functions on $(\Omega, \mathcal{B}, \mu)$, $f = g$ a.e. means that $\mu(f \neq g) = 0$.

Definition 5.49. *A measure space $(\Omega, \mathcal{B}, \mu)$ is **complete** if every subset of a null set is in $\mathcal{B}$, i.e. for all $F \subset \Omega$ such that $F \subset E \in \mathcal{B}$ with $\mu(E) = 0$ implies that $F \in \mathcal{B}$.*

Proposition 5.50 (Completion of a Measure). *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. Set*

$$\begin{aligned}\mathcal{N} = \mathcal{N}^\mu &:= \{N \subset \Omega : \exists\, F \in \mathcal{B} \text{ such that } N \subset F \text{ and } \mu(F) = 0\}, \\ \mathcal{B} = \bar{\mathcal{B}}^\mu &:= \{A \cup N : A \in \mathcal{B} \text{ and } N \in \mathcal{N}\} \text{ and} \\ \bar{\mu}(A \cup N) &:= \mu(A) \text{ for } A \in \mathcal{B} \text{ and } N \in \mathcal{N},\end{aligned}$$

*see Fig. 5.2. Then $\bar{\mathcal{B}}$ is a σ – algebra, $\bar{\mu}$ is a well defined measure on $\bar{\mathcal{B}}$, $\bar{\mu}$ is the unique measure on $\bar{\mathcal{B}}$ which extends μ on $\mathcal{B}$, and $(\Omega, \bar{\mathcal{B}}, \bar{\mu})$ is complete measure space. The σ-algebra, $\bar{\mathcal{B}}$, is called the **completion** of $\mathcal{B}$ relative to μ and $\bar{\mu}$, is called the **completion of** μ.*

Proof. Clearly $\Omega, \emptyset \in \bar{\mathcal{B}}$. Let $A \in \mathcal{B}$ and $N \in \mathcal{N}$ and choose $F \in \mathcal{B}$ such

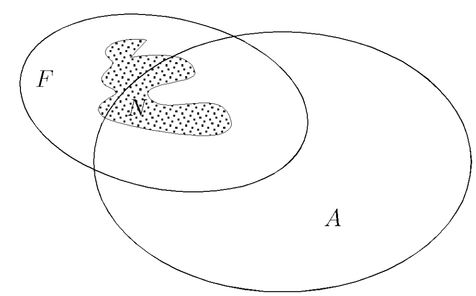


Fig. 5.2. Completing a σ – algebra.

that $N \subset F$ and $\mu(F) = 0$. Since $N^c = (F \setminus N) \cup F^c$,

$$\begin{aligned}(A \cup N)^c &= A^c \cap N^c = A^c \cap (F \setminus N \cup F^c) \\ &= [A^c \cap (F \setminus N)] \cup [A^c \cap F^c]\end{aligned}$$

where $[A^c \cap (F \setminus N)] \in \mathcal{N}$ and $[A^c \cap F^c] \in \mathcal{B}$. Thus $\bar{\mathcal{B}}$ is closed under complements. If $A_i \in \mathcal{B}$ and $N_i \subset F_i \in \mathcal{B}$ such that $\mu(F_i) = 0$ then $\cup(A_i \cup N_i) = (\cup A_i) \cup (\cup N_i) \in \bar{\mathcal{B}}$ since $\cup A_i \in \mathcal{B}$ and $\cup N_i \subset \cup F_i$ and

$\mu(\cup F_i) \leq \sum \mu(F_i) = 0$. Therefore, $\bar{\mathcal{B}}$ is a σ – algebra. Suppose $A \cup N_1 = B \cup N_2$ with $A, B \in \mathcal{B}$ and $N_1, N_2, \in \mathcal{N}$. Then $A \subset A \cup N_1 \subset A \cup N_1 \cup F_2 = B \cup F_2$ which shows that

$$\mu(A) \leq \mu(B) + \mu(F_2) = \mu(B).$$

Similarly, we show that $\mu(B) \leq \mu(A)$ so that $\mu(A) = \mu(B)$ and hence $\bar{\mu}(A \cup N) := \mu(A)$ is well defined. It is left as an exercise to show $\bar{\mu}$ is a measure, i.e. that it is countable additive. ∎

5.8 Appendix Monotone Class Theorems*

This appendix may be safely skipped!

Definition 5.51 (Montone Class). *$\mathcal{C} \subset 2^{\Omega}$ is a* ***monotone class*** *if it is closed under countable increasing unions and countable decreasing intersections.*

Lemma 5.52 (Monotone Class Theorem*). *Suppose $\mathcal{A} \subset 2^{\Omega}$ is an algebra and $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$. Then $\mathcal{C} = \sigma(\mathcal{A})$.*

Proof. For $C \in \mathcal{C}$ let

$$\mathcal{C}(C) = \{B \in \mathcal{C} : C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}\},$$

then $\mathcal{C}(C)$ is a monotone class. Indeed, if $B_n \in \mathcal{C}(C)$ and $B_n \uparrow B$, then $B_n^c \downarrow B^c$ and so

$$\begin{aligned} &\mathcal{C} \ni C \cap B_n \uparrow C \cap B \\ &\mathcal{C} \ni C \cap B_n^c \downarrow C \cap B^c \text{ and} \\ &\mathcal{C} \ni B_n \cap C^c \uparrow B \cap C^c. \end{aligned}$$

Since $\mathcal{C}$ is a monotone class, it follows that $C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}$, i.e. $B \in \mathcal{C}(C)$. This shows that $\mathcal{C}(C)$ is closed under increasing limits and a similar argument shows that $\mathcal{C}(C)$ is closed under decreasing limits. Thus we have shown that $\mathcal{C}(C)$ is a monotone class for all $C \in \mathcal{C}$. If $A \in \mathcal{A} \subset \mathcal{C}$, then $A \cap B, A \cap B^c, B \cap A^c \in \mathcal{A} \subset \mathcal{C}$ for all $B \in \mathcal{A}$ and hence it follows that $\mathcal{A} \subset \mathcal{C}(A) \subset \mathcal{C}$. Since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(A)$ is a monotone class containing $\mathcal{A}$, we conclude that $\mathcal{C}(A) = \mathcal{C}$ for any $A \in \mathcal{A}$. Let $B \in \mathcal{C}$ and notice that $A \in \mathcal{C}(B)$ happens iff $B \in \mathcal{C}(A)$. This observation and the fact that $\mathcal{C}(A) = \mathcal{C}$ for all $A \in \mathcal{A}$ implies $\mathcal{A} \subset \mathcal{C}(B) \subset \mathcal{C}$ for all $B \in \mathcal{C}$. Again since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(B)$ is a monotone class we conclude that $\mathcal{C}(B) = \mathcal{C}$ for all $B \in \mathcal{C}$. That is to say, if $A, B \in \mathcal{C}$ then $A \in \mathcal{C} = \mathcal{C}(B)$ and hence $A \cap B$, $A \cap B^c$, $A^c \cap B \in \mathcal{C}$. So $\mathcal{C}$ is closed under complements (since $\Omega \in \mathcal{A} \subset \mathcal{C}$) and finite intersections and increasing unions from which it easily follows that $\mathcal{C}$ is a σ – algebra. ∎

Random Variables

Notation 6.1 *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ let*

$$f^{-1}\mathcal{E} := f^{-1}(\mathcal{E}) := \{f^{-1}(E)|E \in \mathcal{E}\}.$$

If $\mathcal{G} \subset 2^X$, let

$$f_*\mathcal{G} := \{A \in 2^Y|f^{-1}(A) \in \mathcal{G}\}.$$

Definition 6.2. *Let $\mathcal{E} \subset 2^X$ be a collection of sets, $A \subset X$, $i_A : A \to X$ be the* ***inclusion map*** *($i_A(x) = x$ for all $x \in A$) and*

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}.$$

The following results will be used frequently (often without further reference) in the sequel.

Lemma 6.3 (A key measurability lemma). *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$, then*

$$\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E})). \tag{6.1}$$

In particular, if $A \subset Y$ then

$$(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A), \tag{6.2}$$

(Similar assertion hold with $\sigma(\cdot)$ being replaced by $\mathcal{A}(\cdot)$.)

Proof. Since $\mathcal{E} \subset \sigma(\mathcal{E})$, it follows that $f^{-1}(\mathcal{E}) \subset f^{-1}(\sigma(\mathcal{E}))$. Moreover, by Exercise 6.1 below, $f^{-1}(\sigma(\mathcal{E}))$ is a σ – algebra and therefore,

$$\sigma(f^{-1}(\mathcal{E})) \subset f^{-1}(\sigma(\mathcal{E})).$$

To finish the proof we must show $f^{-1}(\sigma(\mathcal{E})) \subset \sigma(f^{-1}(\mathcal{E}))$, i.e. that $f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))$ for all $B \in \sigma(\mathcal{E})$. To do this we follow the usual measure theoretic mantra, namely let

$$\mathcal{M} := \left\{B \subset Y : f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))\right\} = f_*\sigma(f^{-1}(\mathcal{E})).$$

We will now finish the proof by showing $\sigma(\mathcal{E}) \subset \mathcal{M}$. This is easily achieved by observing that $\mathcal{M}$ is a σ – algebra (see Exercise 6.1) which contains $\mathcal{E}$ and therefore $\sigma(\mathcal{E}) \subset \mathcal{M}$.

Equation (6.2) is a special case of Eq. (6.1). Indeed, $f = i_A : A \to X$ we have

$$(\sigma(\mathcal{E}))_A = i_A^{-1}(\sigma(\mathcal{E})) = \sigma(i_A^{-1}(\mathcal{E})) = \sigma(\mathcal{E}_A).$$

■

Exercise 6.1. If $f : X \to Y$ is a function and $\mathcal{F} \subset 2^Y$ and $\mathcal{B} \subset 2^X$ are σ – algebras (algebras), then $f^{-1}\mathcal{F}$ and $f_*\mathcal{B}$ are σ – algebras (algebras).

Example 6.4. Let $\mathcal{E} = \{(a, b] : -\infty < a < b < \infty\}$ and $\mathcal{B} = \sigma(\mathcal{E})$ be the Borel σ – field on $\mathbb{R}$. Then

$$\mathcal{E}_{(0,1]} = \{(a, b] : 0 \le a < b \le 1\}$$

and we have

$$\mathcal{B}_{(0,1]} = \sigma\left(\mathcal{E}_{(0,1]}\right).$$

In particular, if $A \in \mathcal{B}$ such that $A \subset (0, 1]$, then $A \in \sigma\left(\mathcal{E}_{(0,1]}\right)$.

6.1 Measurable Functions

Definition 6.5. *A* ***measurable space*** *is a pair $(X, \mathcal{M})$, where X is a set and $\mathcal{M}$ is a σ – algebra on X.*

To motivate the notion of a measurable function, suppose $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{R}_+$ is a function. Roughly speaking, we are going to define $\int_X f d\mu$ as a certain limit of sums of the form,

$$\sum_{0<a_1<a_2<a_3<\dots}^{\infty} a_i\mu(f^{-1}(a_i, a_{i+1}]).$$

For this to make sense we will need to require $f^{-1}((a, b]) \in \mathcal{M}$ for all $a < b$. Because of Corollary 6.11 below, this last condition is equivalent to the condition $f^{-1}(\mathcal{B}_\mathbb{R}) \subset \mathcal{M}$.

Definition 6.6. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is* ***measurable*** *of more precisely, $\mathcal{M}/\mathcal{F}$ – measurable or $(\mathcal{M}, \mathcal{F})$ – measurable, if $f^{-1}(\mathcal{F}) \subset \mathcal{M}$, i.e. if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{F}$.*

Remark 6.7. Let $f : X \to Y$ be a function. Given a σ – algebra $\mathcal{F} \subset 2^Y$, the σ – algebra $\mathcal{M} := f^{-1}(\mathcal{F})$ is the smallest σ – algebra on X such that f is $(\mathcal{M}, \mathcal{F})$ - measurable . Similarly, if $\mathcal{M}$ is a σ - algebra on X then

$$\mathcal{F} = f_*\mathcal{M} = \{A \in 2^Y|f^{-1}(A) \in \mathcal{M}\}$$

is the largest σ – algebra on Y such that f is $(\mathcal{M}, \mathcal{F})$ - measurable.

Example 6.8 (Indicator Functions). Let $(X, \mathcal{M})$ be a measurable space and $A \subset X$. Then 1_A is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable iff $A \in \mathcal{M}$. Indeed, $1_A^{-1}(W)$ is either $\emptyset$, X, A or A^c for any $W \subset \mathbb{R}$ with $1_A^{-1}(\{1\}) = A$.

Example 6.9. Suppose $f : X \to Y$ with Y being a finite or countable set and $\mathcal{F} = 2^Y$. Then f is measurable iff $f^{-1}(\{y\}) \in \mathcal{M}$ for all $y \in Y$.

Proposition 6.10. *Suppose that $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ are measurable spaces and further assume $\mathcal{E} \subset \mathcal{F}$ generates $\mathcal{F}$, i.e. $\mathcal{F} = \sigma(\mathcal{E})$. Then a map, $f : X \to Y$ is measurable iff $f^{-1}(\mathcal{E}) \subset \mathcal{M}$.*

Proof. If f is $\mathcal{M}/\mathcal{F}$ measurable, then $f^{-1}(\mathcal{E}) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}$. Conversely if $f^{-1}(\mathcal{E}) \subset \mathcal{M}$ then $\sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}$ and so making use of Lemma 6.3,

$$f^{-1}(\mathcal{F}) = f^{-1}(\sigma(\mathcal{E})) = \sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}.$$

■

Corollary 6.11. *Suppose that $(X, \mathcal{M})$ is a measurable space. Then the following conditions on a function $f : X \to \mathbb{R}$ are equivalent:*

1. *f is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable,*
2. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{Q}$,*
4. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*

Exercise 6.2. Prove Corollary 6.11. **Hint:** See Exercise 3.7.

Exercise 6.3. If $\mathcal{M}$ is the σ – algebra generated by $\mathcal{E} \subset 2^X$, then $\mathcal{M}$ is the union of the σ – algebras generated by countable subsets $\mathcal{F} \subset \mathcal{E}$.

Exercise 6.4. Let $(X, \mathcal{M})$ be a measure space and $f_n : X \to \mathbb{R}$ be a sequence of measurable functions on X. Show that $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \mathbf{in}\ \mathbb{R}\} \in \mathcal{M}$. Similarly show the same holds if $\mathbb{R}$ is replaced by $\mathbb{C}$.

Exercise 6.5. Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{R}})$ – measurable.

Definition 6.12. *Given measurable spaces $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ and a subset $A \subset X$. We say a function $f : A \to Y$ is measurable iff f is $\mathcal{M}_A/\mathcal{F}$ – measurable.*

Proposition 6.13 (Localizing Measurability). *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces and $f : X \to Y$ be a function.*

1. *If f is measurable and $A \subset X$ then $f|_A : A \to Y$ is $\mathcal{M}_A/\mathcal{F}$ – measurable.*
2. *Suppose there exist $A_n \in \mathcal{M}$ such that $X = \cup_{n=1}^\infty A_n$ and $f|A_n$ is $\mathcal{M}_{A_n}/\mathcal{F}$ – measurable for all n, then f is $\mathcal{M}$ – measurable.*

Proof. 1. If $f : X \to Y$ is measurable, $f^{-1}(B) \in \mathcal{M}$ for all $B \in \mathcal{F}$ and therefore

$$f|_A^{-1}(B) = A \cap f^{-1}(B) \in \mathcal{M}_A \text{ for all } B \in \mathcal{F}.$$

2. If $B \in \mathcal{F}$, then

$$f^{-1}(B) = \cup_{n=1}^\infty \left(f^{-1}(B) \cap A_n\right) = \cup_{n=1}^\infty f|_{A_n}^{-1}(B).$$

Since each $A_n \in \mathcal{M}$, $\mathcal{M}_{A_n} \subset \mathcal{M}$ and so the previous displayed equation shows $f^{-1}(B) \in \mathcal{M}$. ■

Lemma 6.14 (Composing Measurable Functions). *Suppose that $(X, \mathcal{M})$, $(Y, \mathcal{F})$ and $(Z, \mathcal{G})$ are measurable spaces. If $f : (X, \mathcal{M}) \to (Y, \mathcal{F})$ and $g : (Y, \mathcal{F}) \to (Z, \mathcal{G})$ are measurable functions then $g \circ f : (X, \mathcal{M}) \to (Z, \mathcal{G})$ is measurable as well.*

Proof. By assumption $g^{-1}(\mathcal{G}) \subset \mathcal{F}$ and $f^{-1}(\mathcal{F}) \subset \mathcal{M}$ so that

$$(g \circ f)^{-1}(\mathcal{G}) = f^{-1}\left(g^{-1}(\mathcal{G})\right) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}.$$

■

Definition 6.15 (σ – Algebras Generated by Functions). *Let X be a set and suppose there is a collection of measurable spaces $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ and functions $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$. Let $\sigma(f_\alpha : \alpha \in I)$ denote the smallest σ – algebra on X such that each f_α is measurable, i.e.*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(\cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)).$$

Example 6.16. Suppose that Y is a finite set, $\mathcal{F} = 2^Y$, and $X = Y^N$ for some $N \in \mathbb{N}$. Let $\pi_i : Y^N \to Y$ be the projection maps, $\pi_i(y_1, \dots, y_N) = y_i$. Then, as the reader should check,

$$\sigma(\pi_1, \dots, \pi_n) = \left\{A \times \Lambda^{N-n} : A \subset \Lambda^n\right\}.$$

Proposition 6.17. *Assuming the notation in Definition 6.15 (so $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$) and additionally let $(Z, \mathcal{M})$ be a measurable space. Then $g : Z \to X$ is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ g$ $\left(Z \xrightarrow{g} X \xrightarrow{f_\alpha} Y_\alpha\right)$ is $(\mathcal{M}, \mathcal{F}_\alpha)$–measurable for all $\alpha \in I$.*

Proof. $(\Rightarrow)$ If g is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in I))$ – measurable, then the composition $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable by Lemma 6.14.

($\Leftarrow$) Since $\sigma(f_\alpha : \alpha \in I) = \sigma(\mathcal{E})$ where $\mathcal{E} := \cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)$, according to Proposition 6.10, it suffices to show $g^{-1}(A) \in \mathcal{M}$ for $A \in f_\alpha^{-1}(\mathcal{F}_\alpha)$. But this is true since if $A = f_\alpha^{-1}(B)$ for some $B \in \mathcal{F}_\alpha$, then $g^{-1}(A) = g^{-1}\left(f_\alpha^{-1}(B)\right) = (f_\alpha \circ g)^{-1}(B) \in \mathcal{M}$ because $f_\alpha \circ g : Z \to Y_\alpha$ is assumed to be measurable. ■

Definition 6.18. *If $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, then the product measure space, $(Y, \mathcal{F})$, is $Y := \prod_{\alpha \in I} Y_\alpha$, $\mathcal{F} := \sigma(\pi_\alpha : \alpha \in I)$ where $\pi_\alpha : Y \to Y_\alpha$ is the α – component projection. We call $\mathcal{F}$ the product σ – algebra and denote it by, $\mathcal{F} = \otimes_{\alpha \in I} \mathcal{F}_\alpha$.*

Let us record an important special case of Proposition 6.17.

Corollary 6.19. *If $(Z, \mathcal{M})$ is a measure space, then $g : Z \to Y = \prod_{\alpha \in I} Y_\alpha$ is $(\mathcal{M}, \mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha)$ – measurable iff $\pi_\alpha \circ g : Z \to Y_\alpha$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$.*

As a special case of the above corollary, if $A = \{1, 2, \ldots, n\}$, then $Y = Y_1 \times \cdots \times Y_n$ and $g = (g_1, \ldots, g_n) : Z \to Y$ is measurable iff each component, $g_i : Z \to Y_i$, is measurable. Here is another closely related result.

Proposition 6.20. *Suppose X is a set, $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, and we are given maps, $f_\alpha : X \to Y_\alpha$, for all $\alpha \in I$. If $f : X \to Y := \prod_{\alpha \in I} Y_\alpha$ is the unique map, such that $\pi_\alpha \circ f = f_\alpha$, then*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(f) = f^{-1}(\mathcal{F})$$

where $\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha$.

Proof. Since $\pi_\alpha \circ f = f_\alpha$ is $\sigma(f_\alpha : \alpha \in I) / \mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ it follows from Corollary 6.19 that $f : X \to Y$ is $\sigma(f_\alpha : \alpha \in I) / \mathcal{F}$ – measurable. Since $\sigma(f)$ is the smallest σ – algebra on X such that f is measurable we may conclude that $\sigma(f) \subset \sigma(f_\alpha : \alpha \in I)$.

Conversely, for each $\alpha \in I$, $f_\alpha = \pi_\alpha \circ f$ is $\sigma(f) / \mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ being the composition of two measurable functions. Since $\sigma(f_\alpha : \alpha \in I)$ is the smallest σ – algebra on X such that each $f_\alpha : X \to Y_\alpha$ is measurable, we learn that $\sigma(f_\alpha : \alpha \in I) \subset \sigma(f)$. ■

Exercise 6.6. Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\mathcal{E}_i$ is a subset of $\mathcal{F}_i$ such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for $i = 1$ and 2. Show $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{E})$ where $\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{E}_i \text{ for } i = 1, 2\}$. **Hints:**

1. First show that if Y is a set and $\mathcal{S}_1$ and $\mathcal{S}_2$ are two non-empty subsets of 2^Y, then $\sigma(\sigma(\mathcal{S}_1) \cup \sigma(\mathcal{S}_2)) = \sigma(\mathcal{S}_1 \cup \mathcal{S}_2)$. (In fact, one has that $\sigma(\cup_{\alpha \in I} \sigma(\mathcal{S}_\alpha)) = \sigma(\cup_{\alpha \in I} \mathcal{S}_\alpha)$ for any collection of non-empty subsets, $\{\mathcal{S}_\alpha\}_{\alpha \in I} \subset 2^Y$.)
2. After this you might start your proof as follows;

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\left(\pi_1^{-1}(\mathcal{F}_1) \cup \pi_2^{-1}(\mathcal{F}_2)\right) = \sigma\left(\pi_1^{-1}(\sigma(\mathcal{E}_2)) \cup \pi_2^{-1}(\sigma(\mathcal{E}_2))\right) = \ldots.$$

Remark 6.21. The reader should convince herself that Exercise 6.6 admits the following extension. If I is any finite or countable index set, $\{(Y_i, \mathcal{F}_i)\}_{i \in I}$ are measurable spaces and $\mathcal{E}_i \subset \mathcal{F}_i$ are such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for all $i \in I$, then

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{E}_j \text{ for all } j \in I\right\}\right)$$

and in particular,

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{F}_j \text{ for all } j \in I\right\}\right).$$

The last fact is easily verified directly without the aid of Exercise 6.6.

Exercise 6.7. Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\emptyset \neq B_i \subset Y_i$ for $i = 1, 2$. Show

$$[\mathcal{F}_1 \otimes \mathcal{F}_2]_{B_1 \times B_2} = [\mathcal{F}_1]_{B_1} \otimes [\mathcal{F}_2]_{B_2}.$$

Hint: you may find it useful to use the result of Exercise 6.6 with

$$\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{F}_i \text{ for } i = 1, 2\}.$$

Definition 6.22. *A function $f : X \to Y$ between two topological spaces is **Borel measurable** if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

Proposition 6.23. *Let X and Y be two topological spaces and $f : X \to Y$ be a continuous function. Then f is Borel measurable.*

Proof. Using Lemma 6.3 and $\mathcal{B}_Y = \sigma(\tau_Y)$,

$$f^{-1}(\mathcal{B}_Y) = f^{-1}(\sigma(\tau_Y)) = \sigma(f^{-1}(\tau_Y)) \subset \sigma(\tau_X) = \mathcal{B}_X.$$

■

Example 6.24. For $i = 1, 2, \ldots, n$, let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be defined by $\pi_i(x) = x_i$. Then each π_i is continuous and therefore $\mathcal{B}_{\mathbb{R}^n} / \mathcal{B}_{\mathbb{R}}$ – measurable.

Lemma 6.25. *Let $\mathcal{E}$ denote the collection of open rectangle in $\mathbb{R}^n$, then $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E})$. We also have that $\mathcal{B}_{\mathbb{R}^n} = \sigma(\pi_1, \ldots, \pi_n) = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}$ and in particular, $A_1 \times \cdots \times A_n \in \mathcal{B}_{\mathbb{R}^n}$ whenever $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$. Therefore $\mathcal{B}_{\mathbb{R}^n}$ may be described as the σ algebra generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$. (Also see Remark 6.21.)*

Proof. Assertion 1. Since $\mathcal{E} \subset \mathcal{B}_{\mathbb{R}^n}$, it follows that $\sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}$. Let

$$\mathcal{E}_0 := \{(a,b) : a,b \in \mathbb{Q}^n \ni a < b\},$$

where, for $a,b \in \mathbb{R}^n$, we write $a < b$ iff $a_i < b_i$ for $i = 1,2,\dots,n$ and let

$$(a,b) = (a_1,b_1) \times \cdots \times (a_n,b_n). \tag{6.3}$$

Since every open set, $V \subset \mathbb{R}^n$, may be written as a (necessarily) countable union of elements from $\mathcal{E}_0$, we have

$$V \in \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}),$$

i.e. $\sigma(\mathcal{E}_0)$ and hence $\sigma(\mathcal{E})$ contains all open subsets of $\mathbb{R}^n$. Hence we may conclude that

$$\mathcal{B}_{\mathbb{R}^n} = \sigma(\text{open sets}) \subset \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}.$$

Assertion 2. Since each $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is continuous, it is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable and therefore, $\sigma(\pi_1,\dots,\pi_n) \subset \mathcal{B}_{\mathbb{R}^n}$. Moreover, if (a,b) is as in Eq. (6.3), then

$$(a,b) = \cap_{i=1}^n \pi_i^{-1}((a_i,b_i)) \in \sigma(\pi_1,\dots,\pi_n).$$

Therefore, $\mathcal{E} \subset \sigma(\pi_1,\dots,\pi_n)$ and $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E}) \subset \sigma(\pi_1,\dots,\pi_n)$.

Assertion 3. If $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1,2,\dots,n$, then

$$A_1 \times \cdots \times A_n = \cap_{i=1}^n \pi_i^{-1}(A_i) \in \sigma(\pi_1,\dots,\pi_n) = \mathcal{B}_{\mathbb{R}^n}.$$

■

Corollary 6.26. *If $(X,\mathcal{M})$ is a measurable space, then*

$$f = (f_1,f_2,\dots,f_n) : X \to \mathbb{R}^n$$

is $(\mathcal{M},\mathcal{B}_{\mathbb{R}^n})$ – measurable iff $f_i : X \to \mathbb{R}$ is $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable for each i. In particular, a function $f : X \to \mathbb{C}$ is $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable iff $\operatorname{Re} f$ *and* $\operatorname{Im} f$ *are $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable.*

Proof. This is an application of Lemma 6.25 and Corollary 6.19 with $Y_i = \mathbb{R}$ for each i. ■

Corollary 6.27. *Let $(X,\mathcal{M})$ be a measurable space and $f,g : X \to \mathbb{C}$ be $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable functions. Then $f \pm g$ and $f \cdot g$ are also $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable.*

Proof. Define $F : X \to \mathbb{C} \times \mathbb{C}$, $A_\pm : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $M : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ by $F(x) = (f(x), g(x))$, $A_\pm(w,z) = w \pm z$ and $M(w,z) = wz$. Then $A_\pm$ and M are continuous and hence $(\mathcal{B}_{\mathbb{C}^2},\mathcal{B}_{\mathbb{C}})$ – measurable. Also F is $(\mathcal{M},\mathcal{B}_{\mathbb{C}^2})$ – measurable since $\pi_1 \circ F = f$ and $\pi_2 \circ F = g$ are $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable. Therefore $A_\pm \circ F = f \pm g$ and $M \circ F = f \cdot g$, being the composition of measurable functions, are also measurable. ■

Lemma 6.28. *Let $\alpha \in \mathbb{C}$, $(X,\mathcal{M})$ be a measurable space and $f : X \to \mathbb{C}$ be a $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable function. Then*

$$F(x) := \begin{cases} \frac{1}{f(x)} & \text{if } f(x) \neq 0 \\ \alpha & \text{if } f(x) = 0 \end{cases}$$

is measurable.

Proof. Define $i : \mathbb{C} \to \mathbb{C}$ by

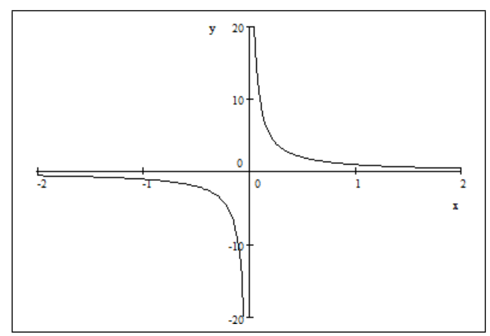
$$i(z) = \begin{cases} \frac{1}{z} & \text{if } z \neq 0 \\ 0 & \text{if } z = 0. \end{cases}$$

For any open set $V \subset \mathbb{C}$ we have

$$i^{-1}(V) = i^{-1}(V \setminus \{0\}) \cup i^{-1}(V \cap \{0\})$$

Because i is continuous except at $z = 0$, $i^{-1}(V \setminus \{0\})$ is an open set and hence in $\mathcal{B}_{\mathbb{C}}$. Moreover, $i^{-1}(V \cap \{0\}) \in \mathcal{B}_{\mathbb{C}}$ since $i^{-1}(V \cap \{0\})$ is either the empty set or the one point set $\{0\}$. Therefore $i^{-1}(\tau_{\mathbb{C}}) \subset \mathcal{B}_{\mathbb{C}}$ and hence $i^{-1}(\mathcal{B}_{\mathbb{C}}) = i^{-1}(\sigma(\tau_{\mathbb{C}})) = \sigma(i^{-1}(\tau_{\mathbb{C}})) \subset \mathcal{B}_{\mathbb{C}}$ which shows that i is Borel measurable. Since $F = i \circ f$ is the composition of measurable functions, F is also measurable. ■

Remark 6.29. For the real case of Lemma 6.28, define i as above but now take z to real. From the plot of i, Figure 6.29, the reader may easily verify that $i^{-1}((-\infty,a])$ is an infinite half interval for all a and therefore i is measurable. See Example 6.34 for another proof of this fact.



We will often deal with functions $f : X \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. When talking about measurability in this context we will refer to the σ – algebra on $\bar{\mathbb{R}}$ defined by

$$\mathcal{B}_{\bar{\mathbb{R}}} := \sigma\left(\{[a, \infty] : a \in \mathbb{R}\}\right). \tag{6.4}$$

Proposition 6.30 (The Structure of $\mathcal{B}_{\bar{\mathbb{R}}}$). *Let $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\bar{\mathbb{R}}}$ be as above, then*

$$\mathcal{B}_{\bar{\mathbb{R}}} = \{A \subset \bar{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}. \tag{6.5}$$

In particular $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ and $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}}$.

Proof. Let us first observe that

$$\{-\infty\} = \cap_{n=1}^{\infty}[-\infty, -n) = \cap_{n=1}^{\infty}[-n, \infty]^c \in \mathcal{B}_{\bar{\mathbb{R}}},$$
$$\{\infty\} = \cap_{n=1}^{\infty}[n, \infty] \in \mathcal{B}_{\bar{\mathbb{R}}} \text{ and } \mathbb{R} = \bar{\mathbb{R}} \setminus \{\pm\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}.$$

Letting $i : \mathbb{R} \to \bar{\mathbb{R}}$ be the inclusion map,

$$\begin{aligned} i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) &= \sigma\left(i^{-1}\left(\left\{[a, \infty] : a \in \bar{\mathbb{R}}\right\}\right)\right) = \sigma\left(\left\{i^{-1}\left([a, \infty]\right) : a \in \bar{\mathbb{R}}\right\}\right) \\ &= \sigma\left(\left\{[a, \infty] \cap \mathbb{R} : a \in \bar{\mathbb{R}}\right\}\right) = \sigma\left(\{[a, \infty) : a \in \mathbb{R}\}\right) = \mathcal{B}_{\mathbb{R}}. \end{aligned}$$

Thus we have shown

$$\mathcal{B}_{\mathbb{R}} = i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) = \{A \cap \mathbb{R} : A \in \mathcal{B}_{\bar{\mathbb{R}}}\}.$$

This implies:

1. $A \in \mathcal{B}_{\bar{\mathbb{R}}} \Longrightarrow A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ and
2. if $A \subset \bar{\mathbb{R}}$ is such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ there exists $B \in \mathcal{B}_{\bar{\mathbb{R}}}$ such that $A \cap \mathbb{R} = B \cap \mathbb{R}$. Because $A \Delta B \subset \{\pm\infty\}$ and $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ we may conclude that $A \in \mathcal{B}_{\bar{\mathbb{R}}}$ as well.

This proves Eq. (6.5). ■

The proofs of the next two corollaries are left to the reader, see Exercises 6.8 and 6.9.

Corollary 6.31. *Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \bar{\mathbb{R}}$ be a function. Then the following are equivalent*

1. *f is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ - measurable,*
2. *$f^{-1}((a, \infty]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
4. *$f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$ and $f^0 : X \to \mathbb{R}$ defined by*

$$f^0(x) := \begin{cases} f(x) & \textit{if} \quad f(x) \in \mathbb{R} \\ 0 & \textit{if}\, f(x) \in \{\pm\infty\} \end{cases}$$

is measurable.

Corollary 6.32. *Let $(X, \mathcal{M})$ be a measurable space, $f, g : X \to \bar{\mathbb{R}}$ be functions and define $f \cdot g : X \to \bar{\mathbb{R}}$ and $(f + g) : X \to \bar{\mathbb{R}}$ using the conventions, $0 \cdot \infty = 0$ and $(f + g)(x) = 0$ if $f(x) = \infty$ and $g(x) = -\infty$ or $f(x) = -\infty$ and $g(x) = \infty$. Then $f \cdot g$ and $f + g$ are measurable functions on X if both f and g are measurable.*

Exercise 6.8. Prove Corollary 6.31 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 6.11. Use Proposition 6.30 to handle item 4.

Exercise 6.9. Prove Corollary 6.32.

Proposition 6.33 (Closure under sups, infs and limits). *Suppose that $(X, \mathcal{M})$ is a measurable space and $f_j : (X, \mathcal{M}) \to \overline{\mathbb{R}}$ for $j \in \mathbb{N}$ is a sequence of $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. Then*

$$\sup{}_j f_j, \quad \inf{}_j f_j, \quad \limsup_{j\to\infty} \; f_j \text{ and } \liminf_{j\to\infty} \; f_j$$

are all $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. (Note that this result is in generally false when $(X, \mathcal{M})$ is a topological space and measurable is replaced by continuous in the statement.)

Proof. Define $g_+(x) := \sup{}_j \; f_j(x)$, then

$$\begin{aligned} \{x : g_+(x) \le a\} &= \{x : f_j(x) \le a \; \forall \, j\} \\ &= \cap_j \{x : f_j(x) \le a\} \in \mathcal{M} \end{aligned}$$

so that g_+ is measurable. Similarly if $g_-(x) = \inf_j f_j(x)$ then

$$\{x : g_-(x) \ge a\} = \cap_j \{x : f_j(x) \ge a\} \in \mathcal{M}.$$

Since

$$\begin{aligned} \limsup_{j\to\infty} \; f_j &= \inf_n \sup \{f_j : j \ge n\} \text{ and} \\ \liminf_{j\to\infty} \; f_j &= \sup_n \inf \{f_j : j \ge n\} \end{aligned}$$

we are done by what we have already proved. ■

Example 6.34. As we saw in Remark 6.29, $i : \mathbb{R} \to \mathbb{R}$ defined by

$$i(z) = \begin{cases} \frac{1}{z} \text{ if } & z \neq 0 \\ 0 \text{ if } & z = 0. \end{cases}$$

is measurable by a simple direct argument. For an alternative argument, let

$$i_n(z) := \frac{z}{z^2 + \frac{1}{n}} \text{ for all } n \in \mathbb{N}.$$

Then i_n is continuous and $\lim_{n\to\infty} i_n(z) = i(z)$ for all $z \in \mathbb{R}$ from which it follows that i is Borel measurable.

Example 6.35. Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of the points in $\mathbb{Q}\cap[0,1]$ and define

$$f(x)=\sum_{n=1}^{\infty}2^{-n}\frac{1}{\sqrt{|x-r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x-r_n|}}=5 \text{ if } x=r_n.$$

Then $f:\mathbb{R}\to\bar{\mathbb{R}}$ is measurable. Indeed, if

$$g_n(x)=\begin{cases}\frac{1}{\sqrt{|x-r_n|}} & \text{if } x\neq r_n\\ 0 & \text{if } x=r_n\end{cases}$$

then $g_n(x)=\sqrt{|i(x-r_n)|}$ is measurable as the composition of measurable is measurable. Therefore $g_n+5\cdot 1_{\{r_n\}}$ is measurable as well. Finally,

$$f(x)=\lim_{N\to\infty}\sum_{n=1}^{N}2^{-n}\frac{1}{\sqrt{|x-r_n|}}$$

is measurable since sums of measurable functions are measurable and limits of measurable functions are measurable. **Moral:** if you can explicitly write a function $f:\bar{\mathbb{R}}\to\bar{\mathbb{R}}$ down then it is going to be measurable.

Definition 6.36. *Given a function $f:X\to\bar{\mathbb{R}}$ let $f_+(x):=\max\{f(x),0\}$ and $f_-(x):=\max(-f(x),0)=-\min(f(x),0)$. Notice that $f=f_+-f_-$.*

Corollary 6.37. *Suppose $(X,\mathcal{M})$ is a measurable space and $f:X\to\bar{\mathbb{R}}$ is a function. Then f is measurable iff $f_\pm$ are measurable.*

Proof. If f is measurable, then Proposition 6.33 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f=f_+-f_-$. ∎

Definition 6.38. *Let $(X,\mathcal{M})$ be a measurable space. A function $\varphi:X\to\mathbb{F}$ ($\mathbb{F}$ denotes either $\mathbb{R}$, $\mathbb{C}$ or $[0,\infty]\subset\bar{\mathbb{R}}$) is a* ***simple function*** *if φ is $\mathcal{M}$ – $\mathcal{B}_{\mathbb{F}}$ measurable and $\varphi(X)$ contains only finitely many elements.*

Any such simple functions can be written as

$$\varphi=\sum_{i=1}^{n}\lambda_i 1_{A_i} \text{ with } A_i\in\mathcal{M} \text{ and } \lambda_i\in\mathbb{F}. \tag{6.6}$$

Indeed, take $\lambda_1,\lambda_2,\dots,\lambda_n$ to be an enumeration of the range of φ and $A_i=\varphi^{-1}(\{\lambda_i\})$. Note that this argument shows that any simple function may be written intrinsically as

$$\varphi=\sum_{y\in\mathbb{F}}y1_{\varphi^{-1}(\{y\})}. \tag{6.7}$$

The next theorem shows that simple functions are "pointwise dense" in the space of measurable functions.

Theorem 6.39 (Approximation Theorem). *Let $f:X\to[0,\infty]$ be measurable and define, see Figure 6.1,*

$$\begin{aligned}\varphi_n(x)&:=\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{f^{-1}\left(\left(\frac{k}{2^n},\frac{k+1}{2^n}\right]\right)}(x)+2^n1_{f^{-1}((2^n,\infty])}(x)\\ &=\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{\left\{\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right\}}(x)+2^n1_{\{f>2^n\}}(x)\end{aligned}$$

then $\varphi_n\le f$ for all n, $\varphi_n(x)\uparrow f(x)$ for all $x\in X$ and $\varphi_n\uparrow f$ uniformly on the sets $X_M:=\{x\in X:f(x)\le M\}$ with $M<\infty$.

Moreover, if $f:X\to\mathbb{C}$ is a measurable function, then there exists simple functions φ_n such that $\lim_{n\to\infty}\varphi_n(x)=f(x)$ for all x and $|\varphi_n|\uparrow|f|$ as $n\to\infty$.

Proof. Since $f^{-1}\left(\left(\frac{k}{2^n},\frac{k+1}{2^n}\right]\right)$ and $f^{-1}((2^n,\infty])$ are in $\mathcal{M}$ as f is measurable, φ_n is a measurable simple function for each n. Because

$$\left(\frac{k}{2^n},\frac{k+1}{2^n}\right]=\left(\frac{2k}{2^{n+1}},\frac{2k+1}{2^{n+1}}\right]\cup\left(\frac{2k+1}{2^{n+1}},\frac{2k+2}{2^{n+1}}\right],$$

if $x\in f^{-1}\left(\left(\frac{2k}{2^{n+1}},\frac{2k+1}{2^{n+1}}\right]\right)$ then $\varphi_n(x)=\varphi_{n+1}(x)=\frac{2k}{2^{n+1}}$ and if $x\in f^{-1}\left(\left(\frac{2k+1}{2^{n+1}},\frac{2k+2}{2^{n+1}}\right]\right)$ then $\varphi_n(x)=\frac{2k}{2^{n+1}}<\frac{2k+1}{2^{n+1}}=\varphi_{n+1}(x)$. Similarly

$$(2^n,\infty]=(2^n,2^{n+1}]\cup(2^{n+1},\infty],$$

and so for $x\in f^{-1}((2^{n+1},\infty])$, $\varphi_n(x)=2^n<2^{n+1}=\varphi_{n+1}(x)$ and for $x\in f^{-1}((2^n,2^{n+1}])$, $\varphi_{n+1}(x)\ge 2^n=\varphi_n(x)$. Therefore $\varphi_n\le\varphi_{n+1}$ for all n. It is clear by construction that $0\le\varphi_n(x)\le f(x)$ for all x and that $0\le f(x)-\varphi_n(x)\le 2^{-n}$ if $x\in X_{2^n}=\{f\le 2^n\}$. Hence we have shown that $\varphi_n(x)\uparrow f(x)$ for all $x\in X$ and $\varphi_n\uparrow f$ uniformly on bounded sets.

For the second assertion, first assume that $f:X\to\mathbb{R}$ is a measurable function and choose $\varphi_n^\pm$ to be non-negative simple functions such that $\varphi_n^\pm\uparrow f_\pm$ as $n\to\infty$ and define $\varphi_n=\varphi_n^+-\varphi_n^-$. Then (using $\varphi_n^+\cdot\varphi_n^-\le f_+\cdot f_-=0$)

$$|\varphi_n|=\varphi_n^++\varphi_n^-\le\varphi_{n+1}^++\varphi_{n+1}^-=|\varphi_{n+1}|$$

and clearly $|\varphi_n|=\varphi_n^++\varphi_n^-\uparrow f_++f_-=|f|$ and $\varphi_n=\varphi_n^+-\varphi_n^-\to f_+-f_-=f$ as $n\to\infty$. Now suppose that $f:X\to\mathbb{C}$ is measurable. We may now choose simple function u_n and v_n such that $|u_n|\uparrow|\operatorname{Re}f|$, $|v_n|\uparrow|\operatorname{Im}f|$, $u_n\to\operatorname{Re}f$ and $v_n\to\operatorname{Im}f$ as $n\to\infty$. Let $\varphi_n=u_n+iv_n$, then

$$|\varphi_n|^2=u_n^2+v_n^2\uparrow|\operatorname{Re}f|^2+|\operatorname{Im}f|^2=|f|^2$$

and $\varphi_n=u_n+iv_n\to\operatorname{Re}f+i\operatorname{Im}f=f$ as $n\to\infty$. ∎

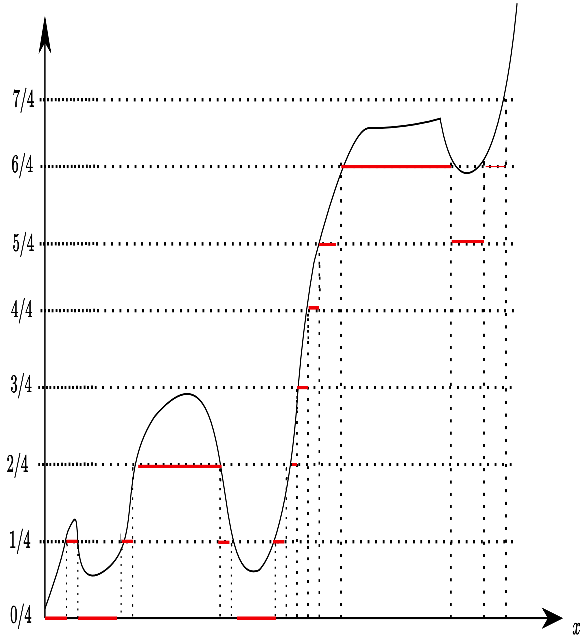


Fig. 6.1. Constructing the simple function, φ_2, approximating a function, $f : X \to [0,\infty]$. The graph of φ_2 is in red.

6.2 Factoring Random Variables

Lemma 6.40. *Suppose that $(\mathbb{Y},\mathcal{F})$ is a measurable space and $Y : \Omega \to \mathbb{Y}$ is a map. Then to every $(\sigma(Y),\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $h : \Omega \to \bar{\mathbb{R}}$, there is a $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function $H : \mathbb{Y} \to \bar{\mathbb{R}}$ such that $h = H \circ Y$. More generally, $\bar{\mathbb{R}}$ may be replaced by any "standard Borel space,"[1] i.e. a space, $(S,\mathcal{B}_S)$ which is measure theoretic isomorphic to a Borel subset of $\mathbb{R}$.*

$$\begin{tikzcd}
(\Omega,\sigma(Y)) \arrow[r, "Y"] \arrow[d, "h"'] & (\mathbb{Y},\mathcal{F}) \arrow[dl, dashed, "H"] \\
(S,\mathcal{B}_S) &
\end{tikzcd}$$

Proof. First suppose that $h = 1_A$ where $A \in \sigma(Y) = Y^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = Y^{-1}(B)$ then $1_A = 1_{Y^{-1}(B)} = 1_B \circ Y$ and hence the lemma is valid in this case with $H = 1_B$. More generally if $h = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ Y$ and hence $h = H \circ Y$ with $H := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$.

For a general $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, h, from $\Omega \to \bar{\mathbb{R}}$, choose simple functions h_n converging to h. Let $H_n : \mathbb{Y} \to \bar{\mathbb{R}}$ be simple functions such that $h_n = H_n \circ Y$. Then it follows that

$$h = \lim_{n\to\infty} h_n = \limsup_{n\to\infty} h_n = \limsup_{n\to\infty} H_n \circ Y = H \circ Y$$

where $H := \limsup_{n\to\infty} H_n$ – a measurable function from $\mathbb{Y}$ to $\bar{\mathbb{R}}$.

For the last assertion we may assume that $S \in \mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_S = (\mathcal{B}_{\mathbb{R}})_S = \{A \cap S : A \in \mathcal{B}_{\mathbb{R}}\}$. Since $i_S : S \to \mathbb{R}$ is measurable, what we have just proved shows there exists, $H : \mathbb{Y} \to \bar{\mathbb{R}}$ which is $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable such that $h = i_S \circ h = H \circ Y$. The only problems with H is that $H(\mathbb{Y})$ may not be contained in S. To fix this, let

$$H_S = \begin{cases} H|_{H^{-1}(S)} & \text{on} \quad H^{-1}(S) \\ * & \text{on } \mathbb{Y} \setminus H^{-1}(S) \end{cases}$$

where $*$ is some fixed arbitrary point in S. It follows from Proposition 6.13 that $H_S : \mathbb{Y} \to S$ is $(\mathcal{F},\mathcal{B}_S)$ – measurable and we still have $h = H_S \circ Y$ as the range of Y must necessarily be in $H^{-1}(S)$. ■

Here is how this lemma will often be used in these notes.

Corollary 6.41. *Suppose that $(\Omega,\mathcal{B})$ is a measurable space, $X_n : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_{\mathbb{R}}$ – measurable functions, and $\mathcal{B}_n := \sigma(X_1,\dots,X_n) \subset \mathcal{B}$ for each $n \in \mathbb{N}$. Then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff there exists $H : \mathbb{R}^n \to \mathbb{R}$ which is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable such that $h = H(X_1,\dots,X_n)$.*

$$\begin{tikzcd}
(\Omega,\mathcal{B}_n = \sigma(Y)) \arrow[r, "Y:=(X_1,\dots,X_n)"] \arrow[d, "h"'] & (\mathbb{R}^n,\mathcal{B}_{\mathbb{R}^n}) \arrow[dl, dashed, "H"] \\
(\mathbb{R},\mathcal{B}_{\mathbb{R}}) &
\end{tikzcd}$$

Proof. By Lemma 6.25 and Corollary 6.19, the map, $Y := (X_1,\dots,X_n) : \Omega \to \mathbb{R}^n$ is $(\mathcal{B},\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}})$ – measurable and by Proposition 6.20, $\mathcal{B}_n = \sigma(X_1,\dots,X_n) = \sigma(Y)$. Thus we may apply Lemma 6.40 to see that there exists a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$, such that $h = H \circ Y = H(X_1,\dots,X_n)$. ■

[1] Standard Borel spaces include almost any measurable space that we will consider in these notes. For example they include all complete seperable metric spaces equipped with the Borel σ – algebra, see Section 9.10.

6.3 Summary of Measurability Statements

It may be worthwhile to gather the statements of the main measurability results of Sections 6.1 and 6.2 in one place. To do this let $(\Omega, \mathcal{B})$, $(X, \mathcal{M})$, and $\{(Y_\alpha, \mathcal{F}_\alpha)\}_{\alpha \in I}$ be measurable spaces and $f_\alpha : \Omega \to Y_\alpha$ be given maps for all $\alpha \in I$. Also let $\pi_\alpha : Y \to Y_\alpha$ be the α – projection map,

$$\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha := \sigma(\pi_\alpha : \alpha \in I)$$

be the product σ – algebra on Y, and $f : \Omega \to Y$ be the unique map determined by $\pi_\alpha \circ f = f_\alpha$ for all $\alpha \in I$. Then the following measurability results hold;

1. For $A \subset \Omega$, the indicator function, 1_A, is $(\mathcal{B}, \mathcal{B}_\mathbb{R})$ – measurable iff $A \in \mathcal{B}$. (Example 6.8).
2. If $\mathcal{E} \subset \mathcal{M}$ generates $\mathcal{M}$ (i.e. $\mathcal{M} = \sigma(\mathcal{E})$), then a map, $g : \Omega \to X$ is $(\mathcal{B}, \mathcal{M})$ – measurable iff $g^{-1}(\mathcal{E}) \subset \mathcal{B}$ (Lemma 6.3 and Proposition 6.10).
3. The notion of measurability may be localized (Proposition 6.13).
4. Composition of measurable functions are measurable (Lemma 6.14).
5. Continuous functions between two topological spaces are also Borel measurable (Proposition 6.23).
6. $\sigma(f) = \sigma(f_\alpha : \alpha \in I)$ (Proposition 6.20).
7. A map, $h : X \to \Omega$ is $(\mathcal{M}, \sigma(f) = \sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Proposition 6.17).
8. A map, $h : X \to Y$ is $(\mathcal{M}, \mathcal{F})$ – measurable iff $\pi_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Corollary 6.19).
9. If $I = \{1, 2, \dots, n\}$, then
$$\otimes_{\alpha \in I} \mathcal{F}_\alpha = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \sigma(\{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathcal{F}_i \text{ for } i \in I\}),$$
this is a special case of Remark 6.21.
10. $\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_\mathbb{R} \otimes \cdots \otimes \mathcal{B}_\mathbb{R}$ (n - times) for all $n \in \mathbb{N}$, i.e. the Borel σ – algebra on $\mathbb{R}^n$ is the same as the product σ – algebra. (Lemma 6.25).
11. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ is closed under the usual pointwise algebraic operations (Corollary 6.32). They are also closed under the countable supremums, infimums, and limits (Proposition 6.33).
12. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_\mathbb{C})$ is closed under the usual pointwise algebraic operations and countable limits. (Corollary 6.27 and Proposition 6.33). The limiting assertion follows by considering the real and imaginary parts of all functions involved.
13. The class of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ and from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_\mathbb{C})$ may be well approximated by measurable simple functions (Theorem 6.39).
14. If $X_i : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_\mathbb{R}$ – measurable maps and $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$, then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff $h = H(X_1, \dots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_\mathbb{R}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$ (Corollary 6.41).
15. We also have the more general factorization Lemma 6.40.

For the most part most of our future measurability issues can be resolved by one or more of the items on this list.

6.4 Distributions / Laws of Random Vectors

The proof of the following proposition is routine and will be left to the reader.

Proposition 6.42. *Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Define a function $\nu : \mathcal{F} \to [0, \infty]$ by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$. Then ν is a measure on $(Y, \mathcal{F})$. (In the future we will denote ν by $f_*\mu$ or $\mu \circ f^{-1}$ or $\mathrm{Law}_\mu(f)$ and call $f_*\mu$ the **push-forward of μ by f** or the **law of f under μ**.*

Definition 6.43. *Suppose that $\{X_i\}_{i=1}^n$ is a sequence of random variables on a probability space, $(\Omega, \mathcal{B}, P)$. The probability measure,*

$$\mu = (X_1, \ldots, X_n)_* P = P \circ (X_1, \ldots, X_n)^{-1} \text{ on } \mathcal{B}_{\mathbb{R}}$$

*(see Proposition 6.42) is called the **joint distribution** (or **law**) of $(X_1, \ldots, X_n)$. To be more explicit,*

$$\mu(B) := P((X_1, \ldots, X_n) \in B) := P(\{\omega \in \Omega : (X_1(\omega), \ldots, X_n(\omega)) \in B\})$$

for all $B \in \mathcal{B}_{\mathbb{R}^n}$.

Corollary 6.44. *The joint distribution, μ is uniquely determined from the knowledge of*

$$P((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n) \text{ for all } A_i \in \mathcal{B}_{\mathbb{R}}$$

or from the knowledge of

$$P(X_1 \le x_1, \ldots, X_n \le x_n) \text{ for all } A_i \in \mathcal{B}_{\mathbb{R}}$$

for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Proof. Apply Proposition 5.15 with $\mathcal{P}$ being the π – systems defined by

$$\mathcal{P} := \{A_1 \times \cdots \times A_n \in \mathcal{B}_{\mathbb{R}^n} : A_i \in \mathcal{B}_{\mathbb{R}}\}$$

for the first case and

$$\mathcal{P} := \{(-\infty, x_1] \times \cdots \times (-\infty, x_n] \in \mathcal{B}_{\mathbb{R}^n} : x_i \in \mathbb{R}\}$$

for the second case. ■

Definition 6.45. *Suppose that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are two finite sequences of random variables on two probability spaces, $(\Omega, \mathcal{B}, P)$ and $(\Omega', \mathcal{B}', P')$ respectively. We write $(X_1, \ldots, X_n) \overset{d}{=} (Y_1, \ldots, Y_n)$ if $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ have the **same distribution / law**, i.e. if*

$$P((X_1, \ldots, X_n) \in B) = P'((Y_1, \ldots, Y_n) \in B) \text{ for all } B \in \mathcal{B}_{\mathbb{R}^n}.$$

More generally, if $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ are two sequences of random variables on two probability spaces, $(\Omega, \mathcal{B}, P)$ and $(\Omega', \mathcal{B}', P')$ we write $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$ iff $(X_1, \ldots, X_n) \overset{d}{=} (Y_1, \ldots, Y_n)$ for all $n \in \mathbb{N}$.

Proposition 6.46. *Let us continue using the notation in Definition 6.45. Further let*

$$X = (X_1, X_2, \ldots) : \Omega \to \mathbb{R}^{\mathbb{N}} \text{ and } Y := (Y_1, Y_2, \ldots) : \Omega' \to \mathbb{R}^{\mathbb{N}}$$

and let $\mathcal{F} := \otimes_{n \in \mathbb{N}} \mathcal{B}_{\mathbb{R}}$ – be the product σ – algebra on $\mathbb{R}^{\mathbb{N}}$. Then $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$ iff $X_* P = Y_* P'$ as measures on $(\mathbb{R}^{\mathbb{N}}, \mathcal{F})$.

Proof. Let

$$\mathcal{P} := \cup_{n=1}^\infty \left\{A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^{\mathbb{N}} : A_i \in \mathcal{B}_{\mathbb{R}} \text{ for } 1 \le i \le n\right\}.$$

Notice that $\mathcal{P}$ is a π – system and it is easy to show $\sigma(\mathcal{P}) = \mathcal{F}$ (see Exercise 6.6). Therefore by Proposition 5.15, $X_* P = Y_* P'$ iff $X_* P = Y_* P'$ on $\mathcal{P}$. Now for $A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^{\mathbb{N}} \in \mathcal{P}$ we have,

$$X_* P\left(A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^{\mathbb{N}}\right) = P((X_1, \ldots, X_n) \in A_1 \times A_2 \times \cdots \times A_n)$$

and hence the condition becomes,

$$P((X_1, \ldots, X_n) \in A_1 \times A_2 \times \cdots \times A_n) = P'((Y_1, \ldots, Y_n) \in A_1 \times A_2 \times \cdots \times A_n)$$

for all $n \in \mathbb{N}$ and $A_i \in \mathcal{B}_{\mathbb{R}}$. Another application of Proposition 5.15 or using Corollary 6.44 allows us to conclude that shows that $X_* P = Y_* P'$ iff $(X_1, \ldots, X_n) \overset{d}{=} (Y_1, \ldots, Y_n)$ for all $n \in \mathbb{N}$. ■

Corollary 6.47. *Continue the notation above and assume that $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$. Further let*

$$X_\pm = \begin{cases} \limsup_{n\to\infty} X_n & \text{if } + \\ \liminf_{n\to\infty} X_n & \text{if } - \end{cases}$$

and define $Y_\pm$ similarly. Then $(X_-, X_+) \overset{d}{=} (Y_-, Y_+)$ as random variables into $\left(\bar{\mathbb{R}}^2, \mathcal{B}_{\bar{\mathbb{R}}} \otimes \mathcal{B}_{\bar{\mathbb{R}}}\right)$. In particular,

$$P\left(\lim_{n\to\infty} X_n \text{ exists in } \mathbb{R}\right) = P'\left(\lim_{n\to\infty} Y \text{ exists in } \mathbb{R}\right). \tag{6.8}$$

Proof. First suppose that $(\Omega', \mathcal{B}', P') = \left(\mathbb{R}^{\mathbb{N}}, \mathcal{F}, P' := X_* P\right)$ where $Y_i(a_1, a_2, \dots) := a_i = \pi_i(a_1, a_2, \dots)$. Then for $C \in \mathcal{B}_{\bar{\mathbb{R}}} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ we have,

$$X^{-1}(\{(Y_-, Y_+) \in C\}) = \{(Y_- \circ X, Y_+ \circ X) \in C\} = \{(X_-, X_+) \in C\},$$

since, for example,

$$Y_- \circ X = \liminf_{n\to\infty} Y_n \circ X = \liminf_{n\to\infty} X_n = X_-.$$

Therefore it follows that

$$P((X_-, X_+) \in C) = P \circ X^{-1}(\{(Y_-, Y_+) \in C\}) = P'(\{(Y_-, Y_+) \in C\}). \quad (6.9)$$

The general result now follows by two applications of this special case.

For the last assertion, take

$$C = \{(x, x) : x \in \mathbb{R}\} \in \mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}} \otimes \mathcal{B}_{\bar{\mathbb{R}}}.$$

Then $(X_-, X_+) \in C$ iff $X_- = X_+ \in \mathbb{R}$ which happens iff $\lim_{n\to\infty} X_n$ exists in $\mathbb{R}$. Similarly, $(Y_-, Y_+) \in C$ iff $\lim_{n\to\infty} Y_n$ exists in $\mathbb{R}$ and therefore Eq. (6.8) holds as a consequence of Eq. (6.9). ■

Exercise 6.10. Let $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ be two sequences of random variables such that $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$. Let $\{S_n\}_{n=1}^\infty$ and $\{T_n\}_{n=1}^\infty$ be defined by, $S_n := X_1 + \dots + X_n$ and $T_n := Y_1 + \dots + Y_n$. Prove the following assertions.

1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^k$ is a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}^k}$ – measurable function, then $f(X_1, \dots, X_n) \overset{d}{=} f(Y_1, \dots, Y_n)$.
2. Use your result in item 1. to show $\{S_n\}_{n=1}^\infty \overset{d}{=} \{T_n\}_{n=1}^\infty$.
   Hint: Apply item 1. with $k = n$ after making a judicious choice for $f : \mathbb{R}^n \to \mathbb{R}^n$.

6.5 Generating All Distributions from the Uniform Distribution

Theorem 6.48. *Given a distribution function, $F : \mathbb{R} \to [0,1]$ let $G : (0,1) \to \mathbb{R}$ be defined (see Figure 6.2) by,*

$$G(y) := \inf\{x : F(x) \ge y\}.$$

Then $G : (0,1) \to \mathbb{R}$ is Borel measurable and $G_* m = \mu_F$ where μ_F is the unique measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$.

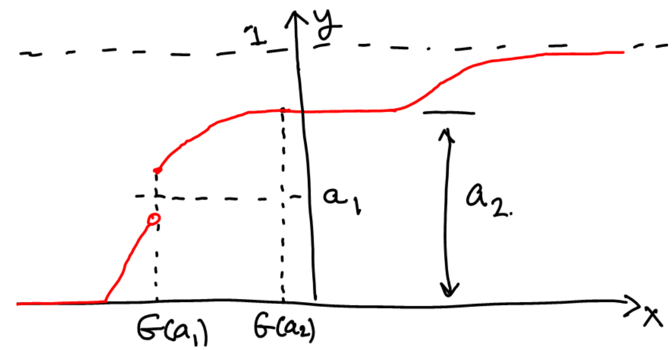


Fig. 6.2. A pictorial definition of G.

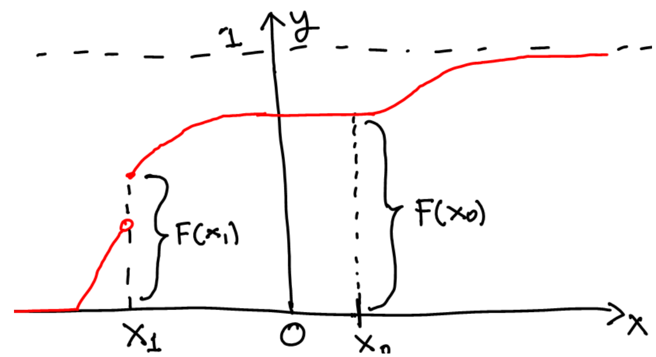


Fig. 6.3. As can be seen from this picture, $G(y) \le x_0$ iff $y \le F(x_0)$ and similarly, $G(y) \le x_1$ iff $y \le x_1$.

Proof. Since $G : (0,1) \to \mathbb{R}$ is a non-decreasing function, G is measurable. We also claim that, for all $x_0 \in \mathbb{R}$, that

$$G^{-1}((0, x_0]) = \{y : G(y) \le x_0\} = (0, F(x_0)] \cap \mathbb{R}, \quad (6.10)$$

see Figure 6.3.

To give a formal proof of Eq. (6.10), $G(y) = \inf\{x : F(x) \ge y\} \le x_0$, there exists $x_n \ge x_0$ with $x_n \downarrow x_0$ such that $F(x_n) \ge y$. By the right continuity of F, it follows that $F(x_0) \ge y$. Thus we have shown

$$\{G \le x_0\} \subset (0, F(x_0)] \cap (0,1).$$

For the converse, if $y \le F(x_0)$ then $G(y) = \inf\{x : F(x) \ge y\} \le x_0$, i.e. $y \in \{G \le x_0\}$. Indeed, $y \in G^{-1}((-\infty, x_0])$ iff $G(y) \le x_0$. Observe that

$$G(F(x_0)) = \inf\{x : F(x) \geq F(x_0)\} \leq x_0$$

and hence $G(y) \leq x_0$ whenever $y \leq F(x_0)$. This shows that

$$(0, F(x_0)] \cap (0,1) \subset G^{-1}((0, x_0]).$$

As a consequence we have $G_* m = \mu_F$. Indeed,

$$\begin{aligned}(G_* m)((-\infty, x]) &= m\left(G^{-1}((-\infty, x])\right) = m(\{y \in (0,1) : G(y) \leq x\}) \\ &= m((0, F(x)] \cap (0,1)) = F(x).\end{aligned}$$

See section 2.5.2 on p. 61 of Resnick for more details. ■

Theorem 6.49 (Durret's Version). *Given a distribution function,* $F : \mathbb{R} \to [0,1]$ *let* $Y : (0,1) \to \mathbb{R}$ *be defined (see Figure 6.4) by,*

$$Y(x) := \sup\{y : F(y) < x\}.$$

Then $Y : (0,1) \to \mathbb{R}$ *is Borel measurable and* $Y_* m = \mu_F$ *where* μ_F *is the unique measure on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *such that* $\mu_F((a,b]) = F(b) - F(a)$ *for all* $-\infty < a < b < \infty$.

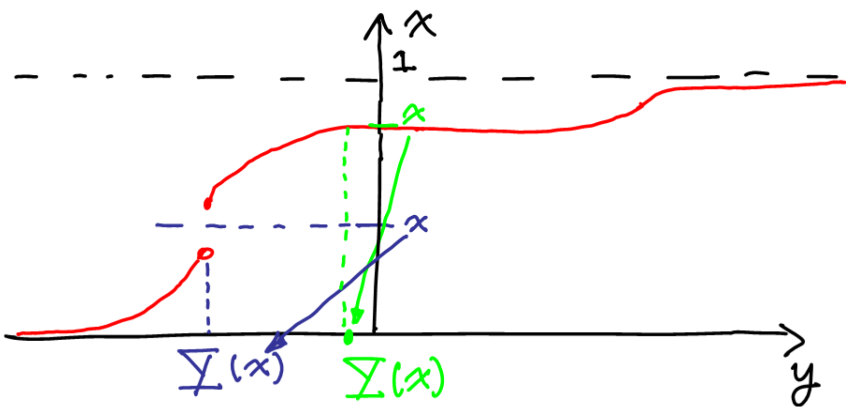


Fig. 6.4. A pictorial definition of $Y(x)$.

Proof. Since $Y : (0,1) \to \mathbb{R}$ is a non-decreasing function, Y is measurable. Also observe, if $y < Y(x)$, then $F(y) < x$ and hence,

$$F(Y(x)-) = \lim_{y \uparrow Y(x)} F(y) \leq x.$$

For $y > Y(x)$, we have $F(y) \geq x$ and therefore,

$$F(Y(x)) = F(Y(x)+) = \lim_{y \downarrow Y(x)} F(y) \geq x$$

and so we have shown

$$F(Y(x)-) \leq x \leq F(Y(x)).$$

We will now show

$$\{x \in (0,1) : Y(x) \leq y_0\} = (0, F(y_0)] \cap (0,1). \tag{6.11}$$

For the inclusion "$\subset$," if $x \in (0,1)$ and $Y(x) \leq y_0$, then $x \leq F(Y(x)) \leq F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely if $x \in (0,1)$ and $x \leq F(y_0)$ then (by definition of $Y(x)$) $y_0 \geq Y(x)$.

From the identity in Eq. (6.11), it follows that Y is measurable and

$$(Y_* m)((-\infty, y_0)) = m\left(Y^{-1}(-\infty, y_0)\right) = m((0, F(y_0)] \cap (0,1)) = F(y_0).$$

Therefore, *Law*$(Y) = \mu_F$ as desired. ■

7

Integration Theory

In this chapter, we will greatly extend the "simple" integral or expectation which was developed in Section 4.3 above. Recall there that if $(\Omega, \mathcal{B}, \mu)$ was measurable space and $\varphi : \Omega \to [0, \infty)$ was a measurable simple function, then we let

$$\mathbb{E}_\mu \varphi := \sum_{\lambda \in [0,\infty)} \lambda \mu (\varphi = \lambda).$$

The conventions being use here is that $0 \cdot \mu(\varphi = 0) = 0$ even when $\mu(\varphi = 0) = \infty$. This convention is necessary in order to make the integral linear – at a minimum we will want $\mathbb{E}_\mu [0] = 0$. Please be careful not blindly apply the $0 \cdot \infty = 0$ convention in other circumstances.

7.1 Integrals of positive functions

Definition 7.1. *Let $L^+ = L^+(\mathcal{B}) = \{f : \Omega \to [0, \infty] : f$ is measurable$\}$. Define*

$$\int_\Omega f(\omega)\, d\mu(\omega) = \int_\Omega f d\mu := \sup \{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le f\}.$$

*We say the $f \in L^+$ is **integrable** if $\int_\Omega f d\mu < \infty$. If $A \in \mathcal{B}$, let*

$$\int_A f(\omega)\, d\mu(\omega) = \int_A f d\mu := \int_\Omega 1_A f \; d\mu.$$

We also use the notation,

$$\mathbb{E} f = \int_\Omega f d\mu \text{ and } \mathbb{E}[f : A] := \int_A f d\mu.$$

Remark 7.2. Because of item 3. of Proposition 4.19, if φ is a non-negative simple function, $\int_\Omega \varphi d\mu = \mathbb{E}_\mu \varphi$ so that $\int_\Omega$ is an extension of $\mathbb{E}_\mu$.

Lemma 7.3. *Let $f, g \in L^+(\mathcal{B})$. Then:*

1. *if $\lambda \ge 0$, then*
$$\int_\Omega \lambda f d\mu = \lambda \int_\Omega f d\mu$$
wherein $\lambda \int_\Omega f d\mu \equiv 0$ if $\lambda = 0$, even if $\int_\Omega f d\mu = \infty$.
2. *if $0 \le f \le g$, then*
$$\int_\Omega f d\mu \le \int_\Omega g d\mu. \tag{7.1}$$
3. *For all $\varepsilon > 0$ and $p > 0$,*
$$\mu(f \ge \varepsilon) \le \frac{1}{\varepsilon^p} \int_\Omega f^p 1_{\{f \ge \varepsilon\}} d\mu \le \frac{1}{\varepsilon^p} \int_\Omega f^p d\mu. \tag{7.2}$$
The inequality in Eq. (7.2) is called Chebyshev's Inequality for $p = 1$ and Markov's inequality for $p = 2$.
4. *If $\int_\Omega f d\mu < \infty$ then $\mu(f = \infty) = 0$ (i.e. $f < \infty$ a.e.) and the set $\{f > 0\}$ is σ – finite.*

Proof. 1. We may assume $\lambda > 0$ in which case,

$$\begin{aligned}
\int_\Omega \lambda f d\mu &= \sup \{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le \lambda f\} \\
&= \sup \left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \lambda^{-1} \varphi \le f\right\} \\
&= \sup \{\mathbb{E}_\mu [\lambda \psi] : \psi \text{ is simple and } \psi \le f\} \\
&= \sup \{\lambda \mathbb{E}_\mu [\psi] : \psi \text{ is simple and } \psi \le f\} \\
&= \lambda \int_\Omega f d\mu.
\end{aligned}$$

2. Since

$$\{\varphi \text{ is simple and } \varphi \le f\} \subset \{\varphi \text{ is simple and } \varphi \le g\},$$

Eq. (7.1) follows from the definition of the integral.
3. Since $1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \frac{1}{\varepsilon} f \le \frac{1}{\varepsilon} f$ we have

$$1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \left(\frac{1}{\varepsilon} f\right)^p \le \left(\frac{1}{\varepsilon} f\right)^p$$

and by monotonicity and the multiplicative property of the integral,

$$\mu(f \ge \varepsilon) = \int_\Omega 1_{\{f \ge \varepsilon\}} d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega 1_{\{f \ge \varepsilon\}} f^p d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega f^p d\mu.$$

4. If $\mu(f=\infty)>0$, then $\varphi_n := n1_{\{f=\infty\}}$ is a simple function such that $\varphi_n \le f$ for all n and hence

$$n\mu(f=\infty) = \mathbb{E}_\mu(\varphi_n) \le \int_\Omega f d\mu$$

for all n. Letting $n\to\infty$ shows $\int_\Omega f d\mu = \infty$. Thus if $\int_\Omega f d\mu < \infty$ then $\mu(f=\infty)=0$.

Moreover,

$$\{f>0\} = \cup_{n=1}^\infty \{f>1/n\}$$

with $\mu(f>1/n) \le n\int_\Omega f d\mu < \infty$ for each n. ■

Theorem 7.4 (Monotone Convergence Theorem). *Suppose $f_n \in L^+$ is a sequence of functions such that $f_n \uparrow f$ (f is necessarily in L^+) then*

$$\int f_n \uparrow \int f \text{ as } n\to\infty.$$

Proof. Since $f_n \le f_m \le f$, for all $n \le m < \infty$,

$$\int f_n \le \int f_m \le \int f$$

from which if follows $\int f_n$ is increasing in n and

$$\lim_{n\to\infty} \int f_n \le \int f. \tag{7.3}$$

For the opposite inequality, let $\varphi : \Omega \to [0,\infty)$ be a simple function such that $0\le\varphi\le f$, $\alpha\in(0,1)$ and $\Omega_n := \{f_n \ge \alpha\varphi\}$. Notice that $\Omega_n \uparrow \Omega$ and $f_n \ge \alpha 1_{\Omega_n}\varphi$ and so by definition of $\int f_n$,

$$\int f_n \ge \mathbb{E}_\mu[\alpha 1_{\Omega_n}\varphi] = \alpha\mathbb{E}_\mu[1_{\Omega_n}\varphi]. \tag{7.4}$$

Then using the identity

$$1_{\Omega_n}\varphi = 1_{\Omega_n}\sum_{y>0} y1_{\{\varphi=y\}} = \sum_{y>0} y1_{\{\varphi=y\}\cap\Omega_n},$$

and the linearity of $\mathbb{E}_\mu$ we have,

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}_\mu[1_{\Omega_n}\varphi] &= \lim_{n\to\infty}\sum_{y>0} y\cdot\mu(\Omega_n\cap\{\varphi=y\}) \\ &= \sum_{y>0} y\lim_{n\to\infty}\mu(\Omega_n\cap\{\varphi=y\}) \text{ (finite sum)} \\ &= \sum_{y>0} y\mu(\{\varphi=y\}) = \mathbb{E}_\mu[\varphi],\end{aligned}$$

wherein we have used the continuity of μ under increasing unions for the third equality. This identity allows us to let $n\to\infty$ in Eq. (7.4) to conclude $\lim_{n\to\infty}\int f_n \ge \alpha\mathbb{E}_\mu[\varphi]$ and since $\alpha\in(0,1)$ was arbitrary we may further conclude, $\mathbb{E}_\mu[\varphi] \le \lim_{n\to\infty}\int f_n$. The latter inequality being true for all simple functions φ with $\varphi \le f$ then implies that

$$\int f = \sup_{0\le\varphi\le f}\mathbb{E}_\mu[\varphi] \le \lim_{n\to\infty}\int f_n,$$

which combined with Eq. (7.3) proves the theorem. ■

Remark 7.5 ("Explicit" Integral Formula). Given $f:\Omega\to[0,\infty]$ measurable, we know from the approximation Theorem 6.39 $\varphi_n\uparrow f$ where

$$\varphi_n := \sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{\left\{\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right\}} + 2^n 1_{\{f>2^n\}}.$$

Therefore by the monotone convergence theorem,

$$\begin{aligned}\int_\Omega f d\mu &= \lim_{n\to\infty}\int_\Omega \varphi_n d\mu \\ &= \lim_{n\to\infty}\left[\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}\mu\left(\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right) + 2^n\mu(f>2^n)\right].\end{aligned}$$

Corollary 7.6. *If $f_n\in L^+$ is a sequence of functions then*

$$\int\sum_{n=1}^\infty f_n = \sum_{n=1}^\infty\int f_n.$$

In particular, if $\sum_{n=1}^\infty\int f_n<\infty$ then $\sum_{n=1}^\infty f_n<\infty$ a.e.

Proof. First off we show that

$$\int(f_1+f_2) = \int f_1 + \int f_2$$

by choosing non-negative simple function φ_n and ψ_n such that $\varphi_n\uparrow f_1$ and $\psi_n\uparrow f_2$. Then $(\varphi_n+\psi_n)$ is simple as well and $(\varphi_n+\psi_n)\uparrow(f_1+f_2)$ so by the monotone convergence theorem,

$$\begin{aligned}\int(f_1+f_2) &= \lim_{n\to\infty}\int(\varphi_n+\psi_n) = \lim_{n\to\infty}\left(\int\varphi_n+\int\psi_n\right) \\ &= \lim_{n\to\infty}\int\varphi_n + \lim_{n\to\infty}\int\psi_n = \int f_1+\int f_2.\end{aligned}$$

Now to the general case. Let $g_N := \sum_{n=1}^{N} f_n$ and $g = \sum_{1}^{\infty} f_n$, then $g_N \uparrow g$ and so again by monotone convergence theorem and the additivity just proved,

$$\sum_{n=1}^{\infty} \int f_n := \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n$$
$$= \lim_{N\to\infty} \int g_N = \int g =: \int \sum_{n=1}^{\infty} f_n.$$

■

Remark 7.7. It is in the proof of Corollary 7.6 (i.e. the linearity of the integral) that we really make use of the assumption that all of our functions are measurable. In fact the definition $\int f d\mu$ makes sense for **all** functions $f : \Omega \to [0,\infty]$ not just measurable functions. Moreover the monotone convergence theorem holds in this generality with no change in the proof. However, in the proof of Corollary 7.6, we use the approximation Theorem 6.39 which relies heavily on the measurability of the functions to be approximated.

Example 7.8 (Sums as Integrals I). Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, $\mu(A) = \#(A)$ for $A \subset \Omega$ is the counting measure on $\mathcal{B}$, and $f : \mathbb{N} \to [0,\infty]$ is a function. Since

$$f = \sum_{n=1}^{\infty} f(n) 1_{\{n\}},$$

it follows from Corollary 7.6 that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^{\infty} \int_{\mathbb{N}} f(n) 1_{\{n\}} d\mu = \sum_{n=1}^{\infty} f(n) \mu(\{n\}) = \sum_{n=1}^{\infty} f(n).$$

Thus the integral relative to counting measure is simply the infinite sum.

Lemma 7.9 (Sums as Integrals II*). *Let Ω be a set and $\rho : \Omega \to [0,\infty]$ be a function, let $\mu = \sum_{\omega\in\Omega} \rho(\omega)\delta_\omega$ on $\mathcal{B} = 2^{\Omega}$, i.e.*

$$\mu(A) = \sum_{\omega\in A} \rho(\omega).$$

If $f : \Omega \to [0,\infty]$ is a function (which is necessarily measurable), then

$$\int_\Omega f d\mu = \sum_\Omega f\rho.$$

Proof. Suppose that $\varphi : \Omega \to [0,\infty)$ is a simple function, then $\varphi = \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}$ and

$$\sum_\Omega \varphi\rho = \sum_{\omega\in\Omega} \rho(\omega) \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}(\omega) = \sum_{z\in[0,\infty)} z \sum_{\omega\in\Omega} \rho(\omega) 1_{\{\varphi=z\}}(\omega)$$
$$= \sum_{z\in[0,\infty)} z\mu(\{\varphi = z\}) = \int_\Omega \varphi d\mu.$$

So if $\varphi : \Omega \to [0,\infty)$ is a simple function such that $\varphi \le f$, then

$$\int_\Omega \varphi d\mu = \sum_\Omega \varphi\rho \le \sum_\Omega f\rho.$$

Taking the sup over φ in this last equation then shows that

$$\int_\Omega f d\mu \le \sum_\Omega f\rho.$$

For the reverse inequality, let $\Lambda \subset\subset \Omega$ be a finite set and $N \in (0,\infty)$. Set $f^N(\omega) = \min\{N, f(\omega)\}$ and let $\varphi_{N,\Lambda}$ be the simple function given by $\varphi_{N,\Lambda}(\omega) := 1_\Lambda(\omega) f^N(\omega)$. Because $\varphi_{N,\Lambda}(\omega) \le f(\omega)$,

$$\sum_\Lambda f^N\rho = \sum_\Omega \varphi_{N,\Lambda}\rho = \int_\Omega \varphi_{N,\Lambda} d\mu \le \int_\Omega f d\mu.$$

Since $f^N \uparrow f$ as $N \to \infty$, we may let $N \to \infty$ in this last equation to concluded

$$\sum_\Lambda f\rho \le \int_\Omega f d\mu.$$

Since Λ is arbitrary, this implies

$$\sum_\Omega f\rho \le \int_\Omega f d\mu.$$

■

Exercise 7.1. Suppose that $\mu_n : \mathcal{B} \to [0,\infty]$ are measures on $\mathcal{B}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in n for all $A \in \mathcal{B}$. Prove that $\mu : \mathcal{B} \to [0,\infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

Proposition 7.10. *Suppose that $f \ge 0$ is a measurable function. Then $\int_\Omega f d\mu = 0$ iff $f = 0$ a.e. Also if $f, g \ge 0$ are measurable functions such that $f \le g$ a.e. then $\int f d\mu \le \int g d\mu$. In particular if $f = g$ a.e. then $\int f d\mu = \int g d\mu$.*

Proof. If $f = 0$ a.e. and $\varphi \le f$ is a simple function then $\varphi = 0$ a.e. This implies that $\mu(\varphi^{-1}(\{y\})) = 0$ for all $y > 0$ and hence $\int_\Omega \varphi d\mu = 0$ and therefore $\int_\Omega f d\mu = 0$. Conversely, if $\int f d\mu = 0$, then by (Lemma 7.3),

$$\mu(f \ge 1/n) \le n \int f d\mu = 0 \text{ for all } n.$$

Therefore, $\mu(f > 0) \le \sum_{n=1}^\infty \mu(f \ge 1/n) = 0$, i.e. $f = 0$ a.e.

For the second assertion let E be the exceptional set where $f > g$, i.e.

$$E := \{\omega \in \Omega : f(\omega) > g(\omega)\}.$$

By assumption E is a null set and $1_{E^c} f \le 1_{E^c} g$ everywhere. Because $g = 1_{E^c} g + 1_E g$ and $1_E g = 0$ a.e.,

$$\int g d\mu = \int 1_{E^c} g d\mu + \int 1_E g d\mu = \int 1_{E^c} g d\mu$$

and similarly $\int f d\mu = \int 1_{E^c} f d\mu$. Since $1_{E^c} f \le 1_{E^c} g$ everywhere,

$$\int f d\mu = \int 1_{E^c} f d\mu \le \int 1_{E^c} g d\mu = \int g d\mu.$$

■

Corollary 7.11. *Suppose that $\{f_n\}$ is a sequence of non-negative measurable functions and f is a measurable function such that $f_n \uparrow f$ off a null set, then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

Proof. Let $E \subset \Omega$ be a null set such that $f_n 1_{E^c} \uparrow f 1_{E^c}$ as $n \to \infty$. Then by the monotone convergence theorem and Proposition 7.10,

$$\int f_n = \int f_n 1_{E^c} \uparrow \int f 1_{E^c} = \int f \text{ as } n \to \infty.$$

■

Lemma 7.12 (Fatou's Lemma). *If $f_n : \Omega \to [0, \infty]$ is a sequence of measurable functions then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n$$

Proof. Define $g_k := \inf_{n \ge k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \le f_n$ for all $k \le n$,

$$\int g_k \le \int f_n \text{ for all } n \ge k$$

and therefore

$$\int g_k \le \liminf_{n\to\infty} \int f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k \to \infty$ to find

$$\int \liminf_{n\to\infty} f_n = \int \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty} \int g_k \le \liminf_{n\to\infty} \int f_n.$$

■

The following Corollary and the next lemma are simple applications of Corollary 7.6.

Corollary 7.13. *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{B}$ is a collection of sets such that $\mu(A_i \cap A_j) = 0$ for all $i \ne j$, then*

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n).$$

Proof. Since

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \int_\Omega 1_{\cup_{n=1}^\infty A_n} d\mu \text{ and}$$
$$\sum_{n=1}^\infty \mu(A_n) = \int_\Omega \sum_{n=1}^\infty 1_{A_n} d\mu$$

it suffices to show

$$\sum_{n=1}^\infty 1_{A_n} = 1_{\cup_{n=1}^\infty A_n} \ \mu - \text{a.e.} \tag{7.5}$$

Now $\sum_{n=1}^\infty 1_{A_n} \ge 1_{\cup_{n=1}^\infty A_n}$ and $\sum_{n=1}^\infty 1_{A_n}(\omega) \ne 1_{\cup_{n=1}^\infty A_n}(\omega)$ iff $\omega \in A_i \cap A_j$ for some $i \ne j$, that is

$$\left\{\omega : \sum_{n=1}^\infty 1_{A_n}(\omega) \ne 1_{\cup_{n=1}^\infty A_n}(\omega)\right\} = \cup_{i<j} A_i \cap A_j$$

and the latter set has measure 0 being the countable union of sets of measure zero. This proves Eq. (7.5) and hence the corollary. ■

Lemma 7.14 (The First Borell – Cantelli Lemma). *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space, $A_n \in \mathcal{B}$, and set*

$$\{A_n \text{ i.o.}\} = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\text{'s}\} = \bigcap_{N=1}^\infty \bigcup_{n \ge N} A_n.$$

If $\sum_{n=1}^\infty \mu(A_n) < \infty$ then $\mu(\{A_n \text{ i.o.}\}) = 0$.

Proof. (First Proof.) Let us first observe that

$$\{A_n \text{ i.o.}\} = \left\{\omega \in \Omega : \sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty\right\}.$$

Hence if $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then

$$\infty > \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \int_{\Omega} 1_{A_n} \, d\mu = \int_{\Omega} \sum_{n=1}^{\infty} 1_{A_n} \, d\mu$$

implies that $\sum_{n=1}^{\infty} 1_{A_n}(\omega) < \infty$ for μ - a.e. ω. That is to say $\mu(\{A_n \text{ i.o.}\}) = 0$.

(Second Proof.) Of course we may give a strictly measure theoretic proof of this fact:

$$\begin{aligned}\mu(A_n \text{ i.o.}) &= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} A_n\right) \\ &\leq \lim_{N\to\infty} \sum_{n\geq N} \mu(A_n)\end{aligned}$$

and the last limit is zero since $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. ■

Example 7.15. Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space (i.e. $P(\Omega) = 1$) and $X_n : \Omega \to \{0, 1\}$ are Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. If $\sum_{n=1}^{\infty} p_n < \infty$, then $P(X_n = 1 \text{ i.o.}) = 0$ and hence $P(X_n = 0 \text{ a.a.}) = 1$. In particular, $P(\lim_{n\to\infty} X_n = 0) = 1$.

7.2 Integrals of Complex Valued Functions

Definition 7.16. *A measurable function $f : \Omega \to \bar{\mathbb{R}}$ is **integrable** if $f_+ := f1_{\{f\geq 0\}}$ and $f_- = -f\ 1_{\{f\leq 0\}}$ are **integrable**. We write $\mathrm{L}^1(\mu; \mathbb{R})$ for the space of real valued integrable functions. For $f \in \mathrm{L}^1(\mu; \mathbb{R})$, let*

$$\int_{\Omega} f d\mu = \int_{\Omega} f_+ d\mu - \int_{\Omega} f_- d\mu.$$

To shorten notation in this chapter we may simply write $\int f d\mu$ or even $\int f$ for $\int_{\Omega} f d\mu$.

Convention: If $f, g : \Omega \to \bar{\mathbb{R}}$ are two measurable functions, let $f + g$ denote the collection of measurable functions $h : \Omega \to \bar{\mathbb{R}}$ such that $h(\omega) = f(\omega) + g(\omega)$ whenever $f(\omega) + g(\omega)$ is well defined, i.e. is not of the form $\infty - \infty$ or $-\infty + \infty$. We use a similar convention for $f - g$. Notice that if $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $h_1, h_2 \in f + g$, then $h_1 = h_2$ a.e. because $|f| < \infty$ and $|g| < \infty$ a.e.

Notation 7.17 (Abuse of notation) *We will sometimes denote the integral $\int_{\Omega} f d\mu$ by $\mu(f)$. With this notation we have $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{B}$.*

Remark 7.18. Since

$$f_{\pm} \leq |f| \leq f_+ + f_-,$$

a measurable function f is **integrable** iff $\int |f| \, d\mu < \infty$. Hence

$$\mathrm{L}^1(\mu; \mathbb{R}) := \left\{f : \Omega \to \bar{\mathbb{R}} : \ f \text{ is measurable and } \int_{\Omega} |f| \, d\mu < \infty\right\}.$$

If $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $f = g$ a.e. then $f_{\pm} = g_{\pm}$ a.e. and so it follows from Proposition 7.10 that $\int f d\mu = \int g d\mu$. In particular if $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ we may define

$$\int_{\Omega} (f + g) \, d\mu = \int_{\Omega} h d\mu$$

where h is any element of $f + g$.

Proposition 7.19. *The map*

$$f \in \mathrm{L}^1(\mu; \mathbb{R}) \to \int_{\Omega} f d\mu \in \mathbb{R}$$

is linear and has the monotonicity property: $\int f d\mu \leq \int g d\mu$ for all $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ such that $f \leq g$ a.e.

Proof. Let $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $a, b \in \mathbb{R}$. By modifying f and g on a null set, we may assume that f, g are real valued functions. We have $af + bg \in \mathrm{L}^1(\mu; \mathbb{R})$ because

$$|af + bg| \leq |a| \, |f| + |b| \, |g| \in \mathrm{L}^1(\mu; \mathbb{R}).$$

If $a < 0$, then

$$(af)_+ = -af_- \text{ and } (af)_- = -af_+$$

so that

$$\int af = -a \int f_- + a \int f_+ = a(\int f_+ - \int f_-) = a \int f.$$

A similar calculation works for $a > 0$ and the case $a = 0$ is trivial so we have shown that

$$\int af = a \int f.$$

Now set $h = f + g$. Since $h = h_+ - h_-$,

$$h_+ - h_- = f_+ - f_- + g_+ - g_-$$

or

$$h_+ + f_- + g_- = h_- + f_+ + g_+.$$

Therefore,

$$\int h_+ + \int f_- + \int g_- = \int h_- + \int f_+ + \int g_+$$

and hence

$$\int h = \int h_+ - \int h_- = \int f_+ + \int g_+ - \int f_- - \int g_- = \int f + \int g.$$

Finally if $f_+ - f_- = f \le g = g_+ - g_-$ then $f_+ + g_- \le g_+ + f_-$ which implies that

$$\int f_+ + \int g_- \le \int g_+ + \int f_-$$

or equivalently that

$$\int f = \int f_+ - \int f_- \le \int g_+ - \int g_- = \int g.$$

The monotonicity property is also a consequence of the linearity of the integral, the fact that $f \le g$ a.e. implies $0 \le g - f$ a.e. and Proposition 7.10. ∎

Definition 7.20. *A measurable function $f : \Omega \to \mathbb{C}$ is* ***integrable*** *if $\int_\Omega |f| \ d\mu < \infty$. Analogously to the real case, let*

$$\mathrm{L}^1(\mu;\mathbb{C}) := \left\{ f : \Omega \to \mathbb{C} : \ f \text{ is measurable and } \int_\Omega |f| \ d\mu < \infty \right\}.$$

denote the complex valued integrable functions. Because, $\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|) \le |f| \le \sqrt{2} \max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|)$, $\int |f| \ d\mu < \infty$ *iff*

$$\int |\mathrm{Re}\, f| \, d\mu + \int |\mathrm{Im}\, f| \, d\mu < \infty.$$

For $f \in \mathrm{L}^1(\mu;\mathbb{C})$ define

$$\int f \ d\mu = \int \mathrm{Re}\, f \ d\mu + i \int \mathrm{Im}\, f \ d\mu.$$

It is routine to show the integral is still linear on $\mathrm{L}^1(\mu;\mathbb{C})$ (prove!). In the remainder of this section, let $\mathrm{L}^1(\mu)$ be either $\mathrm{L}^1(\mu;\mathbb{C})$ or $\mathrm{L}^1(\mu;\mathbb{R})$. If $A \in \mathcal{B}$ and $f \in \mathrm{L}^1(\mu;\mathbb{C})$ or $f : \Omega \to [0,\infty]$ is a measurable function, let

$$\int_A f d\mu := \int_\Omega 1_A f d\mu.$$

Proposition 7.21. *Suppose that $f \in \mathrm{L}^1(\mu;\mathbb{C})$, then*

$$\left| \int_\Omega f d\mu \right| \le \int_\Omega |f| \, d\mu. \tag{7.6}$$

Proof. Start by writing $\int_\Omega f \ d\mu = Re^{i\theta}$ with $R \ge 0$. We may assume that $R = \left|\int_\Omega f d\mu\right| > 0$ since otherwise there is nothing to prove. Since

$$R = e^{-i\theta} \int_\Omega f \ d\mu = \int_\Omega e^{-i\theta} f \ d\mu = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu + i \int_\Omega \mathrm{Im}\left(e^{-i\theta} f\right) d\mu,$$

it must be that $\int_\Omega \mathrm{Im}\left[e^{-i\theta} f\right] d\mu = 0$. Using the monotonicity in Proposition 7.10,

$$\left| \int_\Omega f d\mu \right| = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu \le \int_\Omega \left|\mathrm{Re}\left(e^{-i\theta} f\right)\right| d\mu \le \int_\Omega |f| \, d\mu.$$

∎

Proposition 7.22. *Let $f, g \in \mathrm{L}^1(\mu)$, then*

1. *The set $\{f \ne 0\}$ is σ – finite, in fact $\{|f| \ge \frac{1}{n}\} \uparrow \{f \ne 0\}$ and $\mu(|f| \ge \frac{1}{n}) < \infty$ for all n.*
2. *The following are equivalent*
   a) *$\int_E f = \int_E g$ for all $E \in \mathcal{B}$*
   b) *$\int\limits_\Omega |f - g| = 0$*
   c) *$f = g$ a.e.*

Proof. 1. By Chebyshev's inequality, Lemma 7.3,

$$\mu(|f| \ge \frac{1}{n}) \le n \int_\Omega |f| \, d\mu < \infty$$

for all n.

2. (a) $\Longrightarrow$ (c) Notice that

$$\int_E f = \int_E g \Leftrightarrow \int_E (f - g) = 0$$

for all $E \in \mathcal{B}$. Taking $E = \{\mathrm{Re}(f - g) > 0\}$ and using $1_E \mathrm{Re}(f - g) \ge 0$, we learn that

$$0 = \mathrm{Re} \int_E (f - g) d\mu = \int 1_E \mathrm{Re}(f - g) \Longrightarrow 1_E \mathrm{Re}(f - g) = 0 \text{ a.e.}$$

This implies that $1_E = 0$ a.e. which happens iff

$$\mu(\{\mathrm{Re}(f-g)>0\}) = \mu(E) = 0.$$

Similar $\mu(\mathrm{Re}(f-g)<0) = 0$ so that $\mathrm{Re}(f-g) = 0$ a.e. Similarly, $\mathrm{Im}(f-g) = 0$ a.e and hence $f-g=0$ a.e., i.e. $f=g$ a.e.

(c) $\Longrightarrow$ (b) is clear and so is (b) $\Longrightarrow$ (a) since

$$\left|\int_E f - \int_E g\right| \le \int |f-g| = 0.$$

■

Lemma 7.23 (Integral Comparison I). *Suppose that $h \in \mathrm{L}^1(\mu)$ satisfies*

$$\int_A h d\mu \ge 0 \text{ for all } A \in \mathcal{B}, \tag{7.7}$$

then $h \ge 0$ a.e.

Proof. Since by assumption,

$$0 = \mathrm{Im} \int_A h d\mu = \int_A \mathrm{Im}\, h d\mu \text{ for all } A \in \mathcal{B},$$

we may apply Proposition 7.22 to conclude that $\mathrm{Im}\, h = 0$ a.e. Thus we may now assume that h is real valued. Taking $A = \{h<0\}$ in Eq. (7.7) implies

$$\int_\Omega 1_A |h| \, d\mu = \int_\Omega -1_A h d\mu = -\int_A h d\mu \le 0.$$

However $1_A |h| \ge 0$ and therefore it follows that $\int_\Omega 1_A |h| \, d\mu = 0$ and so Proposition 7.22 implies $1_A |h| = 0$ a.e. which then implies $0 = \mu(A) = \mu(h<0) = 0$.
■

Lemma 7.24 (Integral Comparison II). *Suppose $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space (i.e. there exists $\Omega_n \in \mathcal{B}$ such that $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n) < \infty$ for all n) and $f, g : \Omega \to [0,\infty]$ are $\mathcal{B}$ – measurable functions. Then $f \ge g$ a.e. iff*

$$\int_A f d\mu \ge \int_A g d\mu \text{ for all } A \in \mathcal{B}. \tag{7.8}$$

In particular $f=g$ a.e. iff equality holds in Eq. (7.8).

Proof. It was already shown in Proposition 7.10 that $f \ge g$ a.e. implies Eq. (7.8). For the converse assertion, let $B_n := \{f \le n 1_{\Omega_n}\}$. Then from Eq. (7.8),

$$\infty > n\mu(\Omega_n) \ge \int f 1_{B_n} d\mu \ge \int g 1_{B_n} d\mu$$

from which it follows that both $f1_{B_n}$ and $g1_{B_n}$ are in $\mathrm{L}^1(\mu)$ and hence $h := f1_{B_n} - g1_{B_n} \in \mathrm{L}^1(\mu)$. Using Eq. (7.8) again we know that

$$\int_A h = \int f 1_{B_n \cap A} - \int g 1_{B_n \cap A} \ge 0 \text{ for all } A \in \mathcal{B}.$$

An application of Lemma 7.23 implies $h \ge 0$ a.e., i.e. $f1_{B_n} \ge g1_{B_n}$ a.e. Since $B_n \uparrow \{f<\infty\}$, we may conclude that

$$f 1_{\{f<\infty\}} = \lim_{n\to\infty} f 1_{B_n} \ge \lim_{n\to\infty} g 1_{B_n} = g 1_{\{f<\infty\}} \text{ a.e.}$$

Since $f \ge g$ whenever $f = \infty$, we have shown $f \ge g$ a.e.

If equality holds in Eq. (7.8), then we know that $g \le f$ and $f \le g$ a.e., i.e. $f=g$ a.e. ■

Notice that we can not drop the σ – finiteness assumption in Lemma 7.24. For example, let μ be the measure on $\mathcal{B}$ such that $\mu(A) = \infty$ when $A \ne \emptyset$, $g=3$, and $f=2$. Then equality holds (both sides are infinite unless $A = \emptyset$ when they are both zero) in Eq. (7.8) holds even though $f<g$ **everywhere.**

Definition 7.25. *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $L^1(\mu) = L^1(\Omega,\mathcal{B},\mu)$ denote the set of $\mathrm{L}^1(\mu)$ functions modulo the equivalence relation; $f \sim g$ iff $f=g$ a.e. We make this into a normed space using the norm*

$$\|f-g\|_{L^1} = \int |f-g| \, d\mu$$

and into a metric space using $\rho_1(f,g) = \|f-g\|_{L^1}$.

Warning: in the future we will often not make much of a distinction between $L^1(\mu)$ and $\mathrm{L}^1(\mu)$. On occasion this can be dangerous and this danger will be pointed out when necessary.

Remark 7.26. More generally we may define $L^p(\mu) = L^p(\Omega,\mathcal{B},\mu)$ for $p \in [1,\infty)$ as the set of measurable functions f such that

$$\int_\Omega |f|^p \, d\mu < \infty$$

modulo the equivalence relation; $f \sim g$ iff $f=g$ a.e.

We will see in later that

$$\|f\|_{L^p} = \left(\int |f|^p \, d\mu\right)^{1/p} \text{ for } f \in L^p(\mu)$$

is a norm and $(L^p(\mu), \|\cdot\|_{L^p})$ is a Banach space in this norm and in particular,

$$\|f+g\|_p \le \|f\|_p + \|g\|_p \text{ for all } f,g \in L^p(\mu).$$

Theorem 7.27 (Dominated Convergence Theorem). *Suppose $f_n, g_n, g \in \mathrm{L}^1(\mu)$, $f_n \to f$ a.e., $|f_n| \le g_n \in \mathrm{L}^1(\mu)$, $g_n \to g$ a.e. and $\int_\Omega g_n d\mu \to \int_\Omega g d\mu$. Then $f \in \mathrm{L}^1(\mu)$ and*

$$\int_\Omega f d\mu = \lim_{h\to\infty} \int_\Omega f_n d\mu.$$

(In most typical applications of this theorem $g_n = g \in \mathrm{L}^1(\mu)$ for all n.)

Proof. Notice that $|f| = \lim_{n\to\infty} |f_n| \le \lim_{n\to\infty} |g_n| \le g$ a.e. so that $f \in \mathrm{L}^1(\mu)$. By considering the real and imaginary parts of f separately, it suffices to prove the theorem in the case where f is real. By Fatou's Lemma,

$$\begin{aligned}\int_\Omega (g \pm f) d\mu &= \int_\Omega \liminf_{n\to\infty} (g_n \pm f_n)\, d\mu \le \liminf_{n\to\infty} \int_\Omega (g_n \pm f_n)\, d\mu \\ &= \lim_{n\to\infty} \int_\Omega g_n d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu \right) \\ &= \int_\Omega g d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu \right)\end{aligned}$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\int_\Omega g d\mu \pm \int_\Omega f d\mu \le \int_\Omega g d\mu + \begin{cases} \liminf_{n\to\infty} \int_\Omega f_n d\mu \\ -\limsup_{n\to\infty} \int_\Omega f_n d\mu \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \int_\Omega f_n d\mu \le \int_\Omega f d\mu \le \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

This shows that $\lim_{n\to\infty} \int_\Omega f_n d\mu$ exists and is equal to $\int_\Omega f d\mu$. ■

Exercise 7.2. Give another proof of Proposition 7.21 by first proving Eq. (7.6) with f being a simple function in which case the triangle inequality for complex numbers will do the trick. Then use the approximation Theorem 6.39 along with the dominated convergence Theorem 7.27 to handle the general case.

Corollary 7.28. *Let $\{f_n\}_{n=1}^\infty \subset \mathrm{L}^1(\mu)$ be a sequence such that $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$, then $\sum_{n=1}^\infty f_n$ is convergent a.e. and*

$$\int_\Omega \left(\sum_{n=1}^\infty f_n \right) d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.$$

Proof. The condition $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$ is equivalent to $\sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu)$. Hence $\sum_{n=1}^\infty f_n$ is almost everywhere convergent and if $S_N := \sum_{n=1}^N f_n$, then

$$|S_N| \le \sum_{n=1}^N |f_n| \le \sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu).$$

So by the dominated convergence theorem,

$$\begin{aligned}\int_\Omega \left(\sum_{n=1}^\infty f_n \right) d\mu &= \int_\Omega \lim_{N\to\infty} S_N d\mu = \lim_{N\to\infty} \int_\Omega S_N d\mu \\ &= \lim_{N\to\infty} \sum_{n=1}^N \int_\Omega f_n d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.\end{aligned}$$

■

Example 7.29 (Sums as integrals). Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, μ is counting measure on $\mathcal{B}$ (see Example 7.8), and $f : \mathbb{N} \to \mathbb{C}$ is a function. From Example 7.8 we have $f \in L^1(\mu)$ iff $\sum_{n=1}^\infty |f(n)| < \infty$, i.e. iff the sum, $\sum_{n=1}^\infty f(n)$ is absolutely convergent. Moreover, if $f \in L^1(\mu)$, we may again write

$$f = \sum_{n=1}^\infty f(n)\, 1_{\{n\}}$$

and then use Corollary 7.28 to conclude that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^\infty \int_{\mathbb{N}} f(n)\, 1_{\{n\}} d\mu = \sum_{n=1}^\infty f(n)\, \mu(\{n\}) = \sum_{n=1}^\infty f(n).$$

So again the integral relative to counting measure is simply the infinite sum **provided** the sum is absolutely convergent.

However if $f(n) = (-1)^n \frac{1}{n}$, then

$$\sum_{n=1}^\infty f(n) := \lim_{N\to\infty} \sum_{n=1}^N f(n)$$

is perfectly well defined while $\int_{\mathbb{N}} f d\mu$ is **not**. In fact in this case we have,

$$\int_{\mathbb{N}} f_\pm d\mu = \infty.$$

The point is that when we write $\sum_{n=1}^\infty f(n)$ the ordering of the terms in the sum may matter. On the other hand, $\int_{\mathbb{N}} f d\mu$ knows nothing about the integer ordering.

The following corollary will be routinely be used in the sequel – often without explicit mention.

Corollary 7.30 (Differentiation Under the Integral). *Suppose that $J \subset \mathbb{R}$ is an open interval and $f : J \times \Omega \to \mathbb{C}$ is a function such that*

1. *$\omega \to f(t,\omega)$ is measurable for each $t \in J$.*
2. *$f(t_0,\cdot) \in L^1(\mu)$ for some $t_0 \in J$.*
3. *$\frac{\partial f}{\partial t}(t,\omega)$ exists for all (t,ω).*
4. *There is a function $g \in \mathrm{L}^1(\mu)$ such that $\left|\frac{\partial f}{\partial t}(t,\cdot)\right| \le g$ for each $t \in J$.*

Then $f(t,\cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$ (i.e. $\int_\Omega |f(t,\omega)|\, d\mu(\omega) < \infty$), $t \to \int_\Omega f(t,\omega)d\mu(\omega)$ is a differentiable function on J, and

$$\frac{d}{dt}\int_\Omega f(t,\omega)d\mu(\omega) = \int_\Omega \frac{\partial f}{\partial t}(t,\omega)d\mu(\omega).$$

Proof. By considering the real and imaginary parts of f separately, we may assume that f is real. Also notice that

$$\frac{\partial f}{\partial t}(t,\omega) = \lim_{n\to\infty} n(f(t+n^{-1},\omega) - f(t,\omega))$$

and therefore, for $\omega \to \frac{\partial f}{\partial t}(t,\omega)$ is a sequential limit of measurable functions and hence is measurable for all $t \in J$. By the mean value theorem,

$$|f(t,\omega) - f(t_0,\omega)| \le g(\omega)\,|t-t_0| \text{ for all } t \in J \tag{7.9}$$

and hence

$$|f(t,\omega)| \le |f(t,\omega) - f(t_0,\omega)| + |f(t_0,\omega)| \le g(\omega)\,|t-t_0| + |f(t_0,\omega)|\,.$$

This shows $f(t,\cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$. Let $G(t) := \int_\Omega f(t,\omega)d\mu(\omega)$, then

$$\frac{G(t)-G(t_0)}{t-t_0} = \int_\Omega \frac{f(t,\omega)-f(t_0,\omega)}{t-t_0}d\mu(\omega).$$

By assumption,

$$\lim_{t\to t_0}\frac{f(t,\omega)-f(t_0,\omega)}{t-t_0} = \frac{\partial f}{\partial t}(t,\omega) \text{ for all } \omega \in \Omega$$

and by Eq. (7.9),

$$\left|\frac{f(t,\omega)-f(t_0,\omega)}{t-t_0}\right| \le g(\omega) \text{ for all } t \in J \text{ and } \omega \in \Omega.$$

Therefore, we may apply the dominated convergence theorem to conclude

$$\begin{aligned}\lim_{n\to\infty}\frac{G(t_n)-G(t_0)}{t_n-t_0} &= \lim_{n\to\infty}\int_\Omega \frac{f(t_n,\omega)-f(t_0,\omega)}{t_n-t_0}d\mu(\omega)\\ &= \int_\Omega \lim_{n\to\infty}\frac{f(t_n,\omega)-f(t_0,\omega)}{t_n-t_0}d\mu(\omega)\\ &= \int_\Omega \frac{\partial f}{\partial t}(t_0,\omega)d\mu(\omega)\end{aligned}$$

for **all** sequences $t_n \in J\setminus\{t_0\}$ such that $t_n \to t_0$. Therefore, $\dot G(t_0) = \lim_{t\to t_0}\frac{G(t)-G(t_0)}{t-t_0}$ exists and

$$\dot G(t_0) = \int_\Omega \frac{\partial f}{\partial t}(t_0,\omega)d\mu(\omega).$$

■

Corollary 7.31. *Suppose that $\{a_n\}_{n=0}^\infty \subset \mathbb{C}$ is a sequence of complex numbers such that series*

$$f(z) := \sum_{n=0}^\infty a_n(z-z_0)^n$$

is convergent for $|z-z_0| < R$, where R is some positive number. Then $f : D(z_0,R) \to \mathbb{C}$ is complex differentiable on $D(z_0,R)$ and

$$f'(z) = \sum_{n=0}^\infty na_n(z-z_0)^{n-1} = \sum_{n=1}^\infty na_n(z-z_0)^{n-1}. \tag{7.10}$$

By induction it follows that $f^{(k)}$ exists for all k and that

$$f^{(k)}(z) = \sum_{n=0}^\infty n(n-1)\ldots(n-k+1)a_n(z-z_0)^{n-1}.$$

Proof. Let $\rho < R$ be given and choose $r \in (\rho,R)$. Since $z = z_0 + r \in D(z_0,R)$, by assumption the series $\sum_{n=0}^\infty a_n r^n$ is convergent and in particular $M := \sup_n |a_n r^n| < \infty$. We now apply Corollary 7.30 with $X = \mathbb{N}\cup\{0\}$, μ being counting measure, $\Omega = D(z_0,\rho)$ and $g(z,n) := a_n(z-z_0)^n$. Since

$$\begin{aligned}|g'(z,n)| &= |na_n(z-z_0)^{n-1}| \le n\,|a_n|\,\rho^{n-1}\\ &\le \frac{1}{r}n\left(\frac{\rho}{r}\right)^{n-1}|a_n|\,r^n \le \frac{1}{r}n\left(\frac{\rho}{r}\right)^{n-1}M\end{aligned}$$

and the function $G(n) := \frac{M}{r} n \left(\frac{\rho}{r}\right)^{n-1}$ is summable (by the Ratio test for example), we may use G as our dominating function. It then follows from Corollary 7.30

$$f(z) = \int_X g(z,n) d\mu(n) = \sum_{n=0}^{\infty} a_n (z-z_0)^n$$

is complex differentiable with the differential given as in Eq. (7.10). ■

Definition 7.32 (Moment Generating Function). *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. The **moment generating function** of X is $M_X : \mathbb{R} \to [0,\infty]$ defined by*

$$M_X(t) := \mathbb{E}\left[e^{tX}\right].$$

Proposition 7.33. *Suppose there exists $\varepsilon > 0$ such that $\mathbb{E}\left[e^{\varepsilon |X|}\right] < \infty$, then $M_X(t)$ is a smooth function of $t \in (-\varepsilon, \varepsilon)$ and*

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n \text{ if } |t| \le \varepsilon. \tag{7.11}$$

In particular,

$$\mathbb{E}X^n = \left(\frac{d}{dt}\right)^n |_{t=0} M_X(t) \text{ for all } n \in \mathbb{N}_0. \tag{7.12}$$

Proof. If $|t| \le \varepsilon$, then

$$\mathbb{E}\left[\sum_{n=0}^{\infty} \frac{|t|^n}{n!} |X|^n\right] \le \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{\varepsilon^n}{n!} |X|^n\right] = \mathbb{E}\left[e^{\varepsilon |X|}\right] < \infty.$$

it $e^{tX} \le e^{\varepsilon |X|}$ for all $|t| \le \varepsilon$. Hence it follows from Corollary 7.28 that, for $|t| \le \varepsilon$,

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{t^n}{n!} X^n\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n.$$

Equation (7.12) now is a consequence of Corollary 7.31. ■

Exercise 7.3. Let $d \in \mathbb{N}$, $\Omega = \mathbb{N}_0^d$, $\mathcal{B} = 2^{\Omega}$, $\mu : \mathcal{B} \to \mathbb{N}_0 \cup \{\infty\}$ be counting measure on Ω, and for $x \in \mathbb{R}^d$ and $\omega \in \Omega$, let $x^\omega := x_1^{\omega_1} \dots x_n^{\omega_n}$. Further suppose that $f : \Omega \to \mathbb{C}$ is function and $r_i > 0$ for $1 \le i \le d$ such that

$$\sum_{\omega \in \Omega} |f(\omega)| r^\omega = \int_\Omega |f(\omega)| r^\omega d\mu(\omega) < \infty,$$

where $r := (r_1, \dots, r_d)$. Show;

1. There is a constant, $C < \infty$ such that $|f(\omega)| \le \frac{C}{r^\omega}$ for all $\omega \in \Omega$.
2. Let
$$U := \left\{x \in \mathbb{R}^d : |x_i| < r_i \ \forall\ i\right\} \text{ and } \bar{U} = \left\{x \in \mathbb{R}^d : |x_i| \le r_i \ \forall\ i\right\}$$
Show $\sum_{\omega \in \Omega} |f(\omega) x^\omega| < \infty$ for all $x \in \bar{U}$ and the function, $F : U \to \mathbb{R}$ defined by
$$F(x) = \sum_{\omega \in \Omega} f(\omega) x^\omega \text{ is continuous on } \bar{U}.$$
3. Show, for all $x \in U$ and $1 \le i \le d$, that
$$\frac{\partial}{\partial x_i} F(x) = \sum_{\omega \in \Omega} \omega_i f(\omega) x^{\omega - e_i}$$
where $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ is the i^{th} – standard basis vector on $\mathbb{R}^d$.
4. For any $\alpha \in \Omega$, let $\partial^\alpha := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \dots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}$ and $\alpha! := \prod_{i=1}^d \alpha_i!$ Explain why we may now conclude that
$$\partial^\alpha F(x) = \sum_{\omega \in \Omega} \alpha! f(\omega) x^{\omega - \alpha} \text{ for all } x \in U. \tag{7.13}$$
5. Conclude that $f(\alpha) = \frac{(\partial \alpha F)(0)}{\alpha!}$ for all $\alpha \in \Omega$.
6. If $g : \Omega \to \mathbb{C}$ is another function such that $\sum_{\omega \in \Omega} g(\omega) x^\omega = \sum_{\omega \in \Omega} f(\omega) x^\omega$ for x in a neighborhood of $0 \in \mathbb{R}^d$, then $g(\omega) = f(\omega)$ for all $\omega \in \Omega$.

Solution to Exercise (7.3). We take each item in turn.

1. If no such C existed, then there would exist $\omega(n) \in \Omega$ such that $|f(\omega(n))| r^{\omega(n)} \ge n$ for all $n \in \mathbb{N}$ and therefore, $\sum_{\omega \in \Omega} |f(\omega)| r^\omega \ge n$ for all $n \in \mathbb{N}$ which violates the assumption that $\sum_{\omega \in \Omega} |f(\omega)| r^\omega < \infty$.
2. If $x \in \bar{U}$, then $|x^\omega| \le r^\omega$ and therefore $\sum_{\omega \in \Omega} |f(\omega) x^\omega| \le \sum_{\omega \in \Omega} |f(\omega)| r^\omega < \infty$. The continuity of F now follows by the DCT where we can take $g(\omega) := |f(\omega)| r^\omega$ as the integrable dominating function.
3. For notational simplicity assume that $i = 1$ and let $\rho_i \in (0, r_i)$ be chosen. Then for $|x_i| < \rho_i$, we have,
$$\left|\omega_1 f(\omega) x^{\omega - e_1}\right| \le \omega_1 \rho^{\omega - e_1} \frac{C}{r^\omega} =: g(\omega)$$
where $\rho = (\rho_1, \dots, \rho_d)$. Notice that $g(\omega)$ is summable since,

$$\sum_{\omega\in\Omega} g(\omega) \le \frac{C}{\rho_1}\sum_{\omega_1=0}^{\infty}\omega_1\left(\frac{\rho_1}{r_1}\right)^{\omega_1}\cdot\prod_{i=2}^{d}\sum_{\omega_i=0}^{\infty}\left(\frac{\rho_i}{r_i}\right)^{\omega_i}$$
$$\le \frac{C}{\rho_1}\prod_{i=2}^{d}\frac{1}{1-\frac{\rho_i}{r_i}}\cdot\sum_{\omega_1=0}^{\infty}\omega_1\left(\frac{\rho_1}{r_1}\right)^{\omega_1} < \infty$$

where the last sum is finite as we saw in the proof of Corollary 7.31. Thus we may apply Corollary 7.30 in order to differentiate past the integral (= sum).
4. This is a simple matter of induction. Notice that each time we differentiate, the resulting function is still defined and differentiable on all of U.
5. Setting $x=0$ in Eq. (7.13) shows $(\partial^\alpha F)(0) = \alpha! f(\alpha)$.
6. This follows directly from the previous item since,

$$\alpha! f(\alpha) = \partial^\alpha\left(\sum_{\omega\in\Omega} f(\omega)x^\omega\right)|_{x=0} = \partial^\alpha\left(\sum_{\omega\in\Omega} g(\omega)x^\omega\right)|_{x=0} = \alpha! g(\alpha).$$

7.2.1 Square Integrable Random Variables and Correlations

Suppose that $(\Omega,\mathcal{B},P)$ is a probability space. We say that $X:\Omega\to\mathbb{R}$ is **integrable** if $X\in L^1(P)$ and **square integrable** if $X\in L^2(P)$. When X is integrable we let $a_X := \mathbb{E}X$ be the **mean** of X.

Now suppose that $X,Y:\Omega\to\mathbb{R}$ are two square integrable random variables. Since
$$0\le |X-Y|^2 = |X|^2+|Y|^2-2|X||Y|,$$
it follows that
$$|XY| \le \frac{1}{2}|X|^2+\frac{1}{2}|Y|^2\in L^1(P).$$
In particular by taking $Y=1$, we learn that $|X|\le\frac{1}{2}\left(1+\left|X^2\right|\right)$ which shows that every square integrable random variable is also integrable.

Definition 7.34. *The* ***covariance,*** $\operatorname{Cov}(X,Y)$, *of two square integrable random variables, X and Y, is defined by*

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X-a_X)(Y-a_Y)] = \mathbb{E}[XY]-\mathbb{E}X\cdot\mathbb{E}Y$$

where $a_X:=\mathbb{E}X$ and $a_Y:=\mathbb{E}Y$. The ***variance*** *of X,*

$$\operatorname{Var}(X) := \operatorname{Cov}(X,X) = \mathbb{E}\left[X^2\right]-(\mathbb{E}X)^2 \tag{7.14}$$

We say that X and Y are ***uncorrelated*** *if $\operatorname{Cov}(X,Y)=0$, i.e. $\mathbb{E}[XY]=\mathbb{E}X\cdot\mathbb{E}Y$. More generally we say $\{X_k\}_{k=1}^n\subset L^2(P)$ are* ***uncorrelated*** *iff $\operatorname{Cov}(X_i,X_j)=0$ for all $i\neq j$.*

It follows from Eq. (7.14) that
$$\operatorname{Var}(X)\le\mathbb{E}\left[X^2\right] \text{ for all } X\in L^2(P). \tag{7.15}$$

Lemma 7.35. *The covariance function, $\operatorname{Cov}(X,Y)$ is bilinear in X and Y and $\operatorname{Cov}(X,Y)=0$ if either X or Y is constant. For any constant k, $\operatorname{Var}(X+k)=\operatorname{Var}(X)$ and $\operatorname{Var}(kX)=k^2\operatorname{Var}(X)$. If $\{X_k\}_{k=1}^n$ are uncorrelated $L^2(P)$ – random variables, then*
$$\operatorname{Var}(S_n)=\sum_{k=1}^n\operatorname{Var}(X_k).$$

Proof. We leave most of this simple proof to the reader. As an example of the type of argument involved, let us prove $\operatorname{Var}(X+k)=\operatorname{Var}(X)$;
$$\begin{aligned}\operatorname{Var}(X+k) &= \operatorname{Cov}(X+k,X+k) = \operatorname{Cov}(X+k,X)+\operatorname{Cov}(X+k,k)\\ &= \operatorname{Cov}(X+k,X) = \operatorname{Cov}(X,X)+\operatorname{Cov}(k,X)\\ &= \operatorname{Cov}(X,X) = \operatorname{Var}(X),\end{aligned}$$
wherein we have used the bilinearity of $\operatorname{Cov}(\cdot,\cdot)$ and the property that $\operatorname{Cov}(Y,k)=0$ whenever k is a constant. ■

Exercise 7.4 (A Weak Law of Large Numbers). Assume $\{X_n\}_{n=1}^\infty$ is a sequence if uncorrelated square integrable random variables which are identically distributed, i.e. $X_n\overset{d}{=}X_m$ for all $m,n\in\mathbb{N}$. Let $S_n:=\sum_{k=1}^n X_k$, $\mu:=\mathbb{E}X_k$ and $\sigma^2:=\operatorname{Var}(X_k)$ (these are independent of k). Show;
$$\begin{aligned}\mathbb{E}\left[\frac{S_n}{n}\right] &= \mu,\\ \mathbb{E}\left(\frac{S_n}{n}-\mu\right)^2 &= \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}, \text{ and}\\ P\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right) &\le \frac{\sigma^2}{n\varepsilon^2}\end{aligned}$$
for all $\varepsilon>0$ and $n\in\mathbb{N}$. (Compare this with Exercise 4.13.)

7.2.2 Some Discrete Distributions

Definition 7.36 (Generating Function). *Suppose that $N:\Omega\to\mathbb{N}_0$ is an integer valued random variable on a probability space, $(\Omega,\mathcal{B},P)$. The generating function associated to N is defined by*
$$G_N(z):=\mathbb{E}\left[z^N\right]=\sum_{n=0}^\infty P(N=n)z^n \text{ for } |z|\le 1. \tag{7.16}$$

By Corollary 7.31, it follows that $P(N=n) = \frac{1}{n!} G_N^{(n)}(0)$ so that G_N can be used to completely recover the distribution of N.

Proposition 7.37 (Generating Functions). *The generating function satisfies,*

$$G_N^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right] \text{ for } |z|<1$$

and

$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1)\right],$$

where it is possible that one and hence both sides of this equation are infinite. In particular, $G'(1) := \lim_{z\uparrow 1} G'(z) = \mathbb{E}N$ *and if* $\mathbb{E}N^2 < \infty$,

$$\operatorname{Var}(N) = G''(1) + G'(1) - \left[G'(1)\right]^2. \tag{7.17}$$

Proof. By Corollary 7.31 for $|z|<1$,

$$\begin{aligned} G_N^{(k)}(z) &= \sum_{n=0}^{\infty} P(N=n)\cdot n(n-1)\dots(n-k+1) z^{n-k} \\ &= \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right]. \end{aligned} \tag{7.18}$$

Since, for $z \in (0,1)$,

$$0 \le N(N-1)\dots(N-k+1) z^{N-k} \uparrow N(N-1)\dots(N-k+1) \text{ as } z\uparrow 1,$$

we may apply the MCT to pass to the limit as $z \uparrow 1$ in Eq. (7.18) to find,

$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1)\right].$$

■

Exercise 7.5 (Some Discrete Distributions). Let $p \in (0,1]$ and $\lambda > 0$. In the four parts below, the distribution of N will be described. You should work out the generating function, $G_N(z)$, in each case and use it to verify the given formulas for $\mathbb{E}N$ and $\operatorname{Var}(N)$.

1. Bernoulli(p) : $P(N=1) = p$ and $P(N=0) = 1-p$. You should find $\mathbb{E}N = p$ and $\operatorname{Var}(N) = p - p^2$.
2. Binomial(n,p) : $P(N=k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0,1,\dots,n$. ($P(N=k)$ is the probability of k successes in a sequence of n independent yes/no experiments with probability of success being p.) You should find $\mathbb{E}N = np$ and $\operatorname{Var}(N) = n\left(p-p^2\right)$.
3. Geometric(p) : $P(N=k) = p(1-p)^{k-1}$ for $k \in \mathbb{N}$. ($P(N=k)$ is the probability that the k^{th} – trial is the first time of success out a sequence of independent trials with probability of success being p.) You should find $\mathbb{E}N = 1/p$ and $\operatorname{Var}(N) = \frac{1-p}{p^2}$.
4. Poisson(λ) : $P(N=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ for all $k \in \mathbb{N}_0$. You should find $\mathbb{E}N = \lambda = \operatorname{Var}(N)$.

Exercise 7.6. Let $S_{n,p} \overset{d}{=} \text{Binomial}(n,p)$, $k \in \mathbb{N}$, $p_n = \lambda_n/n$ where $\lambda_n \to \lambda > 0$ as $n \to \infty$. Show that

$$\lim_{n\to\infty} P(S_{n,p_n} = k) = \frac{\lambda^k}{k!} e^{-\lambda} = P(\text{Poisson}(\lambda) = k).$$

Thus we see that for $p = O(1/n)$ and k not too large relative to n that for large n,

$$P(\text{Binomial}(n,p) = k) \cong P(\text{Poisson}(pn) = k) = \frac{(pn)^k}{k!} e^{-pn}.$$

(We will come back to the Poisson distribution and the related Poisson process later on.)

Solution to Exercise (7.6). We have,

$$\begin{aligned} P(S_{n,p_n} = k) &= \binom{n}{k} (\lambda_n/n)^k (1-\lambda_n/n)^{n-k} \\ &= \frac{\lambda_n^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} (1-\lambda_n/n)^{n-k}. \end{aligned}$$

The result now follows since,

$$\lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{n^k} = 1$$

and

$$\begin{aligned} \lim_{n\to\infty} \ln(1-\lambda_n/n)^{n-k} &= \lim_{n\to\infty} (n-k)\ln(1-\lambda_n/n) \\ &= -\lim_{n\to\infty} \left[(n-k)\lambda_n/n\right] = -\lambda. \end{aligned}$$

7.3 Integration on $\mathbb{R}$

Notation 7.38 *If m is Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$, f is a non-negative Borel measurable function and $a < b$ with $a, b \in \bar{\mathbb{R}}$, we will often write $\int_a^b f(x)\,dx$ or $\int_a^b f dm$ for $\int_{(a,b]\cap\mathbb{R}} f dm$.*

Example 7.39. Suppose $-\infty < a < b < \infty$, $f \in C([a,b],\mathbb{R})$ and m be Lebesgue measure on $\mathbb{R}$. Given a partition,

$$\pi = \{a = a_0 < a_1 < \dots < a_n = b\},$$

let
$$\mathrm{mesh}(\pi) := \max\{|a_j - a_{j-1}| : j = 1, \dots, n\}$$
and
$$f_\pi(x) := \sum_{l=0}^{n-1} f(a_l) 1_{(a_l, a_{l+1}]}(x).$$
Then
$$\int_a^b f_\pi \, dm = \sum_{l=0}^{n-1} f(a_l) m((a_l, a_{l+1}]) = \sum_{l=0}^{n-1} f(a_l)(a_{l+1} - a_l)$$
is a Riemann sum. Therefore if $\{\pi_k\}_{k=1}^\infty$ is a sequence of partitions with $\lim_{k\to\infty} \mathrm{mesh}(\pi_k) = 0$, we know that
$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \, dm = \int_a^b f(x) \, dx \tag{7.19}$$
where the latter integral is the Riemann integral. Using the (uniform) continuity of f on $[a,b]$, it easily follows that $\lim_{k\to\infty} f_{\pi_k}(x) = f(x)$ and that $|f_{\pi_k}(x)| \le g(x) := M 1_{(a,b]}(x)$ for all $x \in (a,b]$ where $M := \max_{x\in[a,b]} |f(x)| < \infty$. Since $\int_{\mathbb{R}} g dm = M(b-a) < \infty$, we may apply D.C.T. to conclude,
$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \, dm = \int_a^b \lim_{k\to\infty} f_{\pi_k} \, dm = \int_a^b f \, dm.$$
This equation with Eq. (7.19) shows
$$\int_a^b f \, dm = \int_a^b f(x) \, dx$$
whenever $f \in C([a,b], \mathbb{R})$, i.e. the Lebesgue and the Riemann integral agree on continuous functions. See Theorem 7.70 below for a more general statement along these lines.

Theorem 7.40 (The Fundamental Theorem of Calculus). *Suppose $-\infty < a < b < \infty$, $f \in C((a,b),\mathbb{R}) \cap L^1((a,b),m)$ and $F(x) := \int_a^x f(y) dm(y)$. Then*

1. *$F \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$.*
2. *$F'(x) = f(x)$ for all $x \in (a,b)$.*
3. *If $G \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$ is an anti-derivative of f on (a,b) (i.e. $f = G'|_{(a,b)}$) then*
$$\int_a^b f(x) dm(x) = G(b) - G(a).$$

Proof. Since $F(x) := \int_{\mathbb{R}} 1_{(a,x)}(y) f(y) dm(y)$, $\lim_{x\to z} 1_{(a,x)}(y) = 1_{(a,z)}(y)$ for m – a.e. y and $\left|1_{(a,x)}(y) f(y)\right| \le 1_{(a,b)}(y) |f(y)|$ is an L^1 – function, it follows from the dominated convergence Theorem 7.27 that F is continuous on $[a,b]$. Simple manipulations show,
$$\begin{aligned}
\left|\frac{F(x+h) - F(x)}{h} - f(x)\right| &= \frac{1}{|h|} \begin{cases} \left|\int_x^{x+h} [f(y) - f(x)] \, dm(y)\right| & \text{if } h > 0 \\ \left|\int_{x+h}^x [f(y) - f(x)] \, dm(y)\right| & \text{if } h < 0 \end{cases} \\
&\le \frac{1}{|h|} \begin{cases} \int_x^{x+h} |f(y) - f(x)| \, dm(y) & \text{if } h > 0 \\ \int_{x+h}^x |f(y) - f(x)| \, dm(y) & \text{if } h < 0 \end{cases} \\
&\le \sup\{|f(y) - f(x)| : y \in [x - |h|, x + |h|]\}
\end{aligned}$$
and the latter expression, by the continuity of f, goes to zero as $h \to 0$. This shows $F' = f$ on (a,b).

For the converse direction, we have by assumption that $G'(x) = F'(x)$ for $x \in (a,b)$. Therefore by the mean value theorem, $F - G = C$ for some constant C. Hence
$$\begin{aligned}
\int_a^b f(x) dm(x) &= F(b) = F(b) - F(a) \\
&= (G(b) + C) - (G(a) + C) = G(b) - G(a).
\end{aligned}$$
■

We can use the above results to integrate some non-Riemann integrable functions:

Example 7.41. For all $\lambda > 0$,
$$\int_0^\infty e^{-\lambda x} dm(x) = \lambda^{-1} \text{ and } \int_{\mathbb{R}} \frac{1}{1+x^2} dm(x) = \pi.$$
The proof of these identities are similar. By the monotone convergence theorem, Example 7.39 and the fundamental theorem of calculus for Riemann integrals (or Theorem 7.40 below),
$$\begin{aligned}
\int_0^\infty e^{-\lambda x} dm(x) &= \lim_{N\to\infty} \int_0^N e^{-\lambda x} dm(x) = \lim_{N\to\infty} \int_0^N e^{-\lambda x} dx \\
&= -\lim_{N\to\infty} \frac{1}{\lambda} e^{-\lambda x}|_0^N = \lambda^{-1}
\end{aligned}$$
and
$$\begin{aligned}
\int_{\mathbb{R}} \frac{1}{1+x^2} dm(x) &= \lim_{N\to\infty} \int_{-N}^N \frac{1}{1+x^2} dm(x) = \lim_{N\to\infty} \int_{-N}^N \frac{1}{1+x^2} dx \\
&= \lim_{N\to\infty} \left[\tan^{-1}(N) - \tan^{-1}(-N)\right] = \pi.
\end{aligned}$$

Let us also consider the functions x^{-p}. Using the MCT and the fundamental theorem of calculus,

$$\begin{aligned}\int_{(0,1]} \frac{1}{x^p}\, dm(x) &= \lim_{n\to\infty} \int_0^1 1_{(\frac{1}{n},1]}(x) \frac{1}{x^p} dm(x) \\ &= \lim_{n\to\infty} \int_{\frac{1}{n}}^1 \frac{1}{x^p} dx = \lim_{n\to\infty} \frac{x^{-p+1}}{1-p}\bigg|_{1/n}^1 \\ &= \begin{cases} \frac{1}{1-p} & \text{if } p<1 \\ \infty & \text{if } p>1 \end{cases}\end{aligned}$$

If $p=1$ we find

$$\int_{(0,1]} \frac{1}{x^p}\, dm(x) = \lim_{n\to\infty} \int_{\frac{1}{n}}^1 \frac{1}{x} dx = \lim_{n\to\infty} \ln(x)|_{1/n}^1 = \infty.$$

Exercise 7.7. Show

$$\int_1^\infty \frac{1}{x^p} dm\,(x) = \begin{cases} \infty & \text{if } p \le 1 \\ \frac{1}{p-1} & \text{if } p>1 \end{cases}.$$

Example 7.42 (Integration of Power Series). Suppose $R>0$ and $\{a_n\}_{n=0}^\infty$ is a sequence of complex numbers such that $\sum_{n=0}^\infty |a_n|\, r^n < \infty$ for all $r \in (0,R)$. Then

$$\int_\alpha^\beta \left(\sum_{n=0}^\infty a_n x^n\right) dm(x) = \sum_{n=0}^\infty a_n \int_\alpha^\beta x^n dm(x) = \sum_{n=0}^\infty a_n \frac{\beta^{n+1} - \alpha^{n+1}}{n+1}$$

for all $-R<\alpha<\beta<R$. Indeed this follows from Corollary 7.28 since

$$\begin{aligned}\sum_{n=0}^\infty \int_\alpha^\beta |a_n|\,|x|^n\, dm(x) &\le \sum_{n=0}^\infty \left(\int_0^{|\beta|} |a_n|\,|x|^n\, dm(x) + \int_0^{|\alpha|} |a_n|\,|x|^n\, dm(x)\right) \\ &\le \sum_{n=0}^\infty |a_n| \frac{|\beta|^{n+1} + |\alpha|^{n+1}}{n+1} \le 2r \sum_{n=0}^\infty |a_n|\, r^n < \infty\end{aligned}$$

where $r = \max(|\beta|, |\alpha|)$.

Example 7.43. Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q}\cap[0,1]$ and define

$$f(x) = \sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x-r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x-r_n|}} = 5 \text{ if } x = r_n.$$

Since, By Theorem 7.40,

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{|x-r_n|}}\, dx &= \int_{r_n}^1 \frac{1}{\sqrt{x-r_n}} dx + \int_0^{r_n} \frac{1}{\sqrt{r_n - x}} dx \\ &= 2\sqrt{x-r_n}|_{r_n}^1 - 2\sqrt{r_n - x}|_0^{r_n} = 2\left(\sqrt{1-r_n} - \sqrt{r_n}\right) \\ &\le 4,\end{aligned}$$

we find

$$\int_{[0,1]} f(x) dm(x) = \sum_{n=1}^\infty 2^{-n} \int_{[0,1]} \frac{1}{\sqrt{|x-r_n|}} dx \le \sum_{n=1}^\infty 2^{-n} 4 = 4 < \infty.$$

In particular, $m(f=\infty)=0$, i.e. that $f<\infty$ for almost every $x\in[0,1]$ and this implies that

$$\sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x-r_n|}} < \infty \text{ for a.e. } x \in [0,1].$$

This result is somewhat surprising since the singularities of the summands form a dense subset of $[0,1]$.

Example 7.44. The following limit holds,

$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n dm(x) = 1. \tag{7.20}$$

DCT Proof. To verify this, let $f_n(x) := \left(1-\frac{x}{n}\right)^n 1_{[0,n]}(x)$. Then $\lim_{n\to\infty} f_n(x) = e^{-x}$ for all $x\ge 0$. Moreover by simple calculus[1]

$$1 - x \le e^{-x} \text{ for all } x \in \mathbb{R}.$$

Therefore, for $x<n$, we have

$$0 \le 1 - \frac{x}{n} \le e^{-x/n} \implies \left(1-\frac{x}{n}\right)^n \le \left[e^{-x/n}\right]^n = e^{-x},$$

from which it follows that

$$0 \le f_n(x) \le e^{-x} \text{ for all } x \ge 0.$$

[1] Since $y = 1-x$ is the tangent line to $y = e^{-x}$ at $x=0$ and e^{-x} is convex up, it follows that $1-x \le e^{-x}$ for all $x\in\mathbb{R}$.

From Example 7.41, we know

$$\int_0^\infty e^{-x} dm(x) = 1 < \infty,$$

so that e^{-x} is an integrable function on $[0,\infty)$. Hence by the dominated convergence theorem,

$$\lim_{n\to\infty}\int_0^n \left(1-\frac{x}{n}\right)^n dm(x) = \lim_{n\to\infty}\int_0^\infty f_n(x)dm(x)$$
$$= \int_0^\infty \lim_{n\to\infty} f_n(x)dm(x) = \int_0^\infty e^{-x}dm(x) = 1.$$

MCT Proof. The limit in Eq. (7.20) may also be computed using the monotone convergence theorem. To do this we must show that $n \to f_n(x)$ is increasing in n for each x and for this it suffices to consider $n > x$. But for $n > x$,

$$\frac{d}{dn}\ln f_n(x) = \frac{d}{dn}\left[n\ln\left(1-\frac{x}{n}\right)\right] = \ln\left(1-\frac{x}{n}\right) + \frac{n}{1-\frac{x}{n}}\frac{x}{n^2}$$
$$= \ln\left(1-\frac{x}{n}\right) + \frac{\frac{x}{n}}{1-\frac{x}{n}} = h(x/n)$$

where, for $0 \le y < 1$,

$$h(y) := \ln(1-y) + \frac{y}{1-y}.$$

Since $h(0) = 0$ and

$$h'(y) = -\frac{1}{1-y} + \frac{1}{1-y} + \frac{y}{(1-y)^2} > 0$$

it follows that $h \ge 0$. Thus we have shown, $f_n(x) \uparrow e^{-x}$ as $n \to \infty$ as claimed.

Example 7.45. Suppose that $f_n(x) := n1_{(0,\frac{1}{n}]}(x)$ for $n \in \mathbb{N}$. Then $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$ while

$$\lim_{n\to\infty}\int_{\mathbb{R}} f_n(x)\,dx = \lim_{n\to\infty} 1 = 1 \neq 0 = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)\,dx.$$

The problem is that the best dominating function we can take is

$$g(x) = \sup_n f_n(x) = \sum_{n=1}^\infty n\cdot 1_{(\frac{1}{n+1},\frac{1}{n}]}(x).$$

Notice that

$$\int_{\mathbb{R}} g(x)\,dx = \sum_{n=1}^\infty n\cdot\left(\frac{1}{n}-\frac{1}{n+1}\right) = \sum_{n=1}^\infty \frac{1}{n+1} = \infty.$$

Example 7.46 (Jordan's Lemma). In this example, let us consider the limit;

$$\lim_{n\to\infty}\int_0^\pi \cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}d\theta.$$

Let

$$f_n(\theta) := 1_{(0,\pi]}(\theta)\cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}.$$

Then

$$|f_n| \le 1_{(0,\pi]} \in L^1(m)$$

and

$$\lim_{n\to\infty} f_n(\theta) = 1_{(0,\pi]}(\theta)\,1_{\{\pi\}}(\theta) = 1_{\{\pi\}}(\theta).$$

Therefore by the D.C.T.,

$$\lim_{n\to\infty}\int_0^\pi \cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}d\theta = \int_{\mathbb{R}} 1_{\{\pi\}}(\theta)\,dm(\theta) = m(\{\pi\}) = 0.$$

Example 7.47. Recall from Example 7.41 that

$$\lambda^{-1} = \int_{[0,\infty)} e^{-\lambda x}dm(x) \text{ for all } \lambda > 0.$$

Let $\varepsilon > 0$. For $\lambda \ge 2\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $C_n(\varepsilon) < \infty$ such that

$$0 \le \left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x} = x^n e^{-\lambda x} \le C_n(\varepsilon)e^{-\varepsilon x}.$$

Using this fact, Corollary 7.30 and induction gives

$$n!\lambda^{-n-1} = \left(-\frac{d}{d\lambda}\right)^n \lambda^{-1} = \int_{[0,\infty)} \left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x}dm(x)$$
$$= \int_{[0,\infty)} x^n e^{-\lambda x}dm(x).$$

That is

$$n! = \lambda^n \int_{[0,\infty)} x^n e^{-\lambda x}dm(x). \tag{7.21}$$

Remark 7.48. Corollary 7.30 may be generalized by allowing the hypothesis to hold for $x \in X\setminus E$ where $E \in \mathcal{B}$ is a **fixed** null set, i.e. E must be independent of t. Consider what happens if we formally apply Corollary 7.30 to $g(t) := \int_0^\infty 1_{x\le t}dm(x)$,

$$\dot{g}(t) = \frac{d}{dt}\int_0^\infty 1_{x\le t}dm(x) \overset{?}{=} \int_0^\infty \frac{\partial}{\partial t}1_{x\le t}dm(x).$$

The last integral is zero since $\frac{\partial}{\partial t}1_{x\leq t}=0$ unless $t=x$ in which case it is not defined. On the other hand $g(t)=t$ so that $\dot{g}(t)=1$. (The reader should decide which hypothesis of Corollary 7.30 has been violated in this example.)

Exercise 7.8 (Folland 2.28 on p. 60.). Compute the following limits and justify your calculations:

1. $\lim_{n\to\infty}\int_0^\infty \frac{\sin(\frac{x}{n})}{(1+\frac{x}{n})^n}\,dx$.
2. $\lim_{n\to\infty}\int_0^1 \frac{1+nx^2}{(1+x^2)^n}\,dx$
3. $\lim_{n\to\infty}\int_0^\infty \frac{n\sin(x/n)}{x(1+x^2)}\,dx$
4. For all $a\in\mathbb{R}$ compute,

$$f(a):=\lim_{n\to\infty}\int_a^\infty n(1+n^2x^2)^{-1}dx.$$

Exercise 7.9 (Integration by Parts). Suppose that $f,g:\mathbb{R}\to\mathbb{R}$ are two continuously differentiable functions such that $f'g$, fg', and fg are all Lebesgue integrable functions on $\mathbb{R}$. Prove the following integration by parts formula;

$$\int_{\mathbb{R}} f'(x)\cdot g(x)\,dx=-\int_{\mathbb{R}} f(x)\cdot g'(x)\,dx. \tag{7.22}$$

Similarly show that if Suppose that $f,g:[0,\infty)\to[0,\infty)$ are two continuously differentiable functions such that $f'g$, fg', and fg are all Lebesgue integrable functions on $[0,\infty)$, then

$$\int_0^\infty f'(x)\cdot g(x)\,dx=-f(0)\,g(0)-\int_0^\infty f(x)\cdot g'(x)\,dx. \tag{7.23}$$

Outline: 1. First notice that Eq. (7.22) holds if $f(x)=0$ for $|x|\geq N$ for some $N<\infty$ by undergraduate calculus.

2. Let $\psi:\mathbb{R}\to[0,1]$ be a continuously differentiable function such that $\psi(x)=1$ if $|x|\leq 1$ and $\psi(x)=0$ if $|x|\geq 2$. For any $\varepsilon>0$ let $\psi_\varepsilon(x)=\psi(\varepsilon x)$ Write out the identity in Eq. (7.22) with $f(x)$ being replaced by $f(x)\,\psi_\varepsilon(x)$.

3. Now use the dominated convergence theorem to pass to the limit as $\varepsilon\downarrow 0$ in the identity you found in step 2.

4. A similar outline works to prove Eq. (7.23).

Solution to Exercise (7.9). If f has compact support in $[-N,N]$ for some $N<\infty$, then by undergraduate integration by parts,

$$\begin{aligned}\int_{\mathbb{R}} f'(x)\cdot g(x)\,dx&=\int_{-N}^N f'(x)\cdot g(x)\,dx\\&=f(x)\,g(x)\,|_{-N}^N-\int_{-N}^N f(x)\cdot g'(x)\,dx\\&=-\int_{-N}^N f(x)\cdot g'(x)\,dx=-\int_{\mathbb{R}} f(x)\cdot g'(x)\,dx.\end{aligned}$$

Similarly if f has compact support in $[0,\infty)$, then

$$\begin{aligned}\int_0^\infty f'(x)\cdot g(x)\,dx&=\int_0^N f'(x)\cdot g(x)\,dx\\&=f(x)\,g(x)\,|_0^N-\int_0^N f(x)\cdot g'(x)\,dx\\&=-f(0)\,g(0)-\int_0^N f(x)\cdot g'(x)\,dx\\&=-f(0)-\int_0^\infty f(x)\cdot g'(x)\,dx.\end{aligned}$$

For general f we may apply this identity with $f(x)$ replaced by $\psi_\varepsilon(x)\,f(x)$ to learn,

$$\int_{\mathbb{R}} f'(x)\cdot g(x)\,\psi_\varepsilon(x)\,dx+\int_{\mathbb{R}} f(x)\cdot g(x)\,\psi_\varepsilon'(x)\,dx=-\int_{\mathbb{R}}\psi_\varepsilon(x)\,f(x)\cdot g'(x)\,dx. \tag{7.24}$$

Since $\psi_\varepsilon(x)\to 1$ boundedly and $|\psi_\varepsilon'(x)|=\varepsilon\,|\psi'(\varepsilon x)|\leq C\varepsilon$, we may use the DCT to conclude,

$$\begin{aligned}&\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}} f'(x)\cdot g(x)\,\psi_\varepsilon(x)\,dx=\int_{\mathbb{R}} f'(x)\cdot g(x)\,dx,\\&\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}} f(x)\cdot g'(x)\,\psi_\varepsilon(x)\,dx=\int_{\mathbb{R}} f(x)\cdot g'(x)\,dx,\ \text{and}\\&\left|\int_{\mathbb{R}} f(x)\cdot g(x)\,\psi_\varepsilon'(x)\,dx\right|\leq C\varepsilon\cdot\int_{\mathbb{R}}|f(x)\cdot g(x)|\,dx\to 0\text{ as }\varepsilon\downarrow 0.\end{aligned}$$

Therefore passing to the limit as $\varepsilon\downarrow 0$ in Eq. (7.24) completes the proof of Eq. (7.22). Equation (7.23) is proved in the same way.

Definition 7.49 (Gamma Function). *The* ***Gamma function***, $\Gamma:\mathbb{R}_+\to\mathbb{R}_+$ *is defined by*

$$\Gamma(x):=\int_0^\infty u^{x-1}e^{-u}du \tag{7.25}$$

(The reader should check that $\Gamma(x)<\infty$ for all $x>0$.)

Here are some of the more basic properties of this function.

Example 7.50 (Γ – function properties). Let Γ be the gamma function, then;

1. $\Gamma(1) = 1$ as is easily verified.
2. $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$ as follows by integration by parts;

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty e^{-u}\ u^{x+1}\ \frac{du}{u} = \int_0^\infty u^x \left(-\frac{d}{du}\ e^{-u}\right) du \\ &= x\int_0^\infty u^{x-1}\ e^{-u}\ du = x\ \Gamma(x).\end{aligned}$$

In particular, it follows from items 1. and 2. and induction that

$$\Gamma(n+1) = n! \text{ for all } n \in \mathbb{N}. \tag{7.26}$$

(Equation 7.26was also proved in Eq. (7.21).)

3. $\Gamma(1/2) = \sqrt{\pi}$. This last assertion is a bit trickier. One proof is to make use of the fact (proved below in Lemma 9.29) that

$$\int_{-\infty}^\infty e^{-ar^2} dr = \sqrt{\frac{\pi}{a}} \text{ for all } a > 0. \tag{7.27}$$

Taking $a = 1$ and making the change of variables, $u = r^2$ below implies,

$$\sqrt{\pi} = \int_{-\infty}^\infty e^{-r^2} dr = 2\int_0^\infty u^{-1/2} e^{-u} du = \Gamma(1/2).$$

$$\begin{aligned}\Gamma(1/2) &= 2\int_0^\infty e^{-r^2} dr = \int_{-\infty}^\infty e^{-r^2} dr \\ &= I_1(1) = \sqrt{\pi}.\end{aligned}$$

4. A simple induction argument using items 2. and 3. now shows that

$$\Gamma\left(n+\frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi}$$

where $(-1)!! := 1$ and $(2n-1)!! = (2n-1)(2n-3)\ldots 3\cdot 1$ for $n \in \mathbb{N}$.

7.4 Densities and Change of Variables Theorems

Exercise 7.10 (Measures and Densities). Let $(X, \mathcal{M}, \mu)$ be a measure space and $\rho : X \to [0,\infty]$ be a measurable function. For $A \in \mathcal{M}$, set $\nu(A) := \int_A \rho d\mu$.

1. Show $\nu : \mathcal{M} \to [0,\infty]$ is a measure.
2. Let $f : X \to [0,\infty]$ be a measurable function, show

$$\int_X f d\nu = \int_X f\rho d\mu. \tag{7.28}$$

Hint: first prove the relationship for characteristic functions, then for simple functions, and then for general positive measurable functions.

3. Show that a measurable function $f : X \to \mathbb{C}$ is in $L^1(\nu)$ iff $|f|\rho \in L^1(\mu)$ and if $f \in L^1(\nu)$ then Eq. (7.28) still holds.

Solution to Exercise (7.10). The fact that ν is a measure follows easily from Corollary 7.6. Clearly Eq. (7.28) holds when $f = 1_A$ by definition of ν. It then holds for positive simple functions, f, by linearity. Finally for general $f \in L^+$, choose simple functions, φ_n, such that $0 \le \varphi_n \uparrow f$. Then using MCT twice we find

$$\int_X f d\nu = \lim_{n\to\infty}\int_X \varphi_n d\nu = \lim_{n\to\infty}\int_X \varphi_n \rho d\mu = \int_X \lim_{n\to\infty}\varphi_n\rho d\mu = \int_X f\rho d\mu.$$

By what we have just proved, for all $f : X \to \mathbb{C}$ we have

$$\int_X |f|\, d\nu = \int_X |f|\,\rho d\mu$$

so that $f \in L^1(\mu)$ iff $|f|\,\rho \in L^1(\mu)$. If $f \in L^1(\mu)$ and f is real,

$$\begin{aligned}\int_X f d\nu &= \int_X f_+ d\nu - \int_X f_- d\nu = \int_X f_+\rho d\mu - \int_X f_-\rho d\mu \\ &= \int_X [f_+\rho - f_-\rho]\, d\mu = \int_X f\rho d\mu.\end{aligned}$$

The complex case easily follows from this identity.

Notation 7.51 *It is customary to informally describe ν defined in Exercise 7.10 by writing $d\nu = \rho d\mu$.*

Exercise 7.11 (Abstract Change of Variables Formula). Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Recall that $\nu = f_*\mu : \mathcal{F} \to [0,\infty]$ defined by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$ is a measure on $\mathcal{F}$.

1. Show

$$\int_Y g d\nu = \int_X (g\circ f)\, d\mu \tag{7.29}$$

for all measurable functions $g : Y \to [0,\infty]$. **Hint:** see the hint from Exercise 7.10.

2. Show a measurable function $g : Y \to \mathbb{C}$ is in $L^1(\nu)$ iff $g \circ f \in L^1(\mu)$ and that Eq. (7.29) holds for all $g \in \mathrm{L}^1(\nu)$.

Example 7.52. Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\{X_i\}_{i=1}^n$ are random variables on Ω with $\nu := \mathrm{Law}_P(X_1, \dots, X_n)$, then

$$\mathbb{E}[g(X_1, \dots, X_n)] = \int_{\mathbb{R}^n} g \, d\nu$$

for all $g : \mathbb{R}^n \to \mathbb{R}$ which are Borel measurable and either bounded or non-negative. This follows directly from Exercise 7.11 with $f := (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ and $\mu = P$.

Remark 7.53. As a special case of Example 7.52, suppose that X is a random variable on a probability space, $(\Omega, \mathcal{B}, P)$, and $F(x) := P(X \le x)$. Then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) \, dF(x) \tag{7.30}$$

where $dF(x)$ is shorthand for $d\mu_F(x)$ and μ_F is the unique probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_F((-\infty, x]) = F(x)$ for all $x \in \mathbb{R}$. Moreover if $F : \mathbb{R} \to [0,1]$ happens to be C^1-function, then

$$d\mu_F(x) = F'(x) \, dm(x) \tag{7.31}$$

and Eq. (7.30) may be written as

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) F'(x) \, dm(x). \tag{7.32}$$

To verify Eq. (7.31) it suffices to observe, by the fundamental theorem of calculus, that

$$\mu_F((a,b]) = F(b) - F(a) = \int_a^b F'(x) \, dx = \int_{(a,b]} F' dm.$$

From this equation we may deduce that $\mu_F(A) = \int_A F' dm$ for all $A \in \mathcal{B}_{\mathbb{R}}$. Equation 7.32 now follows from Exercise 7.10.

Exercise 7.12. Let $F : \mathbb{R} \to \mathbb{R}$ be a C^1-function such that $F'(x) > 0$ for all $x \in \mathbb{R}$ and $\lim_{x \to \pm\infty} F(x) = \pm\infty$. (Notice that F is strictly increasing so that $F^{-1} : \mathbb{R} \to \mathbb{R}$ exists and moreover, by the inverse function theorem that F^{-1} is a C^1 – function.) Let m be Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$ and

$$\nu(A) = m(F(A)) = m(\left(F^{-1}\right)^{-1}(A)) = \left(F_*^{-1} m\right)(A)$$

for all $A \in \mathcal{B}_{\mathbb{R}}$. Show $d\nu = F' dm$. Use this result to prove the change of variable formula,

$$\int_{\mathbb{R}} h \circ F \cdot F' dm = \int_{\mathbb{R}} h dm \tag{7.33}$$

which is valid for all Borel measurable functions $h : \mathbb{R} \to [0, \infty]$.

Hint: Start by showing $d\nu = F' dm$ on sets of the form $A = (a,b]$ with $a, b \in \mathbb{R}$ and $a < b$. Then use the uniqueness assertions in Exercise 5.11 to conclude $d\nu = F' dm$ on all of $\mathcal{B}_{\mathbb{R}}$. To prove Eq. (7.33) apply Exercise 7.11 with $g = h \circ F$ and $f = F^{-1}$.

Solution to Exercise (7.12). Let $d\mu = F' dm$ and $A = (a,b]$, then

$$\nu((a,b]) = m(F((a,b])) = m((F(a), F(b)]) = F(b) - F(a)$$

while

$$\mu((a,b]) = \int_{(a,b]} F' dm = \int_a^b F'(x) dx = F(b) - F(a).$$

It follows that both $\mu = \nu = \mu_F$ – where μ_F is the measure described in Theorem 5.33. By Exercise 7.11 with $g = h \circ F$ and $f = F^{-1}$, we find

$$\begin{aligned} \int_{\mathbb{R}} h \circ F \cdot F' dm &= \int_{\mathbb{R}} h \circ F d\nu = \int_{\mathbb{R}} h \circ F d\left(F_*^{-1} m\right) = \int_{\mathbb{R}} (h \circ F) \circ F^{-1} dm \\ &= \int_{\mathbb{R}} h dm. \end{aligned}$$

This result is also valid for all $h \in L^1(m)$.

7.5 Some Common Continuous Distributions

Example 7.54 (Uniform Distribution). Suppose that X has the uniform distribution in $[0,b]$ for some $b \in (0, \infty)$, i.e. $X_* P = \frac{1}{b} \cdot m$ on $[0,b]$. More explicitly,

$$\mathbb{E}[f(X)] = \frac{1}{b} \int_0^b f(x) \, dx \text{ for all bounded measurable } f.$$

The moment generating function for X is;

$$\begin{aligned} M_X(t) &= \frac{1}{b} \int_0^b e^{tx} dx = \frac{1}{bt} \left(e^{tb} - 1\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} (bt)^{n-1} = \sum_{n=0}^{\infty} \frac{b^n}{(n+1)!} t^n. \end{aligned}$$

On the other hand (see Proposition 7.33),

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E} X^n.$$

Thus it follows that

$$\mathbb{E} X^n = \frac{b^n}{n+1}.$$

Of course this may be calculated directly just as easily,

$$\mathbb{E} X^n = \frac{1}{b} \int_0^b x^n dx = \frac{1}{b(n+1)} x^{n+1}|_0^b = \frac{b^n}{n+1}.$$

Definition 7.55. *A random variable $T \geq 0$ is said to be* ***exponential with parameter*** *$\lambda \in [0, \infty)$ provided, $P(T > t) = e^{-\lambda t}$ for all $t \geq 0$. We will write $T \overset{d}{=} E(\lambda)$ for short.*

If $\lambda > 0$, we have

$$P(T > t) = e^{-\lambda t} = \int_t^{\infty} \lambda e^{-\lambda \tau} d\tau$$

from which it follows that $P(T \in (t, t + dt)) = \lambda 1_{t \geq 0} e^{-\lambda t} dt$. Applying Corollary 7.30 repeatedly implies,

$$\mathbb{E} T = \int_0^{\infty} \tau \lambda e^{-\lambda \tau} d\tau = \lambda \left(-\frac{d}{d\lambda} \right) \int_0^{\infty} e^{-\lambda \tau} d\tau = \lambda \left(-\frac{d}{d\lambda} \right) \lambda^{-1} = \lambda^{-1}$$

and more generally that

$$\mathbb{E} T^k = \int_0^{\infty} \tau^k e^{-\lambda \tau} \lambda d\tau = \lambda \left(-\frac{d}{d\lambda} \right)^k \int_0^{\infty} e^{-\lambda \tau} d\tau = \lambda \left(-\frac{d}{d\lambda} \right)^k \lambda^{-1} = k! \lambda^{-k}. \tag{7.34}$$

In particular we see that

$$\mathrm{Var}(T) = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}. \tag{7.35}$$

Alternatively we may compute the moment generating function for T,

$$\begin{aligned} M_T(a) &:= \mathbb{E}\left[e^{aT}\right] = \int_0^{\infty} e^{a\tau} \lambda e^{-\lambda \tau} d\tau \\ &= \int_0^{\infty} e^{a\tau} \lambda e^{-\lambda \tau} d\tau = \frac{\lambda}{\lambda - a} = \frac{1}{1 - a\lambda^{-1}} \end{aligned} \tag{7.36}$$

which is valid for $a < \lambda$. On the other hand (see Proposition 7.33), we know that

$$\mathbb{E}\left[e^{aT}\right] = \sum_{n=0}^{\infty} \frac{a^n}{n!} \mathbb{E}[T^n] \text{ for } |a| < \lambda. \tag{7.37}$$

Comparing this with Eq. (7.36) again shows that Eq. (7.34) is valid.

Here is yet another way to understand and generalize Eq. (7.36). We simply make the change of variables, $u = \lambda \tau$ in the integral in Eq. (7.34) to learn,

$$\mathbb{E} T^k = \lambda^{-k} \int_0^{\infty} u^k e^{-u} d\tau = \lambda^{-k} \Gamma(k+1).$$

This last equation is valid for all $k \in (-1, \infty)$ – in particular k need not be an integer.

Theorem 7.56 (Memoryless property). *A random variable, $T \in (0, \infty]$ has an exponential distribution iff it satisfies the memoryless property:*

$$P(T > s + t | T > s) = P(T > t) \text{ for all } s, t \geq 0,$$

where as usual, $P(A|B) := P(A \cap B) / P(B)$ when $p(B) > 0$. (Note that $T \overset{d}{=} E(0)$ means that $P(T > t) = e^{0t} = 1$ for all $t > 0$ and therefore that $T = \infty$ a.s.)

Proof. (The following proof is taken from [38].) Suppose first that $T \overset{d}{=} E(\lambda)$ for some $\lambda > 0$. Then

$$P(T > s + t | T > s) = \frac{P(T > s + t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t).$$

For the converse, let $g(t) := P(T > t)$, then by assumption,

$$\frac{g(t+s)}{g(s)} = P(T > s + t | T > s) = P(T > t) = g(t)$$

whenever $g(s) \neq 0$ and $g(t)$ is a decreasing function. Therefore if $g(s) = 0$ for some $s > 0$ then $g(t) = 0$ for all $t > s$. Thus it follows that

$$g(t+s) = g(t) g(s) \text{ for all } s, t \geq 0.$$

Since $T > 0$, we know that $g(1/n) = P(T > 1/n) > 0$ for some n and therefore, $g(1) = g(1/n)^n > 0$ and we may write $g(1) = e^{-\lambda}$ for some $0 \leq \lambda < \infty$.

Observe for $p, q \in \mathbb{N}$, $g(p/q) = g(1/q)^p$ and taking $p = q$ then shows, $e^{-\lambda} = g(1) = g(1/q)^q$. Therefore, $g(p/q) = e^{-\lambda p/q}$ so that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{Q}_+ := \mathbb{Q} \cap \mathbb{R}_+$. Given $r, s \in \mathbb{Q}_+$ and $t \in \mathbb{R}$ such that $r \leq t \leq s$ we have, since g is decreasing, that

$$e^{-\lambda r} = g(r) \ge g(t) \ge g(s) = e^{-\lambda s}.$$

Hence letting $s \uparrow t$ and $r \downarrow t$ in the above equations shows that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{R}_+$ and therefore $T \overset{d}{=} E(\lambda)$. ■

Exercise 7.13 (Gamma Distributions). Let X be a positive random variable. For $k, \theta > 0$, we say that $X \overset{d}{=}\mathrm{Gamma}(k,\theta)$ if

$$(X_*P)(dx) = f(x;k,\theta)\,dx \text{ for } x > 0,$$

where

$$f(x;k,\theta) := x^{k-1}\frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \text{ for } x > 0, \text{ and } k,\theta > 0.$$

Find the moment generating function (see Definition 7.32), $M_X(t) = \mathbb{E}\left[e^{tX}\right]$ for $t < \theta^{-1}$. Differentiate your result in t to show

$$\mathbb{E}[X^m] = k(k+1)\dots(k+m-1)\,\theta^m \text{ for all } m \in \mathbb{N}_0.$$

In particular, $\mathbb{E}[X] = k\theta$ and $\mathrm{Var}(X) = k\theta^2$. (Notice that when $k = 1$ and $\theta = \lambda^{-1}$, $X \overset{d}{=} E(\lambda)$.)

7.5.1 Normal (Gaussian) Random Variables

Definition 7.57 (Normal / Gaussian Random Variables). *A random variable, Y,* ***is normal with mean*** *μ* ***standard deviation*** *σ^2 iff*

$$P(Y \in B) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_B e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy \text{ for all } B \in \mathcal{B}_{\mathbb{R}}. \tag{7.38}$$

We will abbreviate this by writing $Y \overset{d}{=} N(\mu,\sigma^2)$. When $\mu = 0$ and $\sigma^2 = 1$ we will simply write N for $N(0,1)$ and if $Y \overset{d}{=} N$, we will say Y is a ***standard normal*** *random variable.*

Observe that Eq. (7.38) is equivalent to writing

$$\mathbb{E}[f(Y)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} f(y)\,e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy$$

for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$. Also observe that $Y \overset{d}{=} N(\mu,\sigma^2)$ is equivalent to $Y \overset{d}{=} \sigma N + \mu$. Indeed, by making the change of variable, $y = \sigma x + \mu$, we find

$$\begin{aligned}\mathbb{E}[f(\sigma N + \mu)] &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(\sigma x + \mu)\,e^{-\frac{1}{2}x^2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(y)\,e^{-\frac{1}{2\sigma^2}(y-\mu)^2}\frac{dy}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} f(y)\,e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy.\end{aligned}$$

Lastly the constant, $\left(2\pi\sigma^2\right)^{-1/2}$ is chosen so that

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{1}{2}y^2}dy = 1,$$

see Example 7.50 and Lemma 9.29.

Exercise 7.14. Suppose that $X \overset{d}{=} N(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ is a C^1 – function such that $Xf(X)$, $f'(X)$ and $f(X)$ are all integrable random variables. Show

$$\begin{aligned}\mathbb{E}[Xf(X)] &= -\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)\,\frac{d}{dx}e^{-\frac{1}{2}x^2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f'(x)\,e^{-\frac{1}{2}x^2}dx = \mathbb{E}[f'(X)].\end{aligned}$$

Example 7.58. Suppose that $X \overset{d}{=} N(0,1)$ and define $\alpha_k := \mathbb{E}\left[X^{2k}\right]$ for all $k \in \mathbb{N}_0$. By Exercise 7.14,

$$\alpha_{k+1} = \mathbb{E}\left[X^{2k+1}\cdot X\right] = (2k+1)\,\alpha_k \text{ with } \alpha_0 = 1.$$

Hence it follows that

$$\alpha_1 = \alpha_0 = 1,\ \alpha_2 = 3\alpha_1 = 3,\ \alpha_3 = 5\cdot 3$$

and by a simple induction argument,

$$\mathbb{E}X^{2k} = \alpha_k = (2k-1)!!, \tag{7.39}$$

where $(-1)!! := 0$. Actually we can use the Γ – function to say more. Namely for any $\beta > -1$,

$$\mathbb{E}|X|^\beta = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |x|^\beta e^{-\frac{1}{2}x^2}dx = \sqrt{\frac{2}{\pi}}\int_0^\infty x^\beta e^{-\frac{1}{2}x^2}dx.$$

Now make the change of variables, $y = x^2/2$ (i.e. $x = \sqrt{2y}$ and $dx = \frac{1}{\sqrt{2}}y^{-1/2}dy$) to learn,

$$\begin{aligned}\mathbb{E}|X|^\beta &= \frac{1}{\sqrt{\pi}}\int_0^\infty (2y)^{\beta/2}\,e^{-y}y^{-1/2}dy \\ &= \frac{1}{\sqrt{\pi}}2^{\beta/2}\int_0^\infty y^{(\beta+1)/2}e^{-y}y^{-1}dy = \frac{1}{\sqrt{\pi}}2^{\beta/2}\Gamma\left(\frac{\beta+1}{2}\right).\end{aligned} \tag{7.40}$$

Exercise 7.15. Suppose that $X \overset{d}{=} N(0,1)$ and $\lambda \in \mathbb{R}$. Show

$$f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right] = \exp\left(-\lambda^2/2\right). \tag{7.41}$$

Hint: Use Corollary 7.30 to show, $f'(\lambda) = i\mathbb{E}\left[Xe^{i\lambda X}\right]$ and then use Exercise 7.14 to see that $f'(\lambda)$ satisfies a simple ordinary differential equation.

Solution to Exercise (7.15). Using Corollary 7.30 and Exercise 7.14,

$$\begin{aligned} f'(\lambda) &= i\mathbb{E}\left[Xe^{i\lambda X}\right] = i\mathbb{E}\left[\frac{d}{dX}e^{i\lambda X}\right] \\ &= i\cdot(i\lambda)\,\mathbb{E}\left[e^{i\lambda X}\right] = -\lambda f(\lambda) \text{ with } f(0) = 1. \end{aligned}$$

Solving for the unique solution of this differential equation gives Eq. (7.41).

Exercise 7.16. Suppose that $X \overset{d}{=} N(0,1)$ and $t \in \mathbb{R}$. Show $\mathbb{E}\left[e^{tX}\right] = \exp\left(t^2/2\right)$. (You could follow the hint in Exercise 7.15 or you could use a completion of the squares argument along with the translation invariance of Lebesgue measure.)

Exercise 7.17. Use Exercise 7.16 and Proposition 7.33 to give another proof that $\mathbb{E}X^{2k} = (2k-1)!!$ when $X \overset{d}{=} N(0,1)$.

Exercise 7.18. Let $X \overset{d}{=} N(0,1)$ and $\alpha \in \mathbb{R}$, find $\rho : \mathbb{R}_+ \to \mathbb{R}_+ := (0,\infty)$ such that

$$\mathbb{E}\left[f\left(|X|^\alpha\right)\right] = \int_{\mathbb{R}_+} f(x)\,\rho(x)\,dx$$

for all continuous functions, $f : \mathbb{R}_+ \to \mathbb{R}$ with compact support in $\mathbb{R}_+$.

Lemma 7.59 (Gaussian tail estimates). *Suppose that X is a standard normal random variable, i.e.*

$$P(X \in A) = \frac{1}{\sqrt{2\pi}}\int_A e^{-x^2/2}dx \text{ for all } A \in \mathcal{B}_{\mathbb{R}},$$

then for all $x \geq 0$,

$$P(X \geq x) \leq \min\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}}e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x}e^{-x^2/2}\right) \leq \frac{1}{2}e^{-x^2/2}. \tag{7.42}$$

Moreover (see [41, Lemma 2.5]),

$$P(X \geq x) \geq \max\left(1 - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right) \tag{7.43}$$

which combined with Eq. (7.42) proves Mill's ratio (see [21]);

$$\lim_{x\to\infty}\frac{P(X \geq x)}{\frac{1}{\sqrt{2\pi}x}e^{-x^2/2}} = 1. \tag{7.44}$$

Proof. See Figure 7.1 where; the green curve is the plot of $P(X \geq x)$, the black is the plot of

$$\min\left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}x}e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x}e^{-x^2/2}\right),$$

the red is the plot of $\frac{1}{2}e^{-x^2/2}$, and the blue is the plot of

$$\max\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right).$$

The formal proof of these estimates for the reader who is not convinced by

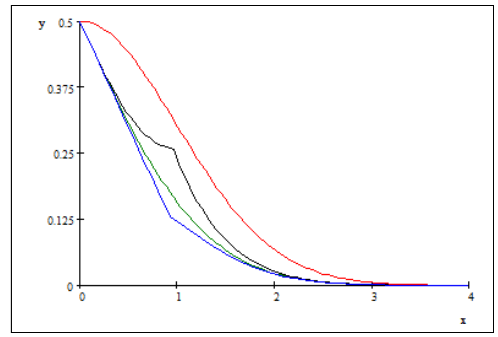


Fig. 7.1. Plots of $P(X \geq x)$ and its estimates.

Figure 7.1 is given below.

We begin by observing that

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy \leq \frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{y}{x}e^{-y^2/2}dy \\ &\leq -\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-y^2/2}|_x^{-\infty} = \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}. \end{aligned} \tag{7.45}$$

If we only want to prove Mill's ratio (7.44), we could proceed as follows. Let $\alpha > 1$, then for $x > 0$,

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy \\ &\geq \frac{1}{\sqrt{2\pi}}\int_x^{\alpha x}\frac{y}{\alpha x}e^{-y^2/2}dy = -\frac{1}{\sqrt{2\pi}}\frac{1}{\alpha x}e^{-y^2/2}|_{y=x}^{y=\alpha x} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\alpha x}e^{-x^2/2}\left[1 - e^{-\alpha^2 x^2/2}\right] \end{aligned}$$

from which it follows,

$$\liminf_{x\to\infty}\left[\sqrt{2\pi}xe^{x^2/2}\cdot P\left(X\geq x\right)\right]\geq 1/\alpha\uparrow 1 \text{ as } \alpha\downarrow 1.$$

The estimate in Eq. (7.45) shows $\limsup_{x\to\infty}\left[\sqrt{2\pi}xe^{x^2/2}\cdot P\left(X\geq x\right)\right]\leq 1.$

To get more precise estimates, we begin by observing,

$$\begin{aligned} P\left(X\geq x\right) &= \frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-y^2/2}dy \qquad (7.46)\\ &\leq \frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-x^2/2}dy\leq \frac{1}{2}-\frac{1}{\sqrt{2\pi}}e^{-x^2/2}x. \end{aligned}$$

This equation along with Eq. (7.45) gives the first equality in Eq. (7.42). To prove the second equality observe that $\sqrt{2\pi}>2$, so

$$\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}\leq\frac{1}{2}e^{-x^2/2} \text{ if } x\geq 1.$$

For $x\leq 1$ we must show,

$$\frac{1}{2}-\frac{x}{\sqrt{2\pi}}e^{-x^2/2}\leq\frac{1}{2}e^{-x^2/2}$$

or equivalently that $f\left(x\right):=e^{x^2/2}-\sqrt{\frac{2}{\pi}}x\leq 1$ for $0\leq x\leq 1$. Since f is convex $\left(f''\left(x\right)=\left(x^2+1\right)e^{x^2/2}>0\right)$, $f\left(0\right)=1$ and $f\left(1\right)\cong 0.85<1$, it follows that $f\leq 1$ on $\left[0,1\right]$. This proves the second inequality in Eq. (7.42).

It follows from Eq. (7.46) that

$$\begin{aligned} P\left(X\geq x\right) &= \frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-y^2/2}dy\\ &\geq \frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x 1dy=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}x \text{ for all } x\geq 0. \end{aligned}$$

So to finish the proof of Eq. (7.43) we must show,

$$\begin{aligned} f\left(x\right) &:= \frac{1}{\sqrt{2\pi}}xe^{-x^2/2}-\left(1+x^2\right)P\left(X\geq x\right)\\ &= \frac{1}{\sqrt{2\pi}}\left[xe^{-x^2/2}-\left(1+x^2\right)\int_x^\infty e^{-y^2/2}dy\right]\leq 0 \text{ for all } 0\leq x<\infty. \end{aligned}$$

This follows by observing that $f\left(0\right)=-1/2<0$, $\lim_{x\uparrow\infty}f\left(x\right)=0$ and

$$\begin{aligned} f'\left(x\right) &= \frac{1}{\sqrt{2\pi}}\left[e^{-x^2/2}\left(1-x^2\right)-2xP\left(X\geq x\right)+\left(1+x^2\right)e^{-x^2/2}\right]\\ &= 2\left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}-xP\left(X\geq y\right)\right)\geq 0, \end{aligned}$$

where the last inequality is a consequence Eq. (7.42). ■

7.6 Stirling's Formula

On occasion one is faced with estimating an integral of the form, $\int_J e^{-G(t)}dt$, where $J=(a,b)\subset\mathbb{R}$ and $G\left(t\right)$ is a C^1 – function with a unique (for simplicity) global minimum at some point $t_0\in J$. The idea is that the majority contribution of the integral will often come from some neighborhood, $\left(t_0-\alpha,t_0+\alpha\right)$, of t_0. Moreover, it may happen that $G\left(t\right)$ can be well approximated on this neighborhood by its Taylor expansion to order 2;

$$G\left(t\right)\cong G\left(t_0\right)+\frac{1}{2}\ddot{G}\left(t_0\right)\left(t-t_0\right)^2.$$

Notice that the linear term is zero since t_0 is a minimum and therefore $\dot{G}\left(t_0\right)=0$. We will further assume that $\ddot{G}\left(t_0\right)\neq 0$ and hence $\ddot{G}\left(t_0\right)>0$. Under these hypothesis we will have,

$$\int_J e^{-G(t)}dt\cong e^{-G(t_0)}\int_{|t-t_0|<\alpha}\exp\left(-\frac{1}{2}\ddot{G}\left(t_0\right)\left(t-t_0\right)^2\right)dt.$$

Making the change of variables, $s=\sqrt{\ddot{G}\left(t_0\right)}\left(t-t_0\right)$, in the above integral then gives,

$$\begin{aligned} \int_J e^{-G(t)}dt &\cong \frac{1}{\sqrt{\ddot{G}\left(t_0\right)}}e^{-G(t_0)}\int_{|s|<\sqrt{\ddot{G}(t_0)}\cdot\alpha}e^{-\frac{1}{2}s^2}ds\\ &= \frac{1}{\sqrt{\ddot{G}\left(t_0\right)}}e^{-G(t_0)}\left[\sqrt{2\pi}-\int_{\sqrt{\ddot{G}(t_0)}\cdot\alpha}^\infty e^{-\frac{1}{2}s^2}ds\right]\\ &= \frac{1}{\sqrt{\ddot{G}\left(t_0\right)}}e^{-G(t_0)}\left[\sqrt{2\pi}-O\left(\frac{1}{\sqrt{\ddot{G}\left(t_0\right)}\cdot\alpha}e^{-\frac{1}{2}\ddot{G}(t_0)\cdot\alpha^2}\right)\right]. \end{aligned}$$

If α is sufficiently large, for example if $\sqrt{\ddot{G}\left(t_0\right)}\cdot\alpha=3$, then the error term is about 0.0037 and we should be able to conclude that

$$\int_J e^{-G(t)}dt\cong\sqrt{\frac{2\pi}{\ddot{G}\left(t_0\right)}}e^{-G(t_0)}. \qquad (7.47)$$

The proof of the next theorem (Stirling's formula for the Gamma function) will illustrate these ideas and what one has to do to carry them out rigorously.

Theorem 7.60 (Stirling's formula). *The Gamma function (see Definition 7.49), satisfies Stirling's formula,*

$$\lim_{x\to\infty}\frac{\Gamma(x+1)}{\sqrt{2\pi}e^{-x}x^{x+1/2}}=1. \tag{7.48}$$

In particular, if $n\in\mathbb{N}$, we have

$$n!=\Gamma(n+1)\sim\sqrt{2\pi}e^{-n}n^{n+1/2}$$

where we write $a_n\sim b_n$ to mean, $\lim_{n\to\infty}\frac{a_n}{b_n}=1$. (See Example 7.65 below for a slightly cruder but more elementary estimate of $n!$)

Proof. (The following proof is an elaboration of the proof found on page 236-237 in Krantz's Real Analysis and Foundations.) We begin with the formula for $\Gamma(x+1)$;

$$\Gamma(x+1)=\int_0^\infty e^{-t}t^x dt=\int_0^\infty e^{-G_x(t)}dt, \tag{7.49}$$

where

$$G_x(t):=t-x\ln t.$$

Then $\dot{G}_x(t)=1-x/t$, $\ddot{G}_x(t)=x/t^2$, G_x has a global minimum (since $\ddot{G}_x>0$) at $t_0=x$ where

$$G_x(x)=x-x\ln x \text{ and } \ddot{G}_x(x)=1/x.$$

So if Eq. (7.47) is valid in this case we should expect,

$$\Gamma(x+1)\cong\sqrt{2\pi x}e^{-(x-x\ln x)}=\sqrt{2\pi}e^{-x}x^{x+1/2}$$

which would give Stirling's formula. The rest of the proof will be spent on rigorously justifying the approximations involved.

Let us begin by making the change of variables $s=\sqrt{\ddot{G}(t_0)}(t-t_0)=\frac{1}{\sqrt{x}}(t-x)$ as suggested above. Then

$$\begin{aligned}G_x(t)-G_x(x)&=(t-x)-x\ln(t/x)=\sqrt{x}s-x\ln\left(\frac{x+\sqrt{x}s}{x}\right)\\&=x\left[\frac{s}{\sqrt{x}}-\ln\left(1+\frac{s}{\sqrt{x}}\right)\right]=s^2q\left(\frac{s}{\sqrt{x}}\right)\end{aligned}$$

where

$$q(u):=\frac{1}{u^2}\left[u-\ln(1+u)\right] \text{ for } u>-1 \text{ with } q(0):=\frac{1}{2}.$$

Setting $q(0)=1/2$ makes q a continuous and in fact smooth function on $(-1,\infty)$, see Figure 7.2. Using the power series expansion for $\ln(1+u)$ we find,

$$q(u)=\frac{1}{2}+\sum_{k=3}^\infty\frac{(-u)^{k-2}}{k} \text{ for } |u|<1. \tag{7.50}$$

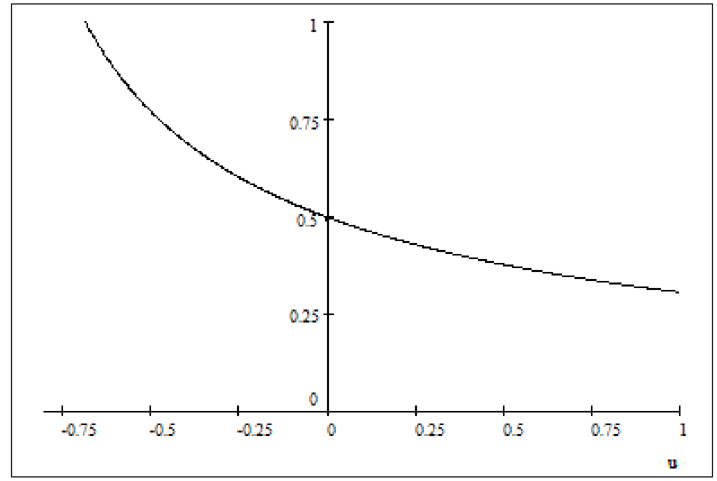


Fig. 7.2. Plot of $q(u)$.

Making the change of variables, $t=x+\sqrt{x}s$ in the second integral in Eq. (7.49) yields,

$$\Gamma(x+1)=e^{-(x-x\ln x)}\sqrt{x}\int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds=x^{x+1/2}e^{-x}\cdot I(x),$$

where

$$I(x)=\int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds=\int_{-\infty}^\infty 1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds. \tag{7.51}$$

From Eq. (7.50) it follows that $\lim_{u\to0}q(u)=1/2$ and therefore,

$$\int_{-\infty}^\infty\lim_{x\to\infty}\left[1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}\right]ds=\int_{-\infty}^\infty e^{-\frac{1}{2}s^2}ds=\sqrt{2\pi}. \tag{7.52}$$

So if there exists a dominating function, $F\in L^1(\mathbb{R},m)$, such that

$$1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}\le F(s) \text{ for all } s\in\mathbb{R} \text{ and } x\ge1,$$

we can apply the DCT to learn that $\lim_{x\to\infty}I(x)=\sqrt{2\pi}$ which will complete the proof of Stirling's formula.

We now construct the desired function F. From Eq. (7.50) it follows that $q(u)\ge1/2$ for $-1<u\le0$. Since $u-\ln(1+u)>0$ for $u\ne0$ ($u-\ln(1+u)$ is convex and has a minimum of 0 at $u=0$) we may conclude that $q(u)>0$ for all $u > -1$ therefore by compactness (on $[0, M]$), $\min_{-1<u\leq M} q(u) = \varepsilon(M) > 0$ for all $M \in (0,\infty)$, see Remark 7.61 for more explicit estimates. Lastly, since $\frac{1}{u}\ln(1+u) \to 0$ as $u \to \infty$, there exists $M < \infty$ ($M = 3$ would due) such that $\frac{1}{u}\ln(1+u) \leq \frac{1}{2}$ for $u \geq M$ and hence,

$$q(u) = \frac{1}{u}\left[1 - \frac{1}{u}\ln(1+u)\right] \geq \frac{1}{2u} \text{ for } u \geq M.$$

So there exists $\varepsilon > 0$ and $M < \infty$ such that (for all $x \geq 1$),

$$\begin{aligned} 1_{s\geq-\sqrt{x}} e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2} &\leq 1_{-\sqrt{x}<s\leq M} e^{-\varepsilon s^2} + 1_{s\geq M} e^{-\sqrt{x}s/2} \\ &\leq 1_{-\sqrt{x}<s\leq M} e^{-\varepsilon s^2} + 1_{s\geq M} e^{-s/2} \\ &\leq e^{-\varepsilon s^2} + e^{-|s|/2} =: F(s) \in L^1(\mathbb{R}, ds). \end{aligned}$$

■

We will sometimes use the following variant of Eq. (7.48);

$$\lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x} = 1 \tag{7.53}$$

To prove this let x go to $x - 1$ in Eq. (7.48) in order to find,

$$1 = \lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{2\pi}e^{-x}\cdot e\cdot(x-1)^{x-1/2}} = \lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x \cdot \sqrt{\frac{x}{x-1}}\cdot e\cdot\left(1-\frac{1}{x}\right)^x}$$

which gives Eq. (7.53) since

$$\lim_{x\to\infty} \sqrt{\frac{x}{x-1}}\cdot e\cdot\left(1-\frac{1}{x}\right)^x = 1.$$

Remark 7.61 (Estimating $q(u)$ by Taylor's Theorem). Another way to estimate $q(u)$ is to use Taylor's theorem with integral remainder. In general if h is C^2 – function on $[0, 1]$, then by the fundamental theorem of calculus and integration by parts,

$$\begin{aligned} h(1) - h(0) &= \int_0^1 \dot{h}(t)\,dt = -\int_0^1 \dot{h}(t)\,d(1-t) \\ &= -\dot{h}(t)(1-t)|_0^1 + \int_0^1 \ddot{h}(t)(1-t)\,dt \\ &= \dot{h}(0) + \frac{1}{2}\int_0^1 \ddot{h}(t)\,d\nu(t) \end{aligned} \tag{7.54}$$

where $d\nu(t) := 2(1-t)\,dt$ which is a probability measure on $[0, 1]$. Applying this to $h(t) = F(a + t(b-a))$ for a C^2 – function on an interval of points between a and b in $\mathbb{R}$ then implies,

$$F(b) - F(a) = (b-a)\dot{F}(a) + \frac{1}{2}(b-a)^2 \int_0^1 \ddot{F}(a + t(b-a))\,d\nu(t). \tag{7.55}$$

(Similar formulas hold to any order.) Applying this result with $F(x) = x - \ln(1+x)$, $a = 0$, and $b = u \in (-1,\infty)$ gives,

$$u - \ln(1+u) = \frac{1}{2}u^2 \int_0^1 \frac{1}{(1+tu)^2}\,d\nu(t),$$

i.e.

$$q(u) = \frac{1}{2}\int_0^1 \frac{1}{(1+tu)^2}\,d\nu(t).$$

From this expression for $q(u)$ it now easily follows that

$$q(u) \geq \frac{1}{2}\int_0^1 \frac{1}{(1+0)^2}\,d\nu(t) = \frac{1}{2} \text{ if } -1 < u \leq 0$$

and

$$q(u) \geq \frac{1}{2}\int_0^1 \frac{1}{(1+u)^2}\,d\nu(t) = \frac{1}{2(1+u)^2}.$$

So an explicit formula for $\varepsilon(M)$ is $\varepsilon(M) = (1+M)^{-2}/2$.

7.6.1 Two applications of Stirling's formula

In this subsection suppose $x \in (0, 1)$ and $S_n \overset{d}{=} \text{Binomial}(n, x)$ for all $n \in \mathbb{N}$, i.e.

$$P_x(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k} \text{ for } 0 \leq k \leq n. \tag{7.56}$$

Recall that $\mathbb{E}S_n = nx$ and $\mathrm{Var}(S_n) = n\sigma^2$ where $\sigma^2 := x(1-x)$. The weak law of large numbers states (Exercise 4.13) that

$$P\left(\left|\frac{S_n}{n} - x\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\sigma^2$$

and therefore, $\frac{S_n}{n}$ is concentrating near its mean value, x, for n large, i.e. $S_n \cong nx$ for n large. The next central limit theorem describes the fluctuations of S_n about nx.

Theorem 7.62 (De Moivre-Laplace Central Limit Theorem). *For all* $-\infty < a < b < \infty,$

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - nx}{\sigma\sqrt{n}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}dy$$
$$= P\left(a \le N \le b\right)$$

where $N \overset{d}{=} N(0,1)$ *. Informally,* $\frac{S_n - nx}{\sigma\sqrt{n}} \overset{d}{\cong} N$ *or equivalently,* $S_n \overset{d}{\cong} nx + \sigma\sqrt{n}\cdot N$ *which if valid in a neighborhood of* nx *whose length is order* $\sqrt{n}$.

Proof. (We are not going to cover all the technical details in this proof as we will give much more general versions of this theorem later.) Starting with the definition of the Binomial distribution we have,

$$p_n := P\left(a \le \frac{S_n - nx}{\sigma\sqrt{n}} \le b\right) = P\left(S_n \in nx + \sigma\sqrt{n}\,[a,b]\right)$$
$$= \sum_{k \in nx + \sigma\sqrt{n}[a,b]} P\left(S_n = k\right)$$
$$= \sum_{k \in nx + \sigma\sqrt{n}[a,b]} \binom{n}{k} x^k \left(1-x\right)^{n-k}.$$

Letting $k = nx + \sigma\sqrt{n}y_k$, i.e. $y_k = (k - nx)/\sigma\sqrt{n}$ we see that $\varDelta y_k = y_{k+1} - y_k = 1/\left(\sigma\sqrt{n}\right)$. Therefore we may write p_n as

$$p_n = \sum_{y_k \in [a,b]} \sigma\sqrt{n}\binom{n}{k} x^k \left(1-x\right)^{n-k} \varDelta y_k. \tag{7.57}$$

So to finish the proof we need to show, for $k = O\left(\sqrt{n}\right)$ $(y_k = O(1))$, that

$$\sigma\sqrt{n}\binom{n}{k} x^k \left(1-x\right)^{n-k} \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_k^2} \text{ as } n \to \infty \tag{7.58}$$

in which case the sum in Eq. (7.57) may be well approximated by the "Riemann sum;"

$$p_n \sim \sum_{y_k \in [a,b]} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_k^2} \varDelta y_k \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}dy \text{ as } n \to \infty.$$

By Stirling's formula,

$$\sigma\sqrt{n}\binom{n}{k} = \sigma\sqrt{n}\frac{1}{k!}\frac{n!}{(n-k)!} \sim \frac{\sigma\sqrt{n}}{\sqrt{2\pi}}\frac{n^{n+1/2}}{k^{k+1/2}\left(n-k\right)^{n-k+1/2}}$$
$$= \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\left(\frac{k}{n}\right)^{k+1/2}\left(1-\frac{k}{n}\right)^{n-k+1/2}}$$
$$= \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k+1/2}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k+1/2}}$$
$$\sim \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\sqrt{x\left(1-x\right)}}\frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}}$$
$$= \frac{1}{\sqrt{2\pi}}\frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}}.$$

In order to shorten the notation, let $z_k := \frac{\sigma}{\sqrt{n}}y_k = O\left(n^{-1/2}\right)$ so that $k = nx + nz_k = n\left(x + z_k\right)$. In this notation we have shown,

$$\sqrt{2\pi}\sigma\sqrt{n}\binom{n}{k}x^k\left(1-x\right)^{n-k} \sim \frac{x^k\left(1-x\right)^{n-k}}{\left(x+z_k\right)^k\left(1-x-z_k\right)^{n-k}}$$
$$= \frac{1}{\left(1+\frac{1}{x}z_k\right)^k\left(1-\frac{1}{1-x}z_k\right)^{n-k}}$$
$$= \frac{1}{\left(1+\frac{1}{x}z_k\right)^{n(x+z_k)}\left(1-\frac{1}{1-x}z_k\right)^{n(1-x-z_k)}} =: q\left(n,k\right). \tag{7.59}$$

Taking logarithms and using Taylor's theorem we learn

$$n\left(x+z_k\right)\ln\left(1+\frac{1}{x}z_k\right)$$
$$= n\left(x+z_k\right)\left(\frac{1}{x}z_k - \frac{1}{2x^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= nz_k + \frac{n}{2x}z_k^2 + O\left(n^{-3/2}\right) \text{ and}$$
$$n\left(1-x-z_k\right)\ln\left(1-\frac{1}{1-x}z_k\right)$$
$$= n\left(1-x-z_k\right)\left(-\frac{1}{1-x}z_k - \frac{1}{2\left(1-x\right)^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= -nz_k + \frac{n}{2\left(1-x\right)}z_k^2 + O\left(n^{-3/2}\right).$$

and then adding these expressions shows,

$$-\ln q(n,k) = \frac{n}{2} z_k^2 \left(\frac{1}{x} + \frac{1}{1-x}\right) + O\left(n^{-3/2}\right)$$
$$= \frac{n}{2\sigma^2} z_k^2 + O\left(n^{-3/2}\right) = \frac{1}{2} y_k^2 + O\left(n^{-3/2}\right).$$

Combining this with Eq. (7.59) shows,

$$\sigma\sqrt{n}\binom{n}{k} x^k (1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} y_k^2 + O\left(n^{-3/2}\right)\right)$$

which gives the desired estimate in Eq. (7.58). ■

The previous central limit theorem has shown that

$$\frac{S_n}{n} \overset{d}{\cong} x + \frac{\sigma}{\sqrt{n}} N$$

which implies the major fluctuations of S_n/n occur within intervals about x of length $O\left(\frac{1}{\sqrt{n}}\right)$. The next result aims to understand the rare events where S_n/n makes a "large" deviation from its mean value, x – in this case a large deviation is something of size $O(1)$ as $n \to \infty$.

Theorem 7.63 (Binomial Large Deviation Bounds). *Let us continue to use the notation in Theorem 7.62. Then for all $y \in (0,x)$,*

$$\lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) = y \ln \frac{x}{y} + (1-y) \ln \frac{1-x}{1-y}.$$

Roughly speaking,

$$P_x\left(\frac{S_n}{n} \le y\right) \approx e^{-nI_x(y)}$$

where $I_x(y)$ is the "rate function,"

$$I_x(y) := y \ln \frac{y}{x} + (1-y) \ln \frac{1-y}{1-x},$$

see Figure 7.3 for the graph of $I_{1/2}$.

Proof. By definition of the binomial distribution,

$$P_x\left(\frac{S_n}{n} \le y\right) = P_x(S_n \le ny) = \sum_{k \le ny} \binom{n}{k} x^k (1-x)^{n-k}.$$

If $a_k \ge 0$, then we have the following crude estimates on $\sum_{k=0}^{m-1} a_k$,

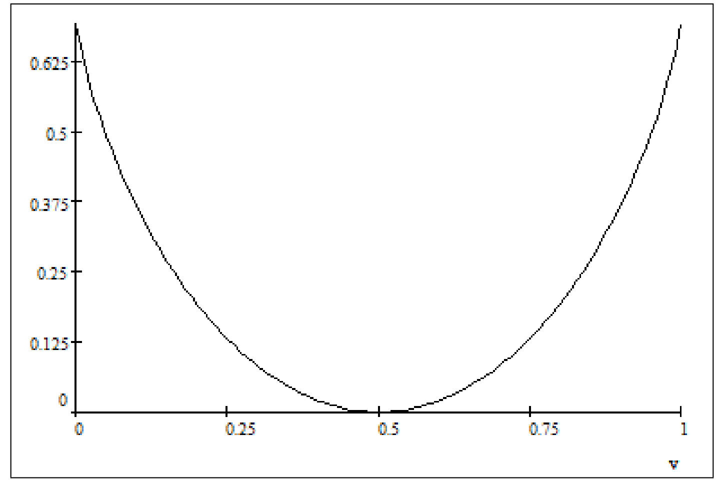


Fig. 7.3. A plot of the rate function, $I_{1/2}$.

$$\max_{k<m} a_k \le \sum_{k=0}^{m-1} a_k \le m \cdot \max_{k<m} a_k. \tag{7.60}$$

In order to apply this with $a_k = \binom{n}{k} x^k (1-x)^{n-k}$ and $m = [ny]$, we need to find the maximum of the a_k for $0 \le k \le ny$. This is easy to do since a_k is increasing for $0 \le k \le ny$ as we now show. Consider,

$$\frac{a_{k+1}}{a_k} = \frac{\binom{n}{k+1} x^{k+1} (1-x)^{n-k-1}}{\binom{n}{k} x^k (1-x)^{n-k}}$$
$$= \frac{k!\,(n-k)! \cdot x}{(k+1)! \cdot (n-k-1)! \cdot (1-x)}$$
$$= \frac{(n-k) \cdot x}{(k+1) \cdot (1-x)}.$$

Therefore, where the latter expression is greater than or equal to 1 iff

$$\frac{a_{k+1}}{a_k} \ge 1 \iff (n-k) \cdot x \ge (k+1) \cdot (1-x)$$
$$\iff nx \ge k + 1 - x \iff k < (n-1)x - 1.$$

Thus for $k < (n-1)x - 1$ we may conclude that $\binom{n}{k} x^k (1-x)^{n-k}$ is increasing in k.

Thus the crude bound in Eq. (7.60) implies,

$$\binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]} \le P_x\left(\frac{S_n}{n} \le y\right) \le [ny] \binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}$$

or equivalently,

$$\begin{aligned}\frac{1}{n} \ln \left[\binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}\right] \\ &\le \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) \\ &\le \frac{1}{n} \ln \left[(ny) \binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}\right].\end{aligned}$$

By Stirling's formula, for k such that k and $n-k$ is large we have,

$$\binom{n}{k} \sim \frac{1}{\sqrt{2\pi}} \frac{n^{n+1/2}}{k^{k+1/2} \cdot (n-k)^{n-k+1/2}} = \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{1}{\left(\frac{k}{n}\right)^{k+1/2} \cdot \left(1-\frac{k}{n}\right)^{n-k+1/2}}$$

and therefore,

$$\frac{1}{n} \ln \binom{n}{k} \sim -\frac{k}{n} \ln\left(\frac{k}{n}\right) - \left(1-\frac{k}{n}\right) \ln\left(1-\frac{k}{n}\right).$$

So taking $k = [ny]$, we learn that

$$\lim_{n\to\infty} \frac{1}{n} \ln \binom{n}{[ny]} = -y \ln y - (1-y) \ln (1-y)$$

and therefore,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) &= -y \ln y - (1-y) \ln (1-y) + y \ln x + (1-y) \ln (1-x) \\ &= y \ln \frac{x}{y} + (1-y) \ln \left(\frac{1-x}{1-y}\right).\end{aligned}$$

■

As a consistency check it is worth noting, by Jensen's inequality described below, that

$$-I_x(y) = y \ln \frac{x}{y} + (1-y) \ln \left(\frac{1-x}{1-y}\right) \le \ln \left(y \frac{x}{y} + (1-y) \frac{1-x}{1-y}\right) = \ln(1) = 0.$$

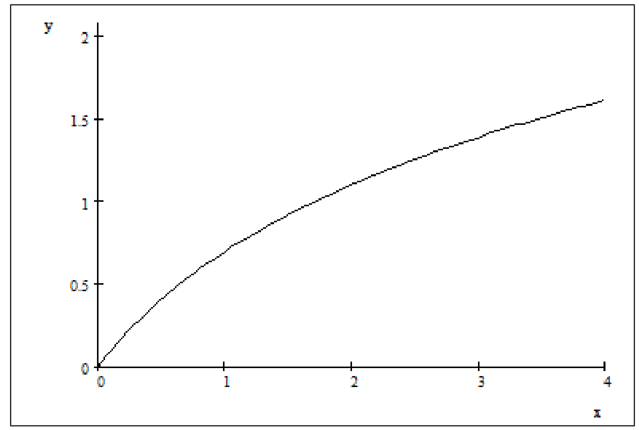
This must be the case since

$$-I_x(y) = \lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) \le \lim_{n\to\infty} \frac{1}{n} \ln 1 = 0.$$

7.6.2 A primitive Stirling type approximation

Theorem 7.64. *Suppose that $f : (0,\infty) \to \mathbb{R}$ is an increasing concave down function (like $f(x) = \ln x$) and let $s_n := \sum_{k=1}^n f(k)$, then*

$$\begin{aligned} s_n - \frac{1}{2}(f(n) + f(1)) &\le \int_1^n f(x)\,dx \\ &\le s_n - \frac{1}{2}[f(n+1) + 2f(1)] + \frac{1}{2} f(2) \\ &\le s_n - \frac{1}{2}[f(n) + 2f(1)] + \frac{1}{2} f(2).\end{aligned}$$



Proof. On the interval, $[k-1,k]$, we have that $f(x)$ is larger than the straight line segment joining $(k-1, f(k-1))$ and $(k, f(k))$ and thus

$$\frac{1}{2}(f(k) + f(k-1)) \le \int_{k-1}^k f(x)\,dx.$$

Summing this equation on $k = 2, \ldots, n$ shows,

$$\begin{aligned} s_n - \frac{1}{2}(f(n) + f(1)) &= \sum_{k=2}^n \frac{1}{2}(f(k) + f(k-1)) \\ &\le \sum_{k=2}^n \int_{k-1}^k f(x)\,dx = \int_1^n f(x)\,dx.\end{aligned}$$

For the upper bound on the integral we observe that $f(x) \le f(k) - f'(k)(x-k)$ for all x and therefore,

$$\int_{k-1}^k f(x)\,dx \le \int_{k-1}^k [f(k) - f'(k)(x-k)]\,dx = f(k) - \frac{1}{2} f'(k).$$

Summing this equation on $k = 2, \ldots, n$ then implies,

$$\int_1^n f(x)\,dx \le \sum_{k=2}^n f(k) - \frac{1}{2}\sum_{k=2}^n f'(k).$$

Since $f''(x) \le 0$, $f'(x)$ is decreasing and therefore $f'(x) \le f'(k-1)$ for $x \in [k-1,k]$ and integrating this equation over $[k-1,k]$ gives

$$f(k) - f(k-1) \le f'(k-1).$$

Summing the result on $k = 3, \ldots, n+1$ then shows,

$$f(n+1) - f(2) \le \sum_{k=2}^n f'(k)$$

and thus ti follows that

$$\begin{aligned}\int_1^n f(x)\,dx &\le \sum_{k=2}^n f(k) - \frac{1}{2}\left(f(n+1) - f(2)\right)\\ &= s_n - \frac{1}{2}\left[f(n+1) + 2f(1)\right] + \frac{1}{2}f(2)\\ &\le s_n - \frac{1}{2}\left[f(n) + 2f(1)\right] + \frac{1}{2}f(2)\end{aligned}$$

■

Example 7.65 (Approximating $n!$). Let us take $f(n) = \ln n$ and recall that

$$\int_1^n \ln x dx = n\ln n - n + 1.$$

Thus we may conlcud that

$$s_n - \frac{1}{2}\ln n \le n\ln n - n + 1 \le s_n - \frac{1}{2}\ln n + \frac{1}{2}\ln 2.$$

Thus it follows that

$$\left(n + \frac{1}{2}\right)\ln n - n + 1 - \ln\sqrt{2} \le s_n \le \left(n + \frac{1}{2}\right)\ln n - n + 1.$$

Exponentiating this identity then implies,

$$\frac{e}{\sqrt{2}} \cdot e^{-n} n^{n+1/2} \le n! \le e \cdot e^{-n} n^{n+1/2}$$

which compares well with Stirling's formula (Theorem 7.60) which states,

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

Observe that

$$\frac{e}{\sqrt{2}} \cong 1.922\,1 \le \sqrt{2\pi} \cong 2.506 \le e \cong 2.718\,3.$$

7.7 Comparison of the Lebesgue and the Riemann Integral*

For the rest of this chapter, let $-\infty < a < b < \infty$ and $f : [a,b] \to \mathbb{R}$ be a bounded function. A partition of $[a,b]$ is a finite subset $\pi \subset [a,b]$ containing $\{a,b\}$. To each partition

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{7.61}$$

of $[a,b]$ let

$$\mathrm{mesh}(\pi) := \max\{|t_j - t_{j-1}| : j = 1, \ldots, n\},$$

$$M_j = \sup\{f(x) : t_j \le x \le t_{j-1}\}, \quad m_j = \inf\{f(x) : t_j \le x \le t_{j-1}\}$$

$$G_\pi = f(a)1_{\{a\}} + \sum_1^n M_j 1_{(t_{j-1},t_j]}, \quad g_\pi = f(a)1_{\{a\}} + \sum_1^n m_j 1_{(t_{j-1},t_j]} \text{ and}$$

$$S_\pi f = \sum M_j(t_j - t_{j-1}) \text{ and } s_\pi f = \sum m_j(t_j - t_{j-1}).$$

Notice that

$$S_\pi f = \int_a^b G_\pi dm \text{ and } s_\pi f = \int_a^b g_\pi dm.$$

The upper and lower Riemann integrals are defined respectively by

$$\overline{\int_a^b} f(x)dx = \inf_\pi S_\pi f \text{ and } \underline{\int_b^a} f(x)dx = \sup_\pi\ s_\pi f.$$

Definition 7.66. *The function f is* ***Riemann integrable*** *iff $\overline{\int_a^b} f = \underline{\int_a^b} f \in \mathbb{R}$ and which case the Riemann integral $\int_a^b f$ is defined to be the common value:*

$$\int_a^b f(x)dx = \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx.$$

The proof of the following Lemma is left to the reader as Exercise 7.29.

Lemma 7.67. *If π' and π are two partitions of $[a,b]$ and $\pi \subset \pi'$ then*

$$\begin{aligned}G_\pi \ge G_{\pi'} \ge f \ge g_{\pi'} \ge g_\pi \text{ and}\\ S_\pi f \ge S_{\pi'} f \ge s_{\pi'} f \ge s_\pi f.\end{aligned}$$

There exists an increasing sequence of partitions $\{\pi_k\}_{k=1}^\infty$ such that $\mathrm{mesh}(\pi_k) \downarrow 0$ and

$$S_{\pi_k} f \downarrow \overline{\int_a^b} f \text{ and } s_{\pi_k} f \uparrow \underline{\int_a^b} f \text{ as } k \to \infty.$$

If we let

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{7.62}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} g dm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{7.63}$$

and

$$\int_{[a,b]} G dm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{7.64}$$

Notation 7.68 *For $x \in [a,b]$, let*

$$H(x) = \limsup_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \sup\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\} \text{ and}$$

$$h(x) = \liminf_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \inf\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\}.$$

Lemma 7.69. *The functions $H, h : [a,b] \to \mathbb{R}$ satisfy:*

1. *$h(x) \le f(x) \le H(x)$ for all $x \in [a,b]$ and $h(x) = H(x)$ iff f is continuous at x.*
2. *If $\{\pi_k\}_{k=1}^\infty$ is any increasing sequence of partitions such that $\mathrm{mesh}(\pi_k) \downarrow 0$ and G and g are defined as in Eq. (7.62), then*
$$G(x) = H(x) \ge f(x) \ge h(x) = g(x) \quad \forall\, x \notin \pi := \cup_{k=1}^\infty \pi_k. \tag{7.65}$$
(Note π is a countable set.)
3. *H and h are Borel measurable.*

Proof. Let $G_k := G_{\pi_k} \downarrow G$ and $g_k := g_{\pi_k} \uparrow g$.

1. It is clear that $h(x) \le f(x) \le H(x)$ for all x and $H(x) = h(x)$ iff $\lim_{y\to x} f(y)$ exists and is equal to $f(x)$. That is $H(x) = h(x)$ iff f is continuous at x.
2. For $x \notin \pi$,
$$G_k(x) \ge H(x) \ge f(x) \ge h(x) \ge g_k(x)\ \forall\, k$$
and letting $k \to \infty$ in this equation implies
$$G(x) \ge H(x) \ge f(x) \ge h(x) \ge g(x)\ \forall\, x \notin \pi. \tag{7.66}$$
Moreover, given $\varepsilon > 0$ and $x \notin \pi$,
$$\sup\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\} \ge G_k(x)$$
for all k large enough, since eventually $G_k(x)$ is the supremum of $f(y)$ over some interval contained in $[x-\varepsilon, x+\varepsilon]$. Again letting $k \to \infty$ implies $\sup_{|y-x|\le\varepsilon} f(y) \ge G(x)$ and therefore, that
$$H(x) = \limsup_{y\to x} f(y) \ge G(x)$$
for all $x \notin \pi$. Combining this equation with Eq. (7.66) then implies $H(x) = G(x)$ if $x \notin \pi$. A similar argument shows that $h(x) = g(x)$ if $x \notin \pi$ and hence Eq. (7.65) is proved.
3. The functions G and g are limits of measurable functions and hence measurable. Since $H = G$ and $h = g$ except possibly on the countable set π, both H and h are also Borel measurable. (You justify this statement.)

■

Theorem 7.70. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{7.67}$$

and the following statements are equivalent:

1. *$H(x) = h(x)$ for m -a.e. x,*
2. *the set*
$$E := \{x \in [a,b] : f \text{ is discontinuous at } x\}$$
is an $\bar{m}$ – null set.
3. *f is Riemann integrable.*

If f is Riemann integrable then f is Lebesgue measurable[2]*, i.e. f is $\mathcal{L}/\mathcal{B}$ – measurable where $\mathcal{L}$ is the Lebesgue σ – algebra and $\mathcal{B}$ is the Borel σ – algebra on $[a,b]$. Moreover if we let $\bar{m}$ denote the completion of m, then*

$$\int_{[a,b]} H dm = \int_a^b f(x)dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{7.68}$$

Proof. Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 7.67 and let G and g be defined as in Lemma 7.69. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (7.67) is a consequence of Eqs. (7.63) and (7.64). From Eq. (7.67), f is Riemann integrable iff

$$\int_{[a,b]} H dm = \int_{[a,b]} h dm$$

[2] f need not be Borel measurable.

and because $h \leq f \leq H$ this happens iff $h(x) = H(x)$ for m - a.e. x. Since $E = \{x : H(x) \neq h(x)\}$, this last condition is equivalent to E being a m – null set. In light of these results and Eq. (7.65), the remaining assertions including Eq. (7.68) are now consequences of Lemma 7.73. ■

Notation 7.71 *In view of this theorem we will often write* $\int_a^b f(x)dx$ *for* $\int_a^b fdm$.

7.8 Measurability on Complete Measure Spaces*

In this subsection we will discuss a couple of measurability results concerning completions of measure spaces.

Proposition 7.72. *Suppose that* $(X, \mathcal{B}, \mu)$ *is a complete measure space*[3] *and* $f : X \to \mathbb{R}$ *is measurable.*

1. *If* $g : X \to \mathbb{R}$ *is a function such that* $f(x) = g(x)$ *for* μ *– a.e.* x, *then* g *is measurable.*
2. *If* $f_n : X \to \mathbb{R}$ *are measurable and* $f : X \to \mathbb{R}$ *is a function such that* $\lim_{n\to\infty} f_n = f$, μ *- a.e., then* f *is measurable as well.*

Proof. 1. Let $E = \{x : f(x) \neq g(x)\}$ which is assumed to be in $\mathcal{B}$ and $\mu(E) = 0$. Then $g = 1_{E^c}f + 1_E g$ since $f = g$ on E^c. Now $1_{E^c}f$ is measurable so g will be measurable if we show $1_E g$ is measurable. For this consider,

$$(1_E g)^{-1}(A) = \begin{cases} E^c \cup (1_E g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\ (1_E g)^{-1}(A) & \text{if } 0 \notin A \end{cases} \tag{7.69}$$

Since $(1_E g)^{-1}(B) \subset E$ if $0 \notin B$ and $\mu(E) = 0$, it follow by completeness of $\mathcal{B}$ that $(1_E g)^{-1}(B) \in \mathcal{B}$ if $0 \notin B$. Therefore Eq. (7.69) shows that $1_E g$ is measurable. 2. Let $E = \{x : \lim\limits_{n\to\infty} f_n(x) \neq f(x)\}$ by assumption $E \in \mathcal{B}$ and $\mu(E) = 0$. Since $g := 1_E f = \lim_{n\to\infty} 1_{E^c} f_n$, g is measurable. Because $f = g$ on E^c and $\mu(E) = 0$, $f = g$ a.e. so by part 1. f is also measurable. ■

The above results are in general false if $(X, \mathcal{B}, \mu)$ is not complete. For example, let $X = \{0, 1, 2\}$, $\mathcal{B} = \{\{0\}, \{1, 2\}, X, \varphi\}$ and $\mu = \delta_0$. Take $g(0) = 0$, $g(1) = 1$, $g(2) = 2$, then $g = 0$ a.e. yet g is not measurable.

Lemma 7.73. *Suppose that* $(X, \mathcal{M}, \mu)$ *is a measure space and* $\bar{\mathcal{M}}$ *is the completion of* $\mathcal{M}$ *relative to* μ *and* $\bar{\mu}$ *is the extension of* μ *to* $\bar{\mathcal{M}}$. *Then a function* $f : X \to \mathbb{R}$ *is* $(\bar{\mathcal{M}}, \mathcal{B} = \mathcal{B}_{\mathbb{R}})$ *– measurable iff there exists a function* $g : X \to \mathbb{R}$ *that is* $(\mathcal{M}, \mathcal{B})$ *– measurable such* $E = \{x : f(x) \neq g(x)\} \in \bar{\mathcal{M}}$ *and* $\bar{\mu}(E) = 0$, *i.e.* $f(x) = g(x)$ *for* $\bar{\mu}$ *– a.e.* x. *Moreover for such a pair* f *and* g, $f \in L^1(\bar{\mu})$ *iff* $g \in L^1(\mu)$ *and in which case*

$$\int_X f d\bar{\mu} = \int_X g d\mu.$$

Proof. Suppose first that such a function g exists so that $\bar{\mu}(E) = 0$. Since g is also $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, we see from Proposition 7.72 that f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable. Conversely if f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, by considering $f_\pm$ we may assume that $f \geq 0$. Choose $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable simple function $\varphi_n \geq 0$ such that $\varphi_n \uparrow f$ as $n \to \infty$. Writing

$$\varphi_n = \sum a_k 1_{A_k}$$

with $A_k \in \bar{\mathcal{M}}$, we may choose $B_k \in \mathcal{M}$ such that $B_k \subset A_k$ and $\bar{\mu}(A_k \setminus B_k) = 0$. Letting

$$\tilde{\varphi}_n := \sum a_k 1_{B_k}$$

we have produced a $(\mathcal{M}, \mathcal{B})$ – measurable simple function $\tilde{\varphi}_n \geq 0$ such that $E_n := \{\varphi_n \neq \tilde{\varphi}_n\}$ has zero $\bar{\mu}$ – measure. Since $\bar{\mu}(\cup_n E_n) \leq \sum_n \bar{\mu}(E_n)$, there exists $F \in \mathcal{M}$ such that $\cup_n E_n \subset F$ and $\mu(F) = 0$. It now follows that

$$1_F \cdot \tilde{\varphi}_n = 1_F \cdot \varphi_n \uparrow g := 1_F f \text{ as } n \to \infty.$$

This shows that $g = 1_F f$ is $(\mathcal{M}, \mathcal{B})$ – measurable and that $\{f \neq g\} \subset F$ has $\bar{\mu}$ – measure zero. Since $f = g$, $\bar{\mu}$ – a.e., $\int_X f d\bar{\mu} = \int_X g d\bar{\mu}$ so to prove Eq. (7.70) it suffices to prove

$$\int_X g d\bar{\mu} = \int_X g d\mu. \tag{7.70}$$

Because $\bar{\mu} = \mu$ on $\mathcal{M}$, Eq. (7.70) is easily verified for non-negative $\mathcal{M}$ – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 6.39 it holds for all $\mathcal{M}$ – measurable functions $g : X \to [0, \infty]$. The rest of the assertions follow in the standard way by considering $(\mathrm{Re}\, g)_\pm$ and $(\mathrm{Im}\, g)_\pm$. ■

7.9 More Exercises

Exercise 7.19. Let μ be a measure on an algebra $\mathcal{A} \subset 2^X$, then $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$ for all $A, B \in \mathcal{A}$.

Exercise 7.20 (From problem 12 on p. 27 of Folland.). Let $(X, \mathcal{M}, \mu)$ be a finite measure space and for $A, B \in \mathcal{M}$ let $\rho(A, B) = \mu(A \Delta B)$ where $A \Delta B = (A \setminus B) \cup (B \setminus A)$. It is clear that $\rho(A, B) = \rho(B, A)$. Show:

[3] Recall this means that if $N \subset X$ is a set such that $N \subset A \in \mathcal{M}$ and $\mu(A) = 0$, then $N \in \mathcal{M}$ as well.

1. ρ satisfies the triangle inequality:
$$\rho(A,C) \le \rho(A,B) + \rho(B,C) \text{ for all } A,B,C \in \mathcal{M}.$$
2. Define $A \sim B$ iff $\mu(A\Delta B) = 0$ and notice that $\rho(A,B) = 0$ iff $A \sim B$. Show "$\sim$ " is an equivalence relation.
3. Let $\mathcal{M}/\sim$ denote $\mathcal{M}$ modulo the equivalence relation, $\sim$, and let $[A] := \{B \in \mathcal{M} : B \sim A\}$. Show that $\bar{\rho}([A],[B]) := \rho(A,B)$ is gives a well defined metric on $\mathcal{M}/\sim$.
4. Similarly show $\tilde{\mu}([A]) = \mu(A)$ is a well defined function on $\mathcal{M}/\sim$ and show $\tilde{\mu} : (\mathcal{M}/\sim) \to \mathbb{R}_+$ is $\bar{\rho}-$ continuous.

Exercise 7.21. Suppose that $\mu_n : \mathcal{M} \to [0,\infty]$ are measures on $\mathcal{M}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in n for all $A \in \mathcal{M}$. Prove that $\mu : \mathcal{M} \to [0,\infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

Exercise 7.22. Now suppose that Λ is some index set and for each $\lambda \in \Lambda$, $\mu_\lambda : \mathcal{M} \to [0,\infty]$ is a measure on $\mathcal{M}$. Define $\mu : \mathcal{M} \to [0,\infty]$ by $\mu(A) = \sum_{\lambda\in\Lambda} \mu_\lambda(A)$ for each $A \in \mathcal{M}$. Show that μ is also a measure.

Exercise 7.23. Let $(X,\mathcal{M},\mu)$ be a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$, show
$$\mu(\{A_n \text{ a.a.}\}) \le \liminf_{n\to\infty} \mu(A_n)$$
and if $\mu(\cup_{m\ge n} A_m) < \infty$ for some n, then
$$\mu(\{A_n \text{ i.o.}\}) \ge \limsup_{n\to\infty} \mu(A_n).$$

Exercise 7.24 (Folland 2.13 on p. 52.). Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of non-negative measurable functions such that $f_n \to f$ pointwise and
$$\lim_{n\to\infty} \int f_n = \int f < \infty.$$
Then
$$\int_E f = \lim_{n\to\infty} \int_E f_n$$
for all measurable sets $E \in \mathcal{M}$. The conclusion need not hold if $\lim_{n\to\infty} \int f_n = \int f$. **Hint:** "Fatou times two."

Exercise 7.25. Give examples of measurable functions $\{f_n\}$ on $\mathbb{R}$ such that f_n decreases to 0 uniformly yet $\int f_n dm = \infty$ for all n. Also give an example of a sequence of measurable functions $\{g_n\}$ on $[0,1]$ such that $g_n \to 0$ while $\int g_n dm = 1$ for all n.

Exercise 7.26. Suppose $\{a_n\}_{n=-\infty}^\infty \subset \mathbb{C}$ is a summable sequence (i.e. $\sum_{n=-\infty}^\infty |a_n| < \infty$), then $f(\theta) := \sum_{n=-\infty}^\infty a_n e^{in\theta}$ is a continuous function for $\theta \in \mathbb{R}$ and
$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta.$$

Exercise 7.27. For any function $f \in L^1(m)$, show $x \in \mathbb{R} \to \int_{(-\infty,x]} f(t)\, dm(t)$ is continuous in x. Also find a finite measure, μ, on $\mathcal{B}_\mathbb{R}$ such that $x \to \int_{(-\infty,x]} f(t)\, d\mu(t)$ is not continuous.

Exercise 7.28. Folland 2.31b and 2.31e on p. 60. (The answer in 2.13b is wrong by a factor of -1 and the sum is on $k = 1$ to ∞. In part (e), s should be taken to be a. You may also freely use the Taylor series expansion
$$(1-z)^{-1/2} = \sum_{n=0}^\infty \frac{(2n-1)!!}{2^n n!} z^n = \sum_{n=0}^\infty \frac{(2n)!}{4^n (n!)^2} z^n \text{ for } |z| < 1.$$

Exercise 7.29. Prove Lemma 7.67.

Functional Forms of the $\pi - \lambda$ Theorem

In this chapter we will develop a very useful function analogue of the $\pi - \lambda$ theorem. The results in this section will be used often in the sequel.

8.1 Multiplicative System Theorems

Notation 8.1 *Let Ω be a set and $\mathbb{H}$ be a subset of the bounded real valued functions on Ω. We say that $\mathbb{H}$ is* ***closed under bounded convergence*** *if; for every sequence, $\{f_n\}_{n=1}^{\infty} \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $|f_n(\omega)| \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f(\omega) := \lim_{n\to\infty} f_n(\omega)$ exists for all $\omega \in \Omega$, then $f \in \mathbb{H}$.*

A subset, $\mathbb{M}$, of $\mathbb{H}$ is called a ***multiplicative system*** *if $\mathbb{M}$ is closed under finite intersections.*

The following result may be found in Dellacherie [11, p. 14]. The style of proof given here may be found in Janson [28, Appendix A., p. 309].

Theorem 8.2 (Dynkin's Multiplicative System Theorem). *Suppose that $\mathbb{H}$ is a vector subspace of bounded functions from Ω to $\mathbb{R}$ which contains the constant functions and is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

Proof. In this proof, we may (and do) assume that $\mathbb{H}$ is the smallest subspace of bounded functions on Ω which contains the constant functions, contains $\mathbb{M}$, and is closed under bounded convergence. (As usual such a space exists by taking the intersection of all such spaces.) The remainder of the proof will be broken into four steps.

Step 1. ($\mathbb{H}$ is an algebra of functions.) For $f \in \mathbb{H}$, let $\mathbb{H}^f := \{g \in \mathbb{H} : gf \in \mathbb{H}\}$. The reader will now easily verify that $\mathbb{H}^f$ is a linear subspace of $\mathbb{H}$, $1 \in \mathbb{H}^f$, and $\mathbb{H}^f$ is closed under bounded convergence. Moreover if $f \in \mathbb{M}$, since $\mathbb{M}$ is a multiplicative system, $\mathbb{M} \subset \mathbb{H}^f$. Hence by the definition of $\mathbb{H}$, $\mathbb{H} = \mathbb{H}^f$, i.e. $fg \in \mathbb{H}$ for all $f \in \mathbb{M}$ and $g \in \mathbb{H}$. Having proved this it now follows for any $f \in \mathbb{H}$ that $\mathbb{M} \subset \mathbb{H}^f$ and therefore as before, $\mathbb{H}^f = \mathbb{H}$. Thus we may conclude that $fg \in \mathbb{H}$ whenever $f, g \in \mathbb{H}$, i.e. $\mathbb{H}$ is an algebra of functions.

Step 2. ($\mathcal{B} := \{A \subset \Omega : 1_A \in \mathbb{H}\}$ is a σ – algebra.) Using the fact that $\mathbb{H}$ is an algebra containing constants, the reader will easily verify that $\mathcal{B}$ is closed under complementation, finite intersections, and contains Ω, i.e. $\mathcal{B}$ is an algebra. Using the fact that $\mathbb{H}$ is closed under bounded convergence, it follows that $\mathcal{B}$ is closed under increasing unions and hence that $\mathcal{B}$ is σ – algebra.

Step 3. ($\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.) Since $\mathbb{H}$ is a vector space and $\mathbb{H}$ contains 1_A for all $A \in \mathcal{B}$, $\mathbb{H}$ contains all $\mathcal{B}$ – measurable simple functions. Since every bounded $\mathcal{B}$ – measurable function may be written as a bounded limit of such simple functions (see Theorem 6.39), it follows that $\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.

Step 4. ($\sigma(\mathbb{M}) \subset \mathcal{B}$.) Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 8.1 below) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f \in \mathbb{M}$ and $a \in \mathbb{R}$, let $F_n := \varphi_n(f - a)$ and $M := \sup_{\omega\in\Omega} |f(\omega) - a|$. By the Weierstrass approximation Theorem 4.36, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$. Since p_l is a polynomial and $\mathbb{H}$ is an algebra, $p_l(f - a) \in \mathbb{H}$ for all l. Moreover, $p_l \circ (f - a) \to F_n$ uniformly as $l \to \infty$, from with it follows that $F_n \in \mathbb{H}$ for all n. Since, $F_n \uparrow 1_{\{f>a\}}$ it follows that $1_{\{f>a\}} \in \mathbb{H}$, i.e. $\{f > a\} \in \mathcal{B}$. As the sets $\{f > a\}$ with $a \in \mathbb{R}$ and $f \in \mathbb{M}$ generate $\sigma(\mathbb{M})$, it follows that $\sigma(\mathbb{M}) \subset \mathcal{B}$.

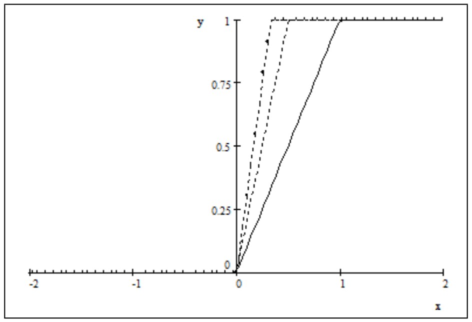


Fig. 8.1. Plots of φ_1, φ_2 and φ_3.

Second proof.* (This proof may safely be skipped.) This proof will make use of Dynkin's $\pi - \lambda$ Theorem 5.14. Let

$$\mathcal{L} := \{A \subset \Omega : 1_A \in \mathbb{H}\}.$$

We then have $\Omega \in \mathcal{L}$ since $1_\Omega = 1 \in \mathbb{H}$, if $A, B \in \mathcal{L}$ with $A \subset B$ then $B \setminus A \in \mathcal{L}$ since $1_{B\setminus A} = 1_B - 1_A \in \mathbb{H}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $A \in \mathcal{L}$ because $1_{A_n} \in \mathbb{H}$ and $1_{A_n} \uparrow 1_A \in \mathbb{H}$. Therefore $\mathcal{L}$ is $\lambda -$ system.

Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 8.1 above) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f_1, f_2, \dots, f_k \in \mathbb{M}$ and $a_1, \dots, a_k \in \mathbb{R}$, let

$$F_n := \prod_{i=1}^k \varphi_n(f_i - a_i)$$

and let

$$M := \sup_{i=1,\dots,k} \sup_\omega |f_i(\omega) - a_i|.$$

By the Weierstrass approximation Theorem 4.36, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$ .Since p_l is a polynomial it is easily seen that $\prod_{i=1}^k p_l \circ (f_i - a_i) \in \mathbb{H}$. Moreover,

$$\prod_{i=1}^k p_l \circ (f_i - a_i) \to F_n \text{ uniformly as } l \to \infty,$$

from with it follows that $F_n \in \mathbb{H}$ for all n. Since,

$$F_n \uparrow \prod_{i=1}^k 1_{\{f_i > a_i\}} = 1_{\cap_{i=1}^k \{f_i > a_i\}}$$

it follows that $1_{\cap_{i=1}^k \{f_i > a_i\}} \in \mathbb{H}$ or equivalently that $\cap_{i=1}^k \{f_i > a_i\} \in \mathcal{L}$. Therefore $\mathcal{L}$ contains the $\pi -$ system, $\mathcal{P}$, consisting of finite intersections of sets of the form, $\{f > a\}$ with $f \in \mathbb{M}$ and $a \in \mathbb{R}$.

As a consequence of the above paragraphs and the $\pi - \lambda$ Theorem 5.14, $\mathcal{L}$ contains $\sigma(\mathcal{P}) = \sigma(\mathbb{M})$. In particular it follows that $1_A \in \mathbb{H}$ for all $A \in \sigma(\mathbb{M})$. Since any positive $\sigma(\mathbb{M}) -$ measurable function may be written as a increasing limit of simple functions (see Theorem 6.39)), it follows that $\mathbb{H}$ contains all non-negative bounded $\sigma(\mathbb{M}) -$ measurable functions. Finally, since any bounded $\sigma(\mathbb{M}) -$ measurable functions may be written as the difference of two such non-negative simple functions, it follows that $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M}) -$ measurable functions. ■

Corollary 8.3. *Suppose $\mathbb{H}$ is a subspace of bounded real valued functions such that $1 \in \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. If $\mathcal{P} \subset 2^\Omega$ is a multiplicative class such that $1_A \in \mathbb{H}$ for all $A \in \mathcal{P}$, then $\mathbb{H}$ contains all bounded $\sigma(\mathcal{P})$ – measurable functions.*

Proof. Let $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$. Then $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system and the proof is completed with an application of Theorem 8.2. ■

Example 8.4. Suppose μ and ν are two probability measure on $(\Omega, \mathcal{B})$ such that

$$\int_\Omega f d\mu = \int_\Omega f d\nu \tag{8.1}$$

for all f in a multiplicative subset, $\mathbb{M}$, of bounded measurable functions on Ω. Then $\mu = \nu$ on $\sigma(\mathbb{M})$. Indeed, apply Theorem 8.2 with $\mathbb{H}$ being the bounded measurable functions on Ω such that Eq. (8.1) holds. In particular if $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$ with $\mathcal{P}$ being a multiplicative class we learn that $\mu = \nu$ on $\sigma(\mathbb{M}) = \sigma(\mathcal{P})$.

Here is a complex version of Theorem 8.2.

Theorem 8.5 (Complex Multiplicative System Theorem). *Suppose $\mathbb{H}$ is a complex linear subspace of the bounded complex functions on Ω, $1 \in \mathbb{H}$, $\mathbb{H}$ is closed under complex conjugation, and $\mathbb{H}$ is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system which is closed under conjugation, then $\mathbb{H}$ contains all bounded complex valued $\sigma(\mathbb{M})$-measurable functions.*

Proof. Let $\mathbb{M}_0 = \operatorname{span}_{\mathbb{C}}(\mathbb{M} \cup \{1\})$ be the complex span of $\mathbb{M}$. As the reader should verify, $\mathbb{M}_0$ is an algebra, $\mathbb{M}_0 \subset \mathbb{H}$, $\mathbb{M}_0$ is closed under complex conjugation and $\sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$. Let

$$\mathbb{H}^{\mathbb{R}} := \{f \in \mathbb{H} : f \text{ is real valued}\} \text{ and}$$
$$\mathbb{M}_0^{\mathbb{R}} := \{f \in \mathbb{M}_0 : f \text{ is real valued}\}.$$

Then $\mathbb{H}^{\mathbb{R}}$ is a real linear space of bounded real valued functions 1 which is closed under bounded convergence and $\mathbb{M}_0^{\mathbb{R}} \subset \mathbb{H}^{\mathbb{R}}$. Moreover, $\mathbb{M}_0^{\mathbb{R}}$ is a multiplicative system (as the reader should check) and therefore by Theorem 8.2, $\mathbb{H}^{\mathbb{R}}$ contains all bounded $\sigma(\mathbb{M}_0^{\mathbb{R}}) -$ measurable real valued functions. Since $\mathbb{H}$ and $\mathbb{M}_0$ are complex linear spaces closed under complex conjugation, for any $f \in \mathbb{H}$ or $f \in \mathbb{M}_0$, the functions $\operatorname{Re} f = \frac{1}{2}(f + \bar{f})$ and $\operatorname{Im} f = \frac{1}{2i}(f - \bar{f})$ are in $\mathbb{H}$ or $\mathbb{M}_0$ respectively. Therefore $\mathbb{M}_0 = \mathbb{M}_0^{\mathbb{R}} + i\mathbb{M}_0^{\mathbb{R}}$, $\sigma(\mathbb{M}_0^{\mathbb{R}}) = \sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$, and $\mathbb{H} = \mathbb{H}^{\mathbb{R}} + i\mathbb{H}^{\mathbb{R}}$. Hence if $f : \Omega \to \mathbb{C}$ is a bounded $\sigma(\mathbb{M}) -$ measurable function, then $f = \operatorname{Re} f + i \operatorname{Im} f \in \mathbb{H}$ since $\operatorname{Re} f$ and $\operatorname{Im} f$ are in $\mathbb{H}^{\mathbb{R}}$. ■

Lemma 8.6. *Suppose that $-\infty < a < b < \infty$ and let Trig$(\mathbb{R}) \subset C(\mathbb{R}, \mathbb{C})$ be the complex linear span of $\{x \to e^{i\lambda x} : \lambda \in \mathbb{R}\}$. Then there exists $f_n \in C_c(\mathbb{R}, [0,1])$ and $g_n \in$ Trig$(\mathbb{R})$ such that $\lim_{n\to\infty} f_n(x) = 1_{(a,b]}(x) = \lim_{n\to\infty} g_n(x)$ for all $x \in \mathbb{R}$.*

Proof. The assertion involving $f_n \in C_c(\mathbb{R}, [0,1])$ was the content of one of your homework assignments. For the assertion involving $g_n \in$Trig$(\mathbb{R})$, it will suffice to show that any $f \in C_c(\mathbb{R})$ may be written as $f(x) = \lim_{n\to\infty} g_n(x)$

for some $\{g_n\} \subset\mathrm{Trig}(\mathbb{R})$ where the limit is uniform for x in compact subsets of $\mathbb{R}$.

So suppose that $f \in C_c(\mathbb{R})$ and $L > 0$ such that $f(x) = 0$ if $|x| \geq L/4$. Then

$$f_L(x) := \sum_{n=-\infty}^{\infty} f(x + nL)$$

is a continuous L – periodic function on $\mathbb{R}$, see Figure 8.2. If $\varepsilon > 0$ is given, we

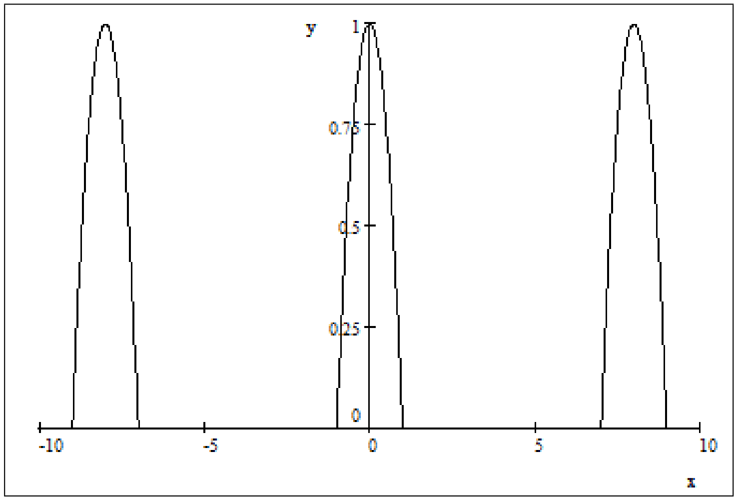


Fig. 8.2. This is plot of $f_8(x)$ where $f(x) = \left(1 - x^2\right) 1_{|x|\leq 1}$. The center hump by itself would be the plot of $f(x)$.

may apply Theorem 4.42 to find $\Lambda \subset\subset \mathbb{Z}$ such that

$$\left| f_L\left(\frac{L}{2\pi}x\right) - \sum_{\alpha\in\Lambda} a_\lambda e^{i\alpha x} \right| \leq \varepsilon \text{ for all } x \in \mathbb{R},$$

wherein we have use the fact that $x \to f_L\left(\frac{L}{2\pi}x\right)$ is a 2π – periodic function of x. Equivalently we have,

$$\max_x \left| f_L(x) - \sum_{\alpha\in\Lambda} a_\lambda e^{i\frac{2\pi\alpha}{L}x} \right| \leq \varepsilon.$$

In particular it follows that $f_L(x)$ is a uniform limit of functions from $\mathrm{Trig}(\mathbb{R})$. Since $\lim_{L\to\infty} f_L(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$, it is easy to conclude there exists $g_n \in\mathrm{Trig}(\mathbb{R})$ such that $\lim_{n\to\infty} g_n(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$. ∎

Corollary 8.7. *Each of the following σ – algebras on $\mathbb{R}^d$ are equal to $\mathcal{B}_{\mathbb{R}^d}$;*

1. $\mathcal{M}_1 := \sigma\left(\cup_{i=1}^n \{x \to f(x_i) : f \in C_c(\mathbb{R})\}\right),$
2. $\mathcal{M}_2 := \sigma\left(x \to f_1(x_1)\dots f_d(x_d) : f_i \in C_c(\mathbb{R})\right)$
3. $\mathcal{M}_3 = \sigma\left(C_c\left(\mathbb{R}^d\right)\right)$, *and*
4. $\mathcal{M}_4 := \sigma\left(\left\{x \to e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\right\}\right).$

Proof. As the functions defining each $\mathcal{M}_i$ are continuous and hence Borel measurable, it follows that $\mathcal{M}_i \subset \mathcal{B}_{\mathbb{R}^d}$ for each i. So to finish the proof it suffices to show $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_i$ for each i.

$\mathcal{M}_1$ case. Let $a, b \in \mathbb{R}$ with $-\infty < a < b < \infty$. By Lemma 8.6, there exists $f_n \in C_c(\mathbb{R})$ such that $\lim_{n\to\infty} f_n = 1_{(a,b]}$. Therefore it follows that $x \to 1_{(a,b]}(x_i)$ is $\mathcal{M}_1$ – measurable for each i. Moreover if $-\infty < a_i < b_i < \infty$ for each i, then we may conclude that

$$x \to \prod_{i=1}^d 1_{(a_i,b_i]}(x_i) = 1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x)$$

is $\mathcal{M}_1$ – measurable as well and hence $(a_1, b_1] \times \dots \times (a_d, b_d] \in \mathcal{M}_1$. As such sets generate $\mathcal{B}_{\mathbb{R}^d}$ we may conclude that $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_1$.

and therefore $\mathcal{M}_1 = \mathcal{B}_{\mathbb{R}^d}$.

$\mathcal{M}_2$ case. As above, we may find $f_{i,n} \to 1_{(a_i,b_i]}$ as $n \to \infty$ for each $1 \leq i \leq d$ and therefore,

$$1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x) = \lim_{n\to\infty} f_{1,n}(x_1)\dots f_{d,n}(x_d) \text{ for all } x \in \mathbb{R}^d.$$

This shows that $1_{(a_1,b_1]\times\dots\times(a_d,b_d]}$ is $\mathcal{M}_2$ – measurable and therefore $(a_1, b_1] \times \dots \times (a_d, b_d] \in \mathcal{M}_2$.

$\mathcal{M}_3$ case. This is easy since $\mathcal{B}_{\mathbb{R}^d} = \mathcal{M}_2 \subset \mathcal{M}_3$.

$\mathcal{M}_4$ case. By Lemma 8.6 here exists $g_n \in\mathrm{Trig}(\mathbb{R})$ such that $\lim_{n\to\infty} g_n = 1_{(a,b]}$. Since $x \to g_n(x_i)$ is in the span $\left\{x \to e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\right\}$ for each n, it follows that $x \to 1_{(a,b]}(x_i)$ is $\mathcal{M}_4$ – measurable for all $-\infty < a < b < \infty$. Therefore, just as in the proof of case 1., we may now conclude that $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_4$. ∎

Corollary 8.8. *Suppose that $\mathbb{H}$ is a subspace of complex valued functions on $\mathbb{R}^d$ which is closed under complex conjugation and bounded convergence. If $\mathbb{H}$ contains any one of the following collection of functions;*

1. $\mathbb{M} := \{x \to f_1(x_1)\dots f_d(x_d) : f_i \in C_c(\mathbb{R})\}$
2. $\mathbb{M} := C_c\left(\mathbb{R}^d\right)$, *or*
3. $\mathbb{M} := \left\{x \to e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\right\}$

then $\mathbb{H}$ contains all bounded complex Borel measurable functions on $\mathbb{R}^d$.

Proof. Observe that if $f \in C_c(\mathbb{R})$ such that $f(x) = 1$ in a neighborhood of 0, then $f_n(x) := f(x/n) \to 1$ as $n \to \infty$. Therefore in cases 1. and 2., $\mathbb{H}$ contains the constant function, 1, since

$$1 = \lim_{n\to\infty} f_n(x_1) \dots f_n(x_d).$$

In case 3, $1 \in \mathbb{M} \subset \mathbb{H}$ as well. The result now follows from Theorem 8.5 and Corollary 8.7. ∎

Proposition 8.9 (Change of Variables Formula). *Suppose that $-\infty < a < b < \infty$ and $u : [a,b] \to \mathbb{R}$ is a continuously differentiable function. Let $[c,d] = u([a,b])$ where $c = \min u([a,b])$ and $d = \max u([a,b])$. (By the intermediate value theorem $u([a,b])$ is an interval.) Then for all bounded measurable functions, $f : [c,d] \to \mathbb{R}$ we have*

$$\int_{u(a)}^{u(b)} f(x)\,dx = \int_a^b f(u(t))\,\dot{u}(t)\,dt. \tag{8.2}$$

Moreover, Eq. (8.2) is also valid if $f : [c,d] \to \mathbb{R}$ is measurable and

$$\int_a^b |f(u(t))|\,|\dot{u}(t)|\,dt < \infty. \tag{8.3}$$

Proof. Let $\mathbb{H}$ denote the space of bounded measurable functions such that Eq. (8.2) holds. It is easily checked that $\mathbb{H}$ is a linear space closed under bounded convergence. Next we show that $\mathbb{M} = C([c,d],\mathbb{R}) \subset \mathbb{H}$ which coupled with Corollary 8.8 will show that $\mathbb{H}$ contains all bounded measurable functions from $[c,d]$ to $\mathbb{R}$.

If $f : [c,d] \to \mathbb{R}$ is a continuous function and let F be an anti-derivative of f. Then by the fundamental theorem of calculus,

$$\begin{aligned}\int_a^b f(u(t))\,\dot{u}(t)\,dt &= \int_a^b F'(u(t))\,\dot{u}(t)\,dt \\ &= \int_a^b \frac{d}{dt} F(u(t))\,dt = F(u(t))|_a^b \\ &= F(u(b)) - F(u(a)) = \int_{u(a)}^{u(b)} F'(x)\,dx = \int_{u(a)}^{u(b)} f(x)\,dx.\end{aligned}$$

Thus $\mathbb{M} \subset \mathbb{H}$ and the first assertion of the proposition is proved.

Now suppose that $f : [c,d] \to \mathbb{R}$ is measurable and Eq. (8.3) holds. For $M < \infty$, let $f_M(x) = f(x) \cdot 1_{|f(x)| \le M}$ – a bounded measurable function. Therefore applying Eq. (8.2) with f replaced by $|f_M|$ shows,

$$\left|\int_{u(a)}^{u(b)} |f_M(x)|\,dx\right| = \left|\int_a^b |f_M(u(t))|\,\dot{u}(t)\,dt\right| \le \int_a^b |f_M(u(t))|\,|\dot{u}(t)|\,dt.$$

Using the MCT, we may let $M \uparrow \infty$ in the previous inequality to learn

$$\left|\int_{u(a)}^{u(b)} |f(x)|\,dx\right| \le \int_a^b |f(u(t))|\,|\dot{u}(t)|\,dt < \infty.$$

Now apply Eq. (8.2) with f replaced by f_M to learn

$$\int_{u(a)}^{u(b)} f_M(x)\,dx = \int_a^b f_M(u(t))\,\dot{u}(t)\,dt.$$

Using the DCT we may now let $M \to \infty$ in this equation to show that Eq. (8.2) remains valid. ∎

Exercise 8.1. Suppose that $u : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function such that $\dot{u}(t) \ge 0$ for all t and $\lim_{t\to\pm\infty} u(t) = \pm\infty$. Show that

$$\int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} f(u(t))\,\dot{u}(t)\,dt \tag{8.4}$$

for all measurable functions $f : \mathbb{R} \to [0,\infty]$. In particular applying this result to $u(t) = at + b$ where $a > 0$ implies,

$$\int_{\mathbb{R}} f(x)\,dx = a \int_{\mathbb{R}} f(at+b)\,dt.$$

Definition 8.10. *The **Fourier transform** or **characteristic function** of a finite measure, μ, on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, is the function, $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^d} e^{i\lambda\cdot x}\,d\mu(x) \text{ for all } \lambda \in \mathbb{R}^d$$

Corollary 8.11. *Suppose that μ and ν are two probability measures on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Then any one of the next three conditions implies that $\mu = \nu$;*

1. $\int_{\mathbb{R}^d} f_1(x_1) \dots f_d(x_d)\,d\nu(x) = \int_{\mathbb{R}^d} f_1(x_1) \dots f_d(x_d)\,d\mu(x)$ *for all* $f_i \in C_c(\mathbb{R})$.
2. $\int_{\mathbb{R}^d} f(x)\,d\nu(x) = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$ *for all* $f \in C_c(\mathbb{R}^d)$.
3. $\hat{\nu} = \hat{\mu}$.

Item 3. asserts that the Fourier transform is injective.

Proof. Let $\mathbb{H}$ be the collection of bounded complex measurable functions from $\mathbb{R}^d$ to $\mathbb{C}$ such that

$$\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f d\nu. \tag{8.5}$$

It is easily seen that $\mathbb{H}$ is a linear space closed under complex conjugation and bounded convergence (by the DCT). Since $\mathbb{H}$ contains one of the multiplicative systems appearing in Corollary 8.8, it contains all bounded Borel measurable functions form $\mathbb{R}^d \to \mathbb{C}$. Thus we may take $f = 1_A$ with $A \in \mathcal{B}_{\mathbb{R}^d}$ in Eq. (8.5) to learn, $\mu(A) = \nu(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$. ∎

In many cases we can replace the condition in item 3. of Corollary 8.11 by;

$$\int_{\mathbb{R}^d} e^{\lambda\cdot x} d\mu(x) = \int_{\mathbb{R}^d} e^{\lambda\cdot x} d\nu(x) \text{ for all } \lambda \in U, \tag{8.6}$$

where U is a neighborhood of $0 \in \mathbb{R}^d$. In order to do this, one must assume at least assume that the integrals involved are finite for all $\lambda \in U$. The idea is to show that Condition 8.6 implies $\hat{\nu} = \hat{\mu}$. You are asked to carry out this argument in Exercise 8.2 making use of the following lemma.

Lemma 8.12 (Analytic Continuation). *Let $\varepsilon > 0$ and $S_\varepsilon := \{x + iy \in \mathbb{C} : |x| < \varepsilon\}$ be an ε strip in $\mathbb{C}$ about the imaginary axis. Suppose that $h : S_\varepsilon \to \mathbb{C}$ is a function such that for each $b \in \mathbb{R}$, there exists $\{c_n(b)\}_{n=0}^\infty \subset \mathbb{C}$ such that*

$$h(z + ib) = \sum_{n=0}^\infty c_n(b) z^n \text{ for all } |z| < \varepsilon. \tag{8.7}$$

If $c_n(0) = 0$ for all $n \in \mathbb{N}_0$, then $h \equiv 0$.

Proof. It suffices to prove the following assertion; if for some $b \in \mathbb{R}$ we know that $c_n(b) = 0$ for all n, then $c_n(y) = 0$ for all n and $y \in (b - \varepsilon, b + \varepsilon)$. We now prove this assertion.

Let us assume that $b \in \mathbb{R}$ and $c_n(b) = 0$ for all $n \in \mathbb{N}_0$. It then follows from Eq. (8.7) that $h(z + ib) = 0$ for all $|z| < \varepsilon$. Thus if $|y - b| < \varepsilon$, we may conclude that $h(x + iy) = 0$ for x in a (possibly very small) neighborhood $(-\delta, \delta)$ of 0. Since

$$\sum_{n=0}^\infty c_n(y) x^n = h(x + iy) = 0 \text{ for all } |x| < \delta,$$

it follows that

$$0 = \frac{1}{n!}\frac{d^n}{dx^n} h(x + iy)|_{x=0} = c_n(y)$$

and the proof is complete. ∎

8.2 Exercises

Exercise 8.2. Suppose $\varepsilon > 0$ and X and Y are two random variables such that $\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{tY}\right] < \infty$ for all $|t| \le \varepsilon$. Show;

1. $\mathbb{E}\left[e^{\varepsilon|X|}\right]$ and $\mathbb{E}\left[e^{\varepsilon|Y|}\right]$ are finite.
2. $\mathbb{E}\left[e^{itX}\right] = \mathbb{E}\left[e^{itY}\right]$ for all $t \in \mathbb{R}$. **Hint:** Consider $h(z) := \mathbb{E}\left[e^{zX}\right] - \mathbb{E}\left[e^{zY}\right]$ for $z \in S_\varepsilon$. Now show for $|z| \le \varepsilon$ and $b \in \mathbb{R}$, that

$$h(z + ib) = \mathbb{E}\left[e^{ibX}e^{zX}\right] - \mathbb{E}\left[e^{ibY}e^{zY}\right] = \sum_{n=0}^\infty c_n(b) z^n \tag{8.8}$$

where

$$c_n(b) := \frac{1}{n!}\left(\mathbb{E}\left[e^{ibX}X^n\right] - \mathbb{E}\left[e^{ibY}Y^n\right]\right). \tag{8.9}$$

3. Conclude from item 2. that $X \stackrel{d}{=} Y$, i.e. that $\mathrm{Law}_P(X) = \mathrm{Law}_P(Y)$.

Exercise 8.3. Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ be a pair of random variables such that

$$\mathbb{E}[f(X) g(Y)] = \mathbb{E}[f(X) g(X)]$$

for every pair of bounded measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$. Show $P(X = Y) = 1$. **Hint:** Let $\mathbb{H}$ denote the bounded Borel measurable functions, $h : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathbb{E}[h(X, Y)] = \mathbb{E}[h(X, X)].$$

Use Theorem 8.2 to show $\mathbb{H}$ is the vector space of all bounded Borel measurable functions. Then take $h(x, y) = 1_{\{x=y\}}$.

Exercise 8.4 (Density of $\mathcal{A}$ – simple functions). Let $(\Omega, \mathcal{B}, P)$ be a probability space and assume that $\mathcal{A}$ is a sub-algebra of $\mathcal{B}$ such that $\mathcal{B} = \sigma(\mathcal{A})$. Let $\mathbb{H}$ denote the bounded measurable functions $f : \Omega \to \mathbb{R}$ such that for every $\varepsilon > 0$ there exists an an $\mathcal{A}$ – simple function, $\varphi : \Omega \to \mathbb{R}$ such that $\mathbb{E}|f - \varphi| < \varepsilon$. Show $\mathbb{H}$ consists of all bounded measurable functions, $f : \Omega \to \mathbb{R}$. **Hint:** let $\mathbb{M}$ denote the collection of $\mathcal{A}$ – simple functions.

Corollary 8.13. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $\{X_n\}_{n=1}^\infty$ is a collection of random variables on Ω, and $\mathcal{B}_\infty := \sigma(X_1, X_2, X_3, \dots)$. Then for all $\varepsilon > 0$ and all bounded $\mathcal{B}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, there exists an $n \in \mathbb{N}$ and a bounded $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}|f - G(X_1, \dots, X_n)| < \varepsilon$. Moreover we may assume that $\sup_{x\in\mathbb{R}^n} |G(x)| \le M := \sup_{\omega\in\Omega} |f(\omega)|$.*

Proof. Apply Exercise 8.4 with $\mathcal{A} := \cup_{n=1}^{\infty} \sigma(X_1, \ldots, X_n)$ in order to find an $\mathcal{A}$ – measurable simple function, φ, such that $\mathbb{E}|f - \varphi| < \varepsilon$. By the definition of $\mathcal{A}$ we know that φ is $\sigma(X_1, \ldots, X_n)$ – measurable for some $n \in \mathbb{N}$. It now follows by the factorization Lemma 6.40 that $\varphi = G(X_1, \ldots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$. If necessary, replace G by $[G \wedge M] \vee (-M)$ in order to insure $\sup_{x \in \mathbb{R}^n} |G(x)| \le M$. ■

Exercise 8.5 (Density of $\mathcal{A}$ in $\mathcal{B} = \sigma(\mathcal{A})$). Keeping the same notation as in Exercise 8.4 but now take $f = 1_B$ for some $B \in \mathcal{B}$ and given $\varepsilon > 0$, write $\varphi = \sum_{i=0}^{n} \lambda_i 1_{A_i}$ where $\lambda_0 = 0$, $\{\lambda_i\}_{i=1}^{n}$ is an enumeration of $\varphi(\Omega) \setminus \{0\}$, and $A_i := \{\varphi = \lambda_i\}$. Show; 1.

$$\mathbb{E}|1_B - \varphi| = P(A_0 \cap B) + \sum_{i=1}^{n} [|1 - \lambda_i| P(B \cap A_i) + |\lambda_i| P(A_i \setminus B)] \tag{8.10}$$

$$\ge P(A_0 \cap B) + \sum_{i=1}^{n} \min\{P(B \cap A_i), P(A_i \setminus B)\}. \tag{8.11}$$

2. Now let $\psi = \sum_{i=0}^{n} \alpha_i 1_{A_i}$ with

$$\alpha_i = \begin{cases} 1 \text{ if } P(A_i \setminus B) \le P(B \cap A_i) \\ 0 \text{ if } P(A_i \setminus B) > P(B \cap A_i) \end{cases}.$$

Then show that

$$\mathbb{E}|1_B - \psi| = P(A_0 \cap B) + \sum_{i=1}^{n} \min\{P(B \cap A_i), P(A_i \setminus B)\} \le \mathbb{E}|1_B - \varphi|.$$

Observe that $\psi = 1_D$ where $D = \cup_{i:\alpha_i=1} A_i \in \mathcal{A}$ and so you have shown; for every $\varepsilon > 0$ there exists a $D \in \mathcal{A}$ such that

$$P(B \Delta D) = \mathbb{E}|1_B - 1_D| < \varepsilon.$$

Exercise 8.6. Suppose that $\{(X_i, \mathcal{B}_i)\}_{i=1}^{n}$ are measurable spaces and for each i, $\mathbb{M}_i$ is a multiplicative system of real bounded measurable functions on X_i such that $\sigma(\mathbb{M}_i) = \mathcal{B}_i$ and there exist $\chi_n \in \mathbb{M}_i$ such that $\chi_n \to 1$ boundedly as $n \to \infty$. Given $f_i : X_i \to \mathbb{R}$ let $f_1 \otimes \cdots \otimes f_n : X_1 \times \cdots \times X_n \to \mathbb{R}$ be defined by

$$(f_1 \otimes \cdots \otimes f_n)(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

Show

$$\mathbb{M}_1 \otimes \cdots \otimes \mathbb{M}_n := \{f_1 \otimes \cdots \otimes f_n : f_i \in \mathbb{M}_i \text{ for } 1 \le i \le n\}$$

is a multiplicative system of bounded measurable functions on $(X := X_1 \times \cdots \times X_n, \mathcal{B} := \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n)$ such that $\sigma(\mathbb{M}_1 \otimes \cdots \otimes \mathbb{M}_n) = \mathcal{B}$.

Solution to Exercise (8.6). I will give the proof in case that $n = 2$. The generalization to higher n is straight forward.

Let $\pi_i : X \to X_i$ be the projection maps, $\pi_1(x_1, x_2) = x_1$ and $\pi_2(x_1, x_2) = x_2$. For $f_i \in \mathbb{M}_i$, $f_i \circ \pi_i : X \to \mathbb{R}$ is the composition of measurable functions and hence measurable. Therefore $f_1 \otimes f_2 = (f_1 \circ \pi_1) \cdot (f_2 \circ \pi_2)$ is a bounded $\mathcal{B}_1 \otimes \mathcal{B}_2$ – measurable function and therefore $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2) \subset \mathcal{B}_1 \otimes \mathcal{B}_2$. Since it is clear that $\mathbb{M}_1 \otimes \mathbb{M}_2$ is a multiplicative system, to finish the proof we must show $\mathcal{B}_1 \otimes \mathcal{B}_2 \subset \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$.

Proof. Let $g \in \mathbb{M}_2$ and let

$$\mathbb{H}_g := \{f \in (\mathcal{B}_1)_b : f \otimes g \text{ is } \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2) \text{ – measurable}\}.$$

You may easily check that $\mathbb{H}_g$ is closed under bounded convergence, $\mathbb{M}_1 \subset \mathbb{H}_g$, and $\mathbb{H}_g$ contains the constant functions. Since $\sigma(\mathbb{M}_1) = \mathcal{B}_1$ it now follows by Dynkin's multiplicative systems Theorem 8.2, that $\mathbb{H}_g = (\mathcal{B}_1)_b$. Thus we have shown that $(\mathcal{B}_1)_b \otimes \mathbb{M}_2$ consists of $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ – measurable functions. By the same logic we may now conclude that $(\mathcal{B}_1)_b \otimes (\mathcal{B}_2)_b$ consists of $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ – measurable functions as well. In particular this shows for any $A_i \in \mathcal{B}_i$ that $1_{A_1 \times A_2} = 1_{A_1} \otimes 1_{A_2}$ is $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ – measurable and therefore $A_1 \times A_2 \in \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ for all $A_i \in \mathcal{B}_i$. As the set $\{A_1 \times A_2 : A_i \in \mathcal{B}_i\}$ generate $\mathcal{B}_1 \otimes \mathcal{B}_2$ we may conclude that $\mathcal{B}_1 \otimes \mathcal{B}_2 \subset \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$. ■

8.3 A Strengthening of the Multiplicative System Theorem*

Notation 8.14 *We say that $\mathbb{H} \subset \ell^{\infty}(\Omega, \mathbb{R})$ is **closed under monotone convergence** if; for every sequence, $\{f_n\}_{n=1}^{\infty} \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $0 \le f_n(\omega) \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f_n(\omega)$ is increasing in n for all $\omega \in \Omega$, then $f := \lim_{n\to\infty} f_n \in \mathbb{H}$.*

Clearly if $\mathbb{H}$ is closed under bounded convergence then it is also closed under monotone convergence. I learned the proof of the converse from Pat Fitzsimmons but this result appears in Sharpe [48, p. 365].

Proposition 8.15. **Let Ω be a set. Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions from Ω to $\mathbb{R}$ which is closed under monotone convergence. Then $\mathbb{H}$ is closed under uniform convergence as well, i.e. $\{f_n\}_{n=1}^{\infty} \subset \mathbb{H}$ with $\sup_{n\in\mathbb{N}} \sup_{\omega\in\Omega} |f_n(\omega)| < \infty$ and $f_n \to f$, then $f \in \mathbb{H}$.*

Proof. Let us first assume that $\{f_n\}_{n=1}^{\infty} \subset \mathbb{H}$ such that f_n converges uniformly to a bounded function, $f : \Omega \to \mathbb{R}$. Let $\|f\|_{\infty} := \sup_{\omega\in\Omega} |f(\omega)|$. Let

$\varepsilon > 0$ be given. By passing to a subsequence if necessary, we may assume $\|f - f_n\|_\infty \le \varepsilon 2^{-(n+1)}$. Let

$$g_n := f_n - \delta_n + M$$

with δ_n and M constants to be determined shortly. We then have

$$g_{n+1} - g_n = f_{n+1} - f_n + \delta_n - \delta_{n+1} \ge -\varepsilon 2^{-(n+1)} + \delta_n - \delta_{n+1}.$$

Taking $\delta_n := \varepsilon 2^{-n}$, then $\delta_n - \delta_{n+1} = \varepsilon 2^{-n}(1 - 1/2) = \varepsilon 2^{-(n+1)}$ in which case $g_{n+1} - g_n \ge 0$ for all n. By choosing M sufficiently large, we will also have $g_n \ge 0$ for all n. Since $\mathbb{H}$ is a vector space containing the constant functions, $g_n \in \mathbb{H}$ and since $g_n \uparrow f + M$, it follows that $f = f + M - M \in \mathbb{H}$. So we have shown that $\mathbb{H}$ is closed under uniform convergence. ■

This proposition immediately leads to the following strengthening of Theorem 8.2.

Theorem 8.16. **Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions on Ω which contains the constant functions and is closed under monotone convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

Proof. Proposition 8.15 reduces this theorem to Theorem 8.2. ■

8.4 The Bounded Approximation Theorem*

This section should be skipped until needed (if ever!).

Notation 8.17 *Given a collection of bounded functions, $\mathbb{M}$, from a set, Ω, to $\mathbb{R}$, let $\mathbb{M}_\uparrow$ ($\mathbb{M}_\downarrow$) denote the the bounded monotone increasing (decreasing) limits of functions from $\mathbb{M}$. More explicitly a bounded function, $f : \Omega \to \mathbb{R}$ is in $\mathbb{M}_\uparrow$ respectively $\mathbb{M}_\downarrow$ iff there exists $f_n \in \mathbb{M}$ such that $f_n \uparrow f$ respectively $f_n \downarrow f$.*

Theorem 8.18 (Bounded Approximation Theorem*). *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that:*

1. *$\sigma(\mathbb{M}) = \mathcal{B}$,*
2. *$1 \in \mathbb{M}$, and*
3. *$|f| \in \mathbb{M}$ for all $f \in \mathbb{M}$.*

Then for every bounded $\sigma(\mathbb{M})$ measurable function, $g : \Omega \to \mathbb{R}$, and every $\varepsilon > 0$, there exists $f \in \mathbb{M}_\downarrow$ and $h \in \mathbb{M}_\uparrow$ such that $f \le g \le h$ and $\mu(h - f) < \varepsilon$.[1]

[1] Bruce: rework the Daniel integral section in the Analysis notes to stick to latticies of bounded functions.

Proof. Let us begin with a few simple observations.

1. $\mathbb{M}$ is a "lattice" – if $f, g \in \mathbb{M}$ then
$$f \vee g = \frac{1}{2}(f + g + |f - g|) \in \mathbb{M}$$
and
$$f \wedge g = \frac{1}{2}(f + g - |f - g|) \in \mathbb{M}.$$
2. If $f, g \in \mathbb{M}_\uparrow$ or $f, g \in \mathbb{M}_\downarrow$ then $f + g \in \mathbb{M}_\uparrow$ or $f + g \in \mathbb{M}_\downarrow$ respectively.
3. If $\lambda \ge 0$ and $f \in \mathbb{M}_\uparrow$ ($f \in \mathbb{M}_\downarrow$), then $\lambda f \in \mathbb{M}_\uparrow$ ($\lambda f \in \mathbb{M}_\downarrow$).
4. If $f \in \mathbb{M}_\uparrow$ then $-f \in \mathbb{M}_\downarrow$ and visa versa.
5. If $f_n \in \mathbb{M}_\uparrow$ and $f_n \uparrow f$ where $f : \Omega \to \mathbb{R}$ is a bounded function, then $f \in \mathbb{M}_\uparrow$. Indeed, by assumption there exists $f_{n,i} \in \mathbb{M}$ such that $f_{n,i} \uparrow f_n$ as $i \to \infty$. By observation (1), $g_n := \max\{f_{ij} : i, j \le n\} \in \mathbb{M}$. Moreover it is clear that $g_n \le \max\{f_k : k \le n\} = f_n \le f$ and hence $g_n \uparrow g := \lim_{n\to\infty} g_n \le f$. Since $f_{ij} \le g$ for all i, j, it follows that $f_n = \lim_{j\to\infty} f_{nj} \le g$ and consequently that $f = \lim_{n\to\infty} f_n \le g \le f$. So we have shown that $g_n \uparrow f \in \mathbb{M}_\uparrow$.

Now let $\mathbb{H}$ denote the collection of bounded measurable functions which satisfy the assertion of the theorem. Clearly, $\mathbb{M} \subset \mathbb{H}$ and in fact it is also easy to see that $\mathbb{M}_\uparrow$ and $\mathbb{M}_\downarrow$ are contained in $\mathbb{H}$ as well. For example, if $f \in \mathbb{M}_\uparrow$, by definition, there exists $f_n \in \mathbb{M} \subset \mathbb{M}_\downarrow$ such that $f_n \uparrow f$. Since $\mathbb{M}_\downarrow \ni f_n \le f \le f \in \mathbb{M}_\uparrow$ and $\mu(f - f_n) \to 0$ by the dominated convergence theorem, it follows that $f \in \mathbb{H}$. As similar argument shows $\mathbb{M}_\downarrow \subset \mathbb{H}$. We will now show $\mathbb{H}$ is a vector sub-space of the bounded $\mathcal{B} = \sigma(\mathbb{M})$ – measurable functions.

$\mathbb{H}$ is closed under addition. If $g_i \in \mathbb{H}$ for $i = 1, 2$, and $\varepsilon > 0$ is given, we may find $f_i \in \mathbb{M}_\downarrow$ and $h_i \in \mathbb{M}_\uparrow$ such that $f_i \le g_i \le h_i$ and $\mu(h_i - f_i) < \varepsilon/2$ for $i = 1, 2$. Since $h = h_1 + h_2 \in \mathbb{M}_\uparrow$, $f := f_1 + f_2 \in \mathbb{M}_\downarrow$, $f \le g_1 + g_2 \le h$, and

$$\mu(h - f) = \mu(h_1 - f_1) + \mu(h_2 - f_2) < \varepsilon,$$

it follows that $g_1 + g_2 \in \mathbb{H}$.

$\mathbb{H}$ is closed under scalar multiplication. If $g \in \mathbb{H}$ then $\lambda g \in \mathbb{H}$ for all $\lambda \in \mathbb{R}$. Indeed suppose that $\varepsilon > 0$ is given and $f \in \mathbb{M}_\downarrow$ and $h \in \mathbb{M}_\uparrow$ such that $f \le g \le h$ and $\mu(h - f) < \varepsilon$. Then for $\lambda \ge 0$, $\mathbb{M}_\downarrow \ni \lambda f \le \lambda g \le \lambda h \in \mathbb{M}_\uparrow$ and

$$\mu(\lambda h - \lambda f) = \lambda \mu(h - f) < \lambda \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, if follows that $\lambda g \in \mathbb{H}$ for $\lambda \ge 0$. Similarly, $\mathbb{M}_\downarrow \ni -h \le -g \le -f \in \mathbb{M}_\uparrow$ and

$$\mu(-f - (-h)) = \mu(h - f) < \varepsilon.$$

which shows $-g \in \mathbb{H}$ as well.

Because of Theorem 8.16, to complete this proof, it suffices to show $\mathbb{H}$ is closed under monotone convergence. So suppose that $g_n \in \mathbb{H}$ and $g_n \uparrow g$, where $g : \Omega \to \mathbb{R}$ is a bounded function. Since $\mathbb{H}$ is a vector space, it follows that $0 \leq \delta_n := g_{n+1} - g_n \in \mathbb{H}$ for all $n \in \mathbb{N}$. So if $\varepsilon > 0$ is given, we can find, $\mathbb{M}_\downarrow \ni u_n \leq \delta_n \leq v_n \in \mathbb{M}_\uparrow$ such that $\mu(v_n - u_n) \leq 2^{-n}\varepsilon$ for all n. By replacing u_n by $u_n \vee 0 \in \mathbb{M}_\downarrow$ (by observation 1.), we may further assume that $u_n \geq 0$. Let

$$v := \sum_{n=1}^{\infty} v_n =\uparrow \lim_{N\to\infty} \sum_{n=1}^{N} v_n \in \mathbb{M}_\uparrow \text{ (using observations 2. and 5.)}$$

and for $N \in \mathbb{N}$, let

$$u^N := \sum_{n=1}^{N} u_n \in \mathbb{M}_\downarrow \text{ (using observation 2).}$$

Then

$$\sum_{n=1}^{\infty} \delta_n = \lim_{N\to\infty} \sum_{n=1}^{N} \delta_n = \lim_{N\to\infty} (g_{N+1} - g_1) = g - g_1$$

and $u^N \leq g - g_1 \leq v$. Moreover,

$$\begin{aligned}\mu\left(v - u^N\right) &= \sum_{n=1}^{N} \mu(v_n - u_n) + \sum_{n=N+1}^{\infty} \mu(v_n) \leq \sum_{n=1}^{N} \varepsilon 2^{-n} + \sum_{n=N+1}^{\infty} \mu(v_n) \\ &\leq \varepsilon + \sum_{n=N+1}^{\infty} \mu(v_n).\end{aligned}$$

However, since

$$\begin{aligned}\sum_{n=1}^{\infty} \mu(v_n) &\leq \sum_{n=1}^{\infty} \mu\left(\delta_n + \varepsilon 2^{-n}\right) = \sum_{n=1}^{\infty} \mu(\delta_n) + \varepsilon\mu(\Omega) \\ &= \sum_{n=1}^{\infty} \mu(g - g_1) + \varepsilon\mu(\Omega) < \infty,\end{aligned}$$

it follows that for $N \in \mathbb{N}$ sufficiently large that $\sum_{n=N+1}^{\infty} \mu(v_n) < \varepsilon$. Therefore, for this N, we have $\mu\left(v - u^N\right) < 2\varepsilon$ and since $\varepsilon > 0$ is arbitrary, if follows that $g - g_1 \in \mathbb{H}$. Since $g_1 \in \mathbb{H}$ and $\mathbb{H}$ is a vector space, we may conclude that $g = (g - g_1) + g_1 \in \mathbb{H}$. ■

9

Multiple and Iterated Integrals

9.1 Iterated Integrals

Notation 9.1 (Iterated Integrals) *If $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are two measure spaces and $f : X\times Y \to \mathbb{C}$ is a $\mathcal{M}\otimes\mathcal{N}$ – measurable function, the* ***iterated integrals*** *of f (when they make sense) are:*

$$\int_X d\mu(x)\int_Y d\nu(y) f(x,y) := \int_X \left[\int_Y f(x,y)d\nu(y)\right] d\mu(x)$$

and

$$\int_Y d\nu(y)\int_X d\mu(x) f(x,y) := \int_Y \left[\int_X f(x,y)d\mu(x)\right] d\nu(y).$$

Notation 9.2 *Suppose that $f : X \to \mathbb{C}$ and $g : Y \to \mathbb{C}$ are functions, let $f\otimes g$ denote the function on $X\times Y$ given by*

$$f\otimes g(x,y) = f(x)g(y).$$

Notice that if f, g are measurable, then $f\otimes g$ is $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{C}})$ – measurable. To prove this let $F(x,y) = f(x)$ and $G(x,y) = g(y)$ so that $f\otimes g = F\cdot G$ will be measurable provided that F and G are measurable. Now $F = f\circ\pi_1$ where $\pi_1 : X\times Y \to X$ is the projection map. This shows that F is the composition of measurable functions and hence measurable. Similarly one shows that G is measurable.

9.2 Tonelli's Theorem and Product Measure

Theorem 9.3. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ-finite measure spaces and f is a nonnegative $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable function, then for each $y\in Y$,*

$$x \to f(x,y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.1}$$

for each $x \in X$,

$$y \to f(x,y) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.2}$$

$$x \to \int_Y f(x,y)d\nu(y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.3}$$

$$y \to \int_X f(x,y)d\mu(x) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.4}$$

and

$$\int_X d\mu(x)\int_Y d\nu(y) f(x,y) = \int_Y d\nu(y)\int_X d\mu(x) f(x,y). \tag{9.5}$$

Proof. Suppose that $E = A\times B \in \mathcal{E} := \mathcal{M}\times\mathcal{N}$ and $f = 1_E$. Then

$$f(x,y) = 1_{A\times B}(x,y) = 1_A(x)1_B(y)$$

and one sees that Eqs. (9.1) and (9.2) hold. Moreover

$$\int_Y f(x,y)d\nu(y) = \int_Y 1_A(x)1_B(y)d\nu(y) = 1_A(x)\nu(B),$$

so that Eq. (9.3) holds and we have

$$\int_X d\mu(x)\int_Y d\nu(y) f(x,y) = \nu(B)\mu(A). \tag{9.6}$$

Similarly,

$$\int_X f(x,y)d\mu(x) = \mu(A)1_B(y) \text{ and}$$
$$\int_Y d\nu(y)\int_X d\mu(x) f(x,y) = \nu(B)\mu(A)$$

from which it follows that Eqs. (9.4) and (9.5) hold in this case as well.

For the moment let us now further assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ and let $\mathbb{H}$ be the collection of all bounded $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable functions on $X\times Y$ such that Eqs. (9.1) – (9.5) hold. Using the fact that measurable functions are closed under pointwise limits and the dominated convergence theorem (the dominating function always being a constant), one easily shows that $\mathbb{H}$ closed under bounded convergence. Since we have just verified that $1_E \in \mathbb{H}$ for all E in the π – class, $\mathcal{E}$, it follows by Corollary 8.3 that $\mathbb{H}$ is the space of all bounded $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable functions on $X\times Y$. Moreover, if $f:X\times Y\to[0,\infty]$ is a $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, let $f_M=M\wedge f$ so that $f_M\uparrow f$ as $M\to\infty$. Then Eqs. (9.1) – (9.5) hold with f replaced by f_M for all $M\in\mathbb{N}$. Repeated use of the monotone convergence theorem allows us to pass to the limit $M\to\infty$ in these equations to deduce the theorem in the case μ and ν are finite measures.

For the σ – finite case, choose $X_n\in\mathcal{M}$, $Y_n\in\mathcal{N}$ such that $X_n\uparrow X$, $Y_n\uparrow Y$, $\mu(X_n)<\infty$ and $\nu(Y_n)<\infty$ for all $m,n\in\mathbb{N}$. Then define $\mu_m(A)=\mu(X_m\cap A)$ and $\nu_n(B)=\nu(Y_n\cap B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ or equivalently $d\mu_m=1_{X_m}d\mu$ and $d\nu_n=1_{Y_n}d\nu$. By what we have just proved Eqs. (9.1) – (9.5) with μ replaced by μ_m and ν by ν_n for all $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable functions, $f:X\times Y\to[0,\infty]$. The validity of Eqs. (9.1) – (9.5) then follows by passing to the limits $m\to\infty$ and then $n\to\infty$ making use of the monotone convergence theorem in the following context. For all $u\in L^+(X,\mathcal{M})$,

$$\int_X u\,d\mu_m=\int_X u1_{X_m}d\mu\uparrow\int_X u\,d\mu \text{ as } m\to\infty,$$

and for all and $v\in L^+(Y,\mathcal{N})$,

$$\int_Y v\,d\mu_n=\int_Y v1_{Y_n}d\mu\uparrow\int_Y v\,d\mu \text{ as } n\to\infty.$$

■

Corollary 9.4. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ – finite measure spaces. Then there exists a unique measure π on $\mathcal{M}\otimes\mathcal{N}$ such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$. Moreover π is given by*

$$\pi(E)=\int_X d\mu(x)\int_Y d\nu(y)1_E(x,y)=\int_Y d\nu(y)\int_X d\mu(x)1_E(x,y) \tag{9.7}$$

for all $E\in\mathcal{M}\otimes\mathcal{N}$ and π is σ – finite.

Proof. Notice that any measure π such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ is necessarily σ – finite. Indeed, let $X_n\in\mathcal{M}$ and $Y_n\in\mathcal{N}$ be chosen so that $\mu(X_n)<\infty$, $\nu(Y_n)<\infty$, $X_n\uparrow X$ and $Y_n\uparrow Y$, then $X_n\times Y_n\in\mathcal{M}\otimes\mathcal{N}$, $X_n\times Y_n\uparrow X\times Y$ and $\pi(X_n\times Y_n)<\infty$ for all n. The uniqueness assertion is a consequence of the combination of Exercises 3.10 and 5.11 Proposition 3.25 with $\mathcal{E}=\mathcal{M}\times\mathcal{N}$. For the existence, it suffices to observe, using the monotone convergence theorem, that π defined in Eq. (9.7) is a measure on $\mathcal{M}\otimes\mathcal{N}$. Moreover this measure satisfies $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ from Eq. (9.6). ■

Notation 9.5 *The measure π is called the product measure of μ and ν and will be denoted by $\mu\otimes\nu$.*

Theorem 9.6 (Tonelli's Theorem). *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ – finite measure spaces and $\pi=\mu\otimes\nu$ is the product measure on $\mathcal{M}\otimes\mathcal{N}$. If $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$, then $f(\cdot,y)\in L^+(X,\mathcal{M})$ for all $y\in Y$, $f(x,\cdot)\in L^+(Y,\mathcal{N})$ for all $x\in X$,*

$$\int_Y f(\cdot,y)d\nu(y)\in L^+(X,\mathcal{M}),\ \int_X f(x,\cdot)d\mu(x)\in L^+(Y,\mathcal{N})$$

and

$$\int_{X\times Y} f\,d\pi=\int_X d\mu(x)\int_Y d\nu(y)f(x,y) \tag{9.8}$$

$$=\int_Y d\nu(y)\int_X d\mu(x)f(x,y). \tag{9.9}$$

Proof. By Theorem 9.3 and Corollary 9.4, the theorem holds when $f=1_E$ with $E\in\mathcal{M}\otimes\mathcal{N}$. Using the linearity of all of the statements, the theorem is also true for non-negative simple functions. Then using the monotone convergence theorem repeatedly along with the approximation Theorem 6.39, one deduces the theorem for general $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$. ■

Example 9.7. In this example we are going to show, $I:=\int_{\mathbb{R}}e^{-x^2/2}dm\,(x)=\sqrt{2\pi}$. To this end we observe, using Tonelli's theorem, that

$$I^2=\left[\int_{\mathbb{R}}e^{-x^2/2}dm\,(x)\right]^2=\int_{\mathbb{R}}e^{-y^2/2}\left[\int_{\mathbb{R}}e^{-x^2/2}dm\,(x)\right]dm\,(y)$$
$$=\int_{\mathbb{R}^2}e^{-\left(x^2+y^2\right)/2}dm^2\,(x,y)$$

where $m^2=m\otimes m$ is "Lebesgue measure" on $\left(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2}=\mathcal{B}_{\mathbb{R}}\otimes\mathcal{B}_{\mathbb{R}}\right)$. From the monotone convergence theorem,

$$I^2=\lim_{R\to\infty}\int_{D_R}e^{-\left(x^2+y^2\right)/2}dm^2\,(x,y)$$

where $D_R=\left\{(x,y):x^2+y^2<R^2\right\}$. Using the change of variables theorem described in Section 9.5 below,[1] we find

$$\int_{D_R}e^{-\left(x^2+y^2\right)/2}d\pi\,(x,y)=\int_{(0,R)\times(0,2\pi)}e^{-r^2/2}r\,dr\,d\theta$$
$$=2\pi\int_0^R e^{-r^2/2}r\,dr=2\pi\left(1-e^{-R^2/2}\right).$$

[1] Alternatively, you can easily show that the integral $\int_{D_R}f\,dm^2$ agrees with the multiple integral in undergraduate analysis when f is continuous. Then use the change of variables theorem from undergraduate analysis.

From this we learn that

$$I^2 = \lim_{R\to\infty} 2\pi \left(1 - e^{-R^2/2}\right) = 2\pi$$

as desired.

9.3 Fubini's Theorem

Notation 9.8 *If $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is any measurable function, let*

$$\bar{\int}_X f d\mu := \begin{cases} \int_X f d\mu & \text{if} \quad \int_X |f| \, d\mu < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 9.9 (Fubini's Theorem). *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ – finite measure spaces, $\pi = \mu \otimes \nu$ is the product measure on $\mathcal{M} \otimes \mathcal{N}$ and $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ – measurable function. Then the following three conditions are equivalent:*

$$\int_{X\times Y} |f| \, d\pi < \infty, \text{ i.e. } f \in L^1(\pi), \tag{9.10}$$

$$\int_X \left(\int_Y |f(x, y)| \, d\nu(y) \right) d\mu(x) < \infty \text{ and} \tag{9.11}$$

$$\int_Y \left(\int_X |f(x, y)| \, d\mu(x) \right) d\nu(y) < \infty. \tag{9.12}$$

If any one (and hence all) of these condition hold, then $f(x, \cdot) \in L^1(\nu)$ for μ-a.e. x, $f(\cdot, y) \in L^1(\mu)$ for ν-a.e. y, $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$, $\bar{\int}_X f(x, \cdot)d\mu(x) \in L^1(\nu)$ and Eqs. (9.8) and (9.9) are still valid after putting a bar over the integral symbols.

Proof. The equivalence of Eqs. (9.10) – (9.12) is a direct consequence of Tonelli's Theorem 9.6. Now suppose $f \in L^1(\pi)$ is a real valued function and let

$$E := \left\{ x \in X : \int_Y |f(x, y)| \, d\nu(y) = \infty \right\}. \tag{9.13}$$

Then by Tonelli's theorem, $x \to \int_Y |f(x, y)| \, d\nu(y)$ is measurable and hence $E \in \mathcal{M}$. Moreover Tonelli's theorem implies

$$\int_X \left[\int_Y |f(x, y)| \, d\nu(y) \right] d\mu(x) = \int_{X\times Y} |f| \, d\pi < \infty$$

which implies that $\mu(E) = 0$. Let $f_\pm$ be the positive and negative parts of f, then

$$\begin{aligned} \bar{\int}_Y f(x, y) \, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y) \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) [f_+(x, y) - f_-(x, y)] \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) f_+(x, y) \, d\nu(y) - \int_Y 1_{E^c}(x) f_-(x, y) \, d\nu(y). \end{aligned} \tag{9.14}$$

Noting that $1_{E^c}(x) f_\pm(x, y) = (1_{E^c} \otimes 1_Y \cdot f_\pm)(x, y)$ is a positive $\mathcal{M} \otimes \mathcal{N}$ – measurable function, it follows from another application of Tonelli's theorem that $x \to \bar{\int}_Y f(x, y) \, d\nu(y)$ is $\mathcal{M}$ – measurable, being the difference of two measurable functions. Moreover

$$\int_X \left| \bar{\int}_Y f(x, y) \, d\nu(y) \right| d\mu(x) \le \int_X \left[\int_Y |f(x, y)| \, d\nu(y) \right] d\mu(x) < \infty,$$

which shows $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$. Integrating Eq. (9.14) on x and using Tonelli's theorem repeatedly implies,

$$\begin{aligned} &\int_X \left[\bar{\int}_Y f(x, y) \, d\nu(y) \right] d\mu(x) \\ &= \int_X d\mu(x) \int_Y d\nu(y) \, 1_{E^c}(x) f_+(x, y) - \int_X d\mu(x) \int_Y d\nu(y) \, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x) \, 1_{E^c}(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x) \, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x) \, f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x) \, f_-(x, y) \\ &= \int_{X\times Y} f_+ d\pi - \int_{X\times Y} f_- d\pi = \int_{X\times Y} (f_+ - f_-) \, d\pi = \int_{X\times Y} f d\pi \end{aligned} \tag{9.15}$$

which proves Eq. (9.8) holds.

Now suppose that $f = u + iv$ is complex valued and again let E be as in Eq. (9.13). Just as above we still have $E \in \mathcal{M}$ and $\mu(E) = 0$ and

$$\begin{aligned} \bar{\int}_Y f(x, y) \, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y) \, d\nu(y) = \int_Y 1_{E^c}(x) [u(x, y) + iv(x, y)] \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) u(x, y) \, d\nu(y) + i \int_Y 1_{E^c}(x) v(x, y) \, d\nu(y). \end{aligned}$$

The last line is a measurable in x as we have just proved. Similarly one shows $\bar{\int}_Y f(\cdot,y)\,d\nu(y) \in L^1(\mu)$ and Eq. (9.8) still holds by a computation similar to that done in Eq. (9.15). The assertions pertaining to Eq. (9.9) may be proved in the same way. ∎

The previous theorems generalize to products of any finite number of σ – finite measure spaces.

Theorem 9.10. *Suppose $\{(X_i,\mathcal{M}_i,\mu_i)\}_{i=1}^n$ are σ – finite measure spaces and $X := X_1\times\cdots\times X_n$. Then there exists a unique measure (π) on $(X,\mathcal{M}_1\otimes\cdots\otimes\mathcal{M}_n)$ such that*

$$\pi(A_1\times\cdots\times A_n)=\mu_1(A_1)\dots\mu_n(A_n) \text{ for all } A_i\in\mathcal{M}_i. \tag{9.16}$$

(This measure and its completion will be denoted by $\mu_1\otimes\cdots\otimes\mu_n$.) If $f: X\to[0,\infty]$ is a $\mathcal{M}_1\otimes\cdots\otimes\mathcal{M}_n$ – measurable function then

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)})\dots\int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1,\dots,x_n) \tag{9.17}$$

where σ is any permutation of $\{1,2,\dots,n\}$. In particular $f\in L^1(\pi)$, iff

$$\int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)})\dots\int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ |f(x_1,\dots,x_n)|<\infty$$

for some (and hence all) permutations, σ. Furthermore, if $f\in L^1(\pi)$, then

$$\int_X f d\pi = \bar{\int}_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)})\dots\bar{\int}_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1,\dots,x_n) \tag{9.18}$$

for all permutations σ.

Proof. (* I would consider skipping this tedious proof.) The proof will be by induction on n with the case $n=2$ being covered in Theorems 9.6 and 9.9. So let $n\ge 3$ and assume the theorem is valid for $n-1$ factors or less. To simplify notation, for $1\le i\le n$, let $X^i=\prod_{j\ne i}X_j$, $\mathcal{M}^i := \otimes_{j\ne i}\mathcal{M}_i$, and $\mu^i := \otimes_{j\ne i}\mu_j$ be the product measure on $(X^i,\mathcal{M}^i)$ which is assumed to exist by the induction hypothesis. Also let $\mathcal{M}:=\mathcal{M}_1\otimes\cdots\otimes\mathcal{M}_n$ and for $x=(x_1,\dots,x_i,\dots,x_n)\in X$ let

$$x^i := (x_1,\dots,\hat{x}_i,\dots,x_n) := (x_1,\dots,x_{i-1},x_{i+1},\dots,x_n).$$

Here is an outline of the argument with some details being left to the reader.

1. If $f: X\to[0,\infty]$ is $\mathcal{M}$ -measurable, then
$$(x_1,\dots,\hat{x}_i,\dots,x_n)\to\int_{X_i} f(x_1,\dots,x_i,\dots,x_n)\,d\mu_i(x_i)$$
is $\mathcal{M}^i$ -measurable. Thus by the induction hypothesis, the right side of Eq. (9.17) is well defined.
2. If $\sigma\in S_n$ (the permutations of $\{1,2,\dots,n\}$) we may define a measure π on $(X,\mathcal{M})$ by;
$$\pi(A) := \int_{X_{\sigma 1}} d\mu_{\sigma 1}(x_{\sigma 1})\dots\int_{X_{\sigma n}} d\mu_{\sigma n}(x_{\sigma n})\,1_A(x_1,\dots,x_n). \tag{9.19}$$
It is easy to check that π is a measure which satisfies Eq. (9.16). Using the σ – finiteness assumptions and the fact that
$$\mathcal{P} := \{A_1\times\cdots\times A_n : A_i\in\mathcal{M}_i \text{ for } 1\le i\le n\}$$
is a π – system such that $\sigma(\mathcal{P})=\mathcal{M}$, it follows from Exercise 5.1 that there is only one such measure satisfying Eq. (9.16). Thus the formula for π in Eq. (9.19) is independent of $\sigma\in S_n$.
3. From Eq. (9.19) and the usual simple function approximation arguments we may conclude that Eq. (9.17) is valid.
Now suppose that $f\in L^1(X,\mathcal{M},\pi)$.
4. Using step 1 it is easy to check that
$$(x_1,\dots,\hat{x}_i,\dots,x_n)\to\bar{\int}_{X_i} f(x_1,\dots,x_i,\dots,x_n)\,d\mu_i(x_i)$$
is $\mathcal{M}^i$ – measurable. Indeed,
$$(x_1,\dots,\hat{x}_i,\dots,x_n)\to\int_{X_i} |f(x_1,\dots,x_i,\dots,x_n)|\,d\mu_i(x_i)$$
is $\mathcal{M}^i$ – measurable and therefore
$$E := \left\{(x_1,\dots,\hat{x}_i,\dots,x_n) : \int_{X_i}|f(x_1,\dots,x_i,\dots,x_n)|\,d\mu_i(x_i)<\infty\right\}\in\mathcal{M}^i.$$
Now let $u := \operatorname{Re} f$ and $v := \operatorname{Im} f$ and $u_\pm$ and $v_\pm$ are the positive and negative parts of u and v respectively, then
$$\begin{aligned}\bar{\int}_{X_i} f(x)\,d\mu_i(x_i) &= \int_{X_i} 1_E(x^i)\,f(x)\,d\mu_i(x_i)\\ &= \int_{X_i} 1_E(x^i)\,u(x)\,d\mu_i(x_i) + i\int_{X_i} 1_E(x^i)\,v(x)\,d\mu_i(x_i).\end{aligned}$$
Both of these later terms are $\mathcal{M}^i$ – measurable since, for example,
$$\int_{X_i} 1_E(x^i)\,u(x)\,d\mu_i(x_i) = \int_{X_i} 1_E(x^i)\,u_+(x)\,d\mu_i(x_i) - \int_{X_i} 1_E(x^i)\,u_-(x)\,d\mu_i(x_i)$$
which is $\mathcal{M}^i$ – measurable by step 1.

5. It now follows by induction that the right side of Eq. (9.18) is well defined.
6. Let $i := \sigma n$ and $T : X \to X_i \times X^i$ be the obvious identification;

$$T\left(x_i, (x_1, \dots, \hat{x}_i, \dots, x_n)\right) = (x_1, \dots, x_n).$$

One easily verifies T is $\mathcal{M}/\mathcal{M}_i \otimes \mathcal{M}^i$ – measurable (use Corollary 6.19 repeatedly) and that $\pi \circ T^{-1} = \mu_i \otimes \mu^i$ (see Exercise 5.1).
7. Let $f \in L^1(\pi)$. Combining step 6. with the abstract change of variables Theorem (Exercise 7.11) implies

$$\int_X f d\pi = \int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right). \tag{9.20}$$

By Theorem 9.9, we also have

$$\begin{aligned}\int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right) &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f \circ T(x_i, x^i) \\ &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f(x_1, \dots, x_n).\end{aligned} \tag{9.21}$$

Then by the induction hypothesis,

$$\int_{X^i} d\mu^i(x_i) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \dots, x_n) = \prod_{j \neq i} \bar{\int}_{X_j} d\mu_j(x_j) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \dots, x_n) \tag{9.22}$$

where the ordering the integrals in the last product are inconsequential. Combining Eqs. (9.20) – (9.22) completes the proof.

■

Convention: We are now going to drop the bar above the integral sign with the understanding that $\int_X f d\mu = 0$ whenever $f : X \to \mathbb{C}$ is a measurable function such that $\int_X |f|\, d\mu = \infty$. However if f is a non-negative function (i.e. $f : X \to [0, \infty]$) non-integrable function we will interpret $\int_X f d\mu$ to be infinite.

Example 9.11. In this example we will show

$$\lim_{M \to \infty} \int_0^M \frac{\sin x}{x} dx = \pi/2. \tag{9.23}$$

To see this write $\frac{1}{x} = \int_0^\infty e^{-tx} dt$ and use Fubini-Tonelli to conclude that

$$\begin{aligned}\int_0^M \frac{\sin x}{x} dx &= \int_0^M \left[\int_0^\infty e^{-tx} \sin x\ dt\right] dx \\ &= \int_0^\infty \left[\int_0^M e^{-tx} \sin x\ dx\right] dt \\ &= \int_0^\infty \frac{1}{1+t^2}\left(1 - te^{-Mt} \sin M - e^{-Mt} \cos M\right) dt \\ &\to \int_0^\infty \frac{1}{1+t^2} dt = \frac{\pi}{2} \text{ as } M \to \infty,\end{aligned}$$

wherein we have used the dominated convergence theorem (for instance, take $g(t) := \frac{1}{1+t^2}\left(1 + te^{-t} + e^{-t}\right))$ to pass to the limit.

The next example is a refinement of this result.

Example 9.12. We have

$$\int_0^\infty \frac{\sin x}{x} e^{-\Lambda x} dx = \frac{1}{2}\pi - \arctan \Lambda \text{ for all } \Lambda > 0 \tag{9.24}$$

and for$\Lambda, M \in [0, \infty)$,

$$\left|\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx - \frac{1}{2}\pi + \arctan \Lambda\right| \le C \frac{e^{-M\Lambda}}{M} \tag{9.25}$$

where $C = \max_{x \ge 0} \frac{1+x}{1+x^2} = \frac{1}{2\sqrt{2}-2} \cong 1.2$. In particular Eq. (9.23) is valid.

To verify these assertions, first notice that by the fundamental theorem of calculus,

$$|\sin x| = \left|\int_0^x \cos y dy\right| \le \left|\int_0^x |\cos y|\, dy\right| \le \left|\int_0^x 1 dy\right| = |x|$$

so $\left|\frac{\sin x}{x}\right| \le 1$ for all $x \neq 0$. Making use of the identity

$$\int_0^\infty e^{-tx} dt = 1/x$$

and Fubini's theorem,

$$\begin{aligned}
\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx &= \int_0^M dx \sin x\, e^{-\Lambda x} \int_0^\infty e^{-tx} dt \\
&= \int_0^\infty dt \int_0^M dx \sin x\, e^{-(\Lambda+t)x} \\
&= \int_0^\infty \frac{1-(\cos M + (\Lambda+t)\sin M)\, e^{-M(\Lambda+t)}}{(\Lambda+t)^2+1} dt \\
&= \int_0^\infty \frac{1}{(\Lambda+t)^2+1} dt - \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt \\
&= \frac{1}{2}\pi - \arctan \Lambda - \varepsilon(M,\Lambda) \qquad (9.26)
\end{aligned}$$

where

$$\varepsilon(M,\Lambda) = \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt.$$

Since

$$\left| \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} \right| \le \frac{1+(\Lambda+t)}{(\Lambda+t)^2+1} \le C,$$

$$|\varepsilon(M,\Lambda)| \le \int_0^\infty e^{-M(\Lambda+t)} dt = C\frac{e^{-M\Lambda}}{M}.$$

This estimate along with Eq. (9.26) proves Eq. (9.25) from which Eq. (9.23) follows by taking $\Lambda \to \infty$ and Eq. (9.24) follows (using the dominated convergence theorem again) by letting $M \to \infty$.

Lemma 9.13. *Suppose that X is a random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ is a C^1 – functions such that $\lim_{x\to-\infty} \varphi(x) = 0$ and either $\varphi'(x) \ge 0$ for all x or $\int_{\mathbb{R}} |\varphi'(x)|\, dx < \infty$. Then*

$$\mathbb{E}[\varphi(X)] = \int_{-\infty}^\infty \varphi'(y)\, P(X > y)\, dy.$$

Similarly if $X \ge 0$ and $\varphi : [0,\infty) \to \mathbb{R}$ is a C^1 – function such that $\varphi(0) = 0$ and either $\varphi' \ge 0$ or $\int_0^\infty |\varphi'(x)|\, dx < \infty$, then

$$\mathbb{E}[\varphi(X)] = \int_0^\infty \varphi'(y)\, P(X > y)\, dy.$$

Proof. By the fundamental theorem of calculus for all $M < \infty$ and $x \in \mathbb{R}$,

$$\varphi(x) = \varphi(-M) + \int_{-M}^x \varphi'(y)\, dy. \qquad (9.27)$$

Under the stated assumptions on φ, we may use either the monotone or the dominated convergence theorem to let $M \to \infty$ in Eq. (9.27) to find,

$$\varphi(x) = \int_{-\infty}^x \varphi'(y)\, dy = \int_{\mathbb{R}} 1_{y<x} \varphi'(y)\, dy \text{ for all } x \in \mathbb{R}.$$

Therefore,

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\left[\int_{\mathbb{R}} 1_{y<X}\varphi'(y)\, dy\right] = \int_{\mathbb{R}} \mathbb{E}[1_{y<X}]\, \varphi'(y)\, dy = \int_{-\infty}^\infty \varphi'(y)\, P(X>y)\, dy,$$

where we applied Fubini's theorem for the second equality. The proof of the second assertion is similar and will be left to the reader. ■

Example 9.14. Here are a couple of examples involving Lemma 9.13.

1. Suppose X is a random variable, then

$$\mathbb{E}\left[e^X\right] = \int_{-\infty}^\infty P(X>y)\, e^y dy = \int_0^\infty P(X > \ln u)\, du, \qquad (9.28)$$

   where we made the change of variables, $u = e^y$, to get the second equality.
2. If $X \ge 0$ and $p \ge 1$, then

$$\mathbb{E}X^p = p\int_0^\infty y^{p-1} P(X>y)\, dy. \qquad (9.29)$$

9.4 Fubini's Theorem and Completions*

Notation 9.15 *Given $E \subset X \times Y$ and $x \in X$, let*

$${}_xE := \{y \in Y : (x,y) \in E\}.$$

Similarly if $y \in Y$ is given let

$$E_y := \{x \in X : (x,y) \in E\}.$$

If $f : X \times Y \to \mathbb{C}$ is a function let $f_x = f(x,\cdot)$ and $f^y := f(\cdot,y)$ so that $f_x : Y \to \mathbb{C}$ and $f^y : X \to \mathbb{C}$.

Theorem 9.16. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are **complete** σ – finite measure spaces. Let $(X\times Y,\mathcal{L},\lambda)$ be the completion of $(X\times Y,\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$. If f is $\mathcal{L}$ – measurable and (a) $f \ge 0$ or (b) $f \in L^1(\lambda)$ then f_x is $\mathcal{N}$ – measurable for μ a.e. x and f^y is $\mathcal{M}$ – measurable for ν a.e. y and in case (b) $f_x \in L^1(\nu)$ and $f^y \in L^1(\mu)$ for μ a.e. x and ν a.e. y respectively. Moreover,*

$$\left(x \to \int_Y f_x d\nu\right) \in L^1(\mu) \ \ and \ \left(y \to \int_X f^y d\mu\right) \in L^1(\nu)$$

and

$$\int_{X\times Y} f d\lambda = \int_Y d\nu \int_X d\mu\, f = \int_X d\mu \int_Y d\nu\, f.$$

Proof. If $E \in \mathcal{M}\otimes\mathcal{N}$ is a $\mu\otimes\nu$ null set (i.e. $(\mu\otimes\nu)(E)=0$), then

$$0 = (\mu\otimes\nu)(E) = \int\limits_X \nu({}_xE)d\mu(x) = \int\limits_X \mu(E_y)d\nu(y).$$

This shows that

$$\mu(\{x : \nu({}_xE) \neq 0\}) = 0 \text{ and } \nu(\{y : \mu(E_y) \neq 0\}) = 0,$$

i.e. $\nu({}_xE)=0$ for μ a.e. x and $\mu(E_y)=0$ for ν a.e. y. If h is $\mathcal{L}$ measurable and $h=0$ for λ – a.e., then there exists $E\in\mathcal{M}\otimes\mathcal{N}$ such that $\{(x,y) : h(x,y)\neq 0\}\subset E$ and $(\mu\otimes\nu)(E)=0$. Therefore $|h(x,y)|\leq 1_E(x,y)$ and $(\mu\otimes\nu)(E)=0$. Since

$$\{h_x \neq 0\} = \{y\in Y : h(x,y)\neq 0\} \subset {}_xE \text{ and}$$
$$\{h_y \neq 0\} = \{x\in X : h(x,y)\neq 0\} \subset E_y$$

we learn that for μ a.e. x and ν a.e. y that $\{h_x\neq 0\}\in\mathcal{M}$, $\{h_y\neq 0\}\in\mathcal{N}$, $\nu(\{h_x\neq 0\})=0$ and a.e. and $\mu(\{h_y\neq 0\})=0$. This implies $\int_Y h(x,y)d\nu(y)$ exists and equals 0 for μ a.e. x and similarly that $\int_X h(x,y)d\mu(x)$ exists and equals 0 for ν a.e. y. Therefore

$$0 = \int_{X\times Y} h d\lambda = \int_Y \left(\int_X h d\mu\right) d\nu = \int_X \left(\int_Y h d\nu\right) d\mu.$$

For general $f\in L^1(\lambda)$, we may choose $g\in L^1(\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$ such that $f(x,y)=g(x,y)$ for $\lambda-$ a.e. (x,y). Define $h:=f-g$. Then $h=0$, $\lambda-$ a.e. Hence by what we have just proved and Theorem 9.6 $f=g+h$ has the following properties:

1. For μ a.e. x, $y\to f(x,y)=g(x,y)+h(x,y)$ is in $L^1(\nu)$ and
$$\int_Y f(x,y)d\nu(y) = \int_Y g(x,y)d\nu(y).$$
2. For ν a.e. y, $x\to f(x,y)=g(x,y)+h(x,y)$ is in $L^1(\mu)$ and
$$\int_X f(x,y)d\mu(x) = \int_X g(x,y)d\mu(x).$$

From these assertions and Theorem 9.6, it follows that

$$\begin{aligned}\int_X d\mu(x)\int_Y d\nu(y) f(x,y) &= \int_X d\mu(x)\int_Y d\nu(y) g(x,y)\\ &= \int_Y d\nu(y)\int_Y d\nu(x) g(x,y)\\ &= \int_{X\times Y} g(x,y)d(\mu\otimes\nu)(x,y)\\ &= \int_{X\times Y} f(x,y)d\lambda(x,y).\end{aligned}$$

Similarly it is shown that

$$\int_Y d\nu(y)\int_X d\mu(x) f(x,y) = \int_{X\times Y} f(x,y)d\lambda(x,y).$$

■

9.5 Lebesgue Measure on $\mathbb{R}^d$ and the Change of Variables Theorem

Notation 9.17 *Let*

$$m^d := \overbrace{m\otimes\cdots\otimes m}^{d\ times} \ on\ \mathcal{B}_{\mathbb{R}^d} = \overbrace{\mathcal{B}_{\mathbb{R}}\otimes\cdots\otimes\mathcal{B}_{\mathbb{R}}}^{d\ times}$$

be the d – fold product of Lebesgue measure m on $\mathcal{B}_{\mathbb{R}}$. We will also use m^d to denote its completion and let $\mathcal{L}_d$ be the completion of $\mathcal{B}_{\mathbb{R}^d}$ relative to m^d. A subset $A\in\mathcal{L}_d$ is called a Lebesgue measurable set and m^d is called d – dimensional Lebesgue measure, or just Lebesgue measure for short.

Definition 9.18. *A function $f:\mathbb{R}^d\to\mathbb{R}$ is* ***Lebesgue measurable*** *if $f^{-1}(\mathcal{B}_{\mathbb{R}})\subset\mathcal{L}_d$.*

Notation 9.19 *I will often be sloppy in the sequel and write m for m^d and dx for $dm(x)=dm^d(x)$, i.e.*

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d} f dm^d.$$

Hopefully the reader will understand the meaning from the context.

Theorem 9.20. *Lebesgue measure m^d is translation invariant. Moreover m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d)=1$.*

Proof. Let $A = J_1 \times \cdots \times J_d$ with $J_i \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}^d$. Then

$$x + A = (x_1 + J_1) \times (x_2 + J_2) \times \cdots \times (x_d + J_d)$$

and therefore by translation invariance of m on $\mathcal{B}_{\mathbb{R}}$ we find that

$$m^d(x + A) = m(x_1 + J_1) \dots m(x_d + J_d) = m(J_1) \dots m(J_d) = m^d(A)$$

and hence $m^d(x + A) = m^d(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$ since it holds for A in a multiplicative system which generates $\mathcal{B}_{\mathbb{R}^d}$. From this fact we see that the measure $m^d(x + \cdot)$ and $m^d(\cdot)$ have the same null sets. Using this it is easily seen that $m(x + A) = m(A)$ for all $A \in \mathcal{L}_d$. The proof of the second assertion is Exercise 9.13. ■

Exercise 9.1. In this problem you are asked to show there is no reasonable notion of Lebesgue measure on an infinite dimensional Hilbert space. To be more precise, suppose H is an infinite dimensional Hilbert space and m is a **countably additive** measure on $\mathcal{B}_H$ which is invariant under translations and satisfies, $m(B_0(\varepsilon)) > 0$ for all $\varepsilon > 0$. Show $m(V) = \infty$ for all non-empty open subsets $V \subset H$.

Theorem 9.21 (Change of Variables Theorem). *Let $\Omega \subset_o \mathbb{R}^d$ be an open set and $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ be a C^1 – diffeomorphism,[2] see Figure 9.1. Then for any Borel measurable function, $f : T(\Omega) \to [0, \infty]$,*

$$\int_\Omega f(T(x)) \, |\det T'(x)| dx = \int_{T(\Omega)} f(y) \, dy, \tag{9.30}$$

where $T'(x)$ is the linear transformation on $\mathbb{R}^d$ defined by $T'(x)v := \frac{d}{dt}|_0 T(x + tv)$. More explicitly, viewing vectors in $\mathbb{R}^d$ as columns, $T'(x)$ may be represented by the matrix

$$T'(x) = \begin{bmatrix} \partial_1 T_1(x) & \dots & \partial_d T_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 T_d(x) & \dots & \partial_d T_d(x) \end{bmatrix}, \tag{9.31}$$

i.e. the i - j – matrix entry of $T'(x)$ is given by $T'(x)_{ij} = \partial_i T_j(x)$ where $T(x) = (T_1(x), \dots, T_d(x))^{\mathrm{tr}}$ and $\partial_i = \partial/\partial x_i$.

Remark 9.22. Theorem 9.21 is best remembered as the statement: if we make the change of variables $y = T(x)$, then $dy = |\det T'(x)| dx$. As usual, you must also change the limits of integration appropriately, i.e. if x ranges through Ω then y must range through $T(\Omega)$.

[2] That is $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ is a continuously differentiable bijection and the inverse map $T^{-1} : T(\Omega) \to \Omega$ is also continuously differentiable.

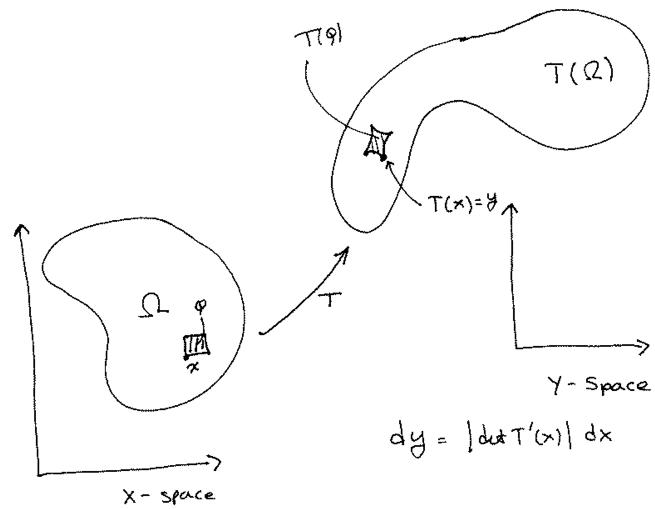


Fig. 9.1. The geometric setup of Theorem 9.21.

Note: you may skip the rest of this section!

Proof. The proof will be by induction on d. The case $d = 1$ was essentially done in Exercise 7.12. Nevertheless, for the sake of completeness let us give a proof here. Suppose $d = 1$, $a < \alpha < \beta < b$ such that $[a, b]$ is a compact subinterval of Ω. Then $|\det T'| = |T'|$ and

$$\int_{[a,b]} 1_{T((\alpha,\beta])}(T(x)) \, |T'(x)| \, dx = \int_{[a,b]} 1_{(\alpha,\beta]}(x) \, |T'(x)| \, dx = \int_\alpha^\beta |T'(x)| \, dx.$$

If $T'(x) > 0$ on $[a, b]$, then

$$\begin{aligned} \int_\alpha^\beta |T'(x)| \, dx &= \int_\alpha^\beta T'(x) \, dx = T(\beta) - T(\alpha) \\ &= m(T((\alpha, \beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y) \, dy \end{aligned}$$

while if $T'(x) < 0$ on $[a, b]$, then

$$\begin{aligned} \int_\alpha^\beta |T'(x)| \, dx &= -\int_\alpha^\beta T'(x) \, dx = T(\alpha) - T(\beta) \\ &= m(T((\alpha, \beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y) \, dy. \end{aligned}$$

Combining the previous three equations shows

$$\int_{[a,b]} f\left(T\left(x\right)\right)\left|T'\left(x\right)\right| dx = \int_{T([a,b])} f\left(y\right) dy \tag{9.32}$$

whenever f is of the form $f = 1_{T((\alpha,\beta])}$ with $a < \alpha < \beta < b$. An application of Dynkin's multiplicative system Theorem 8.16 then implies that Eq. (9.32) holds for every bounded measurable function $f : T\left([a,b]\right) \to \mathbb{R}$. (Observe that $\left|T'\left(x\right)\right|$ is continuous and hence bounded for x in the compact interval, $[a,b]$.) Recall that $\Omega = \sum_{n=1}^{N}\left(a_n,b_n\right)$ where $a_n, b_n \in \mathbb{R}\cup\{\pm\infty\}$ for $n = 1,2,\dots < N$ with $N = \infty$ possible. Hence if $f : T\left(\Omega\right) \to \mathbb{R}_+$ is a Borel measurable function and $a_n < \alpha_k < \beta_k < b_n$ with $\alpha_k \downarrow a_n$ and $\beta_k \uparrow b_n$, then by what we have already proved and the monotone convergence theorem

$$\begin{aligned}
\int_\Omega 1_{(a_n,b_n)} \cdot (f\circ T)\cdot|T'|dm &= \int_\Omega \left(1_{T((a_n,b_n))}\cdot f\right)\circ T\cdot|T'|dm \\
&= \lim_{k\to\infty}\int_\Omega \left(1_{T([\alpha_k,\beta_k])}\cdot f\right)\circ T\cdot|T'|\ dm \\
&= \lim_{k\to\infty}\int_{T(\Omega)} 1_{T([\alpha_k,\beta_k])}\cdot f\ dm \\
&= \int_{T(\Omega)} 1_{T((a_n,b_n))}\cdot f\ dm.
\end{aligned}$$

Summing this equality on n, then shows Eq. (9.30) holds.

To carry out the induction step, we now suppose $d > 1$ and suppose the theorem is valid with d being replaced by $d-1$. For notational compactness, let us write vectors in $\mathbb{R}^d$ as row vectors rather than column vectors. Nevertheless, the matrix associated to the differential, $T'\left(x\right)$, will always be taken to be given as in Eq. (9.31).

Case 1. Suppose $T\left(x\right)$ has the form

$$T\left(x\right) = \left(x_i, T_2\left(x\right),\dots,T_d\left(x\right)\right) \tag{9.33}$$

or

$$T\left(x\right) = \left(T_1\left(x\right),\dots,T_{d-1}\left(x\right),x_i\right) \tag{9.34}$$

for some $i \in \{1,\dots,d\}$. For definiteness we will assume T is as in Eq. (9.33), the case of T in Eq. (9.34) may be handled similarly. For $t \in \mathbb{R}$, let $i_t : \mathbb{R}^{d-1} \to \mathbb{R}^d$ be the inclusion map defined by

$$i_t\left(w\right) := w_t := \left(w_1,\dots,w_{i-1},t,w_{i+1},\dots,w_{d-1}\right),$$

Ω_t be the (possibly empty) open subset of $\mathbb{R}^{d-1}$ defined by

$$\Omega_t := \left\{w \in \mathbb{R}^{d-1} : \left(w_1,\dots,w_{i-1},t,w_{i+1},\dots,w_{d-1}\right) \in \Omega\right\}$$

and $T_t : \Omega_t \to \mathbb{R}^{d-1}$ be defined by

$$T_t\left(w\right) = \left(T_2\left(w_t\right),\dots,T_d\left(w_t\right)\right),$$

see Figure 9.2. Expanding $\det T'\left(w_t\right)$ along the first row of the matrix $T'\left(w_t\right)$

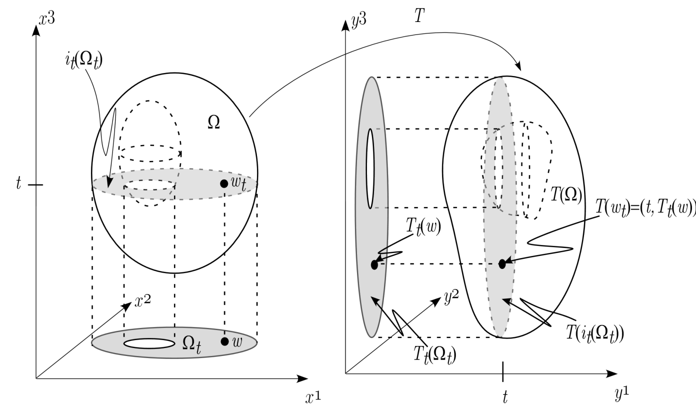


Fig. 9.2. In this picture $d = i = 3$ and Ω is an egg-shaped region with an egg-shaped hole. The picture indicates the geometry associated with the map T and slicing the set Ω along planes where $x_3 = t$.

shows

$$\left|\det T'\left(w_t\right)\right| = \left|\det T_t'\left(w\right)\right|.$$

Now by the Fubini-Tonelli Theorem and the induction hypothesis,

$$\begin{aligned}\int_\Omega f\circ T|\det T'|dm &= \int_{\mathbb{R}^d} 1_\Omega\cdot f\circ T|\det T'|dm\\ &= \int_{\mathbb{R}^d} 1_\Omega(w_t)(f\circ T)(w_t)|\det T'(w_t)|dwdt\\ &= \int_{\mathbb{R}}\left[\int_{\Omega_t}(f\circ T)(w_t)|\det T'(w_t)|dw\right]dt\\ &= \int_{\mathbb{R}}\left[\int_{\Omega_t} f(t,T_t(w))|\det T_t'(w)|dw\right]dt\\ &= \int_{\mathbb{R}}\left[\int_{T_t(\Omega_t)} f(t,z)\,dz\right]dt = \int_{\mathbb{R}}\left[\int_{\mathbb{R}^{d-1}} 1_{T(\Omega)}(t,z)\,f(t,z)\,dz\right]dt\\ &= \int_{T(\Omega)} f(y)\,dy\end{aligned}$$

wherein the last two equalities we have used Fubini-Tonelli along with the identity;

$$T(\Omega) = \sum_{t\in\mathbb{R}} T(i_t(\Omega)) = \sum_{t\in\mathbb{R}}\{(t,z): z\in T_t(\Omega_t)\}.$$

Case 2. (Eq. (9.30) is true locally.) Suppose that $T:\Omega\to\mathbb{R}^d$ is a general map as in the statement of the theorem and $x_0\in\Omega$ is an arbitrary point. We will now show there exists an open neighborhood $W\subset\Omega$ of x_0 such that

$$\int_W f\circ T|\det T'|dm = \int_{T(W)} fdm$$

holds for all Borel measurable function, $f:T(W)\to[0,\infty]$. Let M_i be the 1-i minor of $T'(x_0)$, i.e. the determinant of $T'(x_0)$ with the first row and i^{th} – column removed. Since

$$0\neq\det T'(x_0) = \sum_{i=1}^d(-1)^{i+1}\,\partial_i T_j(x_0)\cdot M_i,$$

there must be some i such that $M_i\neq 0$. Fix an i such that $M_i\neq 0$ and let,

$$S(x) := (x_i, T_2(x),\dots,T_d(x)). \tag{9.35}$$

Observe that $|\det S'(x_0)| = |M_i|\neq 0$. Hence by the inverse function Theorem, there exist an open neighborhood W of x_0 such that $W\subset_o\Omega$ and $S(W)\subset_o\mathbb{R}^d$ and $S:W\to S(W)$ is a C^1 – diffeomorphism. Let $R:S(W)\to T(W)\subset_o\mathbb{R}^d$ to be the C^1 – diffeomorphism defined by

$$R(z) := T\circ S^{-1}(z)\ \text{ for all } z\in S(W).$$

Because

$$(T_1(x),\dots,T_d(x)) = T(x) = R(S(x)) = R((x_i,T_2(x),\dots,T_d(x)))$$

for all $x\in W$, if

$$(z_1,z_2,\dots,z_d) = S(x) = (x_i,T_2(x),\dots,T_d(x))$$

then

$$R(z) = \left(T_1\left(S^{-1}(z)\right),z_2,\dots,z_d\right). \tag{9.36}$$

Observe that S is a map of the form in Eq. (9.33), R is a map of the form in Eq. (9.34), $T'(x) = R'(S(x))\,S'(x)$ (by the chain rule) and (by the multiplicative property of the determinant)

$$|\det T'(x)| = |\det R'(S(x))|\,|\det S'(x)|\ \ \forall\, x\in W.$$

So if $f:T(W)\to[0,\infty]$ is a Borel measurable function, two applications of the results in Case 1. shows,

$$\begin{aligned}\int_W f\circ T\cdot|\det T'|dm &= \int_W (f\circ R\cdot|\det R'|)\circ S\cdot|\det S'|\ dm\\ &= \int_{S(W)} f\circ R\cdot|\det R'|dm = \int_{R(S(W))} fdm\\ &= \int_{T(W)} fdm\end{aligned}$$

and Case 2. is proved.

Case 3. (General Case.) Let $f:\Omega\to[0,\infty]$ be a general non-negative Borel measurable function and let

$$K_n := \{x\in\Omega:\operatorname{dist}(x,\Omega^c)\ge 1/n \text{ and } |x|\le n\}.$$

Then each K_n is a compact subset of Ω and $K_n\uparrow\Omega$ as $n\to\infty$. Using the compactness of K_n and case 2, for each $n\in\mathbb{N}$, there is a finite open cover $\mathcal{W}_n$ of K_n such that $W\subset\Omega$ and Eq. (9.30) holds with Ω replaced by W for each $W\in\mathcal{W}_n$. Let $\{W_i\}_{i=1}^\infty$ be an enumeration of $\cup_{n=1}^\infty\mathcal{W}_n$ and set $\tilde W_1 = W_1$ and $\tilde W_i := W_i\setminus(W_1\cup\cdots\cup W_{i-1})$ for all $i\ge 2$. Then $\Omega = \sum_{i=1}^\infty\tilde W_i$ and by repeated use of case 2.,

$$\begin{aligned}\int\limits_{\Omega} f\circ T|\det T'|dm &= \sum_{i=1}^{\infty}\int\limits_{\Omega} 1_{\tilde{W}_i}\cdot (f\circ T)\cdot |\det T'|dm \\ &= \sum_{i=1}^{\infty}\int\limits_{W_i}\left[\left(1_{T(\tilde{W}_i)} f\right)\circ T\right]\cdot |\det T'|dm \\ &= \sum_{i=1}^{\infty}\int\limits_{T(W_i)} 1_{T(\tilde{W}_i)}\cdot f\ dm = \sum_{i=1}^{n}\int\limits_{T(\Omega)} 1_{T(\tilde{W}_i)}\cdot f\ dm \\ &= \int\limits_{T(\Omega)} f dm.\end{aligned}$$

■

Remark 9.23. When $d=1$, one often learns the change of variables formula as

$$\int_a^b f(T(x))\, T'(x)\, dx = \int_{T(a)}^{T(b)} f(y)\, dy \tag{9.37}$$

where $f:[a,b]\to\mathbb{R}$ is a continuous function and T is C^1 – function defined in a neighborhood of $[a,b]$. If $T'>0$ on (a,b) then $T((a,b))=(T(a),T(b))$ and Eq. (9.37) is implies Eq. (9.30) with $\Omega=(a,b)$. On the other hand if $T'<0$ on (a,b) then $T((a,b))=(T(b),T(a))$ and Eq. (9.37) is equivalent to

$$\int_{(a,b)} f(T(x))\,(-|T'(x)|)\, dx = -\int_{T(b)}^{T(a)} f(y)\, dy = -\int_{T((a,b))} f(y)\, dy$$

which is again implies Eq. (9.30). On the other hand Eq. (9.37) is more general than Eq. (9.30) since it does not require T to be injective. The standard proof of Eq. (9.37) is as follows. For $z\in T([a,b])$, let

$$F(z) := \int_{T(a)}^{z} f(y)\, dy.$$

Then by the chain rule and the fundamental theorem of calculus,

$$\begin{aligned}\int_a^b f(T(x))\, T'(x)\, dx &= \int_a^b F'(T(x))\, T'(x)\, dx = \int_a^b \frac{d}{dx}\left[F(T(x))\right] dx \\ &= F(T(x))\,|_a^b = \int_{T(a)}^{T(b)} f(y)\, dy.\end{aligned}$$

An application of Dynkin's multiplicative systems theorem now shows that Eq. (9.37) holds for all bounded measurable functions f on (a,b). Then by the usual truncation argument, it also holds for all positive measurable functions on (a,b).

Exercise 9.2. Continuing the setup in Theorem 9.21, show that $f\in L^1\left(T(\Omega), m^d\right)$ iff

$$\int\limits_{\Omega} |f\circ T|\, |\det T'|dm < \infty$$

and if $f\in L^1\left(T(\Omega), m^d\right)$, then Eq. (9.30) holds.

Example 9.24. Continuing the setup in Theorem 9.21, if $A\in\mathcal{B}_\Omega$, then

$$\begin{aligned}m(T(A)) &= \int_{\mathbb{R}^d} 1_{T(A)}(y)\, dy = \int_{\mathbb{R}^d} 1_{T(A)}(Tx)\, |\det T'(x)|\, dx \\ &= \int_{\mathbb{R}^d} 1_A(x)\, |\det T'(x)|\, dx\end{aligned}$$

wherein the second equality we have made the change of variables, $y=T(x)$. Hence we have shown

$$d(m\circ T) = |\det T'(\cdot)|\, dm.$$

Taking $T\in GL(d,\mathbb{R}) = GL(\mathbb{R}^d)$ – the space of $d\times d$ invertible matrices in the previous example implies $m\circ T = |\det T|\, m$, i.e.

$$m(T(A)) = |\det T|\, m(A) \text{ for all } A\in\mathcal{B}_{\mathbb{R}^d}. \tag{9.38}$$

This equation also shows that $m\circ T$ and m have the same null sets and hence the equality in Eq. (9.38) is valid for any $A\in\mathcal{L}_d$. In particular we may conclude that m is invariant under those $T\in GL(d,\mathbb{R})$ with $|\det(T)|=1$. For example if T is a rotation (i.e. $T^{\mathrm{tr}}T=I$), then $\det T=\pm 1$ and hence m is invariant under all rotations. This is not obvious from the definition of m^d as a product measure!

Example 9.25. Suppose that $T(x)=x+b$ for some $b\in\mathbb{R}^d$. In this case $T'(x)=I$ and therefore it follows that

$$\int\limits_{\mathbb{R}^d} f(x+b)\, dx = \int\limits_{\mathbb{R}^d} f(y)\, dy$$

for all measurable $f:\mathbb{R}^d\to[0,\infty]$ or for any $f\in L^1(m)$. In particular Lebesgue measure is invariant under translations.

Example 9.26 (Polar Coordinates). Suppose $T:(0,\infty)\times(0,2\pi)\to\mathbb{R}^2$ is defined by

$$x = T(r,\theta) = (r\cos\theta, r\sin\theta),$$

i.e. we are making the change of variable,

$$x_1 = r\cos\theta \text{ and } x_2 = r\sin\theta \text{ for } 0 < r < \infty \text{ and } 0 < \theta < 2\pi.$$

In this case

$$T'(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and therefore

$$dx = |\det T'(r,\theta)|\, drd\theta = rdrd\theta.$$

Observing that

$$\mathbb{R}^2 \setminus T\left((0,\infty)\times(0,2\pi)\right) = \ell := \{(x,0) : x \geq 0\}$$

has m^2 – measure zero, it follows from the change of variables Theorem 9.21 that

$$\int_{\mathbb{R}^2} f(x)dx = \int_0^{2\pi} d\theta \int_0^\infty dr\; r\cdot f(r(\cos\theta,\sin\theta)) \tag{9.39}$$

for any Borel measurable function $f : \mathbb{R}^2 \to [0,\infty]$.

Example 9.27 (Holomorphic Change of Variables). Suppose that $f : \Omega \subset_o \mathbb{C} \cong \mathbb{R}^2 \to \mathbb{C}$ is an injective holomorphic function such that $f'(z) \neq 0$ for all $z \in \Omega$. We may express f as

$$f(x+iy) = U(x,y) + iV(x,y)$$

for all $z = x+iy \in \Omega$. Hence if we make the change of variables,

$$w = u+iv = f(x+iy) = U(x,y) + iV(x,y)$$

then

$$dudv = \left|\det \begin{bmatrix} U_x & U_y \\ V_x & V_y \end{bmatrix}\right| dxdy = |U_xV_y - U_yV_x|\, dxdy.$$

Recalling that U and V satisfy the Cauchy Riemann equations, $U_x = V_y$ and $U_y = -V_x$ with $f' = U_x + iV_x$, we learn

$$U_xV_y - U_yV_x = U_x^2 + V_x^2 = |f'|^2.$$

Therefore

$$dudv = |f'(x+iy)|^2\, dxdy.$$

Example 9.28. In this example we will evaluate the integral

$$I := \iint_\Omega \left(x^4 - y^4\right) dxdy$$

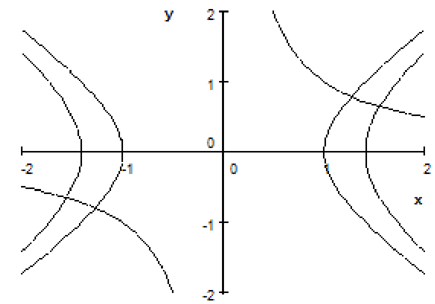


Fig. 9.3. The region Ω consists of the two curved rectangular regions shown.

where

$$\Omega = \left\{(x,y) : 1 < x^2 - y^2 < 2,\; 0 < xy < 1\right\},$$

see Figure 9.3. We are going to do this by making the change of variables,

$$(u,v) := T(x,y) = \left(x^2 - y^2, xy\right),$$

in which case

$$dudv = \left|\det \begin{bmatrix} 2x & -2y \\ y & x \end{bmatrix}\right| dxdy = 2\left(x^2+y^2\right) dxdy$$

Notice that

$$\left(x^4 - y^4\right) = \left(x^2 - y^2\right)\left(x^2+y^2\right) = u\left(x^2+y^2\right) = \frac{1}{2}ududv.$$

The function T is not injective on Ω but it is injective on each of its connected components. Let D be the connected component in the first quadrant so that $\Omega = -D \cup D$ and $T(\pm D) = (1,2)\times(0,1)$. The change of variables theorem then implies

$$I_\pm := \iint_{\pm D} \left(x^4 - y^4\right) dxdy = \frac{1}{2}\iint_{(1,2)\times(0,1)} ududv = \frac{1}{2}\frac{u^2}{2}\Big|_1^2 \cdot 1 = \frac{3}{4}$$

and therefore $I = I_+ + I_- = 2\cdot(3/4) = 3/2$.

Exercise 9.3 (Spherical Coordinates). Let $T : (0,\infty)\times(0,\pi)\times(0,2\pi) \to \mathbb{R}^3$ be defined by

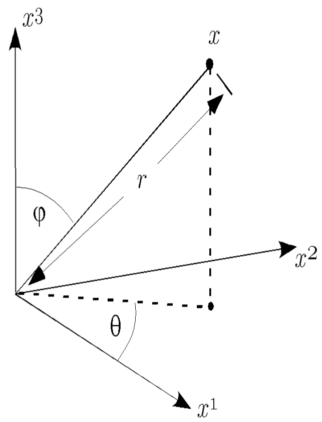


Fig. 9.4. The relation of x to (r, ϕ, θ) in spherical coordinates.

$$\begin{aligned} T(r, \varphi, \theta) &= (r \sin \varphi \cos \theta, r \sin \varphi \sin \theta, r \cos \varphi) \\ &= r (\sin \varphi \cos \theta, \sin \varphi \sin \theta, \cos \varphi), \end{aligned}$$

see Figure 9.4. By making the change of variables $x = T(r, \varphi, \theta)$, show

$$\int_{\mathbb{R}^3} f(x) dx = \int_0^\pi d\varphi \int_0^{2\pi} d\theta \int_0^\infty dr \; r^2 \sin \varphi \cdot f(T(r, \varphi, \theta))$$

for any Borel measurable function, $f : \mathbb{R}^3 \to [0, \infty]$.

Lemma 9.29. *Let $a > 0$ and*

$$I_d(a) := \int_{\mathbb{R}^d} e^{-a|x|^2} dm(x).$$

Then $I_d(a) = (\pi/a)^{d/2}$.

Proof. By Tonelli's theorem and induction,

$$\begin{aligned} I_d(a) &= \int_{\mathbb{R}^{d-1} \times \mathbb{R}} e^{-a|y|^2} e^{-at^2} m_{d-1}(dy) \; dt \\ &= I_{d-1}(a) I_1(a) = I_1^d(a). \qquad (9.40) \end{aligned}$$

So it suffices to compute:

$$I_2(a) = \int_{\mathbb{R}^2} e^{-a|x|^2} dm(x) = \int_{\mathbb{R}^2 \setminus \{0\}} e^{-a(x_1^2 + x_2^2)} dx_1 dx_2.$$

Using polar coordinates, see Eq. (9.39), we find,

$$\begin{aligned} I_2(a) &= \int_0^\infty dr \; r \int_0^{2\pi} d\theta \; e^{-ar^2} = 2\pi \int_0^\infty r e^{-ar^2} dr \\ &= 2\pi \lim_{M \to \infty} \int_0^M r e^{-ar^2} dr = 2\pi \lim_{M \to \infty} \frac{e^{-ar^2}}{-2a} \int_0^M = \frac{2\pi}{2a} = \pi/a. \end{aligned}$$

This shows that $I_2(a) = \pi/a$ and the result now follows from Eq. (9.40). ■

9.6 The Polar Decomposition of Lebesgue Measure*

Let

$$S^{d-1} = \{x \in \mathbb{R}^d : |x|^2 := \sum_{i=1}^d x_i^2 = 1\}$$

be the unit sphere in $\mathbb{R}^d$ equipped with its Borel σ − algebra, $\mathcal{B}_{S^{d-1}}$ and $\Phi : \mathbb{R}^d \setminus \{0\} \to (0, \infty) \times S^{d-1}$ be defined by $\Phi(x) := (|x|, |x|^{-1} x)$. The inverse map, $\Phi^{-1} : (0, \infty) \times S^{d-1} \to \mathbb{R}^d \setminus \{0\}$, is given by $\Phi^{-1}(r, \omega) = r\omega$. Since Φ and Φ^{-1} are continuous, they are both Borel measurable. For $E \in \mathcal{B}_{S^{d-1}}$ and $a > 0$, let

$$E_a := \{r\omega : r \in (0, a] \text{ and } \omega \in E\} = \Phi^{-1}((0, a] \times E) \in \mathcal{B}_{\mathbb{R}^d}.$$

Definition 9.30. *For $E \in \mathcal{B}_{S^{d-1}}$, let $\sigma(E) := d \cdot m(E_1)$. We call σ the surface measure on S^{d-1}.*

It is easy to check that σ is a measure. Indeed if $E \in \mathcal{B}_{S^{d-1}}$, then $E_1 = \Phi^{-1}((0, 1] \times E) \in \mathcal{B}_{\mathbb{R}^d}$ so that $m(E_1)$ is well defined. Moreover if $E = \sum_{i=1}^\infty E_i$, then $E_1 = \sum_{i=1}^\infty (E_i)_1$ and

$$\sigma(E) = d \cdot m(E_1) = \sum_{i=1}^\infty m((E_i)_1) = \sum_{i=1}^\infty \sigma(E_i).$$

The intuition behind this definition is as follows. If $E \subset S^{d-1}$ is a set and $\varepsilon > 0$ is a small number, then the volume of

$$(1, 1 + \varepsilon] \cdot E = \{r\omega : r \in (1, 1 + \varepsilon] \text{ and } \omega \in E\}$$

should be approximately given by $m((1, 1 + \varepsilon] \cdot E) \cong \sigma(E)\varepsilon$, see Figure 9.5 below. On the other hand

$$m((1, 1 + \varepsilon] E) = m(E_{1+\varepsilon} \setminus E_1) = \left\{(1 + \varepsilon)^d - 1\right\} m(E_1).$$

Therefore we expect the area of E should be given by

$$\sigma(E) = \lim_{\varepsilon \downarrow 0} \frac{\left\{(1 + \varepsilon)^d - 1\right\} m(E_1)}{\varepsilon} = d \cdot m(E_1).$$

The following theorem is motivated by Example 9.26 and Exercise 9.3.

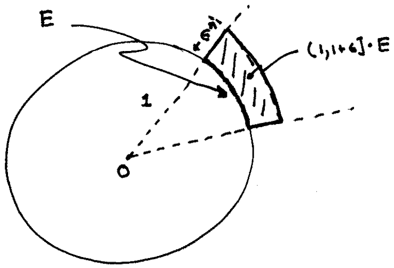


Fig. 9.5. Motivating the definition of surface measure for a sphere.

Theorem 9.31 (Polar Coordinates). *If $f : \mathbb{R}^d \to [0,\infty]$ is a $(\mathcal{B}_{R^d}, \mathcal{B})$–measurable function then*

$$\int_{\mathbb{R}^d} f(x)dm(x) = \int_{(0,\infty)\times S^{d-1}} f(r\omega)r^{d-1}\; drd\sigma(\omega). \tag{9.41}$$

In particular if $f : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable then

$$\int_{\mathbb{R}^d} f(|x|)dx = \int_0^\infty f(r)dV(r) \tag{9.42}$$

where $V(r) = m\left(B(0,r)\right) = r^d m\left(B(0,1)\right) = d^{-1}\sigma\left(S^{d-1}\right) r^d$.

Proof. By Exercise 7.11,

$$\int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d\setminus\{0\}} \left(f\circ\Phi^{-1}\right)\circ\Phi\; dm = \int_{(0,\infty)\times S^{d-1}} \left(f\circ\Phi^{-1}\right)\; d\left(\Phi_* m\right) \tag{9.43}$$

and therefore to prove Eq. (9.41) we must work out the measure $\Phi_* m$ on $\mathcal{B}_{(0,\infty)}\otimes \mathcal{B}_{S^{d-1}}$ defined by

$$\Phi_* m(A) := m\left(\Phi^{-1}(A)\right)\; \forall\; A \in \mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}. \tag{9.44}$$

If $A = (a,b]\times E$ with $0 < a < b$ and $E \in \mathcal{B}_{S^{d-1}}$, then

$$\Phi^{-1}(A) = \{r\omega : r \in (a,b] \text{ and } \omega \in E\} = bE_1 \setminus aE_1$$

wherein we have used $E_a = aE_1$ in the last equality. Therefore by the basic scaling properties of m and the fundamental theorem of calculus,

$$\begin{aligned}\left(\Phi_* m\right)\left((a,b]\times E\right) &= m\left(bE_1\setminus aE_1\right) = m(bE_1) - m(aE_1)\\ &= b^d m(E_1) - a^d m(E_1) = d\cdot m(E_1)\int_a^b r^{d-1}dr. \end{aligned}\tag{9.45}$$

Letting $d\rho(r) = r^{d-1}dr$, i.e.

$$\rho(J) = \int_J r^{d-1}dr\; \forall\; J\in\mathcal{B}_{(0,\infty)}, \tag{9.46}$$

Eq. (9.45) may be written as

$$\left(\Phi_* m\right)\left((a,b]\times E\right) = \rho((a,b])\cdot\sigma(E) = (\rho\otimes\sigma)\left((a,b]\times E\right). \tag{9.47}$$

Since

$$\mathcal{E} = \{(a,b]\times E : 0 < a < b \text{ and } E\in\mathcal{B}_{S^{d-1}}\},$$

is a π class (in fact it is an elementary class) such that $\sigma(\mathcal{E}) = \mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}$, it follows from the $\pi-\lambda$ Theorem and Eq. (9.47) that $\Phi_* m = \rho\otimes\sigma$. Using this result in Eq. (9.43) gives

$$\int_{\mathbb{R}^d} f dm = \int_{(0,\infty)\times S^{d-1}} \left(f\circ\Phi^{-1}\right)\; d\left(\rho\otimes\sigma\right)$$

which combined with Tonelli's Theorem 9.6 proves Eq. (9.43). ■

Corollary 9.32. *The surface area $\sigma(S^{d-1})$ of the unit sphere $S^{d-1}\subset\mathbb{R}^d$ is*

$$\sigma(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \tag{9.48}$$

where Γ is the gamma function is as in Example 7.47 and 7.50.

Proof. Using Theorem 9.31 we find

$$I_d(1) = \int_0^\infty dr\; r^{d-1}e^{-r^2}\int_{S^{d-1}} d\sigma = \sigma(S^{d-1})\int_0^\infty r^{d-1}e^{-r^2}dr.$$

We simplify this last integral by making the change of variables $u = r^2$ so that $r = u^{1/2}$ and $dr = \frac{1}{2}u^{-1/2}du$. The result is

$$\begin{aligned}\int_0^\infty r^{d-1}e^{-r^2}dr &= \int_0^\infty u^{\frac{d-1}{2}}e^{-u}\frac{1}{2}u^{-1/2}du\\ &= \frac{1}{2}\int_0^\infty u^{\frac{d}{2}-1}e^{-u}du = \frac{1}{2}\Gamma(d/2).\end{aligned}\tag{9.49}$$

Combing the the last two equations with Lemma 9.29 which states that $I_d(1) = \pi^{d/2}$, we conclude that

$$\pi^{d/2} = I_d(1) = \frac{1}{2}\sigma(S^{d-1})\Gamma(d/2)$$

which proves Eq. (9.48). ■

9.7 More Spherical Coordinates*

In this section we will define spherical coordinates in all dimensions. Along the way we will develop an explicit method for computing surface integrals on spheres. As usual when $n = 2$ define spherical coordinates $(r, \theta) \in (0, \infty) \times [0, 2\pi)$ so that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = T_2(\theta, r).$$

For $n = 3$ we let $x_3 = r\cos\varphi_1$ and then

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = T_2(\theta, r\sin\varphi_1),$$

as can be seen from Figure 9.6, so that

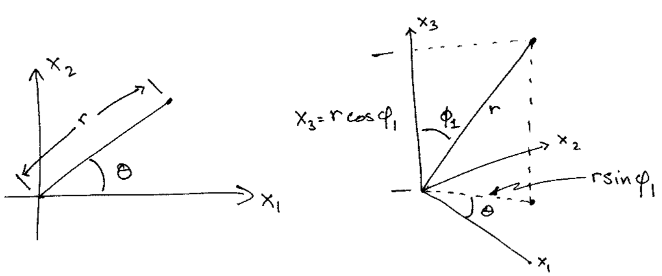


Fig. 9.6. Setting up polar coordinates in two and three dimensions.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} T_2(\theta, r\sin\varphi_1) \\ r\cos\varphi_1 \end{pmatrix} = \begin{pmatrix} r\sin\varphi_1\cos\theta \\ r\sin\varphi_1\sin\theta \\ r\cos\varphi_1 \end{pmatrix} =: T_3(\theta, \varphi_1, r,).$$

We continue to work inductively this way to define

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r\sin\varphi_{n-1},) \\ r\cos\varphi_{n-1} \end{pmatrix} = T_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r).$$

So for example,

$$\begin{aligned} x_1 &= r\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_2\cos\varphi_1 \\ x_4 &= r\cos\varphi_2 \end{aligned}$$

and more generally,

$$\begin{aligned} x_1 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\cos\varphi_1 \\ &\vdots \\ x_{n-2} &= r\sin\varphi_{n-2}\sin\varphi_{n-3}\cos\varphi_{n-4} \\ x_{n-1} &= r\sin\varphi_{n-2}\cos\varphi_{n-3} \\ x_n &= r\cos\varphi_{n-2}. \end{aligned} \tag{9.50}$$

By the change of variables formula,

$$\begin{aligned} &\int_{\mathbb{R}^n} f(x)dm(x) \\ &= \int_0^\infty dr \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} d\varphi_1\dots d\varphi_{n-2}d\theta \begin{bmatrix} \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) \\ \times f(T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r)) \end{bmatrix} \end{aligned} \tag{9.51}$$

where

$$\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) := \left|\det T_n'(\theta, \varphi_1, \dots, \varphi_{n-2}, r)\right|.$$

Proposition 9.33. *The Jacobian, Δ_n is given by*

$$\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) = r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1. \tag{9.52}$$

If f is a function on rS^{n-1} – the sphere of radius r centered at 0 inside of $\mathbb{R}^n$, then

$$\begin{aligned} &\int_{rS^{n-1}} f(x)d\sigma(x) = r^{n-1}\int_{S^{n-1}} f(r\omega)d\sigma(\omega) \\ &= \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} f(T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r))\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r)d\varphi_1\dots d\varphi_{n-2}d\theta \end{aligned} \tag{9.53}$$

Proof. We are going to compute Δ_n inductively. Letting $\rho := r\sin\varphi_{n-1}$ and writing $\frac{\partial T_n}{\partial\xi}$ for $\frac{\partial T_n}{\partial\xi}(\theta, \varphi_1, \dots, \varphi_{n-2}, \rho)$ we have

$$\begin{aligned} &\Delta_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r) \\ &\quad = \left|\begin{bmatrix} \frac{\partial T_n}{\partial\theta} & \frac{\partial T_n}{\partial\varphi_1} & \dots & \frac{\partial T_n}{\partial\varphi_{n-2}} & \frac{\partial T_n}{\partial\rho} r\cos\varphi_{n-1} & \frac{\partial T_n}{\partial\rho}\sin\varphi_{n-1} \\ 0 & 0 & \dots & 0 & -r\sin\varphi_{n-1} & \cos\varphi_{n-1} \end{bmatrix}\right| \\ &\quad = r\left(\cos^2\varphi_{n-1} + \sin^2\varphi_{n-1}\right)\Delta_n(, \theta, \varphi_1, \dots, \varphi_{n-2}, \rho) \\ &\quad = r\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r\sin\varphi_{n-1}), \end{aligned}$$

i.e.

$$\Delta_{n+1}(\theta,\varphi_1,\dots,\varphi_{n-2},\varphi_{n-1},r)=r\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r\sin\varphi_{n-1}). \tag{9.54}$$

To arrive at this result we have expanded the determinant along the bottom row. Staring with $\Delta_2(\theta,r)=r$ already derived in Example 9.26, Eq. (9.54) implies,

$$\begin{aligned}\Delta_3(\theta,\varphi_1,r)&=r\Delta_2(\theta,r\sin\varphi_1)=r^2\sin\varphi_1\\ \Delta_4(\theta,\varphi_1,\varphi_2,r)&=r\Delta_3(\theta,\varphi_1,r\sin\varphi_2)=r^3\sin^2\varphi_2\sin\varphi_1\\ &\vdots\\ \Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r)&=r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1\end{aligned}$$

which proves Eq. (9.52). Equation (9.53) now follows from Eqs. (9.41), (9.51) and (9.52). ■

As a simple application, Eq. (9.53) implies

$$\begin{aligned}\sigma(S^{n-1})&=\int_{0\le\varphi_i\le\pi,0\le\theta\le2\pi}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1d\varphi_1\dots d\varphi_{n-2}d\theta\\ &=2\pi\prod_{k=1}^{n-2}\gamma_k=\sigma(S^{n-2})\gamma_{n-2}\end{aligned} \tag{9.55}$$

where $\gamma_k:=\int_0^\pi\sin^k\varphi d\varphi$. If $k\ge1$, we have by integration by parts that,

$$\begin{aligned}\gamma_k&=\int_0^\pi\sin^k\varphi d\varphi=-\int_0^\pi\sin^{k-1}\varphi\ d\cos\varphi=2\delta_{k,1}+(k-1)\int_0^\pi\sin^{k-2}\varphi\ \cos^2\varphi d\varphi\\ &=2\delta_{k,1}+(k-1)\int_0^\pi\sin^{k-2}\varphi\ \left(1-\sin^2\varphi\right)d\varphi=2\delta_{k,1}+(k-1)\left[\gamma_{k-2}-\gamma_k\right]\end{aligned}$$

and hence γ_k satisfies $\gamma_0=\pi$, $\gamma_1=2$ and the recursion relation

$$\gamma_k=\frac{k-1}{k}\gamma_{k-2}\text{ for }k\ge2.$$

Hence we may conclude

$$\gamma_0=\pi,\ \gamma_1=2,\ \gamma_2=\frac{1}{2}\pi,\ \gamma_3=\frac{2}{3}2,\ \gamma_4=\frac{3}{4}\frac{1}{2}\pi,\ \gamma_5=\frac{4}{5}\frac{2}{3}2,\ \gamma_6=\frac{5}{6}\frac{3}{4}\frac{1}{2}\pi$$

and more generally by induction that

$$\gamma_{2k}=\pi\frac{(2k-1)!!}{(2k)!!}\text{ and }\gamma_{2k+1}=2\frac{(2k)!!}{(2k+1)!!}.$$

Indeed,

$$\gamma_{2(k+1)+1}=\frac{2k+2}{2k+3}\gamma_{2k+1}=\frac{2k+2}{2k+3}2\frac{(2k)!!}{(2k+1)!!}=2\frac{[2(k+1)]!!}{(2(k+1)+1)!!}$$

and

$$\gamma_{2(k+1)}=\frac{2k+1}{2k+1}\gamma_{2k}=\frac{2k+1}{2k+2}\pi\frac{(2k-1)!!}{(2k)!!}=\pi\frac{(2k+1)!!}{(2k+2)!!}.$$

The recursion relation in Eq. (9.55) may be written as

$$\sigma(S^n)=\sigma\left(S^{n-1}\right)\gamma_{n-1} \tag{9.56}$$

which combined with $\sigma\left(S^1\right)=2\pi$ implies

$$\begin{aligned}\sigma\left(S^1\right)&=2\pi,\\ \sigma(S^2)&=2\pi\cdot\gamma_1=2\pi\cdot2,\\ \sigma(S^3)&=2\pi\cdot2\cdot\gamma_2=2\pi\cdot2\cdot\frac{1}{2}\pi=\frac{2^2\pi^2}{2!!},\\ \sigma(S^4)&=\frac{2^2\pi^2}{2!!}\cdot\gamma_3=\frac{2^2\pi^2}{2!!}\cdot2\frac{2}{3}=\frac{2^3\pi^2}{3!!}\\ \sigma(S^5)&=2\pi\cdot2\cdot\frac{1}{2}\pi\cdot\frac{2}{3}2\cdot\frac{3}{4}\frac{1}{2}\pi=\frac{2^3\pi^3}{4!!},\\ \sigma(S^6)&=2\pi\cdot2\cdot\frac{1}{2}\pi\cdot\frac{2}{3}2\cdot\frac{3}{4}\frac{1}{2}\pi\cdot\frac{4}{5}\frac{2}{3}2=\frac{2^4\pi^3}{5!!}\end{aligned}$$

and more generally that

$$\sigma(S^{2n})=\frac{2\left(2\pi\right)^n}{(2n-1)!!}\text{ and }\sigma(S^{2n+1})=\frac{(2\pi)^{n+1}}{(2n)!!} \tag{9.57}$$

which is verified inductively using Eq. (9.56). Indeed,

$$\sigma(S^{2n+1})=\sigma(S^{2n})\gamma_{2n}=\frac{2\left(2\pi\right)^n}{(2n-1)!!}\pi\frac{(2n-1)!!}{(2n)!!}=\frac{(2\pi)^{n+1}}{(2n)!!}$$

and

$$\sigma(S^{(n+1)})=\sigma(S^{2n+2})=\sigma(S^{2n+1})\gamma_{2n+1}=\frac{(2\pi)^{n+1}}{(2n)!!}2\frac{(2n)!!}{(2n+1)!!}=\frac{2\left(2\pi\right)^{n+1}}{(2n+1)!!}.$$

Using

$$(2n)!!=2n\left(2(n-1)\right)\dots(2\cdot1)=2^nn!$$

we may write $\sigma(S^{2n+1})=\frac{2\pi^{n+1}}{n!}$ which shows that Eqs. (9.41) and (9.57 are in agreement. We may also write the formula in Eq. (9.57) as

$$\sigma(S^n)=\begin{cases}\frac{2(2\pi)^{n/2}}{(n-1)!!}\text{ for }n\text{ even}\\ \frac{(2\pi)^{\frac{n+1}{2}}}{(n-1)!!}\text{ for }n\text{ odd.}\end{cases}$$

9.8 Gaussian Random Vectors

Definition 9.34 (Gaussian Random Vectors). *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X : \Omega \to \mathbb{R}^d$ be a random vector. We say that X is Gaussian if there exists an $d \times d$ – symmetric matrix Q and a vector $\mu \in \mathbb{R}^d$ such that*

$$\mathbb{E}\left[e^{i\lambda \cdot X}\right] = \exp\left(-\frac{1}{2}Q\lambda \cdot \lambda + i\mu \cdot \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d. \tag{9.58}$$

We will write $X \overset{d}{=} N(Q, \mu)$ to denote a Gaussian random vector such that Eq. (9.58) holds.

Notice that if there exists a random variable satisfying Eq. (9.58) then its law is uniquely determined by Q and μ because of Corollary 8.11. In the exercises below your will develop some basic properties of Gaussian random vectors – see Theorem 9.38 for a summary of what you will prove.

Exercise 9.4. Show that Q must be non-negative in Eq. (9.58).

Definition 9.35. *Given a Gaussian random vector, X, we call the pair, (Q, μ) appearing in Eq. (9.58) the* ***characteristics*** *of X. We will also abbreviate the statement that X is a Gaussian random vector with characteristics (Q, μ) by writing $X \overset{d}{=} N(Q, \mu)$.*

Lemma 9.36. *Suppose that $X \overset{d}{=} N(Q, \mu)$ and $A : \mathbb{R}^d \to \mathbb{R}^m$ is a $m \times d$ – real matrix and $\alpha \in \mathbb{R}^m$, then $AX + \alpha \overset{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$. In short we might abbreviate this by saying, $AN(Q, \mu) + \alpha \overset{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$.*

Proof. Let $\xi \in \mathbb{R}^m$, then

$$\begin{aligned}
\mathbb{E}\left[e^{i\xi \cdot (AX+\alpha)}\right] &= e^{i\xi\cdot\alpha}\mathbb{E}\left[e^{iA^{\mathrm{tr}}\xi \cdot X}\right] = e^{i\xi\cdot\alpha}\exp\left(-\frac{1}{2}QA^{\mathrm{tr}}\xi \cdot A^{\mathrm{tr}}\xi + i\mu \cdot A^{\mathrm{tr}}\xi\right) \\
&= e^{i\xi\cdot\alpha}\exp\left(-\frac{1}{2}AQA^{\mathrm{tr}}\xi \cdot \xi + iA\mu \cdot \xi\right) \\
&= \exp\left(-\frac{1}{2}AQA^{\mathrm{tr}}\xi \cdot \xi + i(A\mu + \alpha) \cdot \xi\right)
\end{aligned}$$

from which it follows that $AX + \alpha \overset{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$. ■

Exercise 9.5. Let P be the probability measure on $\Omega := \mathbb{R}^d$ defined by

$$dP(x) := \left(\frac{1}{2\pi}\right)^{d/2} e^{-\frac{1}{2}x\cdot x}dx = \prod_{i=1}^{d}\left(\frac{1}{\sqrt{2\pi}}e^{-x_i^2/2}dx_i\right).$$

Show that $N : \Omega \to \mathbb{R}^d$ defined by $N(x) = x$ is Gaussian and satisfies Eq. (9.58) with $Q = I$ and $\mu = 0$. Also show

$$\mu_i = \mathbb{E}N_i \text{ and } \delta_{ij} = \mathrm{Cov}(N_i, N_j) \text{ for all } 1 \le i, j \le d. \tag{9.59}$$

Hint: use Exercise 7.15 and (of course) Fubini's theorem.

Exercise 9.6. Let A be any real $m \times d$ matrix and $\mu \in \mathbb{R}^m$ and set $X := AN + \mu$ where $\Omega = \mathbb{R}^d$, P, and N are as in Exercise 9.5. Show that X is Gaussian by showing Eq. (9.58) holds with $Q = AA^{\mathrm{tr}}$ (A^{tr} is the transpose of the matrix A) and $\mu = \mu$. Also show that

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \mathrm{Cov}(X_i, X_j) \text{ for all } 1 \le i, j \le m. \tag{9.60}$$

Remark 9.37 (Spectral Theorem). Recall that if Q is a real symmetric $d \times d$ matrix, then the spectral theorem asserts there exists an orthonormal basis, $\{u\}_{j=1}^d$, such that $Qu_j = \lambda_j u_j$ for some $\lambda_j \in \mathbb{R}$. Moreover, $\lambda_j \ge 0$ for all j is equivalent to Q being non-negative. When $Q \ge 0$ we may define $Q^{1/2}$ by

$$Q^{1/2}u_j := \sqrt{\lambda_j}u_j \text{ for } 1 \le j \le d.$$

Notice that $Q^{1/2} \ge 0$ and $Q = \left(Q^{1/2}\right)^2$ and $Q^{1/2}$ is still symmetric. If Q is positive definite, we may also define, $Q^{-1/2}$ by

$$Q^{-1/2}u_j := \frac{1}{\sqrt{\lambda_j}}u_j \text{ for } 1 \le j \le d$$

so that $Q^{-1/2} = \left[Q^{1/2}\right]^{-1}$.

Exercise 9.7. Suppose that Q is a positive definite (for simplicity) $d \times d$ real matrix and $\mu \in \mathbb{R}^d$ and let $\Omega = \mathbb{R}^d$, P, and N be as in Exercise 9.5. By Exercise 9.6 we know that $X = Q^{1/2}N + \mu$ is a Gaussian random vector satisfying Eq. (9.58). Use the multi-dimensional change of variables formula to show

$$\mathrm{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}}\exp\left(-\frac{1}{2}Q^{-1}(y-\mu)\cdot(y-\mu)\right)dy.$$

Let us summarize some of what the preceding exercises have shown.

Theorem 9.38. *To each positive definite $d \times d$ real symmetric matrix Q and $\mu \in \mathbb{R}^d$ there exist Gaussian random vectors, X, satisfying Eq. (9.58). Moreover for such an X,*

$$\mathrm{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}}\exp\left(-\frac{1}{2}Q^{-1}(y-\mu)\cdot(y-\mu)\right)dy$$

where Q and μ may be computed from X using,

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \mathrm{Cov}(X_i, X_j) \text{ for all } 1 \le i, j \le m. \tag{9.61}$$

When Q is degenerate, i.e. $\mathrm{Nul}(Q) \neq \{0\}$, then $X = Q^{1/2}N + \mu$ is still a Gaussian random vectors satisfying Eq. (9.58). However now the $\mathrm{Law}_P(X)$ is a measure on $\mathbb{R}^d$ which is concentrated on the non-trivial subspace, $\mathrm{Nul}(Q)^{\perp}$ – the details of this are left to the reader for now.

Exercise 9.8 (Gaussian random vectors are "highly" integrable.). Suppose that $X : \Omega \to \mathbb{R}^d$ is a Gaussian random vector, say $X \overset{d}{=} N(Q,\mu)$. Let $\|x\| := \sqrt{x \cdot x}$ and $m := \max\{Qx \cdot x : \|x\| = 1\}$ be the largest eigenvalue[3] of Q. Then $\mathbb{E}\left[e^{\varepsilon\|X\|^2}\right] < \infty$ for every $\varepsilon < \frac{1}{2m}$.

Because of Eq. (9.61), for all $\lambda \in \mathbb{R}^d$ we have

$$\mu \cdot \lambda = \sum_{i=1}^{d} \mathbb{E}X_i \cdot \lambda_i = \mathbb{E}(\lambda \cdot X)$$

and

$$\begin{aligned} Q\lambda \cdot \lambda &= \sum_{i,j} Q_{ij}\lambda_i\lambda_j = \sum_{i,j} \lambda_i\lambda_j \operatorname{Cov}(X_i, X_j) \\ &= \operatorname{Cov}\left(\sum_i \lambda_i X_i, \sum_j \lambda_j X_j\right) = \operatorname{Var}(\lambda \cdot X). \end{aligned}$$

Therefore we may reformulate the definition of a Gaussian random vector as follows.

Definition 9.39 (Gaussian Random Vectors). *Let $(\Omega, \mathcal{B}, P)$ be a probability space. A random vector, $X : \Omega \to \mathbb{R}^d$, is Gaussian iff for all $\lambda \in \mathbb{R}^d$,*

$$\mathbb{E}\left[e^{i\lambda \cdot X}\right] = \exp\left(-\frac{1}{2}\operatorname{Var}(\lambda \cdot X) + i\mathbb{E}(\lambda \cdot X)\right). \tag{9.62}$$

In short, X is a Gaussian random vector iff $\lambda \cdot X$ is a Gaussian random variable for all $\lambda \in \mathbb{R}^d$.

Remark 9.40. To conclude that a random vector, $X : \Omega \to \mathbb{R}^d$, is Gaussian it is **not** enough to check that each of its components, $\{X_i\}_{i=1}^d$, are Gaussian random variables. The following simple counter example was provided by Nate Eldredge. Let $(X,Y) : \Omega \to \mathbb{R}^2$ be a Random vector such that $(X,Y)_* P = \mu \otimes \nu$ where $d\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}dx$ and $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$. Then $(X, YX) : \Omega \to \mathbb{R}^2$ is a random vector such that both components, X and YX, are Gaussian random variables but (X, YX) is **not** a Gaussian random vector.

Exercise 9.9. Prove the assertion made in Remark 9.40. **Hint:** explicitly compute $\mathbb{E}\left[e^{i(\lambda_1 X + \lambda_2 XY)}\right]$.

[3] For those who know about operator norms observe that $m = \|Q\|$ in this case.

9.8.1 *Gaussian measures with possibly degenerate covariances

The main aim of this subsection is to explicitly describe Gaussian measures with possibly degenerate covariances, Q. The case where $Q > 0$ has already been done in Theorem 9.38.

Remark 9.41. Recall that if Q is a real symmetric $N \times N$ matrix, then the spectral theorem asserts there exists an orthonormal basis, $\{u\}_{j=1}^N$, such that $Qu_j = \lambda_j u_j$ for some $\lambda_j \in \mathbb{R}$. Moreover, $\lambda_j \geq 0$ for all j is equivalent to Q being non-negative. Hence if $Q \geq 0$ and $f : \{\lambda_j : j = 1, 2, \ldots, N\} \to \mathbb{R}$, we may define $f(Q)$ to be the unique linear transformation on $\mathbb{R}^N$ such that $f(Q)u_j = \lambda_j u_j$.

Example 9.42. When $Q \geq 0$ and $f(x) := \sqrt{x}$, we write $Q^{1/2}$ or $\sqrt{Q}$ for $f(Q)$. Notice that $Q^{1/2} \geq 0$ and $Q = Q^{1/2}Q^{1/2}$.

Example 9.43. When Q is symmetric and

$$f(x) = \begin{cases} 1/x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

we will denote $f(Q)$ by Q^{-1}. As the notation suggests, $f(Q)$ is the inverse of Q when Q is invertible which happens iff $\lambda_i \neq 0$ for all i. When Q is not invertible,

$$Q^{-1} := f(Q) = Q|_{\mathrm{Ran}(Q)}^{-1} P, \tag{9.63}$$

where $P : \mathbb{R}^N \to \mathbb{R}^N$ be orthogonal projection onto the $\mathrm{Ran}(Q)$. Observe that $P = g(Q)$ where $g(x) = 1_{x \neq 0}$.

Lemma 9.44. *For any $Q \geq 0$ we can find a matrix, A, such that $Q = AA^{\mathrm{tr}}$. In fact it suffices to take $A = Q^{1/2}$.*

Proposition 9.45. *Suppose $X \overset{d}{=} N(Q,c)$ (see Definition 9.35) where $c \in \mathbb{R}^N$ and Q is a positive semi-definite $N \times N$ real matrix. If $\mu = \mu_{(Q,c)} = P \circ X^{-1}$, then*

$$\int_{\mathbb{R}^N} f(x)\, d\mu(x) = \frac{1}{Z}\int_{c+\mathrm{Ran}(Q)} f(x) \exp\left(-\frac{1}{2}Q^{-1}(x-c)\cdot(x-c)\right) dx$$

where dx is now "Lebesgue measure" on $c + \mathrm{Ran}(Q)$, Q^{-1} is defined as in Eq. (9.63), and $Z := \sqrt{\det\left(2\pi Q|_{\mathrm{Ran}(Q)}\right)}$.

Proof. Let $k = \dim \mathrm{Ran}(Q)$ and choose a linear transformation, $U : \mathbb{R}^k \to \mathbb{R}^N$, such that $\mathrm{Ran}(U) = \mathrm{Ran}(Q)$ and $U : \mathbb{R}^k \to \mathrm{Ran}(Q)$ is an isometric isomorphism. Letting $A := Q^{1/2}U$, we have

$$AA^{\rm tr} = Q^{1/2}UU^{\rm tr}Q^{1/2} = Q^{1/2}P_{\operatorname{Ran}(Q)}Q^{1/2} = Q.$$

Therefore, if $Y = N(I_{k\times k}, 0)$, then $X = AY + c \overset{d}{=} N(Q, c)$ by Lemma 9.36. Observe that $X - c = Q^{1/2}UY$ takes values in $\operatorname{Ran}(Q)$ and hence the Law of $(X - c)$ is a probability measure on $\mathbb{R}^N$ which is concentrated on $\operatorname{Ran}(Q)$. From this it follows that $\mu = P \circ X^{-1}$ is a probability measure on measure on $\mathbb{R}^N$ which is concentrated on the affine space, $c + \operatorname{Ran}(Q)$. At any rate from Theorem 9.38 we have

$$\begin{aligned}\int_{\mathbb{R}^N} f(x)\,d\mu(x) &= \int_{\mathbb{R}^k} f(Ay+c)\left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}|y|^2}dy \\ &= \int_{\mathbb{R}^k} f\left(Q^{1/2}Uy+c\right)\left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}|y|^2}dy.\end{aligned}$$

Since

$$Q^{1/2}Uy + c = UU^{\rm tr}Q^{1/2}Uy + c,$$

we may make the change of variables, $z = U^{\rm tr}Q^{1/2}Uy$, using

$$dz = \sqrt{\det Q|_{\operatorname{Ran}(Q)}}dy = \sqrt{\prod_{i:\lambda_i\neq 0}\lambda_i}dy$$

and

$$\begin{aligned}|y|^2 &= \left|\left(U^{\rm tr}Q^{1/2}U\right)^{-1}z\right|^2 = \left|U^{\rm tr}Q^{-1/2}Uz\right|^2 = \left(Q^{-1/2}Uz, Q^{-1/2}Uz\right)_{\mathbb{R}^N} \\ &= \left(Q^{-1}Uz, Uz\right)_{\mathbb{R}^N},\end{aligned}$$

to find

$$\begin{aligned}\int_{\mathbb{R}^N} f(x)\,d\mu(x) &= \int_{\mathbb{R}^k} f(Uz+c)\left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}\left|\left(U^{\rm tr}Q^{1/2}U\right)^{-1}z\right|^2}dz \\ &= \int_{\mathbb{R}^k} f(Uz+c)\left(\frac{1}{2\pi}\right)^{k/2}\frac{1}{\sqrt{\det Q|_{\operatorname{Ran}(Q)}}}e^{-\frac{1}{2}\left(Q^{-1}Uz,Uz\right)_{\mathbb{R}^N}}dz \\ &= \int_{\mathbb{R}^k} f(Uz+c)\frac{1}{\sqrt{\det\left(2\pi Q|_{\operatorname{Ran}(Q)}\right)}}e^{-\frac{1}{2}\left(Q^{-1}Uz,Uz\right)_{\mathbb{R}^N}}dz \\ &= \int_{\mathbb{R}^k} f(Uz+c)\frac{1}{\sqrt{\det\left(2\pi Q|_{\operatorname{Ran}(Q)}\right)}}e^{-\frac{1}{2}\left(Q^{-1}(Uz+c-c),(Uz+c-c)\right)_{\mathbb{R}^N}}dz.\end{aligned}$$

This completes the proof, since $x = Uz + c \in c + \operatorname{Ran}(Q)$ is by definition distributed as Lebesgue measure on $c+\operatorname{Ran}(Q)$ when z is distributed as Lebesgue measure on $\mathbb{R}^k$. ∎

9.9 Kolmogorov's Extension Theorems

In this section we will extend the results of Section 5.5 to spaces which are not simply products of discrete spaces. We begin with a couple of results involving the topology on $\mathbb{R}^N$.

9.9.1 Regularity and compactness results

Theorem 9.46 (Inner-Outer Regularity). *Suppose μ is a probability measure on $\left(\mathbb{R}^N, \mathcal{B}_{\mathbb{R}^N}\right)$, then for all $B \in \mathcal{B}_{\mathbb{R}^N}$ we have*

$$\mu(B) = \inf\{\mu(V) : B \subset V \text{ and } V \text{ is open}\} \tag{9.64}$$

and

$$\mu(B) = \sup\{\mu(K) : K \subset B \text{ with } K \text{ compact}\}. \tag{9.65}$$

Proof. In this proof, C, and C_i will always denote a closed subset of $\mathbb{R}^N$ and V, V_i will always be open subsets of $\mathbb{R}^N$. Let $\mathcal{F}$ be the collection of sets, $A \in \mathcal{B}$, such that for all $\varepsilon > 0$ there exists an open set V and a closed set, C, such that $C \subset A \subset V$ and $\mu(V \setminus C) < \varepsilon$. The key point of the proof is to show $\mathcal{F} = \mathcal{B}$ for this certainly implies Equation (9.64) and also that

$$\mu(B) = \sup\{\mu(C) : C \subset B \text{ with } C \text{ closed}\}. \tag{9.66}$$

Moreover, by MCT, we know that if C is closed and $K_n := C \cap \left\{x \in \mathbb{R}^N : |x| \le n\right\}$, then $\mu(K_n) \uparrow \mu(C)$. This observation along with Eq. (9.66) shows Eq. (9.65) is valid as well.

To prove $\mathcal{F} = \mathcal{B}$, it suffices to show $\mathcal{F}$ is a $\sigma-$algebra which contains all closed subsets of $\mathbb{R}^N$. To the prove the latter assertion, given a closed subset, $C \subset \mathbb{R}^N$, and $\varepsilon > 0$, let

$$C_\varepsilon := \cup_{x\in C}B(x,\varepsilon)$$

where $B(x,\varepsilon) := \left\{y \in \mathbb{R}^N : |y - x| < \varepsilon\right\}$. Then C_ε is an open set and $C_\varepsilon \downarrow C$ as $\varepsilon \downarrow 0$. (You prove.) Hence by the DCT, we know that $\mu(C_\varepsilon \setminus C) \downarrow 0$ form which it follows that $C \in \mathcal{F}$.

We will now show that $\mathcal{F}$ is an algebra. Clearly $\mathcal{F}$ contains the empty set and if $A \in \mathcal{F}$ with $C \subset A \subset V$ and $\mu(V \setminus C) < \varepsilon$, then $V^c \subset A^c \subset C^c$ with $\mu(C^c \setminus V^c) = \mu(V \setminus C) < \varepsilon$. This shows $A^c \in \mathcal{F}$. Similarly if $A_i \in \mathcal{F}$ for $i = 1, 2$ and $C_i \subset A_i \subset V_i$ with $\mu(V_i \setminus C_i) < \varepsilon$, then

$$C := C_1 \cup C_2 \subset A_1 \cup A_2 \subset V_1 \cup V_2 =: V$$

and

$$\mu (V \setminus C) \le \mu (V_1 \setminus C) + \mu (V_2 \setminus C)$$
$$\le \mu (V_1 \setminus C_1) + \mu (V_2 \setminus C_2) < 2\varepsilon.$$

This implies that $A_1 \cup A_2 \in \mathcal{F}$ and we have shown $\mathcal{F}$ is an algebra.

We now show that $\mathcal{F}$ is a σ – algebra. To do this it suffices to show $A := \sum_{n=1}^{\infty} A_n \in \mathcal{F}$ if $A_n \in \mathcal{F}$ with $A_n \cap A_m = \emptyset$ for $m \neq n$. Let $C_n \subset A_n \subset V_n$ with $\mu (V_n \setminus C_n) < \varepsilon 2^{-n}$ for all n and let $C^N := \cup_{n \le N} C_n$ and $V := \cup_{n=1}^{\infty} V_n$. Then $C^N \subset A \subset V$ and

$$\begin{aligned}\mu \left(V \setminus C^N\right) &\le \sum_{n=0}^{\infty} \mu \left(V_n \setminus C^N\right) \le \sum_{n=0}^{N} \mu (V_n \setminus C_n) + \sum_{n=N+1}^{\infty} \mu (V_n) \\ &\le \sum_{n=0}^{N} \varepsilon 2^{-n} + \sum_{n=N+1}^{\infty} \left[\mu (A_n) + \varepsilon 2^{-n}\right] \\ &= \varepsilon + \sum_{n=N+1}^{\infty} \mu (A_n).\end{aligned}$$

The last term is less that 2ε for N sufficiently large because $\sum_{n=1}^{\infty} \mu (A_n) = \mu (A) < \infty$. ■

Notation 9.47 *Let $I := [0,1]$, $Q = I^{\mathbb{N}}$, $\pi_j : Q \to I$ be the projection map, $\pi_j (x) = x_j$ (where $x = (x_1, x_2, \dots, x_j, \dots)$ for all $j \in \mathbb{N}$, and $\mathcal{B}_Q := \sigma (\pi_j : j \in \mathbb{N})$ be the product σ – algebra on Q. Let us further say that a sequence $\{x(m)\}_{m=1}^{\infty} \subset Q$, where $x(m) = (x_1(m), x_2(m), \dots)$, converges to $x \in Q$ iff $\lim_{m\to\infty} x_j (m) = x_j$ for all $j \in \mathbb{N}$. (This is just pointwise convergence.)*

Lemma 9.48 (Baby Tychonoff's Theorem). *The infinite dimensional cube, Q, is compact, i.e. every sequence $\{x(m)\}_{m=1}^{\infty} \subset Q$ has a convergent subsequence,$\{x(m_k)\}_{k=1}^{\infty}$.*

Proof. Since I is compact, it follows that for each $j \in \mathbb{N}$, $\{x_j (m)\}_{m=1}^{\infty}$ has a convergent subsequence. It now follow by Cantor's diagonalization method, that there is a subsequence, $\{m_k\}_{k=1}^{\infty}$, of $\mathbb{N}$ such that $\lim_{k\to\infty} x_j (m_k) \in I$ exists for all $j \in \mathbb{N}$. ■

Corollary 9.49 (Finite Intersection Property). *Suppose that $K_m \subset Q$ are sets which are, (i) closed under taking sequential limits[4], and (ii) have the finite intersection property, (i.e. $\cap_{m=1}^{n} K_m \neq \emptyset$ for all $m \in \mathbb{N}$), then $\cap_{m=1}^{\infty} K_m \neq \emptyset$.*

Proof. By assumption, for each $n \in \mathbb{N}$, there exists $x(n) \in \cap_{m=1}^{n} K_m$. Hence by Lemma 9.48 there exists a subsequence, $x(n_k)$, such that $x := \lim_{k\to\infty} x(n_k)$ exists in Q. Since $x(n_k) \in \cap_{m=1}^{n} K_m$ for all k large, and each K_m is closed under sequential limits, it follows that $x \in K_m$ for all m. Thus we have shown, $x \in \cap_{m=1}^{\infty} K_m$ and hence $\cap_{m=1}^{\infty} K_m \neq \emptyset$. ■

[4] For example, if $K_m = K'_m \times Q$ with K'_m being a closed subset of I^m, then K_m is closed under sequential limits.

9.9.2 Kolmogorov's Extension Theorem and Infinite Product Measures

Theorem 9.50 (Kolmogorov's Extension Theorem). *Let $I := [0,1]$. For each $n \in \mathbb{N}$, let μ_n be a probability measure on $(I^n, \mathcal{B}_{I^n})$ such that $\mu_{n+1} (A \times I) = \mu_n (A)$. Then there exists a unique measure, P on $(Q, \mathcal{B}_Q)$ such that*

$$P(A \times Q) = \mu_n (A) \tag{9.67}$$

for all $A \in \mathcal{B}_{I^n}$ and $n \in \mathbb{N}$.

Proof. Let $\mathcal{A} := \cup \mathcal{B}_n$ where $\mathcal{B}_n := \{A \times Q : A \in \mathcal{B}_{I^n}\} = \sigma (\pi_1, \dots, \pi_n)$, where $\pi_i (x) = x_i$ if $x = (x_1, x_2, \dots) \in Q$. Then define P on $\mathcal{A}$ by Eq. (9.67) which is easily seen (Exercise 9.10) to be a well defined finitely additive measure on $\mathcal{A}$. So to finish the proof it suffices to show if $B_n \in \mathcal{A}$ is a decreasing sequence such that

$$\inf_n P(B_n) = \lim_{n\to\infty} P(B_n) = \varepsilon > 0,$$

then $B := \cap B_n \neq \emptyset$.

To simplify notation, we may reduce to the case where $B_n \in \mathcal{B}_n$ for all n. To see this is permissible, Let us choose $1 \le n_1 < n_2 < n_3 < \dots$. such that $B_k \in \mathcal{B}_{n_k}$ for all k. (This is possible since $\mathcal{B}_n$ is increasing in n.) We now define a new decreasing sequence of sets, $\left\{\tilde{B}_k\right\}_{k=1}^{\infty}$ as follows,

$$\left(\tilde{B}_1, \tilde{B}_2, \dots\right) = \left(\overbrace{Q, \dots, Q}^{n_1 - 1 \text{ times}}, \overbrace{B_1, \dots, B_1}^{n_2 - n_1 \text{ times}}, \overbrace{B_2, \dots, B_2}^{n_3 - n_2 \text{ times}}, \overbrace{B_3, \dots, B_3}^{n_4 - n_3 \text{ times}}, \dots\right).$$

We then have $\tilde{B}_n \in \mathcal{B}_n$ for all n, $\lim_{n\to\infty} P\left(\tilde{B}_n\right) = \varepsilon > 0$, and $B = \cap_{n=1}^{\infty} \tilde{B}_n$. Hence we may replace B_n by $\tilde{B}_n$ if necessary so as to have $B_n \in \mathcal{B}_n$ for all n.

Since $B_n \in \mathcal{B}_n$, there exists $B'_n \in \mathcal{B}_{I^n}$ such that $B_n = B'_n \times Q$ for all n. Using the regularity Theorem 9.46, there are compact sets, $K'_n \subset B'_n \subset I^n$, such that $\mu_n (B'_n \setminus K'_n) \le \varepsilon 2^{-n-1}$ for all $n \in \mathbb{N}$. Let $K_n := K'_n \times Q$, then $P(B_n \setminus K_n) \le \varepsilon 2^{-n-1}$ for all n. Moreover,

$$\begin{aligned}P(B_n \setminus [\cap_{m=1}^{n} K_m]) &= P(\cup_{m=1}^{n} [B_n \setminus K_m]) \le \sum_{m=1}^{n} P(B_n \setminus K_m) \\ &\le \sum_{m=1}^{n} P(B_m \setminus K_m) \le \sum_{m=1}^{n} \varepsilon 2^{-m-1} \le \varepsilon/2.\end{aligned}$$

So, for all $n \in \mathbb{N}$,

$$P\left(\cap_{m=1}^{n} K_m\right) = P\left(B_n\right) - P\left(B_n \setminus \left[\cap_{m=1}^{n} K_m\right]\right) \geq \varepsilon - \varepsilon/2 = \varepsilon/2,$$

and in particular, $\cap_{m=1}^{n} K_m \neq \emptyset$. An application of Corollary 9.49 now implies, $\emptyset \neq \cap_n K_n \subset \cap_n B_n$. ■

Exercise 9.10. Show that Eq. (9.67) defines a well defined finitely additive measure on $\mathcal{A} := \cup \mathcal{B}_n$.

The next result is an easy corollary of Theorem 9.50.

Theorem 9.51. *Suppose $\{(X_n, \mathcal{M}_n)\}_{n\in\mathbb{N}}$ are standard Borel spaces (see Appendix 9.10 below), $X := \prod_{n\in\mathbb{N}} X_n$, $\pi_n : X \to X_n$ be the n^{th} – projection map, $\mathcal{B}_n := \sigma\left(\pi_k : k \leq n\right)$, $\mathcal{B} = \sigma(\pi_n : n \in \mathbb{N})$, and $T_n := X_{n+1} \times X_{n+2} \times \ldots.$ Further suppose that for each $n \in \mathbb{N}$ we are given a probability measure, μ_n on $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ such that*

$$\mu_{n+1}\left(A \times X_{n+1}\right) = \mu_n\left(A\right) \text{ for all } n \in \mathbb{N} \text{ and } A \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n.$$

Then there exists a unique probability measure, P, on $(X, \mathcal{B})$ such that $P\left(A \times T_n\right) = \mu_n\left(A\right)$ for all $A \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$.

Proof. Since each $(X_n, \mathcal{M}_n)$ is measure theoretic isomorphic to a Borel subset of I, we may assume that $X_n \in \mathcal{B}_I$ and $\mathcal{M}_n = (\mathcal{B}_I)_{X_n}$ for all n. Given $A \in \mathcal{B}_{I^n}$, let $\bar{\mu}_n\left(A\right) := \mu_n\left(A \cap \left[X_1 \times \cdots \times X_n\right]\right)$ – a probability measure on $\mathcal{B}_{I^n}$. Furthermore,

$$\begin{aligned}
\bar{\mu}_{n+1}\left(A \times I\right) &= \mu_{n+1}\left(\left[A \times I\right] \cap \left[X_1 \times \cdots \times X_{n+1}\right]\right) \\
&= \mu_{n+1}\left(\left(A \cap \left[X_1 \times \cdots \times X_n\right]\right) \times X_{n+1}\right) \\
&= \mu_n\left(\left(A \cap \left[X_1 \times \cdots \times X_n\right]\right)\right) = \bar{\mu}_n\left(A\right).
\end{aligned}$$

Hence by Theorem 9.50, there is a unique probability measure, $\bar{P}$, on $I^{\mathbb{N}}$ such that

$$\bar{P}\left(A \times I^{\mathbb{N}}\right) = \bar{\mu}_n\left(A\right) \text{ for all } n \in \mathbb{N} \text{ and } A \in \mathcal{B}_{I^n}.$$

We will now check that $P := \bar{P}|_{\otimes_{n=1}^{\infty}\mathcal{M}_n}$ is the desired measure. First off we have

$$\begin{aligned}
\bar{P}\left(X\right) &= \lim_{n\to\infty} \bar{P}\left(X_1 \times \cdots \times X_n \times I^{\mathbb{N}}\right) = \lim_{n\to\infty} \bar{\mu}_n\left(X_1 \times \cdots \times X_n\right) \\
&= \lim_{n\to\infty} \mu_n\left(X_1 \times \cdots \times X_n\right) = 1.
\end{aligned}$$

Secondly, if $A \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$, we have

$$\begin{aligned}
P\left(A \times T_n\right) &= \bar{P}\left(A \times T_n\right) = \bar{P}\left(\left(A \times I^{\mathbb{N}}\right) \cap X\right) \\
&= \bar{P}\left(A \times I^{\mathbb{N}}\right) = \bar{\mu}_n\left(A\right) = \mu_n\left(A\right).
\end{aligned}$$

■

Here is an example of this theorem in action.

Theorem 9.52 (Infinite Product Measures). *Suppose that $\{\nu_n\}_{n=1}^{\infty}$ are a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $\mathcal{B} := \otimes_{n\in\mathbb{N}}\mathcal{B}_{\mathbb{R}}$ is the product σ – algebra on $\mathbb{R}^{\mathbb{N}}$. Then there exists a unique probability measure, ν, on $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}\right)$, such that*

$$\nu\left(A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^{\mathbb{N}}\right) = \nu_1\left(A_1\right)\ldots\nu_n\left(A_n\right) \ \forall \ A_i \in \mathcal{B}_{\mathbb{R}} \ \&\ n \in \mathbb{N}. \tag{9.68}$$

Moreover, this measure satisfies,

$$\int_{\mathbb{R}^{\mathbb{N}}} f\left(x_1, \ldots, x_n\right) d\nu\left(x\right) = \int_{\mathbb{R}^n} f\left(x_1, \ldots, x_n\right) d\nu_1\left(x_1\right)\ldots d\nu_n\left(x_n\right) \tag{9.69}$$

for all $n \in \mathbb{N}$ and $f : \mathbb{R}^n \to \mathbb{R}$ which are bounded and measurable or non-negative and measurable.

Proof. The measure ν is created by apply Theorem 9.51 with $\mu_n := \nu_1 \otimes \cdots \otimes \nu_n$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n} = \otimes_{k=1}^{n}\mathcal{B}_{\mathbb{R}})$ for each $n \in \mathbb{N}$. Observe that

$$\mu_{n+1}\left(A \times \mathbb{R}\right) = \mu_n\left(A\right) \cdot \nu_{n+1}\left(\mathbb{R}\right) = \mu_n\left(A\right),$$

so that $\{\mu_n\}_{n=1}^{\infty}$ satisfies the needed consistency conditions. Thus there exists a unique measure ν on $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}\right)$ such that

$$\nu\left(A \times \mathbb{R}^{\mathbb{N}}\right) = \mu_n\left(A\right) \text{ for all } A \in \mathcal{B}_{\mathbb{R}^n} \text{ and } n \in \mathbb{N}.$$

Taking $A = A_1 \times A_2 \times \cdots \times A_n$ with $A_i \in \mathcal{B}_{\mathbb{R}}$ then gives Eq. (9.68). For this measure, it follows that Eq. (9.69) holds when $f = 1_{A_1\times\cdots\times A_n}$. Thus by an application of Theorem 8.2 with $\mathbb{M} = \{1_{A_1\times\cdots\times A_n} : A_i \in \mathcal{B}_{\mathbb{R}}\}$ and $\mathbb{H}$ being the set of bounded measurable functions, $f : \mathbb{R}^n \to \mathbb{R}$, for which Eq. (9.69) shows that Eq. (9.69) holds for all bounded and measurable functions, $f : \mathbb{R}^n \to \mathbb{R}$. The statement involving non-negative functions follows by a simple limiting argument involving the MCT. ■

It turns out that the existence of infinite product measures require no topological restrictions on the measure spaces involved. See Corollary 17.57 below.

9.10 Appendix: Standard Borel Spaces*

For more information along the lines of this section, see Royden [46] and Parthasarathy [40].

Definition 9.53. *Two measurable spaces, $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are said to be* ***isomorphic*** *if there exists a bijective map, $f : X \to Y$ such that $f(\mathcal{M}) = \mathcal{N}$ and $f^{-1}(\mathcal{N}) = \mathcal{M}$, i.e. both f and f^{-1} are measurable. In this case we say f is a measure theoretic isomorphism and we will write $X \cong Y$.*

Definition 9.54. *A measurable space, $(X, \mathcal{M})$ is said to be a* ***standard Borel space*** *if $(X, \mathcal{M}) \cong (B, \mathcal{B}_B)$ where B is a Borel subset of $\left((0,1), \mathcal{B}_{(0,1)}\right)$.*

Definition 9.55 (Polish spaces). *A* ***Polish space*** *is a separable topological space (X, τ) which admits a complete metric, ρ, such that $\tau = \tau_\rho$.*

The main goal of this chapter is to prove every Borel subset of a Polish space is a standard Borel space, see Corollary 9.65 below. Along the way we will show a number of spaces, including $[0,1]$, $(0,1]$, $[0,1]^d$, $\mathbb{R}^d$, $\{0,1\}^{\mathbb{N}}$, and $\mathbb{R}^{\mathbb{N}}$, are all (measure theoretic) isomorphic to $(0,1)$. Moreover we also will see that a countable product of standard Borel spaces is again a standard Borel space, see Corollary 9.62.

*On first reading, you may wish to skip the rest of this section.

Lemma 9.56. *Suppose $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are measurable spaces such that $X = \sum_{n=1}^\infty X_n$, $Y = \sum_{n=1}^\infty Y_n$, with $X_n \in \mathcal{M}$ and $Y_n \in \mathcal{N}$. If $(X_n, \mathcal{M}_{X_n})$ is isomorphic to $(Y_n, \mathcal{N}_{Y_n})$ for all n then $X \cong Y$. Moreover, if $(X_n, \mathcal{M}_n)$ and $(Y_n, \mathcal{N}_n)$ are isomorphic measure spaces, then $(X := \prod_{n=1}^\infty X_n, \otimes_{n=1}^\infty \mathcal{M}_n)$ are $(Y := \prod_{n=1}^\infty Y_n, \otimes_{n=1}^\infty \mathcal{N}_n)$ are isomorphic.*

Proof. For each $n \in \mathbb{N}$, let $f_n : X_n \to Y_n$ be a measure theoretic isomorphism. Then define $f : X \to Y$ by $f = f_n$ on X_n. Clearly, $f : X \to Y$ is a bijection and if $B \in \mathcal{N}$, then

$$f^{-1}(B) = \cup_{n=1}^\infty f^{-1}(B \cap Y_n) = \cup_{n=1}^\infty f_n^{-1}(B \cap Y_n) \in \mathcal{M}.$$

This shows f is measurable and by similar considerations, f^{-1} is measurable as well. Therefore, $f : X \to Y$ is the desired measure theoretic isomorphism.

For the second assertion, let $f_n : X_n \to Y_n$ be a measure theoretic isomorphism of all $n \in \mathbb{N}$ and then define

$$f(x) = (f_1(x_1), f_2(x_2), \dots) \text{ with } x = (x_1, x_2, \dots) \in X.$$

Again it is clear that f is bijective and measurable, since

$$f^{-1}\left(\prod_{n=1}^\infty B_n\right) = \prod_{n=1}^\infty f_n^{-1}(B_n) \in \otimes_{n=1}^\infty \mathcal{N}_n$$

for all $B_n \in \mathcal{M}_n$ and $n \in \mathbb{N}$. Similar reasoning shows that f^{-1} is measurable as well. ■

Proposition 9.57. *Let $-\infty < a < b < \infty$. The following measurable spaces equipped with there Borel σ – algebras are all isomorphic; $(0,1)$, $[0,1]$, $(0,1]$, $[0,1)$, (a,b), $[a,b]$, $(a,b]$, $[a,b)$, $\mathbb{R}$, and $(0,1) \cup \Lambda$ where Λ is a finite or countable subset of $\mathbb{R} \setminus (0,1)$.*

Proof. It is easy to see by that any bounded open, closed, or half open interval is isomorphic to any other such interval using an affine transformation. Let us now show $(-1,1) \cong [-1,1]$. To prove this it suffices, by Lemma 9.56,to observe that

$$(-1,1) = \{0\} \cup \sum_{n=0}^\infty \left((-2^{-n}, -2^{-n}] \cup [2^{-n-1}, 2^{-n})\right)$$

and

$$[-1,1] = \{0\} \cup \sum_{n=0}^\infty \left([-2^{-n}, -2^{-n-1}) \cup (2^{-n-1}, 2^{-n}]\right).$$

Similarly $(0,1)$ is isomorphic to $(0,1]$ because

$$(0,1) = \sum_{n=0}^\infty [2^{-n-1}, 2^{-n}) \text{ and } (0,1] = \sum_{n=0}^\infty (2^{-n-1}, 2^{-n}].$$

The assertion involving $\mathbb{R}$ can be proved using the bijection, $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$.

If $\Lambda = \{1\}$, then by Lemma 9.56 and what we have already proved, $(0,1) \cup \{1\} = (0,1] \cong (0,1)$. Similarly if $N \in \mathbb{N}$ with $N \geq 2$ and $\Lambda = \{2, \dots, N+1\}$, then

$$(0,1) \cup \Lambda \cong (0,1] \cup \Lambda = (0, 2^{-N+1}] \cup \left[\sum_{n=1}^{N-1} (2^{-n}, 2^{-n-1}]\right] \cup \Lambda$$

while

$$(0,1) = \left(0, 2^{-N+1}\right) \cup \left[\sum_{n=1}^{N-1} \left(2^{-n}, 2^{-n-1}\right)\right] \cup \left\{2^{-n} : n = 1, 2, \dots, N\right\}$$

and so again it follows from what we have proved and Lemma 9.56 that $(0,1) \cong (0,1) \cup \Lambda$. Finally if $\Lambda = \{2, 3, 4, \dots\}$ is a countable set, we can show $(0,1) \cong (0,1) \cup \Lambda$ with the aid of the identities,

$$(0,1) = \left[\sum_{n=1}^\infty \left(2^{-n}, 2^{-n-1}\right)\right] \cup \left\{2^{-n} : n \in \mathbb{N}\right\}$$

and

$$(0,1) \cup \Lambda \cong (0,1] \cup \Lambda = \left[\sum_{n=1}^\infty (2^{-n}, 2^{-n-1}]\right] \cup \Lambda.$$

■

Notation 9.58 *Suppose $(X, \mathcal{M})$ is a measurable space and A is a set. Let $\pi_a : X^A \to X$ denote projection operator onto the a^{th} – component of X^A (i.e. $\pi_a(\omega) = \omega(a)$ for all $a \in A$) and let $\mathcal{M}^{\otimes A} := \sigma(\pi_a : a \in A)$ be the product σ – algebra on X^A.*

Lemma 9.59. *If $\varphi : A \to B$ is a bijection of sets and $(X, \mathcal{M})$ is a measurable space, then $\left(X^A, \mathcal{M}^{\otimes A}\right) \cong \left(X^B, \mathcal{M}^{\otimes B}\right)$.*

Proof. The map $f : X^B \to X^A$ defined by $f(\omega) = \omega \circ \varphi$ for all $\omega \in X^B$ is a bijection with $f^{-1}(\alpha) = \alpha \circ \varphi^{-1}$. If $a \in A$ and $\omega \in X^B$, we have

$$\pi_a^{X^A} \circ f(\omega) = f(\omega)(a) = \omega(\varphi(a)) = \pi_{\varphi(a)}^{X^B}(\omega),$$

where $\pi_a^{X^A}$ and $\pi_b^{X^B}$ are the projection operators on X^A and X^B respectively. Thus $\pi_a^{X^A} \circ f = \pi_{\varphi(a)}^{X^B}$ for all $a \in A$ which shows f is measurable. Similarly, $\pi_b^{X^B} \circ f^{-1} = \pi_{\varphi^{-1}(b)}^{X^A}$ showing f^{-1} is measurable as well. ∎

Proposition 9.60. *Let $\Omega := \{0,1\}^{\mathbb{N}}$, $\pi_i : \Omega \to \{0,1\}$ be projection onto the i^{th} component, and $\mathcal{B} := \sigma(\pi_1, \pi_2, \dots)$ be the product σ – algebra on Ω. Then $(\Omega, \mathcal{B}) \cong \left((0,1), \mathcal{B}_{(0,1)}\right)$.*

Proof. We will begin by using a specific binary digit expansion of a point $x \in [0,1)$ to construct a map from $[0,1) \to \Omega$. To this end, let $r_1(x) = x$,

$$\gamma_1(x) := 1_{x \geq 2^{-1}} \text{ and } r_2(x) := x - 2^{-1}\gamma_1(x) \in (0, 2^{-1}),$$

then let $\gamma_2 := 1_{r_2 \geq 2^{-2}}$ and $r_3 = r_2 - 2^{-2}\gamma_2 \in \left(0, 2^{-2}\right)$. Working inductively, we construct $\{\gamma_k(x), r_k(x)\}_{k=1}^{\infty}$ such that $\gamma_k(x) \in \{0,1\}$, and

$$r_{k+1}(x) = r_k(x) - 2^{-k}\gamma_k(x) = x - \sum_{j=1}^{k} 2^{-j}\gamma_j(x) \in \left(0, 2^{-k}\right) \tag{9.70}$$

for all k. Let us now define $g : [0,1) \to \Omega$ by $g(x) := (\gamma_1(x), \gamma_2(x), \dots)$. Since each component function, $\pi_j \circ g = \gamma_j : [0,1) \to \{0,1\}$, is measurable it follows that g is measurable.

By construction,

$$x = \sum_{j=1}^{k} 2^{-j}\gamma_j(x) + r_{k+1}(x)$$

and $r_{k+1}(x) \to 0$ as $k \to \infty$, therefore

$$x = \sum_{j=1}^{\infty} 2^{-j}\gamma_j(x) \text{ and } r_{k+1}(x) = \sum_{j=k+1}^{\infty} 2^{-j}\gamma_j(x). \tag{9.71}$$

Hence if we define $f : \Omega \to [0,1]$ by $f = \sum_{j=1}^{\infty} 2^{-j}\pi_j$, then $f(g(x)) = x$ for all $x \in [0,1)$. This shows g is injective, f is surjective, and f in injective on the range of g.

We now claim that $\Omega_0 := g([0,1))$, the range of g, consists of those $\omega \in \Omega$ such that $\omega_i = 0$ for infinitely many i. Indeed, if there exists an $k \in \mathbb{N}$ such that $\gamma_j(x) = 1$ for all $j \geq k$, then (by Eq. (9.71)) $r_{k+1}(x) = 2^{-k}$ which would contradict Eq. (9.70). Hence $g([0,1)) \subset \Omega_0$. Conversely if $\omega \in \Omega_0$ and $x = f(\omega) \in [0,1)$, it is not hard to show inductively that $\gamma_j(x) = \omega_j$ for all j, i.e. $g(x) = \omega$. For example, if $\omega_1 = 1$ then $x \geq 2^{-1}$ and hence $\gamma_1(x) = 1$. Alternatively, if $\omega_1 = 0$, then

$$x = \sum_{j=2}^{\infty} 2^{-j}\omega_j < \sum_{j=2}^{\infty} 2^{-j} = 2^{-1}$$

so that $\gamma_1(x) = 0$. Hence it follows that $r_2(x) = \sum_{j=2}^{\infty} 2^{-j}\omega_j$ and by similar reasoning we learn $r_2(x) \geq 2^{-2}$ iff $\omega_2 = 1$, i.e. $\gamma_2(x) = 1$ iff $\omega_2 = 1$. The full induction argument is now left to the reader.

Since single point sets are in $\mathcal{B}$ and

$$\Lambda := \Omega \setminus \Omega_0 = \cup_{n=1}^{\infty}\{\omega \in \Omega : \omega_j = 1 \text{ for } j \geq n\}$$

is a countable set, it follows that $\Lambda \in \mathcal{B}$ and therefore $\Omega_0 = \Omega \setminus \Lambda \in \mathcal{B}$. Hence we may now conclude that $g : \left([0,1), \mathcal{B}_{[0,1)}\right) \to (\Omega_0, \mathcal{B}_{\Omega_0})$ is a measurable bijection with measurable inverse given by $f|_{\Omega_0}$, i.e. $\left([0,1), \mathcal{B}_{[0,1)}\right) \cong (\Omega_0, \mathcal{B}_{\Omega_0})$. An application of Lemma 9.56 and Proposition 9.57 now implies

$$\Omega = \Omega_0 \cup \Lambda \cong [0,1) \cup \mathbb{N} \cong [0,1) \cong (0,1).$$

∎

Corollary 9.61. *The following spaces are all isomorphic to $\left((0,1), \mathcal{B}_{(0,1)}\right)$; $(0,1)^d$ and $\mathbb{R}^d$ for any $d \in \mathbb{N}$ and $[0,1]^{\mathbb{N}}$ and $\mathbb{R}^{\mathbb{N}}$ where both of these spaces are equipped with their natural product σ – algebras, .*

Proof. In light of Lemma 9.56 and Proposition 9.57 we know that $(0,1)^d \cong \mathbb{R}^d$ and $(0,1)^{\mathbb{N}} \cong [0,1]^{\mathbb{N}} \cong \mathbb{R}^{\mathbb{N}}$. So, using Proposition 9.60, it suffices to show $(0,1)^d \cong \Omega \cong (0,1)^{\mathbb{N}}$ and to do this it suffices to show $\Omega^d \cong \Omega$ and $\Omega^{\mathbb{N}} \cong \Omega$.

To reduce the problem further, let us observe that $\Omega^d \cong \{0,1\}^{\mathbb{N} \times \{1,2,\dots,d\}}$ and $\Omega^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}^2}$. For example, let $g : \Omega^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}^2}$ be defined by $g(\omega)(i,j) = \omega(i)(j)$ for all $\omega \in \Omega^{\mathbb{N}} = \left[\{0,1\}^{\mathbb{N}}\right]^{\mathbb{N}}$. Then g is a bijection and since $\pi_{(i,j)}^{\{0,1\}^{\mathbb{N}^2}} \circ g(\omega) = \pi_j^{\Omega}\left(\pi_i^{\Omega^{\mathbb{N}}}(\omega)\right)$, it follows that g is measurable. The inverse, $g^{-1} : \{0,1\}^{\mathbb{N}^2} \to \Omega^{\mathbb{N}}$, to g is given by $g^{-1}(\alpha)(i)(j) = \alpha(i,j)$. To see this map is measurable, we have $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1} : \{0,1\}^{\mathbb{N}^2} \to \Omega = \{0,1\}^{\mathbb{N}}$ is given $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1}(\alpha) = g^{-1}(\alpha)(i)(\cdot) = \alpha(i,\cdot)$ and hence

$$\pi_j^{\Omega} \circ \pi_i^{\Omega^{\mathbb{N}}} \circ g(\alpha) = \alpha(i,j) = \pi_{i,j}^{\{0,1\}^{\mathbb{N}^2}}(\alpha)$$

from which it follows that $\pi_j^{\Omega} \circ \pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1} = \pi^{\{0,1\}^{\mathbb{N}^2}}$ is measurable for all $i,j \in \mathbb{N}$ and hence $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1}$ is measurable for all $i \in \mathbb{N}$ and hence g^{-1} is measurable. This shows $\Omega^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}^2}$. The proof that $\Omega^d \cong \{0,1\}^{\mathbb{N}\times\{1,2,\dots,d\}}$ is analogous.

We may now complete the proof with a couple of applications of Lemma 9.59. Indeed $\mathbb{N}$, $\mathbb{N}\times\{1,2,\dots,d\}$, and $\mathbb{N}^2$ all have the same cardinality and therefore,

$$\{0,1\}^{\mathbb{N}\times\{1,2,\dots,d\}} \cong \{0,1\}^{\mathbb{N}^2} \cong \{0,1\}^{\mathbb{N}} = \Omega.$$

■

Corollary 9.62. *Suppose that $(X_n,\mathcal{M}_n)$ for $n\in\mathbb{N}$ are standard Borel spaces, then $X := \prod_{n=1}^\infty X_n$ equipped with the product σ – algebra, $\mathcal{M} := \otimes_{n=1}^\infty \mathcal{M}_n$ is again a standard Borel space.*

Proof. Let $A_n \in \mathcal{B}_{[0,1]}$ be Borel sets on $[0,1]$ such that there exists a measurable isomorpohism, $f_n : X_n \to A_n$. Then $f : X \to A := \prod_{n=1}^\infty A_n$ defined by $f(x_1,x_2,\dots) = (f_1(x_1), f_2(x_2),\dots)$ is easily seen to me a measure theoretic isomorphism when A is equipped with the product σ – algebra, $\otimes_{n=1}^\infty \mathcal{B}_{A_n}$. So according to Corollary 9.61, to finish the proof it suffice to show $\otimes_{n=1}^\infty \mathcal{B}_{A_n} = \mathcal{M}_A$ where $\mathcal{M} := \otimes_{n=1}^\infty \mathcal{B}_{[0,1]}$ is the product σ – algebra on $[0,1]^{\mathbb{N}}$.

The σ – algebra, $\otimes_{n=1}^\infty \mathcal{B}_{A_n}$, is generated by sets of the form, $B := \prod_{n=1}^\infty B_n$ where $B_n \in \mathcal{B}_{A_n} \subset \mathcal{B}_{[0,1]}$. On the other hand, the σ – algebra, $\mathcal{M}_A$ is generated by sets of the form, $A \cap \tilde{B}$ where $\tilde{B} := \prod_{n=1}^\infty \tilde{B}_n$ with $\tilde{B}_n \in \mathcal{B}_{[0,1]}$. Since

$$A \cap \tilde{B} = \prod_{n=1}^\infty \left(\tilde{B}_n \cap A_n\right) = \prod_{n=1}^\infty B_n$$

where $B_n = \tilde{B}_n \cap A_n$ is the generic element in $\mathcal{B}_{A_n}$, we see that $\otimes_{n=1}^\infty \mathcal{B}_{A_n}$ and $\mathcal{M}_A$ can both be generated by the same collections of sets, we may conclude that $\otimes_{n=1}^\infty \mathcal{B}_{A_n} = \mathcal{M}_A$. ■

Our next goal is to show that any Polish space with its Borel σ – algebra is a standard Borel space.

Notation 9.63 *Let $Q := [0,1]^{\mathbb{N}}$ denote the (infinite dimensional)* ***unit cube*** *in $\mathbb{R}^{\mathbb{N}}$. For $a,b\in Q$ let*

$$d(a,b) := \sum_{n=1}^\infty \frac{1}{2^n}|a_n - b_n| = \sum_{n=1}^\infty \frac{1}{2^n}|\pi_n(a) - \pi_n(b)|. \tag{9.72}$$

Exercise 9.11. Show d is a metric and that the Borel σ – algebra on (Q,d) is the same as the product σ – algebra.

Solution to Exercise (9.11). It is easily seen that d is a metric on Q which, by Eq. (9.72) is measurable relative to the product σ – algebra, $\mathcal{M}$.. Therefore, $\mathcal{M}$ contains all open balls and hence contains the Borel σ – algebra, $\mathcal{B}$. Conversely, since

$$|\pi_n(a) - \pi_n(b)| \le 2^n d(a,b),$$

each of the projection operators, $\pi_n : Q \to [0,1]$ is continuous. Therefore each π_n is $\mathcal{B}$ – measurable and hence $\mathcal{M} = \sigma(\{\pi_n\}_{n=1}^\infty) \subset \mathcal{B}$.

Theorem 9.64. *To every separable metric space (X,ρ), there exists a continuous injective map $G : X \to Q$ such that $G : X \to G(X) \subset Q$ is a homeomorphism. Moreover if the metric, ρ, is also complete, then $G(X)$ is a G_δ –set, i.e. the $G(X)$ is the countable intersection of open subsets of (Q,d). In short, any separable metrizable space X is homeomorphic to a subset of (Q,d) and if X is a Polish space then X is homeomorphic to a G_δ – subset of (Q,d).*

Proof. (This proof follows that in Rogers and Williams [44, Theorem 82.5 on p. 106.].) By replacing ρ by $\frac{\rho}{1+\rho}$ if necessary, we may assume that $0 \le \rho < 1$. Let $D = \{a_n\}_{n=1}^\infty$ be a countable dense subset of X and define

$$G(x) = (\rho(x,a_1), \rho(x,a_2), \rho(x,a_3), \dots) \in Q$$

and

$$\gamma(x,y) = d(G(x), G(y)) = \sum_{n=1}^\infty \frac{1}{2^n}|\rho(x,a_n) - \rho(y,a_n)|$$

for $x,y\in X$. To prove the first assertion, we must show G is injective and γ is a metric on X which is compatible with the topology determined by ρ.

If $G(x) = G(y)$, then $\rho(x,a) = \rho(y,a)$ for all $a \in D$. Since D is a dense subset of X, we may choose $\alpha_k \in D$ such that

$$0 = \lim_{k\to\infty} \rho(x,\alpha_k) = \lim_{k\to\infty}\rho(y,\alpha_k) = \rho(y,x)$$

and therefore $x = y$. A simple argument using the dominated convergence theorem shows $y \to \gamma(x,y)$ is ρ – continuous, i.e. $\gamma(x,y)$ is small if $\rho(x,y)$ is small. Conversely,

$$\begin{aligned}\rho(x,y) &\le \rho(x,a_n) + \rho(y,a_n) = 2\rho(x,a_n) + \rho(y,a_n) - \rho(x,a_n) \\ &\le 2\rho(x,a_n) + |\rho(x,a_n) - \rho(y,a_n)| \le 2\rho(x,a_n) + 2^n\gamma(x,y).\end{aligned}$$

Hence if $\varepsilon > 0$ is given, we may choose n so that $2\rho(x,a_n) < \varepsilon/2$ and so if $\gamma(x,y) < 2^{-(n+1)}\varepsilon$, it will follow that $\rho(x,y) < \varepsilon$. This shows $\tau_\gamma = \tau_\rho$. Since $G : (X,\gamma) \to (Q,d)$ is isometric, G is a homeomorphism.

Now suppose that (X,ρ) is a complete metric space. Let $S := G(X)$ and σ be the metric on S defined by $\sigma(G(x), G(y)) = \rho(x,y)$ for all $x,y \in X$. Then (S,σ) is a complete metric (being the isometric image of a complete metric space) and by what we have just prove, $\tau_\sigma = \tau_{d_S}$. Consequently, if $u \in S$ and $\varepsilon > 0$ is given, we may find $\delta'(\varepsilon)$ such that $B_\sigma(u, \delta'(\varepsilon)) \subset B_d(u,\varepsilon)$. Taking $\delta(\varepsilon) = \min(\delta'(\varepsilon), \varepsilon)$, we have $\operatorname{diam}_d(B_d(u,\delta(\varepsilon))) < \varepsilon$ and $\operatorname{diam}_\sigma(B_d(u,\delta(\varepsilon))) < \varepsilon$ where

$$\operatorname{diam}_\sigma(A) := \{\sup \sigma(u,v) : u,v \in A\} \text{ and}$$
$$\operatorname{diam}_d(A) := \{\sup d(u,v) : u,v \in A\}.$$

Let $\bar{S}$ denote the closure of S inside of (Q,d) and for each $n \in \mathbb{N}$ let

$$\mathcal{N}_n := \{N \in \tau_d : \operatorname{diam}_d(N) \vee \operatorname{diam}_\sigma(N \cap S) < 1/n\}$$

and let $U_n := \cup \mathcal{N}_n \in \tau_d$. From the previous paragraph, it follows that $S \subset U_n$ and therefore $S \subset \bar{S} \cap (\cap_{n=1}^\infty U_n)$.

Conversely if $u \in \bar{S} \cap (\cap_{n=1}^\infty U_n)$ and $n \in \mathbb{N}$, there exists $N_n \in \mathcal{N}_n$ such that $u \in N_n$. Moreover, since $N_1 \cap \dots \cap N_n$ is an open neighborhood of $u \in \bar{S}$, there exists $u_n \in N_1 \cap \dots \cap N_n \cap S$ for each $n \in \mathbb{N}$. From the definition of $\mathcal{N}_n$, we have $\lim_{n\to\infty} d(u,u_n) = 0$ and $\sigma(u_n,u_m) \le \max(n^{-1}, m^{-1}) \to 0$ as $m,n \to \infty$. Since (S,σ) is complete, it follows that $\{u_n\}_{n=1}^\infty$ is convergent in (S,σ) to some element $u_0 \in S$. Since (S,d_S) has the same topology as (S,σ) it follows that $d(u_n,u_0) \to 0$ as well and thus that $u = u_0 \in S$. We have now shown, $S = \bar{S} \cap (\cap_{n=1}^\infty U_n)$. This completes the proof because we may write $\bar{S} = \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ where $S_{1/n} := \{u \in Q : d(u,\bar{S}) < 1/n\}$ and therefore, $S = \left(\bigcap_{n=1}^\infty U_n\right) \cap \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ is a G_δ set. ∎

Corollary 9.65. *Every Polish space, X, with its Borel σ – algebra is a standard Borel space. Consequently any Borel subset of X is also a standard Borel space.*

Proof. Theorem 9.64 shows that X is homeomorphic to a measurable (in fact a G_δ) subset Q_0 of (Q,d) and hence $X \cong Q_0$. Since Q is a standard Borel space so is Q_0 and hence so is X. ∎

9.11 More Exercises

Exercise 9.12. Let $(X_j, \mathcal{M}_j, \mu_j)$ for $j = 1,2,3$ be σ – finite measure spaces. Let $F : (X_1 \times X_2) \times X_3 \to X_1 \times X_2 \times X_3$ be defined by

$$F((x_1,x_2),x_3) = (x_1,x_2,x_3).$$

1. Show F is $((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$ – measurable and F^{-1} is $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3)$ – measurable. That is
$$F : ((X_1 \times X_2) \times X_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3) \to (X_1 \times X_2 \times X_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$$
is a "measure theoretic isomorphism."
2. Let $\pi := F_*[(\mu_1 \otimes \mu_2) \otimes \mu_3]$, i.e. $\pi(A) = [(\mu_1 \otimes \mu_2) \otimes \mu_3](F^{-1}(A))$ for all $A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$. Then π is the unique measure on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ such that
$$\pi(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$
for all $A_i \in \mathcal{M}_i$. We will write $\pi := \mu_1 \otimes \mu_2 \otimes \mu_3$.
3. Let $f : X_1 \times X_2 \times X_3 \to [0,\infty]$ be a $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, \mathcal{B}_{\mathbb{R}})$ – measurable function. Verify the identity,
$$\int_{X_1 \times X_2 \times X_3} f d\pi = \int_{X_3} d\mu_3(x_3) \int_{X_2} d\mu_2(x_2) \int_{X_1} d\mu_1(x_1) f(x_1,x_2,x_3),$$
makes sense and is correct.
4. (Optional.) Also show the above identity holds for any one of the six possible orderings of the iterated integrals.

Exercise 9.13. Prove the second assertion of Theorem 9.20. That is show m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$. **Hint:** Look at the proof of Theorem 5.34.

Exercise 9.14. (Part of Folland Problem 2.46 on p. 69.) Let $X = [0,1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$ be the Borel σ – field on X, m be Lebesgue measure on $[0,1]$ and ν be counting measure, $\nu(A) = \#(A)$. Finally let $D = \{(x,x) \in X^2 : x \in X\}$ be the diagonal in X^2. Show

$$\int_X \left[\int_X 1_D(x,y) d\nu(y)\right] dm(x) \neq \int_X \left[\int_X 1_D(x,y) dm(x)\right] d\nu(y)$$

by explicitly computing both sides of this equation.

Exercise 9.15. Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

Exercise 9.16. Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$ should be $\mathcal{M} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ in this problem.)

Exercise 9.17. Folland Problem 2.55 on p. 77. (Explicit integrations.)

Exercise 9.18. Folland Problem 2.56 on p. 77. Let $f \in L^1((0,a), dm)$, $g(x) = \int_x^a \frac{f(t)}{t} dt$ for $x \in (0,a)$, show $g \in L^1((0,a), dm)$ and

$$\int_0^a g(x) dx = \int_0^a f(t) dt.$$

Exercise 9.19. Show $\int_0^\infty \left|\frac{\sin x}{x}\right| dm(x) = \infty$. So $\frac{\sin x}{x} \notin L^1([0,\infty), m)$ and $\int_0^\infty \frac{\sin x}{x} dm(x)$ is not defined as a Lebesgue integral.

Exercise 9.20. Folland Problem 2.57 on p. 77.

Exercise 9.21. Folland Problem 2.58 on p. 77.

Exercise 9.22. Folland Problem 2.60 on p. 77. Properties of the Γ – function.

Exercise 9.23. Folland Problem 2.61 on p. 77. Fractional integration.

Exercise 9.24. Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on S^{n-1}.

Exercise 9.25. Folland Problem 2.64 on p. 80. On the integrability of $|x|^a \left|\log |x|\right|^b$ for x near 0 and x near ∞ in $\mathbb{R}^n$.

Exercise 9.26. Show, using Problem 9.24 that

$$\int_{S^{d-1}} \omega_i \omega_j d\sigma(\omega) = \frac{1}{d}\delta_{ij}\sigma\left(S^{d-1}\right).$$

Hint: show $\int_{S^{d-1}} \omega_i^2 d\sigma(\omega)$ is independent of i and therefore

$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d}\sum_{j=1}^{d}\int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

10

Independence

As usual, $(\Omega, \mathcal{B}, P)$ will be some fixed probability space. Recall that for $A, B \in \mathcal{B}$ with $P(B) > 0$ we let

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

which is to be read as; **the probability of A given B.**

Definition 10.1. *We say that A is independent of B is $P(A|B) = P(A)$ or equivalently that*

$$P(A \cap B) = P(A) P(B).$$

We further say a finite sequence of collection of sets, $\{\mathcal{C}_i\}_{i=1}^n$, are independent if

$$P(\cap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

for all $A_i \in \mathcal{C}_i$ and $J \subset \{1, 2, \ldots, n\}$.

10.1 Basic Properties of Independence

If $\{\mathcal{C}_i\}_{i=1}^n$, are independent classes then so are $\{\mathcal{C}_i \cup \{\Omega\}\}_{i=1}^n$. Moreover, if we assume that $\Omega \in \mathcal{C}_i$ for each i, then $\{\mathcal{C}_i\}_{i=1}^n$, are independent iff

$$P\left(\cap_{j=1}^n A_j\right) = \prod_{j=1}^n P(A_j) \text{ for all } (A_1, \ldots, A_n) \in \mathcal{C}_1 \times \cdots \times \mathcal{C}_n.$$

Theorem 10.2. *Suppose that $\{\mathcal{C}_i\}_{i=1}^n$ is a finite sequence of independent π – classes. Then $\{\sigma(\mathcal{C}_i)\}_{i=1}^n$ are also independent.*

Proof. As mentioned above, we may always assume without loss of generality that $\Omega \in \mathcal{C}_i$. Fix, $A_j \in \mathcal{C}_j$ for $j = 2, 3, \ldots, n$. We will begin by showing that

$$Q(A) := P(A \cap A_2 \cap \cdots \cap A_n) = P(A) P(A_2) \ldots P(A_n) \text{ for all } A \in \sigma(\mathcal{C}_1). \tag{10.1}$$

Since $Q(\cdot)$ and $P(A_2) \ldots P(A_n) P(\cdot)$ are both finite measures agreeing on Ω and A in the π – system $\mathcal{C}_1$, Eq. (10.1) is a direct consequence of Proposition 5.15. Since $(A_2, \ldots, A_n) \in \mathcal{C}_2 \times \cdots \times \mathcal{C}_n$ were arbitrary we may now conclude that $\sigma(\mathcal{C}_1), \mathcal{C}_2, \ldots, \mathcal{C}_n$ are independent.

By applying the result we have just proved to the sequence, $\mathcal{C}_2, \ldots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ shows that $\sigma(\mathcal{C}_2), \mathcal{C}_3, \ldots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ are independent. Similarly we show inductively that

$$\sigma(\mathcal{C}_j), \mathcal{C}_{j+1}, \ldots, \mathcal{C}_n, \sigma(\mathcal{C}_1), \ldots, \sigma(\mathcal{C}_{j-1})$$

are independent for each $j = 1, 2, \ldots, n$. The desired result occurs at $j = n$. ■

Definition 10.3. *Let $(\Omega, , \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i : \Omega \to S_i$ be a measurable map for $1 \le i \le n$. The maps $\{Y_i\}_{i=1}^n$ are P - independent iff $\{\mathcal{C}_i\}_{i=1}^n$ are P – independent, where $\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i) = \sigma(Y_i) \subset \mathcal{B}$ for $1 \le i \le n$.*

Theorem 10.4 (Independence and Product Measures). *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i : \Omega \to S_i$ be a measurable map for $1 \le i \le n$. Further let $\mu_i := P \circ Y_i^{-1} = \mathrm{Law}_P(Y_i)$. Then $\{Y_i\}_{i=1}^n$ are independent iff*

$$\mathrm{Law}_P(Y_1, \ldots, Y_n) = \mu_1 \otimes \cdots \otimes \mu_n,$$

where $(Y_1, \ldots, Y_n) : \Omega \to S_1 \times \cdots \times S_n$ and

$$\mathrm{Law}_P(Y_1, \ldots, Y_n) = P \circ (Y_1, \ldots, Y_n)^{-1} : \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n \to [0, 1]$$

is the joint law of $Y_1, \ldots, Y_n$.

Proof. Recall that the general element of $\mathcal{C}_i$ is of the form $A_i = Y_i^{-1}(B_i)$ with $B_i \in \mathcal{S}_i$. Therefore for $A_i = Y_i^{-1}(B_i) \in \mathcal{C}_i$ we have

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P((Y_1, \ldots, Y_n) \in B_1 \times \cdots \times B_n) \\ &= ((Y_1, \ldots, Y_n)_* P)(B_1 \times \cdots \times B_n). \end{aligned}$$

If $(Y_1, \ldots, Y_n)_* P = \mu_1 \otimes \cdots \otimes \mu_n$ it follows that

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= \mu_1 \otimes \cdots \otimes \mu_n (B_1 \times \cdots \times B_n) \\ &= \mu_1(B_1) \cdots \mu(B_n) = P(Y_1 \in B_1) \cdots P(Y_n \in B_n) \\ &= P(A_1) \ldots P(A_n) \end{aligned}$$

and therefore $\{\mathcal{C}_i\}$ are P – independent and hence $\{Y_i\}$ are P – independent.

Conversely if $\{Y_i\}$ are P – independent, i.e. $\{\mathcal{C}_i\}$ are P – independent, then

$$\begin{aligned} P\left((Y_1,\dots,Y_n)\in B_1\times\dots\times B_n\right) &= P\left(A_1\cap\dots\cap A_n\right) \\ &= P\left(A_1\right)\dots P\left(A_n\right) \\ &= P\left(Y_1\in B_1\right)\cdots P\left(Y_n\in B_n\right) \\ &= \mu_1\left(B_1\right)\cdots\mu\left(B_n\right) \\ &= \mu_1\otimes\dots\otimes\mu_n\left(B_1\times\dots\times B_n\right). \end{aligned}$$

Since

$$\pi := \{B_1\times\dots\times B_n : B_i\in\mathcal{S}_i \text{ for } 1\le i\le n\}$$

is a π – system which generates $\mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$ and

$$(Y_1,\dots,Y_n)_* P = \mu_1\otimes\dots\otimes\mu_n \text{ on } \pi,$$

it follows that $(Y_1,\dots,Y_n)_* P = \mu_1\otimes\dots\otimes\mu_n$ on all of $\mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$. ■

Remark 10.5. When have a collection of not necessarily independent random functions, $Y_i : \Omega\to S_i$, like in Theorem 10.4 it is **not** in general possible to recover the joint distribution, $\pi := \mathrm{Law}_P(Y_1,\dots,Y_n)$, from the individual distributions, $\mu_i = \mathrm{Law}_P(Y_i)$ for all $1\le i\le n$. For example suppose that $S_i = \mathbb{R}$ for $i=1,2$. μ is a probability measure on $(\mathbb{R},\mathcal{B}_\mathbb{R})$, and (Y_1,Y_2) have joint distribution, π, given by,

$$\pi(C) = \int_\mathbb{R} 1_C(x,x)\,d\mu(x) \text{ for all } C\in\mathcal{B}_\mathbb{R}.$$

If we let $\mu_i = \mathrm{Law}_P(Y_i)$, then for all $A\in\mathcal{B}_\mathbb{R}$ we have

$$\begin{aligned} \mu_1(A) &= P(Y_1\in A) = P((Y_1,Y_2)\in A\times\mathbb{R}) \\ &= \pi(A\times\mathbb{R}) = \int_\mathbb{R} 1_{A\times\mathbb{R}}(x,x)\,d\mu(x) = \mu(A). \end{aligned}$$

Similarly we show that $\mu_2=\mu$. On the other hand if μ is not concentrated on one point, $\mu\otimes\mu$ is another probability measure on $\left(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2}\right)$ with the same **marginals** as π, i.e. $\pi(A\times\mathbb{R}) = \mu(A) = \pi(\mathbb{R}\times A)$ for all $A\in\mathcal{B}_\mathbb{R}$.

Lemma 10.6. *Let $(\Omega,,\mathcal{B},P)$ be a probability space, $\{(S_i,\mathcal{S}_i)\}_{i=1}^n$ and $\{(T_i,\mathcal{T}_i)\}_{i=1}^n$ be two collection of measurable spaces, $F_i : S_i\to T_i$ be a measurable map for each i and $Y_i:\Omega\to S_i$ be a collection of P – independent measurable maps. Then $\{F_i\circ Y_i\}_{i=1}^n$ are also P – independent.*

Proof. Notice that

$$\sigma(F_i\circ Y_i) = (F_i\circ Y_i)^{-1}(\mathcal{T}_i) = Y_i^{-1}\left(F_i^{-1}(\mathcal{T}_i)\right)\subset Y_i^{-1}(\mathcal{S}_i) = \mathcal{C}_i.$$

The fact that $\{\sigma(F_i\circ Y_i)\}_{i=1}^n$ is independent now follows easily from the assumption that $\{\mathcal{C}_i\}$ are P – independent. ■

Example 10.7. If $\Omega := \prod_{i=1}^n S_i$, $\mathcal{B} := \mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$, $Y_i(s_1,\dots,s_n) = s_i$ for all $(s_1,\dots,s_n)\in\Omega$, and $\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i)$ for all i. Then the probability measures, P, on $(\Omega,\mathcal{B})$ for which $\{\mathcal{C}_i\}_{i=1}^n$ are independent are precisely the product measures, $P=\mu_1\otimes\dots\otimes\mu_n$ where μ_i is a probability measure on $(S_i,\mathcal{S}_i)$ for $1\le i\le n$. Notice that in this setting,

$$\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i) = \{S_1\times\dots\times S_{i-1}\times B\times S_{i+1}\times\dots\times S_n : B\in\mathcal{S}_i\}\subset\mathcal{B}.$$

Proposition 10.8. *Suppose that $(\Omega,\mathcal{B},P)$ is a probability space and $\{Z_j\}_{j=1}^n$ are independent integrable random variables. Then $\prod_{j=1}^n Z_j$ is also integrable and*

$$\mathbb{E}\left[\prod_{j=1}^n Z_j\right] = \prod_{j=1}^n \mathbb{E}Z_j.$$

Proof. Let $\mu_j := P\circ Z_j^{-1} : \mathcal{B}_\mathbb{R}\to[0,1]$ be the law of Z_j for each j. Then we know $(Z_1,\dots,Z_n)_* P = \mu_1\otimes\dots\otimes\mu_n$. Therefore by Example 7.52 and Tonelli's theorem,

$$\begin{aligned} \mathbb{E}\left[\prod_{j=1}^n |Z_j|\right] &= \int_{\mathbb{R}^n}\left[\prod_{j=1}^n |z_j|\right] d\left(\otimes_{j=1}^n\mu_j\right)(z) \\ &= \prod_{j=1}^n\int_{\mathbb{R}^n} |z_j|\,d\mu_j(z_j) = \prod_{j=1}^n \mathbb{E}|Z_j| < \infty \end{aligned}$$

which shows that $\prod_{j=1}^n Z_j$ is integrable. Thus again by Example 7.52 and Fubini's theorem,

$$\begin{aligned} \mathbb{E}\left[\prod_{j=1}^n Z_j\right] &= \int_{\mathbb{R}^n}\left[\prod_{j=1}^n z_j\right] d\left(\otimes_{j=1}^n\mu_j\right)(z) \\ &= \prod_{j=1}^n\int_\mathbb{R} z_j\,d\mu_j(z_j) = \prod_{j=1}^n \mathbb{E}Z_j. \end{aligned}$$

■

Theorem 10.9. *Let $(\Omega,,\mathcal{B},P)$ be a probability space, $\{(S_i,\mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i:\Omega\to S_i$ be a measurable map for $1\le i\le n$. Further let $\mu_i := P\circ Y_i^{-1} = \mathrm{Law}_P(Y_i)$ and $\pi := P\circ(Y_1,\dots,Y_n)^{-1} : \mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$ be the joint distribution of*

$$(Y_1,\dots,Y_n):\Omega\to S_1\times\dots\times S_n.$$

Then the following are equivalent,

1. *$\{Y_i\}_{i=1}^n$ are independent,*
2. *$\pi = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$*
3. *for all bounded measurable functions, $f : (S_1 \times \cdots \times S_n, \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,*
$$\mathbb{E} f(Y_1, \dots, Y_n) = \int_{S_1 \times \cdots \times S_n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n), \qquad (10.2)$$
(where the integrals may be taken in any order),
4. *$\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)\right] = \prod_{i=1}^n \mathbb{E}\left[f_i(Y_i)\right]$ for all bounded (or non-negative) measurable functions, $f_i : S_i \to \mathbb{R}$ or $\mathbb{C}$.*

Proof. ($1 \iff 2$) has already been proved in Theorem 10.4. The fact that ($2. \implies 3.$) now follows from Exercise 7.11 and Fubini's theorem. Similarly, ($3. \implies 4.$) follows from Exercise 7.11 and Fubini's theorem after taking $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$. Lastly for ($4. \implies 1.$), let $A_i \in \mathcal{S}_i$ and take $f_i := 1_{A_i}$ in 4. to learn,

$$P\left(\cap_{i=1}^n \{Y_i \in A_i\}\right) = \mathbb{E}\left[\prod_{i=1}^n 1_{A_i}(Y_i)\right] = \prod_{i=1}^n \mathbb{E}\left[1_{A_i}(Y_i)\right] = \prod_{i=1}^n P(Y_i \in A_i)$$

which shows that the $\{Y_i\}_{i=1}^n$ are independent. ■

Corollary 10.10. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space and $\{Y_j : \Omega \to \mathbb{R}\}_{j=1}^n$ is a sequence of random variables with countable ranges, say $\Lambda \subset \mathbb{R}$. Then $\{Y_j\}_{j=1}^n$ are independent iff*

$$P\left(\cap_{j=1}^n \{Y_j = y_j\}\right) = \prod_{j=1}^n P(Y_j = y_j) \qquad (10.3)$$

for all choices of $y_1, \dots, y_n \in \Lambda$.

Proof. If the $\{Y_j\}$ are independent then clearly Eq. (10.3) holds by definition as $\{Y_j = y_j\} \in Y_j^{-1}(\mathcal{B}_{\mathbb{R}})$. Conversely if Eq. (10.3) holds and $f_i : \mathbb{R} \to [0, \infty)$ are measurable functions then,

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)\right] &= \sum_{y_1, \dots, y_n \in \Lambda} \prod_{i=1}^n f_i(y_i) \cdot P\left(\cap_{j=1}^n \{Y_j = y_j\}\right) \\
&= \sum_{y_1, \dots, y_n \in \Lambda} \prod_{i=1}^n f_i(y_i) \cdot \prod_{j=1}^n P(Y_j = y_j) \\
&= \prod_{i=1}^n \sum_{y_i \in \Lambda} f_i(y_i) \cdot P(Y_j = y_j) \\
&= \prod_{i=1}^n \mathbb{E}\left[f_i(Y_i)\right]
\end{aligned}$$

wherein we have used Tonelli's theorem for sum in the third equality. It now follows that $\{Y_i\}$ are independent using item 4. of Theorem 10.9. ■

Exercise 10.1. Suppose that $\Omega = (0,1]$, $\mathcal{B} = \mathcal{B}_{(0,1]}$, and $P = m$ is Lebesgue measure on $\mathcal{B}$. Let $Y_i(\omega) := \omega_i$ be the i^{th} – digit in the base two expansion of ω. To be more precise, the $Y_i(\omega) \in \{0,1\}$ is chosen so that

$$\omega = \sum_{i=1}^{\infty} Y_i(\omega)\, 2^{-i} \text{ for all } \omega_i \in \{0,1\}.$$

As long as $\omega \neq k2^{-n}$ for some $0 < k \le n$, the above equation uniquely determines the $\{Y_i(\omega)\}$. Owing to the fact that $\sum_{l=n+1}^{\infty} 2^{-l} = 2^{-n}$, if $\omega = k2^{-n}$, there is some ambiguity in the definitions of the $Y_i(\omega)$ for large i which you may resolve anyway you choose. Show the random variables, $\{Y_i\}_{i=1}^n$, are i.i.d. for each $n \in \mathbb{N}$ with $P(Y_i = 1) = 1/2 = P(Y_i = 0)$ for all i.

Hint: the idea is that knowledge of $(Y_1(\omega), \dots, Y_n(\omega))$ is equivalent to knowing for which $k \in \mathbb{N}_0 \cap [0, 2^n)$ that $\omega \in (2^{-n}k, 2^{-n}(k+1)]$ and that this knowledge in no way helps you predict the value of $Y_{n+1}(\omega)$. More formally, you might start by showing,

$$P\left(\{Y_{n+1} = 1\} | (2^{-n}k, 2^{-n}(k+1)]\right) = \frac{1}{2} = P\left(\{Y_{n+1} = 0\} | (2^{-n}k, 2^{-n}(k+1)]\right).$$

See Section 10.9 if you need some more help with this exercise.

Exercise 10.2. Let X, Y be two random variables on $(\Omega, \mathcal{B}, P)$.

1. Show that X and Y are independent iff $\mathrm{Cov}(f(X), g(Y)) = 0$ (i.e. $f(X)$ and $g(Y)$ are **uncorrelated**) for bounded measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$.
2. If $X, Y \in L^2(P)$ and X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.
3. Show by example that if $X, Y \in L^2(P)$ and $\mathrm{Cov}(X, Y) = 0$ does not necessarily imply that X and Y are independent. **Hint:** try taking $(X, Y) = (X, ZX)$ where X and Z are independent simple random variables such that $\mathbb{E}Z = 0$ similar to Remark 9.40.

Solution to Exercise (10.2). 1. Since

$$\mathrm{Cov}(f(X), g(Y)) = \mathbb{E}\left[f(X)\, g(Y)\right] - \mathbb{E}\left[f(X)\right] \mathbb{E}\left[g(Y)\right]$$

it follows that $\mathrm{Cov}(f(X), g(Y)) = 0$ iff

$$\mathbb{E}\left[f(X)\, g(Y)\right] = \mathbb{E}\left[f(X)\right] \mathbb{E}\left[g(Y)\right]$$

from which item 1. easily follows.

2. Let $f_M(x) = x 1_{|x| \le M}$, then by independence,

$$\mathbb{E}\left[f_M(X) f_M(Y)\right] = \mathbb{E}\left[f_M(X)\right] \mathbb{E}\left[f_M(Y)\right]. \tag{10.4}$$

Since

$$|f_M(X) f_M(Y)| \le |XY| \le \frac{1}{2}\left(X^2+Y^2\right) \in L^1(P),$$
$$|f_M(X)| \le |X| \le \frac{1}{2}\left(1+X^2\right) \in L^1(P), \text{ and}$$
$$|f_M(Y)| \le |Y| \le \frac{1}{2}\left(1+Y^2\right) \in L^1(P),$$

we may use the DCT three times to pass to the limit as $M \to \infty$ in Eq. (10.4) to learn that $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$, i.e. $\operatorname{Cov}(X,Y) = 0$.

3. Let X and Z be independent with $P(Z = \pm 1) = \frac{1}{2}$ and take $Y = XZ$. Then $\mathbb{E}Z = 0$ and

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \mathbb{E}\left[X^2 Z\right] - \mathbb{E}[X]\,\mathbb{E}[XZ] \\ &= \mathbb{E}\left[X^2\right] \cdot \mathbb{E}Z - \mathbb{E}[X]\,\mathbb{E}[X]\,\mathbb{E}Z = 0.\end{aligned}$$

On the other hand it should be intuitively clear that X and Y are not independent since knowledge of X typically will give some information about Y. To verify this assertion let us suppose that X is a discrete random variable with $P(X=0) = 0$. Then

$$P(X = x, Y = y) = P(X = x, xZ = y) = P(X = x) \cdot P(X = y/x)$$

while

$$P(X = x)\,P(Y = y) = P(X = x) \cdot P(XZ = y).$$

Thus for X and Y to be independent we would have to have,

$$P(xX = y) = P(XZ = y) \text{ for all } x, y.$$

This is clearly not going to be true in general. For example, suppose that $P(X=1) = \frac{1}{2} = P(X=0)$. Taking $x = y = 1$ in the previously displayed equation would imply

$$\frac{1}{2} = P(X=1) = P(XZ=1) = P(X=1, Z=1) = P(X=1)\,P(Z=1) = \frac{1}{4}$$

which is false.

Exercise 10.3 (A correlation inequality). Suppose that X is a random variable and $f, g : \mathbb{R} \to \mathbb{R}$ are two increasing functions such that both $f(X)$ and $g(X)$ are square integrable, i.e. $\mathbb{E}|f(X)|^2 + \mathbb{E}|g(X)|^2 < \infty$. Show $\operatorname{Cov}(f(X), g(X)) \ge 0$. **Hint:** let Y be another random variable which has the same law as X and is independent of X. Then consider

$$\mathbb{E}\left[(f(Y) - f(X)) \cdot (g(Y) - g(X))\right].$$

Let us now specialize to the case where $S_i = \mathbb{R}^{m_i}$ and $\mathcal{S}_i = \mathcal{B}_{\mathbb{R}^{m_i}}$ for some $m_i \in \mathbb{N}$.

Theorem 10.11. *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $m_j \in \mathbb{N}$, $S_j = \mathbb{R}^{m_j}$, $\mathcal{S}_j = \mathcal{B}_{\mathbb{R}^{m_j}}$, $Y_j : \Omega \to S_j$ be random vectors, and $\mu_j := \mathrm{Law}_P(Y_j) = P \circ Y_j^{-1} : \mathcal{S}_j \to [0,1]$ for $1 \le j \le n$. The the following are equivalent;*

1. *$\{Y_j\}_{j=1}^n$ are independent,*
2. $\mathrm{Law}_P(Y_1, \dots, Y_n) = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n$
3. *for all bounded measurable functions, $f : (S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,*
$$\mathbb{E} f(Y_1, \dots, Y_n) = \int_{S_1 \times \dots \times S_n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n), \tag{10.5}$$
(where the integrals may be taken in any order),
4. *$\mathbb{E}\left[\prod_{j=1}^n f_j(Y_j)\right] = \prod_{j=1}^n \mathbb{E}\left[f_j(Y_j)\right]$ for all bounded (or non-negative) measurable functions, $f_j : S_j \to \mathbb{R}$ or $\mathbb{C}$.*
5. *$P\left(\cap_{j=1}^n \{Y_j \le y_j\}\right) = \prod_{j=1}^n P(\{Y_j \le y_j\})$ for all $y_j \in S_j$, where we say that $Y_j \le y_j$ iff $(Y_j)_k \le (y_j)_k$ for $1 \le k \le m_j$.*
6. *$\mathbb{E}\left[\prod_{j=1}^n f_j(Y_j)\right] = \prod_{j=1}^n \mathbb{E}\left[f_j(Y_j)\right]$ for all $f_j \in C_c(S_j, \mathbb{R})$,*
7. *$\mathbb{E}\left[e^{i \sum_{j=1}^n \lambda_j \cdot Y_j}\right] = \prod_{j=1}^n \mathbb{E}\left[e^{i \lambda_j \cdot Y_j}\right]$ for all $\lambda_j \in S_j = \mathbb{R}^{m_j}$.*

Proof. The equivalence of 1. – 4. has already been proved in Theorem 10.9. It is also clear that item 4. implies both or items 5. –7. upon noting that item 5. may be written as,

$$\mathbb{E}\left[\prod_{j=1}^n 1_{(-\infty, y_j]}(Y_j)\right] = \prod_{j=1}^n \mathbb{E}\left[1_{(-\infty, y_j]}(Y_j)\right]$$

where $(-\infty, y_j] := (-\infty, (y_j)_1] \times \dots \times (-\infty, (y_j)_{m_j}]$. The proofs that either 5. or 6. or 7. implies item 3. is a simple application of the multiplicative system theorem in the form of either Corollary 8.3 or Corollary 8.8. In each case, let $\mathbb{H}$ denote the linear space of bounded measurable functions such that Eq. (10.5) holds. To complete the proof I will simply give you the multiplicative system, $\mathbb{M}$, to use in each of the cases. To describe $\mathbb{M}$, let $N = m_1 + \dots + m_n$ and

$$\begin{aligned} y &= (y_1, \dots, y_n) = \left(y^1, y^2, \dots, y^N\right) \in \mathbb{R}^N \text{ and} \\ \lambda &= (\lambda_1, \dots, \lambda_n) = \left(\lambda^1, \lambda^2, \dots, \lambda^N\right) \in \mathbb{R}^N \end{aligned}$$

For showing 5. $\implies$ 3.take $\mathbb{M} = \left\{1_{(-\infty, y]} : y \in \mathbb{R}^N\right\}$.

For showing 6. $\implies$ 3. take $\mathbb{M}$ to be a those functions on $\mathbb{R}^N$ which are of the form, $f(y) = \prod_{l=1}^N f_l(y^l)$ with each $f_l \in C_c(\mathbb{R})$.

For showing 7. $\implies$ 3. take $\mathbb{M}$ to be the functions of the form,

$$f(y) = \exp\left(i \sum_{j=1}^n \lambda_j \cdot y_j\right) = \exp(i\lambda \cdot y).$$

■

Definition 10.12. *A collection of subsets of $\mathcal{B}$, $\{\mathcal{C}_t\}_{t\in T}$ is said to be independent iff $\{\mathcal{C}_t\}_{t\in\Lambda}$ are independent for all finite subsets, $\Lambda \subset T$. More explicitly, we are requiring*

$$P(\cap_{t\in\Lambda} A_t) = \prod_{t\in\Lambda} P(A_t)$$

whenever Λ is a finite subset of T and $A_t \in \mathcal{C}_t$ for all $t \in \Lambda$.

Corollary 10.13. *If $\{\mathcal{C}_t\}_{t\in T}$ is a collection of independent classes such that each $\mathcal{C}_t$ is a π – system, then $\{\sigma(\mathcal{C}_t)\}_{t\in T}$ are independent as well.*

Definition 10.14. *A collections of random variables, $\{X_t : t \in T\}$ are* ***independent*** *iff $\{\sigma(X_t) : t \in T\}$ are independent.*

Example 10.15. Suppose that $\{\mu_n\}_{n=1}^\infty$ is any sequence of probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Let $\Omega = \mathbb{R}^\mathbb{N}$, $\mathcal{B} := \otimes_{n=1}^\infty \mathcal{B}_\mathbb{R}$ be the product σ – algebra on Ω, and $P := \otimes_{n=1}^\infty \mu_n$ be the product measure. Then the random variables, $\{Y_n\}_{n=1}^\infty$ defined by $Y_n(\omega) = \omega_n$ for all $\omega \in \Omega$ are independent with $\mathrm{Law}_P(Y_n) = \mu_n$ for each n.

Lemma 10.16 (Independence of groupings). *Suppose that $\{\mathcal{B}_t : t \in T\}$ is an independent family of σ – fields. Suppose further that $\{T_s\}_{s\in S}$ is a partition of T (i.e. $T = \sum_{s\in S} T_s$) and let*

$$\mathcal{B}_{T_s} = \vee_{t\in T_s} \mathcal{B}_t = \sigma(\cup_{t\in T_s} \mathcal{B}_t).$$

Then $\{\mathcal{B}_{T_s}\}_{s\in S}$ is again independent family of σ fields.

Proof. Let

$$\mathcal{C}_s = \{\cap_{\alpha\in K} B_\alpha : B_\alpha \in \mathcal{B}_\alpha,\ K \subset\subset T_s\}.$$

It is now easily checked that $\mathcal{B}_{T_s} = \sigma(\mathcal{C}_s)$ and that $\{\mathcal{C}_s\}_{s\in S}$ is an independent family of π – systems. Therefore $\{\mathcal{B}_{T_s}\}_{s\in S}$ is an independent family of σ – algebras by Corollary 10.13. ■

Corollary 10.17. *Suppose that $\{Y_n\}_{n=1}^\infty$ is a sequence of independent random variables (or vectors) and $\Lambda_1, \dots, \Lambda_m$ is a collection of pairwise disjoint subsets of $\mathbb{N}$. Further suppose that $f_i : \mathbb{R}^{\Lambda_i} \to \mathbb{R}$ is a measurable function for each $1 \le i \le m$, then $Z_i := f_i\left(\{Y_l\}_{l\in\Lambda_i}\right)$ is again a collection of independent random variables.*

Proof. Notice that $\sigma(Z_i) \subset \sigma\left(\{Y_l\}_{l\in\Lambda_i}\right) = \sigma(\cup_{l\in\Lambda_i} \sigma(Y_l))$. Since $\{\sigma(Y_l)\}_{l=1}^\infty$ are independent by assumption, it follows from Lemma 10.16 that $\left\{\sigma\left(\{Y_l\}_{l\in\Lambda_i}\right)\right\}_{i=1}^m$ are independent and therefore so is $\{\sigma(Z_i)\}_{i=1}^m$, i.e. $\{Z_i\}_{i=1}^m$ are independent. ■

Definition 10.18 (i.i.d.). *A sequences of random variables, $\{X_n\}_{n=1}^\infty$, on a probability space, $(\Omega, \mathcal{B}, P)$, are* ***i.i.d. (= independent and identically distributed)*** *if they are independent and $(X_n)_* P = (X_k)_* P$ for all k, n. That is we should have*

$$P(X_n \in A) = P(X_k \in A) \text{ for all } k, n \in \mathbb{N} \text{ and } A \in \mathcal{B}_\mathbb{R}.$$

Observe that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables iff

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{j=1}^n P(X_i \in A_i) = \prod_{j=1}^n P(X_1 \in A_i) = \prod_{j=1}^n \mu(A_i) \tag{10.6}$$

where $\mu = (X_1)_* P$. The identity in Eq. (10.6) is to hold for all $n \in \mathbb{N}$ and all $A_i \in \mathcal{B}_\mathbb{R}$. If we choose $\mu_n = \mu$ in Example 10.15, the $\{Y_n\}_{n=1}^\infty$ there are i.i.d. with $\mathrm{Law}_P(Y_n) = P \circ Y_n^{-1} = \mu$ for all $n \in \mathbb{N}$.

The following theorem follows immediately from the definitions and Theorem 10.11.

Theorem 10.19. *Let $\mathbb{X} := \{X_t : t \in T\}$ be a collection of random variables. Then the following are equivalent:*

1. *The collection $\mathbb{X}$ is independent,*
2. $$P(\cap_{t\in\Lambda} \{X_t \in A_t\}) = \prod_{t\in\Lambda} P(X_t \in A_t)$$
   for all finite subsets, $\Lambda \subset T$, and all $\{A_t\}_{t\in\Lambda} \subset \mathcal{B}_\mathbb{R}$.
3. $$P(\cap_{t\in\Lambda} \{X_t \le x_t\}) = \prod_{t\in\Lambda} P(X_t \le x_t)$$
   for all finite subsets, $\Lambda \subset T$, and all $\{x_t\}_{t\in\Lambda} \subset \mathbb{R}$.

4. *For all $\Gamma \subset\subset T$ and $f_t : \mathbb{R}^n \to \mathbb{R}$ which are bounded an measurable for all $t \in \Gamma$,*

$$\mathbb{E}\left[\prod_{t\in\Gamma} f_t(X_t)\right] = \prod_{t\in\Gamma} \mathbb{E} f_t(X_t) = \int_{\mathbb{R}^\Gamma} \prod_{t\in\Gamma} f_t(x_t) \prod_{t\in\Gamma} d\mu_t(x_t).$$

5. $\mathbb{E}\left[\prod_{t\in\Gamma} \exp\left(e^{i\lambda_t \cdot X_t}\right)\right] = \prod_{t\in\Gamma} \hat{\mu}_t(\lambda).$
6. *For all $\Gamma \subset\subset T$ and $f : (\mathbb{R}^n)^\Gamma \to \mathbb{R}$,*

$$\mathbb{E}[f(X_\Gamma)] = \int_{(\mathbb{R}^n)^\Gamma} f(x) \prod_{t\in\Gamma} d\mu_t(x_t).$$

7. *For all $\Gamma \subset\subset T$, $\mathrm{Law}_P(X_\Gamma) = \otimes_{t\in\Gamma}\mu_t$.*
8. $\mathrm{Law}_P(X) = \otimes_{t\in T}\mu_t$.

Moreover, if $\mathcal{B}_t$ is a sub-σ - algebra of $\mathcal{B}$ for $t \in T$, then $\{\mathcal{B}_t\}_{t\in T}$ are independent iff for all $\Gamma \subset\subset T$,

$$\mathbb{E}\left[\prod_{t\in\Gamma} X_t\right] = \prod_{t\in\Gamma} \mathbb{E} X_t \text{ for all } X_t \in L^\infty(\Omega, \mathcal{B}_t, P).$$

Proof. The equivalence of 1. and 2. follows almost immediately form the definition of independence and the fact that $\sigma(X_t) = \{\{X_t \in A\} : A \in \mathcal{B}_\mathbb{R}\}$. Clearly 2. implies 3. holds. Finally, 3. implies 2. is an application of Corollary 10.13 with $\mathcal{C}_t := \{\{X_t \le a\} : a \in \mathbb{R}\}$ and making use the observations that $\mathcal{C}_t$ is a π − system for all t and that $\sigma(\mathcal{C}_t) = \sigma(X_t)$. The remaining equivalence are also easy to check. ■

Definition 10.20 (Conditional Independence). *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $\mathcal{B}_i \subset \mathcal{B}$ be a sub-sigma algebra of $\mathcal{B}$ for $i = 1, 2, 3$. We say that $\mathcal{B}_1$ is **independent of** $\mathcal{B}_3$ **conditioned** on $\mathcal{B}_2$ (written $\mathcal{B}_1 \overset{\mathcal{B}_2}{\perp\!\!\!\perp} \mathcal{B}_3$) provided,*

$$P(A \cap B|\mathcal{B}_2) = P(A|\mathcal{B}_2) \cdot P(B|\mathcal{B}_2) \ \text{a.s.}$$

for all $A \in \mathcal{B}_1$ and $B \in \mathcal{B}_3$. This can be equivalently stated as

$$\mathbb{E}(f \cdot g|\mathcal{B}_2) = \mathbb{E}(f|\mathcal{B}_2) \cdot \mathbb{E}(g|\mathcal{B}_2) \ \text{a.s.}$$

*for all $f \in (\mathcal{B}_1)_b$ and $g \in (\mathcal{B}_3)_b$, where $\mathcal{B}_b$ denotes the **bounded** B − **measurable functions.** If X, Y, Z are measurable functions on $(\Omega, \mathcal{B})$, we say that X is independent of Z conditioned on Y (written as $X \overset{Y}{\perp\!\!\!\perp} Z$) provided $\sigma(X) \overset{\sigma(Y)}{\perp\!\!\!\perp} \sigma(Z)$.*

Example 10.21. Let X and Y be two i.i.d. random variables such that $P(X = 1) = 1/2 = P(Y = 1)$ and $P(X = 2) = 1/2 = P(Y = 2)$. Then

$$\mathbb{E}[Y|X = Y] = \mathbb{E}[X|X = Y] = \frac{\frac{1}{4}(1+2)}{\frac{1}{4}+\frac{1}{4}} = \frac{3}{2}$$

and

$$\mathbb{E}[XY|X = Y] = \frac{\frac{1}{4}(1+4)}{\frac{1}{4}+\frac{1}{4}} = \frac{5}{4}.$$

Notice that

$$\mathbb{E}[XY|X = Y] = \frac{5}{4} \ne \frac{9}{4} = \mathbb{E}[Y|X = Y] \cdot \mathbb{E}[X|X = Y].$$

So independence does not necessarily imply conditional independence!

See Exercise 10.6 and Theorem 17.4 for a couple more examples involving conditional independence.

Exercise 10.4. Suppose $\mathbb{M}_i \subset (\mathcal{B}_i)_b$ for $i = 1$ and $i = 3$ are multiplicative systems such that $\mathcal{B}_i = \sigma(\mathbb{M}_i)$. Show $\mathcal{B}_1 \overset{\mathcal{B}_2}{\perp\!\!\!\perp} \mathcal{B}_3$ iff

$$\mathbb{E}(f \cdot g|\mathcal{B}_2) = \mathbb{E}(f|\mathcal{B}_2) \cdot \mathbb{E}(g|\mathcal{B}_2) \ \text{a.s.} \ \forall\, f \in \mathbb{M}_1 \text{ and } g \in \mathbb{M}_3. \tag{10.7}$$

Hint: Do this by two applications of the functional form of the multiplicative systems theorem, see Theorems 8.16 and 8.5 of Chapter 8. For the first application, fix an $f \in \mathbb{M}_1$ and let

$$\mathbb{H} := \{g \in (\mathcal{B}_3)_b : \mathbb{E}(f \cdot g|\mathcal{B}_2) = \mathbb{E}(f|\mathcal{B}_2) \cdot \mathbb{E}(g|\mathcal{B}_2) \ \text{a.s.}\}.$$

(See the proof of Theorem 17.4 if you get stuck.)

10.2 Examples of Independence

10.2.1 An Example of Ranks

Lemma 10.22 (No Ties). *Suppose that X and Y are independent random variables on a probability space $(\Omega, \mathcal{B}, P)$. If $F(x) := P(X \le x)$ is continuous, then $P(X = Y) = 0$.*

Proof. Let $\mu(A) := P(X \in A)$ and $\nu(A) = P(Y \in A)$. Because F is continuous, $\mu(\{y\}) = F(y) - F(y-) = 0$, and hence

$$\begin{aligned}P(X=Y) &= \mathbb{E}\left[1_{\{X=Y\}}\right] = \int_{\mathbb{R}^2} 1_{\{x=y\}} d(\mu\otimes\nu)(x,y) \\ &= \int_{\mathbb{R}} d\nu(y) \int_{\mathbb{R}} d\mu(x) 1_{\{x=y\}} = \int_{\mathbb{R}} \mu(\{y\}) d\nu(y) \\ &= \int_{\mathbb{R}} 0 \, d\nu(y) = 0.\end{aligned}$$

Second Proof. For sake of comparison, lets give a proof where we do not allow ourselves to use Fubini's theorem. To this end let $\left\{a_l := \frac{l}{N}\right\}_{l=-\infty}^{\infty}$ (or for the moment any sequence such that, $a_l < a_{l+1}$ for all $l \in \mathbb{Z}$, $\lim_{l\to\pm\infty} a_l = \pm\infty$). Then

$$\{(x,x) : x \in \mathbb{R}\} \subset \cup_{l\in\mathbb{Z}} [(a_l, a_{l+1}] \times (a_l, a_{l+1}]]$$

and therefore,

$$\begin{aligned}P(X=Y) &\le \sum_{l\in\mathbb{Z}} P(X \in (a_l, a_{l+1}], Y \in (a_l, a_{l+1}]) = \sum_{l\in\mathbb{Z}} [F(a_{l+1}) - F(a_l)]^2 \\ &\le \sup_{l\in\mathbb{Z}} [F(a_{l+1}) - F(a_l)] \sum_{l\in\mathbb{Z}} [F(a_{l+1}) - F(a_l)] = \sup_{l\in\mathbb{Z}} [F(a_{l+1}) - F(a_l)].\end{aligned}$$

Since F is continuous and $F(\infty+) = 1$ and $F(\infty-) = 0$, it is easily seen that F is uniformly continuous on $\mathbb{R}$. Therefore, if we choose $a_l = \frac{l}{N}$, we have

$$P(X=Y) \le \limsup_{N\to\infty} \sup_{l\in\mathbb{Z}} \left[F\left(\frac{l+1}{N}\right) - F\left(\frac{l}{N}\right)\right] = 0.$$

■

Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with common continuous distribution function, F. So by Lemma 10.22 we know that

$$P(X_i = X_j) = 0 \text{ for all } i \ne j.$$

Let R_n denote the "**rank**" of X_n in the list $(X_1, \dots, X_n)$, i.e.

$$R_n := \sum_{j=1}^{n} 1_{X_j \ge X_n} = \#\{j \le n : X_j \ge X_n\}.$$

Thus $R_n = k$ if X_n is the k^{th} – largest element in the list, $(X_1, \dots, X_n)$. For example if $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, \dots) = (9, -8, 3, 7, 23, 0, -11, \dots)$, we have $R_1 = 1$, $R_2 = 2$, $R_3 = 2$, $R_4 = 2$, $R_5 = 1$, $R_6 = 5$, and $R_7 = 7$. Observe that rank order, from lowest to highest, of $(X_1, X_2, X_3, X_4, X_5)$ is $(X_2, X_3, X_4, X_1, X_5)$. This can be determined by the values of R_i for $i = 1, 2, \dots, 5$ as follows. Since $R_5 = 1$, we must have X_5 in the last slot, i.e. $(*, *, *, *, X_5)$. Since $R_4 = 2$, we know out of the remaining slots, X_4 must be in the second from the far most right, i.e. $(*, *, X_4, *, X_5)$. Since $R_3 = 2$, we know that X_3 is again the second from the right of the remaining slots, i.e. we now know, $(*, X_3, X_4, *, X_5)$. Similarly, $R_2 = 2$ implies $(X_2, X_3, X_4, *, X_5)$ and finally $R_1 = 1$ gives, $(X_2, X_3, X_4, X_1, X_5)$ $(= (-8, 4, 7, 9, 23)$ in the example). As another example, if $R_i = i$ for $i = 1, 2, \dots, n$, then $X_n < X_{n-1} < \dots < X_1$.

Theorem 10.23 (Renyi Theorem). *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. and assume that $F(x) := P(X_n \le x)$ is continuous. The $\{R_n\}_{n=1}^{\infty}$ is an independent sequence,*

$$P(R_n = k) = \frac{1}{n} \text{ for } k = 1, 2, \dots, n,$$

and the events, $A_n = \{X_n \text{ is a record}\} = \{R_n = 1\}$ are independent as n varies and

$$P(A_n) = P(R_n = 1) = \frac{1}{n}.$$

Proof. By Problem 6 on p. 110 of Resnick or by Fubini's theorem, $(X_1, \dots, X_n)$ and $(X_{\sigma 1}, \dots, X_{\sigma n})$ have the same distribution for any permutation σ.

Since F is continuous, it now follows that up to a set of measure zero,

$$\Omega = \sum_{\sigma} \{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}$$

and therefore

$$1 = P(\Omega) = \sum_{\sigma} P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}).$$

Since $P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\})$ is independent of σ we may now conclude that

$$P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}) = \frac{1}{n!}$$

for all σ. As observed before the statement of the theorem, to each realization $(\varepsilon_1, \dots, \varepsilon_n)$, (here $\varepsilon_i \in \mathbb{N}$ with $\varepsilon_i \le i$) of $(R_1, \dots, R_n)$ there is a uniquely determined permutation, $\sigma = \sigma(\varepsilon_1, \dots, \varepsilon_n)$, such that $X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}$. (Notice that there are $n!$ permutations of $\{1, 2, \dots, n\}$ and there are also $n!$ choices for the $\{(\varepsilon_1, \dots, \varepsilon_n) : 1 \le \varepsilon_i \le i\}$.) From this it follows that

$$\{(R_1, \dots, R_n) = (\varepsilon_1, \dots, \varepsilon_n)\} = \{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}$$

and therefore,

$$P(\{(R_1, \dots, R_n) = (\varepsilon_1, \dots, \varepsilon_n)\}) = P(X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}) = \frac{1}{n!}.$$

Since

$$\begin{aligned} P(\{R_n = \varepsilon_n\}) &= \sum_{(\varepsilon_1,\dots\varepsilon_{n-1})} P(\{(R_1,\dots,R_n) = (\varepsilon_1,\dots,\varepsilon_n)\}) \\ &= \sum_{(\varepsilon_1,\dots\varepsilon_{n-1})} \frac{1}{n!} = (n-1)! \cdot \frac{1}{n!} = \frac{1}{n} \end{aligned}$$

we have shown that

$$P(\{(R_1,\dots,R_n) = (\varepsilon_1,\dots,\varepsilon_n)\}) = \frac{1}{n!} = \prod_{j=1}^{n} \frac{1}{j} = \prod_{j=1}^{n} P(\{R_j = \varepsilon_j\}).$$

■

10.3 Gaussian Random Vectors

As you saw in Exercise 10.2, uncorrelated random variables are typically not independent. However, if the random variables involved are jointly Gaussian (see Definition 9.34), then independence and uncorrelated are actually the same thing!

Lemma 10.24. *Suppose that $Z = (X,Y)^{\mathrm{tr}}$ is a Gaussian random vector with $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$. Then X is independent of Y iff $\operatorname{Cov}(X_i, Y_j) = 0$ for all $1 \le i \le k$ and $1 \le j \le l$. This lemma also holds more generally. Namely if $\{X^l\}_{l=1}^n$ is a sequence of random vectors such that $(X^1,\dots,X^n)$ is a Gaussian random vector. Then $\{X^l\}_{l=1}^n$ are independent iff $\operatorname{Cov}\left(X_i^l, X_k^{l'}\right) = 0$ for all $l \ne l'$ and i and k.*

Proof. We know by Exercise 10.2 that if X_i and Y_j are independent, then $\operatorname{Cov}(X_i, Y_j) = 0$. For the converse direction, if $\operatorname{Cov}(X_i, Y_j) = 0$ for all $1 \le i \le k$ and $1 \le j \le l$ and $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then

$$\begin{aligned} \operatorname{Var}(x \cdot X + y \cdot Y) &= \operatorname{Var}(x \cdot X) + \operatorname{Var}(y \cdot Y) + 2\operatorname{Cov}(x \cdot X, y \cdot Y) \\ &= \operatorname{Var}(x \cdot X) + \operatorname{Var}(y \cdot Y). \end{aligned}$$

Therefore using the fact that (X,Y) is a Gaussian random vector,

$$\begin{aligned} \mathbb{E}\left[e^{ix\cdot X} e^{iy \cdot Y}\right] &= \mathbb{E}\left[e^{i(x\cdot X + y\cdot Y)}\right] \\ &= \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X + y\cdot Y) + \mathbb{E}(x\cdot X + y \cdot Y)\right) \\ &= \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X) + i\mathbb{E}(x\cdot X) - \frac{1}{2}\operatorname{Var}(y\cdot Y) + i\mathbb{E}(y\cdot Y)\right) \\ &= \mathbb{E}\left[e^{ix\cdot X}\right] \cdot \mathbb{E}\left[e^{iy\cdot Y}\right], \end{aligned}$$

and because x and y were arbitrary, we may conclude from Theorem 10.11 that X and Y are independent. ■

Corollary 10.25. *Suppose that $X : \Omega \to \mathbb{R}^k$ and $Y : \Omega \to \mathbb{R}^l$ are two independent random Gaussian vectors, then (X,Y) is also a Gaussian random vector. This corollary generalizes to multiple independent random Gaussian vectors.*

Proof. Let $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then

$$\begin{aligned} \mathbb{E}\left[e^{i(x,y)\cdot(X,Y)}\right] =& \mathbb{E}\left[e^{i(x\cdot X + y \cdot Y)}\right] = \mathbb{E}\left[e^{ix\cdot X} e^{iy\cdot Y}\right] = \mathbb{E}\left[e^{ix\cdot X}\right] \cdot \mathbb{E}\left[e^{iy\cdot Y}\right] \\ =& \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X) + i\mathbb{E}(x\cdot X)\right) \\ & \times \exp\left(-\frac{1}{2}\operatorname{Var}(y\cdot Y) + i\mathbb{E}(y\cdot Y)\right) \\ =& \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X) + i\mathbb{E}(x\cdot X) - \frac{1}{2}\operatorname{Var}(y\cdot Y) + i\mathbb{E}(y\cdot Y)\right) \\ =& \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X + y\cdot Y) + i\mathbb{E}(x\cdot X + y\cdot Y)\right) \end{aligned}$$

which shows that (X,Y) is again Gaussian. ■

Notation 10.26 *Suppose that $\{X_i\}_{i=1}^n$ is a collection of $\mathbb{R}$ – valued variables or $\mathbb{R}^d$ – valued random vectors. We will write $X_1 \overset{\perp\!\perp}{+} X_2 \overset{\perp\!\perp}{+} \dots \overset{\perp\!\perp}{+} X_n$ for $X_1 + \dots + X_n$ under the additional assumption that the $\{X_i\}_{i=1}^n$ are independent.*

Corollary 10.27. *Suppose that $\{X_i\}_{i=1}^n$ are independent Gaussian random variables, then $S_n := \sum_{i=1}^n X_i$ is a Gaussian random variables with :*

$$\operatorname{Var}(S_n) = \sum_{i=1}^n \operatorname{Var}(X_i) \text{ and } \mathbb{E}S_n = \sum_{i=1}^n \mathbb{E}X_i, \tag{10.8}$$

i.e.

$$X_1 \overset{\perp\!\perp}{+} X_2 \overset{\perp\!\perp}{+} \dots \overset{\perp\!\perp}{+} X_n \overset{d}{=} N\left(\sum_{i=1}^n \operatorname{Var}(X_i), \sum_{i=1}^n \mathbb{E}X_i\right).$$

In particular if $\{X_i\}_{i=1}^\infty$ are i.i.d. Gaussian random variables with $\mathbb{E}X_i = \mu$ and $\sigma^2 = \operatorname{Var}(X_i)$, then

$$\frac{S_n}{n} - \mu \overset{d}{=} N\left(0, \frac{\sigma^2}{n}\right) \text{ and} \tag{10.9}$$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{d}{=} N(0,1). \tag{10.10}$$

Equation (10.10) is a very special case of the central limit theorem while Eq. (10.9) leads to a very special case of the strong law of large numbers, see Corollary 10.28.

Proof. The fact that S_n, $\frac{S_n}{n}-\mu$, and $\frac{S_n-n\mu}{\sigma\sqrt{n}}$ are all Gaussian follows from Corollary 10.27 and Lemma 9.36 or by direct calculation. The formulas for the variances and means of these random variables are routine to compute. ■

Recall the first Borel Cantelli-Lemma 7.14 states that if $\{A_n\}_{n=1}^\infty$ are measurable sets, then

$$\sum_{n=1}^\infty P(A_n)<\infty \implies P(\{A_n \text{ i.o.}\})=0. \tag{10.11}$$

Corollary 10.28. *Let $\{X_i\}_{i=1}^\infty$ be i.i.d. Gaussian random variables with $\mathbb{E}X_i=\mu$ and $\sigma^2=\operatorname{Var}(X_i)$. Then $\lim_{n\to\infty}\frac{S_n}{n}=\mu$ a.s. and moreover for every $\alpha<\frac{1}{2}$, there exists $N_\alpha:\Omega\to\mathbb{N}\cup\{\infty\}$, such that $P(N_\alpha=\infty)=0$ and*

$$\left|\frac{S_n}{n}-\mu\right|\le n^{-\alpha} \text{ for } n\ge N_\alpha.$$

In particular, $\lim_{n\to\infty}\frac{S_n}{n}=\mu$ a.s.

Proof. Let $Z\overset{d}{=}N(0,1)$ so that $\frac{\sigma}{\sqrt{n}}Z\overset{d}{=}N\left(0,\frac{\sigma^2}{n}\right)$. From the Eq. (10.9) and Eq. (7.42),

$$\begin{aligned}P\left(\left|\frac{S_n}{n}-\mu\right|\ge\varepsilon\right)&=P\left(\left|\frac{\sigma}{\sqrt{n}}Z\right|\ge\varepsilon\right)=P\left(|Z|\ge\frac{\sqrt{n}\varepsilon}{\sigma}\right)\\&\le\exp\left(-\frac{1}{2}\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)^2\right)=\exp\left(-\frac{\varepsilon^2}{2\sigma^2}n\right).\end{aligned}$$

Taking $\varepsilon=n^{-\alpha}$ with $1-2\alpha>0$, it follows that

$$\sum_{n=1}^\infty P\left(\left|\frac{S_n}{n}-\mu\right|\ge n^{-\alpha}\right)\le\sum_{n=1}^\infty\exp\left(-\frac{1}{2\sigma^2}n^{1-2\alpha}\right)<\infty$$

and so by the first Borel-Cantelli lemma,

$$P\left(\left\{\left|\frac{S_n}{n}-\mu\right|\ge n^{-\alpha}\text{ i.o.}\right\}\right)=0.$$

Therefore, P – a.s., $\left|\frac{S_n}{n}-\mu\right|\le n^{-\alpha}$ a.a., and in particular $\lim_{n\to\infty}\frac{S_n}{n}=\mu$ a.s. ■

Theorem 10.29. *Suppose that $Z=(X,Y)^{\mathrm{tr}}$ is a mean zero Gaussian random vector with $X\in\mathbb{R}^k$ and $Y\in\mathbb{R}^l$. Let $Q=Q_X:=\mathbb{E}[XX^{\mathrm{tr}}]$ and then let*

$$W:=Y-\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1}X$$

where $Q^{-1}=Q|_{\operatorname{Ran}(Q)}^{-1}P$ is as in Example 9.43 below. Then $(X,W)^{\mathrm{tr}}$ is again a Gaussian random vector and moreover W is independent of X. The covariance matrix for W is

$$\mathbb{E}\left[WW^{\mathrm{tr}}\right]=\mathbb{E}\left[YY^{\mathrm{tr}}\right]-\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1}\mathbb{E}\left[XY^{\mathrm{tr}}\right]. \tag{10.12}$$

Proof. Let Λ be any $k\times l$ matrix and let $W:=Y-\Lambda X$. Since

$$\begin{pmatrix}X\\W\end{pmatrix}=\begin{pmatrix}I&0\\-\Lambda&I\end{pmatrix}\begin{pmatrix}X\\Y\end{pmatrix},$$

according to Lemma 9.36 $(X,W)^{\mathrm{tr}}$ is still Gaussian. So according to Lemma 10.24, in order to make W independent of X it suffices to choose Λ so that W and X are uncorrelated, i.e.

$$\begin{aligned}0=\operatorname{Cov}(W_j,X_i)&=\operatorname{Cov}\left(Y_j-\sum_k\Lambda_{jk}X_k,X_i\right)\\&=\mathbb{E}[Y_jX_i]-\sum_k\Lambda_{jk}\mathbb{E}(X_kX_i).\end{aligned}$$

In matrix notation, we want to choose Λ so that

$$\mathbb{E}\left[YX^{\mathrm{tr}}\right]=\Lambda\mathbb{E}\left[XX^{\mathrm{tr}}\right]. \tag{10.13}$$

In the case $Q:=\mathbb{E}[XX^{\mathrm{tr}}]$ is non-degenerate, we see that $\Lambda:=\mathbb{E}[YX^{\mathrm{tr}}]Q^{-1}$ is the desired solution. In fact this works for general Q where Q^{-1} is defined in Example 9.43. To see this is correct, recall

$$v\cdot Qv=v\cdot\mathbb{E}\left[XX^{\mathrm{tr}}v\right]=\mathbb{E}\left[(v\cdot X)^2\right]$$

from which it follows that

$$\operatorname{Nul}(Q)=\left\{v\in\mathbb{R}^k:v\cdot X=0\right\}.$$

Hence it follows that

$$\mathbb{E}\left[YX^{\mathrm{tr}}\right]v=\Lambda\mathbb{E}\left[XX^{\mathrm{tr}}\right]v \text{ for all } v\in\operatorname{Nul}(Q)$$

no matter how Λ is chosen. On the other hand if $v\in\operatorname{Ran}(Q)=\operatorname{Nul}(Q)^\perp$,

$$\Lambda\mathbb{E}\left[XX^{\mathrm{tr}}\right]v=\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1}Qv=\mathbb{E}\left[YX^{\mathrm{tr}}\right]v$$

as desired.

To prove Eq. (10.12) let $B:=\mathbb{E}\left[YX^{\mathrm{tr}}\right]$ so that

$$W:=Y-BQ^{-1}X.$$

We then have

$$\begin{aligned}\mathbb{E}\left[WW^{\mathrm{tr}}\right]&=\mathbb{E}\left[\left(Y-BQ^{-1}X\right)\left(Y-BQ^{-1}X\right)^{\mathrm{tr}}\right]\\&=\mathbb{E}\left[\left(Y-BQ^{-1}X\right)\left(Y^{\mathrm{tr}}-X^{\mathrm{tr}}Q^{-1}B^{\mathrm{tr}}\right)\right]\\&=\mathbb{E}\left[YY^{\mathrm{tr}}-YX^{\mathrm{tr}}Q^{-1}B^{\mathrm{tr}}-BQ^{-1}XY^{\mathrm{tr}}+BQ^{-1}XX^{\mathrm{tr}}Q^{-1}B^{\mathrm{tr}}\right]\\&=\mathbb{E}\left[YY^{\mathrm{tr}}\right]-BQ^{-1}B^{\mathrm{tr}}-BQ^{-1}B^{\mathrm{tr}}+BQ^{-1}QQ^{-1}B^{\mathrm{tr}}\\&=\mathbb{E}\left[YY^{\mathrm{tr}}\right]-BQ^{-1}B^{\mathrm{tr}}\\&=\mathbb{E}\left[YY^{\mathrm{tr}}\right]-\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1}\mathbb{E}\left[XY^{\mathrm{tr}}\right].\end{aligned}$$

■

Corollary 10.30. *Suppose that $Z=(X,Y)^{\mathrm{tr}}$ is a mean zero Gaussian random vector with $X\in\mathbb{R}^k$ and $Y\in\mathbb{R}^l$,*

$$\begin{aligned}A&:=\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1},\\Q_W&:=\mathbb{E}\left[YY^{\mathrm{tr}}\right]-\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1}\mathbb{E}\left[XY^{\mathrm{tr}}\right],\end{aligned}$$

and suppose $W\overset{d}{=}N\left(Q_W,0\right)$. If $f:\mathbb{R}^k\times\mathbb{R}^l\to\mathbb{R}$ is a bounded measurable function, then

$$\mathbb{E}\left[f\left(X,Y\right)|X\right]=\mathbb{E}\left[f\left(x,Ax+W\right)\right]|_{x=X}.$$

As an important special case, if $x\in\mathbb{R}^k$ and $y\in\mathbb{R}^l$, then

$$\mathbb{E}\left[e^{i(x\cdot X+y\cdot Y)}|X\right]=e^{i(x\cdot X+y\cdot AX)}e^{-\frac{1}{2}\operatorname{Var}(y\cdot W)}=e^{i(x\cdot X+y\cdot AX)}e^{-\frac{1}{2}Q_Wy\cdot y}. \tag{10.14}$$

Proof. Using the notation in Theorem 10.29,

$$\mathbb{E}\left[f\left(X,Y\right)|X\right]=\mathbb{E}\left[f\left(X,AX+W\right)|X\right]$$

where $W\overset{d}{=}N\left(Q_W,0\right)$ and W is independent of X. The result now follows by an application of Exercise 14.4. Let us now specialize to the case where $f\left(X,Y\right)=e^{i(x\cdot X+y\cdot Y)}$ in which case

$$\begin{aligned}\mathbb{E}\left[e^{i(x\cdot X+y\cdot Y)}|X\right]&=\mathbb{E}\left[e^{i\left(x\cdot x'+y\cdot\left(Ax'+W\right)\right)}|X\right]|_{x'=X}=e^{i(x\cdot X+y\cdot AX)}\mathbb{E}\left[e^{iy\cdot W}\right]\\&=e^{i(x\cdot X+y\cdot AX)}e^{-\frac{1}{2}\operatorname{Var}(y\cdot W)}=e^{i(x\cdot X+y\cdot AX)}e^{-\frac{1}{2}Q_Wy\cdot y}.\end{aligned}$$

■

Exercise 10.5. Suppose now that $(X,Y,Z)^{\mathrm{tr}}$ is a mean zero Gaussian random vector with $X\in\mathbb{R}^k$, $Y\in\mathbb{R}^l$, and $Z\in\mathbb{R}^m$. Show for all $y\in\mathbb{R}^l$ and $z\in\mathbb{R}^m$ that

$$\begin{aligned}&\mathbb{E}\left[\exp\left(i\left(y\cdot Y+z\cdot Z\right)\right)|X\right]\\&\quad=\exp\left(-\operatorname{Cov}\left(y\cdot W_1,z\cdot W_2\right)\right)\cdot\mathbb{E}\left[\exp\left(iy\cdot Y\right)|X\right]\cdot\mathbb{E}\left[\exp\left(iz\cdot Z\right)|X\right].\end{aligned}$$

In performing these computations please use the following definitions,

$$\begin{aligned}Q&:=Q_X:=\mathbb{E}\left[XX^{\mathrm{tr}}\right],\\A&:=\mathbb{E}\left[\begin{bmatrix}Y\\Z\end{bmatrix}X^{\mathrm{tr}}\right]Q^{-1}=\begin{bmatrix}\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1}\\\mathbb{E}\left[ZX^{\mathrm{tr}}\right]Q^{-1}\end{bmatrix}=:\begin{bmatrix}A_1\\A_2\end{bmatrix},\end{aligned}$$

and

$$W:=\begin{bmatrix}W_1\\W_2\end{bmatrix}=\begin{bmatrix}Y\\Z\end{bmatrix}-AX=\begin{bmatrix}Y-A_1X\\Z-A_2X\end{bmatrix}.$$

Exercise 10.6. Keeping the same notation as in Exercise 10.5, show $Y\overset{X}{\perp\!\!\!\perp}Z$ (see Definition 10.20) iff

$$\mathbb{E}\left[YZ^{\mathrm{tr}}\right]=\mathbb{E}\left[YX^{\mathrm{tr}}\right]Q^{-1}\mathbb{E}\left[XZ^{\mathrm{tr}}\right].$$

where

$$Q=Q_X:=\mathbb{E}\left[XX^{\mathrm{tr}}\right].$$

10.4 Summing independent random variables

Exercise 10.7. Suppose that $X\overset{d}{=}N\left(0,a^2\right)$ and $Y\overset{d}{=}N\left(0,b^2\right)$ and X and Y are independent. Show by direct computation using the formulas for the distributions of X and Y that $X+Y=N\left(0,a^2+b^2\right)$.

Solution to Exercise (10.7). If $f:\mathbb{R}\to\mathbb{R}$ be a bounded measurable function, then

$$\mathbb{E}\left[f\left(X+Y\right)\right]=\frac{1}{Z}\int_{\mathbb{R}^2}f\left(x+y\right)e^{-\frac{1}{2a^2}x^2}e^{-\frac{1}{2b^2}y^2}dxdy,$$

where $Z=2\pi ab$. Let us make the change of variables, $(x,z)=(x,x+y)$ and observe that $dxdy=dxdz$ (you check). Therefore we have,

$$\mathbb{E}\left[f\left(X+Y\right)\right]=\frac{1}{Z}\int_{\mathbb{R}^2}f\left(z\right)e^{-\frac{1}{2a^2}x^2}e^{-\frac{1}{2b^2}(z-x)^2}dxdz$$

which shows, $\operatorname{Law}_P\left(X+Y\right)(dz)=\rho\left(z\right)dz$ where

$$\rho(z) = \frac{1}{Z}\int_{\mathbb{R}} e^{-\frac{1}{2a^2}x^2} e^{-\frac{1}{2b^2}(z-x)^2} dx. \tag{10.15}$$

Working the exponent, for any $c \in \mathbb{R}$, we have

$$\begin{aligned}\frac{1}{a^2}x^2 + \frac{1}{b^2}(z-x)^2 &= \frac{1}{a^2}x^2 + \frac{1}{b^2}\left(x^2 - 2xz + z^2\right) \\ &= \left(\frac{1}{a^2} + \frac{1}{b^2}\right)x^2 - \frac{2}{b^2}xz + \frac{1}{b^2}z^2 \\ &= \left(\frac{1}{a^2} + \frac{1}{b^2}\right)\left[(x-cz)^2 + 2cxz - c^2z^2\right] - \frac{2}{b^2}xz + \frac{1}{b^2}z^2.\end{aligned}$$

Let us now choose (to complete the squares) c such that where c must be chosen so that

$$c\left(\frac{1}{a^2} + \frac{1}{b^2}\right) = \frac{1}{b^2} \implies c = \frac{a^2}{a^2+b^2},$$

in which case,

$$\frac{1}{a^2}x^2 + \frac{1}{b^2}(z-x)^2 = \left(\frac{1}{a^2} + \frac{1}{b^2}\right)\left[(x-cz)^2\right] + \left[\frac{1}{b^2} - c^2\left(\frac{1}{a^2} + \frac{1}{b^2}\right)\right]z^2$$

where,

$$\frac{1}{b^2} - c^2\left(\frac{1}{a^2} + \frac{1}{b^2}\right) = \frac{1}{b^2}(1-c) = \frac{1}{a^2+b^2}.$$

So making the change of variables, $x \to x - cz$, in the integral in Eq. (10.15) implies,

$$\begin{aligned}\rho(z) &= \frac{1}{Z}\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{1}{a^2} + \frac{1}{b^2}\right)w^2 - \frac{1}{2}\frac{1}{a^2+b^2}z^2\right)dw \\ &= \frac{1}{\tilde{Z}}\exp\left(-\frac{1}{2}\frac{1}{a^2+b^2}z^2\right)\end{aligned}$$

where,

$$\begin{aligned}\frac{1}{\tilde{Z}} &= \frac{1}{Z}\cdot\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{1}{a^2} + \frac{1}{b^2}\right)w^2\right)dw = \frac{1}{2\pi ab}\sqrt{2\pi\left(\frac{1}{a^2} + \frac{1}{b^2}\right)^{-1}} \\ &= \frac{1}{2\pi ab}\sqrt{2\pi\frac{a^2b^2}{a^2+b^2}} = \frac{1}{\sqrt{2\pi\left(a^2+b^2\right)}}.\end{aligned}$$

Thus it follows that $X \overset{\perp\perp}{+} Y \overset{d}{=} N\left(a^2+b^2, 0\right)$.

Exercise 10.8. Show that the sum, $N_1 + N_2$, of two independent Poisson random variables, N_1 and N_2, with parameters λ_1 and λ_2 respectively is again a Poisson random variable with parameter $\lambda_1 + \lambda_2$. (You could use generating functions or do this by hand.) In short $\mathrm{Poi}(\lambda_1) \overset{\perp\perp}{+} \mathrm{Poi}(\lambda_2) \overset{d}{=} \mathrm{Poi}(\lambda_1 + \lambda_2)$.

Solution to Exercise (10.8). Let $z \in \mathbb{C}$, then by independence,

$$\begin{aligned}\mathbb{E}\left[z^{N_1+N_2}\right] &= \mathbb{E}\left[z^{N_1}z^{N_2}\right] = \mathbb{E}\left[z^{N_1}\right]\mathbb{E}\left[z^{N_2}\right] \\ &= e^{\lambda_1(z-1)}\cdot e^{\lambda_2(z-1)} = e^{(\lambda_1+\lambda_2)(z-1)}\end{aligned}$$

from which it follows that $N_1 + N_2 \overset{d}{=} \mathrm{Poisson}(\lambda_1 + \lambda_2)$.

Example 10.31 (Gamma Distribution Sums). We will show here that $\mathrm{Gamma}(k,\theta) \overset{\perp\perp}{+} \mathrm{Gamma}(l,\theta) = \mathrm{Gamma}(k+l,\theta)$. In Exercise 7.13 you showed if $k, \theta > 0$ then

$$\mathbb{E}\left[e^{tX}\right] = (1-\theta t)^{-k} \text{ for } t < \theta^{-1}$$

where X is a positive random variable with $X \overset{d}{=} \mathrm{Gamma}(k,\theta)$, i.e.

$$(X_*P)(dx) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}dx \text{ for } x > 0.$$

Suppose that X and Y are independent Random variables with $X \overset{d}{=} \mathrm{Gamma}(k,\theta)$ and $Y \overset{d}{=} \mathrm{Gamma}(l,\theta)$ for some $l > 0$. It now follows that

$$\begin{aligned}\mathbb{E}\left[e^{t(X+Y)}\right] &= \mathbb{E}\left[e^{tX}e^{tY}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] \\ &= (1-\theta t)^{-k}(1-\theta t)^{-l} = (1-\theta t)^{-(k+l)}.\end{aligned}$$

Therefore it follows from Exercise 8.2 that $X + Y \overset{d}{=} \mathrm{Gamma}(k+l,\theta)$.

Example 10.32 (Exponential Distribution Sums). If $\{T_k\}_{k=1}^n$ are independent random variables such that $T_k \overset{d}{=} E(\lambda_k)$ for all k, then

$$T_1 \overset{\perp\perp}{+} T_2 \overset{\perp\perp}{+} \dots \overset{\perp\perp}{+} T_n = \mathrm{Gamma}\left(n, \lambda^{-1}\right).$$

This follows directly from Example 10.31 using $E(\lambda) = \mathrm{Gamma}\left(1, \lambda^{-1}\right)$ and induction. We will verify this directly later on in Corollary 11.8.

Example 10.31 may also be verified using brute force. To this end, suppose that $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function, then

$$\mathbb{E}[f(X+Y)] = \int_{\mathbb{R}_+^2} f(x+y)\, x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} y^{l-1} \frac{e^{-y/\theta}}{\theta^l \Gamma(l)} dxdy$$
$$= \frac{1}{\theta^{k+l}\Gamma(k)\Gamma(l)} \int_{\mathbb{R}_+^2} f(x+y)\, x^{k-1} y^{l-1} e^{-(x+y)/\theta} dxdy.$$

Let us now make the change of variables, $x = x$ and $z = x + y$, so that $dxdy = dxdz$, to find,

$$\mathbb{E}[f(X+Y)] = \frac{1}{\theta^{k+l}\Gamma(k)\Gamma(l)} \int 1_{0\le x\le z<\infty} f(z)\, x^{k-1} (z-x)^{l-1} e^{-z/\theta} dxdz. \tag{10.16}$$

To finish the proof we must now do that x integral and show,

$$\int_0^z x^{k-1} (z-x)^{l-1} dx = z^{k+l-1} \frac{\Gamma(k)\Gamma(l)}{\Gamma(k+l)}.$$

(In fact we already know this must be correct from our Laplace transform computations above.) First make the change of variable, $x = zt$ to find,

$$\int_0^z x^{k-1} (z-x)^{l-1} dx = z^{k+l-1} B(k,l)$$

where $B(k,l)$ is the **beta − function** defined by;

$$B(k,l) := \int_0^1 t^{k-1} (1-t)^{l-1} dt \text{ for } \operatorname{Re} k, \operatorname{Re} l > 0. \tag{10.17}$$

Combining these results with Eq. (10.16) then shows,

$$\mathbb{E}[f(X+Y)] = \frac{B(k,l)}{\theta^{k+l}\Gamma(k)\Gamma(l)} \int_0^\infty f(z)\, z^{k+l-1} e^{-z/\theta} dz. \tag{10.18}$$

Since we already know that

$$\int_0^\infty z^{k+l-1} e^{-z/\theta} dz = \theta^{k+l}\Gamma(k+l)$$

it follows by taking $f = 1$ in Eq. (10.18) that

$$1 = \frac{B(k,l)}{\theta^{k+l}\Gamma(k)\Gamma(l)} \theta^{k+l}\Gamma(k+l)$$

which implies,

$$B(k,l) = \frac{\Gamma(k)\Gamma(l)}{\Gamma(k+l)}. \tag{10.19}$$

Therefore, using this back in Eq. (10.18) implies

$$\mathbb{E}[f(X+Y)] = \frac{1}{\theta^{k+l}\Gamma(k+l)} \int_0^\infty f(z)\, z^{k+l-1} e^{-z/\theta} dz$$

from which it follows that $X + Y \stackrel{d}{=} \text{Gamma}(k+l, \theta)$.

Let us pause to give a direct verification of Eq. (10.19). By definition of the gamma function,

$$\Gamma(k)\Gamma(l) = \int_{\mathbb{R}_+^2} x^{k-1} e^{-x} y^{l-1} e^{-y} dxdy = \int_{\mathbb{R}_+^2} x^{k-1} y^{l-1} e^{-(x+y)} dxdy.$$
$$= \int_{0\le x\le z<\infty} x^{k-1} (z-x)^{l-1} e^{-z} dxdz$$

Making the change of variables, $x = x$ and $z = x + y$ it follows,

$$\Gamma(k)\Gamma(l) = \int_{0\le x\le z<\infty} x^{k-1} (z-x)^{l-1} e^{-z} dxdz.$$

Now make the change of variables, $x = zt$ to find,

$$\Gamma(k)\Gamma(l) = \int_0^\infty dz e^{-z} \int_0^1 dt\, (zt)^{k-1} (z - tz)^{l-1} z$$
$$= \int_0^\infty e^{-z} z^{k+l-1} dz \cdot \int_0^1 t^{k-1} (1-t)^{l-1} dt$$
$$= \Gamma(k+l)\, B(k,l).$$

Definition 10.33 (Beta distribution). *The β – distribution is*

$$d\mu_{x,y}(t) = \frac{t^{x-1} (1-t)^{y-1} dt}{B(x,y)}.$$

Observe that

$$\int_0^1 t d\mu_{x,y}(t) = \frac{B(x+1,y)}{B(x,y)} = \frac{\frac{\Gamma(x+1)\Gamma(y)}{\Gamma(x+y+1)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}} = \frac{x}{x+y}$$

and

$$\int_0^1 t^2 d\mu_{x,y}(t) = \frac{B(x+2,y)}{B(x,y)} = \frac{\frac{\Gamma(x+2)\Gamma(y)}{\Gamma(x+y+2)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}} = \frac{(x+1)\,x}{(x+y+1)(x+y)}.$$

10.5 A Strong Law of Large Numbers

Theorem 10.34 (A simple form of the strong law of large numbers). *If $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables such that $\mathbb{E}\left[|X_n|^4\right] < \infty$, then*

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu \text{ a.s.}$$

where $S_n := \sum_{k=1}^n X_k$ and $\mu := \mathbb{E}X_n = \mathbb{E}X_1$.

Exercise 10.9. Use the following outline to give a proof of Theorem 10.34.

1. First show that $x^p \le 1 + x^4$ for all $x \ge 0$ and $1 \le p \le 4$. Use this to conclude;
$$\mathbb{E}|X_n|^p \le 1 + \mathbb{E}|X_n|^4 < \infty \text{ for } 1 \le p \le 4.$$
Thus $\gamma := \mathbb{E}\left[|X_n - \mu|^4\right]$ and the standard deviation (σ^2) of X_n defined by,
$$\sigma^2 := \mathbb{E}\left[X_n^2\right] - \mu^2 = \mathbb{E}\left[(X_n - \mu)^2\right] < \infty,$$
are finite constants independent of n.
2. Show for all $n \in \mathbb{N}$ that
$$\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^4\right] = \frac{1}{n^4}\left(n\gamma + 3n(n-1)\sigma^4\right)$$
$$= \frac{1}{n^2}\left[n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4\right].$$
(Thus $\frac{S_n}{n} \to \mu$ in $L^4(P)$.)
3. Use item 2. and Chebyshev's inequality to show
$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \le \frac{n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4}{\varepsilon^4 n^2}.$$
4. Use item 3. and the first Borel Cantelli Lemma 7.14 to conclude $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ a.s.

10.6 A Central Limit Theorem

In this section we will give a preliminary a couple versions of the central limit theorem following [32, Chapter 2.14]. Let us set up some notation. Given a square integrable random variable Y, let

$$\bar{Y} := \frac{Y - \mathbb{E}Y}{\sigma(Y)} \text{ where } \sigma(Y) := \sqrt{\mathbb{E}(Y - \mathbb{E}Y)^2} = \sqrt{\operatorname{Var}(Y)}.$$

Let us also recall that if $Z = N\left(0, \sigma^2\right)$, then $Z \overset{d}{=} \sqrt{\sigma} N(0,1)$ and so by Eq. (7.40) with $\beta = 3$ we have,

$$\mathbb{E}\left|Z^3\right| = \sigma^3 \mathbb{E}|N(0,1)|^3 = \sqrt{8/\pi}\sigma^3. \tag{10.20}$$

Theorem 10.35 (A CLT proof w/o Fourier). *Suppose that $\{X_k\}_{k=1}^{\infty} \subset L^3(P)$ is a sequence of independent random variables such that*

$$C := \sup_k \mathbb{E}|X_k - \mathbb{E}X_k|^3 < \infty$$

Then for every function, $f \in C^3(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}} \left|f^{(3)}(x)\right| < \infty$ we have

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \frac{M}{3!}\left(1 + \sqrt{8/\pi}\right)\frac{C}{\sigma(S_n)^3} \cdot n, \tag{10.21}$$

where $S_n := X_1 + \cdots + X_n$ and $N \overset{d}{=} N(0,1)$. In particular if we further assume that

$$\delta := \liminf_{n\to\infty} \frac{1}{n}\sigma(S_n)^2 = \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \operatorname{Var}(X_i) > 0, \tag{10.22}$$

Then it follows that

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| = O\left(\frac{1}{\sqrt{n}}\right) \text{ as } n \to \infty \tag{10.23}$$

which is to say, $\bar{S}_n$ is "close" in distribution to N, which we abbreviate by $\bar{S}_n \overset{d}{\cong} N$ for large n.

(It should be noted that the estimate in Eq. (10.21) is valid for any finite collection of random variables, $\{X_k\}_{k=1}^n$.)

Proof. Let $n \in \mathbb{N}$ be fixed and then Let $\{Y_k, N_k\}_{k=1}^{\infty}$ be a collection of independent random variables such that

$$Y_k \overset{d}{=} \bar{X}_k = \frac{X_k - \mathbb{E}X_k}{\sigma(S_n)} \text{ and } N_k \overset{d}{=} N(0, \operatorname{Var}(Y_k)) \text{ for } 1 \le k \le n.$$

Let $S_n^Y = Y_1 + \cdots + Y_n \overset{d}{=} \bar{S}_n$ and $T_n := N_1 + \cdots + N_n$. Since

$$\sum_{k=1}^n \operatorname{Var}(N_k) = \sum_{k=1}^n \operatorname{Var}(Y_k) = \frac{1}{\sigma(S_n)^2}\sum_{k=1}^n \operatorname{Var}(X_k - \mathbb{E}X_k)$$
$$= \frac{1}{\sigma(S_n)^2}\sum_{k=1}^n \operatorname{Var}(X_k) = 1,$$

it follows by Corollary 10.27) that $T_n \stackrel{d}{=} N(0,1)$.

To compare $\mathbb{E}f\left(\bar{S}_n\right)$ with $\mathbb{E}f(N)$ we may compare $\mathbb{E}f\left(S_n^Y\right)$ with $\mathbb{E}f(T_n)$ which we will do by interpolating between S_n^Y and T_n. To this end, for $0 \le k \le n$, let

$$V_k := N_1 + \cdots + N_k + Y_{k+1} + \cdots + Y_n$$

with the convention that $V_n = T_n$ and $V_0 = S_n^Y$. Then by a telescoping series argument, it follows that

$$f(T_n) - f\left(S_n^Y\right) = f(V_n) - f(V_0) = \sum_{k=1}^{n} \left[f(V_k) - f(V_{k-1})\right]. \tag{10.24}$$

We now make use of Taylor's theorem with integral remainder the form,

$$f(x+\Delta) - f(x) = f'(x)\,\Delta + \frac{1}{2}f''(x)\,\Delta^2 + r(x,\Delta)\,\Delta^3 \tag{10.25}$$

where

$$r(x,\Delta) := \frac{1}{2}\int_0^1 f'''(x+t\Delta)(1-t)^2\,dt.$$

Taking Eq. (10.24) with Δ replaced by δ and subtracting the results then implies

$$f(x+\Delta) - f(x+\delta) = f'(x)(\Delta-\delta) + \frac{1}{2}f''(x)\left(\Delta^2-\delta^2\right) + \rho(x,\Delta,\delta), \tag{10.26}$$

where

$$|\rho(x,\Delta,\delta)| = \left|r(x,\Delta)\,\Delta^3 - r(x,\delta)\,\delta^3\right| \le \frac{M}{3!}\left[|\Delta|^3 + |\delta|^3\right], \tag{10.27}$$

wherein we have used the simple estimate, $|r(x,\Delta)| \vee |r(x,\delta)| \le M/3!$.

If we define

$$U_k := N_1 + \cdots + N_{k-1} + Y_{k+1} + \cdots + Y_n,$$

then $V_k = U_k + N_k$ and $V_{k-1} = U_k + Y_k$. Hence, using Eq. (10.26) with $x = U_k$, $\Delta = N_k$and $\delta = Y_k$, it follows that

$$\begin{aligned} f(V_k) - f(V_{k-1}) &= f(U_k + N_k) - f(U_k + Y_k) \\ &= f'(U_k)(N_k - Y_k) + \frac{1}{2}f''(U_k)\left(N_k^2 - Y_k^2\right) + R_k \end{aligned} \tag{10.28}$$

where

$$|R_k| \le \frac{M}{3!}\left[|N_k|^3 + |Y_k|^3\right]. \tag{10.29}$$

Taking expectations of Eq. (10.28) using; Eq. (10.29), $\mathbb{E}N_k = 0 = \mathbb{E}Y_k$, $\mathbb{E}N_k^2 = \mathbb{E}Y_k^2$, and the fact that U_k is independent of both Y_k and N_k, we find

$$|\mathbb{E}\left[f(V_k) - f(V_{k-1})\right]| = |\mathbb{E}R_k| \le \frac{M}{3!}\mathbb{E}\left[|N_k|^3 + |Y_k|^3\right].$$

Making use of Eq. (10.20) it follows that

$$\mathbb{E}|N_k|^3 = \sqrt{8/\pi}\cdot\mathrm{Var}(N_k)^{3/2} = \sqrt{8/\pi}\cdot\mathrm{Var}(Y_k)^{3/2} = \sqrt{8/\pi}\cdot\left(\mathbb{E}Y_k^2\right)^{3/2} \le \sqrt{8/\pi}\cdot\mathbb{E}|Y_k|^3,$$

wherein we have used Jensen's (or Hölder's) inequality (see Chapter 12 below) for the last inequality. Combining these estimates with Eq. (10.24) shows,

$$\begin{aligned} \left|\mathbb{E}\left[f(T_n) - f\left(S_n^Y\right)\right]\right| &= \left|\sum_{k=1}^n \mathbb{E}R_k\right| \le \sum_{k=1}^n \mathbb{E}|R_k| \\ &\le \frac{M}{3!}\sum_{k=1}^n \mathbb{E}\left[|N_k|^3 + |Y_k|^3\right] \\ &\le \frac{M}{3!}\left(1+\sqrt{8/\pi}\right)\sum_{k=1}^n \mathbb{E}\left[|Y_k|^3\right]. \end{aligned} \tag{10.30}$$

Since

$$\mathbb{E}|Y_k|^3 = \mathbb{E}\left|\left(\frac{X_k - \mathbb{E}X_k}{\sigma(S_n)}\right)^3\right| \le \frac{C}{\sigma(S_n)^3} \text{ and}$$
$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| = \left|\mathbb{E}\left[f(T_n) - f\left(S_n^Y\right)\right]\right|,$$

we see that Eq. (10.21) now follows from Eq. (10.30). ∎

Corollary 10.36. *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables in $L^3(P)$, $C := \mathbb{E}|X_1 - \mathbb{E}X_1|^3 < \infty$, $S_n := X_1 + \cdots + X_n$, and $N \stackrel{d}{=} N(0,1)$. Then for every function, $f \in C^3(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}}\left|f^{(3)}(x)\right| < \infty$ we have*

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \frac{M}{3!\sqrt{n}}\left(1+\sqrt{8/\pi}\right)\frac{C}{\mathrm{Var}(X_1)^{3/2}}. \tag{10.31}$$

(This is a specialized form of the "Berry–Esseen theorem.")

By a slight modification of the proof of Theorem 10.35 we have the following central limit theorem.

Theorem 10.37 (A CLT proof w/o Fourier). *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables in $L^2(P)$, $S_n := X_1 + \cdots + X_n$, and $N \stackrel{d}{=} N(0,1)$. Then for every function, $f \in C^2(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}}\left|f^{(2)}(x)\right| < \infty$ and f'' being uniformly continuous on $\mathbb{R}$ we have,*

$$\lim_{n\to\infty}\mathbb{E}f\left(\bar{S}_n\right) = \mathbb{E}f(N).$$

Proof. In this proof we use the following form of Taylor's theorem;

$$f(x+\Delta)-f(x)=f'(x)\,\Delta+\frac{1}{2}f''(x)\,\Delta^2+r(x,\Delta)\,\Delta^2 \tag{10.32}$$

where

$$r(x,\Delta)=\int_0^1\left[f''(x+t\Delta)-f''(x)\right](1-t)\,dt.$$

Taking Eq. (10.32) with Δ replaced by δ and subtracting the results then implies

$$f(x+\Delta)-f(x+\delta)=f'(x)(\Delta-\delta)+\frac{1}{2}f''(x)\left(\Delta^2-\delta^2\right)+\rho(x,\Delta,\delta)$$

where now,

$$\rho(x,\Delta,\delta)=r(x,\Delta)\,\Delta^2-r(x,\delta)\,\delta^2.$$

Since f'' is uniformly continuous it follows that

$$\varepsilon(\Delta):=\frac{1}{2}\sup\left\{|f''(x+t\Delta)-f(x)|:x\in\mathbb{R}\text{ and }0\le t\le 1\right\}\to 0$$

Thus we may conclude that

$$|r(x,\Delta)|\le\int_0^1|f''(x+t\Delta)-f''(x)|(1-t)\,dt\le\int_0^1 2\varepsilon(\Delta)(1-t)\,dt=\varepsilon(\Delta).$$

and therefore that

$$|\rho(x,\Delta,\delta)|\le\varepsilon(\Delta)\,\Delta^2+\varepsilon(\delta)\,\delta^2.$$

So working just as in the proof of Theorem 10.35 we may conclude,

$$\left|\mathbb{E}f(N)-\mathbb{E}f\left(\bar{S}_n\right)\right|\le\sum_{k=1}^n\mathbb{E}|R_k|$$

where now,

$$|R_k|=\varepsilon(N_k)\,N_k^2+\varepsilon(Y_k)\,Y_k^2.$$

Since the $\{Y_k\}_{k=1}^n$ and the $\{N_k\}_{k=1}^n$ are *i.i.d.* now it follows that

$$\left|\mathbb{E}f(N)-\mathbb{E}f\left(\bar{S}_n\right)\right|\le n\cdot\mathbb{E}\left[\varepsilon(N_1)\,N_1^2+\varepsilon(Y_1)\,Y_1^2\right].$$

Since $\operatorname{Var}(S_n)=n\cdot\operatorname{Var}(X_1)$, we have $Y_1=\frac{X_1-\mathbb{E}X_1}{\sqrt{n}\sigma(X_1)}$, $\operatorname{Var}(N_1)=\operatorname{Var}(Y_1)=\frac{1}{n}$ and therefore $N_1\overset{d}{=}\sqrt{\frac{1}{n}}N$. Combining these observations shows,

$$\left|\mathbb{E}f(N)-\mathbb{E}f\left(\bar{S}_n\right)\right|\le\mathbb{E}\left[\varepsilon\left(\sqrt{\frac{1}{n}}N\right)N^2+\varepsilon\left(\frac{X_1-\mathbb{E}X_1}{\sqrt{n}\sigma(X_1)}\right)\frac{(X_1-\mathbb{E}X_1)^2}{\sigma^2(X_1)}\right]$$

which goes to zero as $n\to\infty$ by the DCT. ∎

Lemma 10.38. *Suppose that $\{W\}\cup\{W_n\}_{n=1}^\infty$ is a collection of random variables such that $\lim_{n\to\infty}\mathbb{E}f(W_n)=\mathbb{E}f(W)$ for all $f\in C_c^\infty(\mathbb{R})$, then $\lim_{n\to\infty}\mathbb{E}f(W_n)=\mathbb{E}f(W)$ for all bounded continuous functions, $f:\mathbb{R}\to\mathbb{R}$.*

Proof. According to Theorem 21.29 below it suffices to show $\lim_{n\to\infty}\mathbb{E}f(W_n)=\mathbb{E}f(W)$ for all $f\in C_c(\mathbb{R})$. For such a function, $f\in C_c(\mathbb{R})$, we may find[1] $f_k\in C_c^\infty(\mathbb{R})$ with all supports being contained in a compact subset of $\mathbb{R}$ such that $\varepsilon_k:=\sup_{x\in\mathbb{R}}|f(x)-f_k(x)|\to 0$ as $k\to\infty$. We then have,

$$\begin{aligned}|\mathbb{E}f(W)-\mathbb{E}f(W_n)|\le&|\mathbb{E}f(W)-\mathbb{E}f_k(W)|\\&+|\mathbb{E}f_k(W)-\mathbb{E}f_k(W_n)|+|\mathbb{E}f_k(W_n)-\mathbb{E}f(W_n)|\\\le&\mathbb{E}|f(W)-f_k(W)|\\&+|\mathbb{E}f_k(W)-\mathbb{E}f_k(W_n)|+\mathbb{E}|f_k(W_n)-f(W_n)|\\\le&2\varepsilon_k+|\mathbb{E}f_k(W)-\mathbb{E}f_k(W_n)|.\end{aligned}$$

Therefore it follows that

$$\begin{aligned}\limsup_{n\to\infty}|\mathbb{E}f(W)-\mathbb{E}f(W_n)|&\le 2\varepsilon_k+\limsup_{n\to\infty}|\mathbb{E}f_k(W)-\mathbb{E}f_k(W_n)|\\&=2\varepsilon_k\to 0\text{ as }k\to\infty.\end{aligned}$$

∎

Corollary 10.39. *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of independent random variables, then under the hypothesis on this sequence in either of Theorem 10.35 or Theorem 10.37 we have that $\lim_{n\to\infty}\mathbb{E}f\left(\bar{S}_n\right)=\mathbb{E}f(N(0,1))$ for all $f:\mathbb{R}\to\mathbb{R}$ which are bounded and continuous.*

For more on the methods employed in this section the reader is advised to look up "Stein's method." In Chapters 22 and 23 below, we will relax the assumptions in the above theorem. The proofs later will be based in the characteristic functional or equivalently the Fourier transform.

10.7 The Second Borel-Cantelli Lemma

Lemma 10.40. *If $0\le x\le\frac{1}{2}$, then*

$$e^{-2x}\le 1-x\le e^{-x}. \tag{10.33}$$

Moreover, the upper bound in Eq. (10.33) is valid for all $x\in\mathbb{R}$.

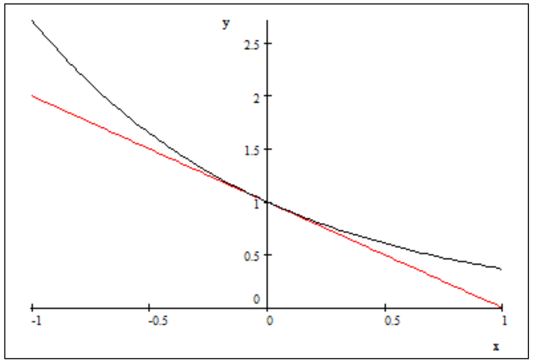


Fig. 10.1. A graph of $1-x$ and e^{-x} showing that $1-x \le e^{-x}$ for all x.

Proof. The upper bound follows by the convexity of e^{-x}, see Figure 10.1. For the lower bound we use the convexity of $\varphi(x) = e^{-2x}$ to conclude that the line joining $(0,1) = (0,\varphi(0))$ and $\left(1/2, e^{-1}\right) = (1/2, \varphi(1/2))$ lies above $\varphi(x)$ for $0 \le x \le 1/2$. Then we use the fact that the line $1-x$ lies above this line to conclude the lower bound in Eq. (10.33), see Figure 10.2. See Example 12.54

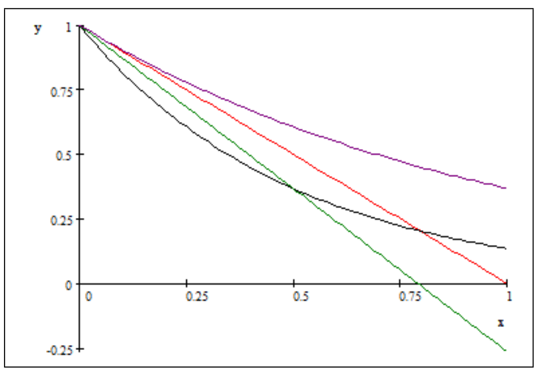


Fig. 10.2. A graph of $1-x$ (in red), the line joining $(0,1)$ and $\left(1/2, e^{-1}\right)$ (in green), e^{-x} (in purple), and e^{-2x} (in black) showing that $e^{-2x} \le 1-x \le e^{-x}$ for all $x \in [0,1/2]$.

below for a more formal proof of this lemma. ∎

For $\{a_n\}_{n=1}^\infty \subset [0,1]$, let

$$\prod_{n=1}^\infty (1-a_n) := \lim_{N\to\infty} \prod_{n=1}^N (1-a_n).$$

The limit exists since, $\prod_{n=1}^N (1-a_n)$ decreases as N increases.

Exercise 10.10. Show; if $\{a_n\}_{n=1}^\infty \subset [0,1)$, then

$$\prod_{n=1}^\infty (1-a_n) = 0 \iff \sum_{n=1}^\infty a_n = \infty.$$

The implication, $\Longleftarrow$, holds even if $a_n = 1$ is allowed.

Solution to Exercise (10.10). By Eq. (10.33) we always have,

$$\prod_{n=1}^N (1-a_n) \le \prod_{n=1}^N e^{-a_n} = \exp\left(-\sum_{n=1}^N a_n\right)$$

which upon passing to the limit as $N \to \infty$ gives

$$\prod_{n=1}^\infty (1-a_n) \le \exp\left(-\sum_{n=1}^\infty a_n\right).$$

Hence if $\sum_{n=1}^\infty a_n = \infty$ then $\prod_{n=1}^\infty (1-a_n) = 0$.

Conversely, suppose that $\sum_{n=1}^\infty a_n < \infty$. In this case $a_n \to 0$ as $n \to \infty$ and so there exists an $m \in \mathbb{N}$ such that $a_n \in [0,1/2]$ for all $n \ge m$. Therefore by Eq. (10.33), for any $N \ge m$,

$$\begin{aligned}\prod_{n=1}^N (1-a_n) &= \prod_{n=1}^m (1-a_n) \cdot \prod_{n=m+1}^N (1-a_n) \\ &\ge \prod_{n=1}^m (1-a_n) \cdot \prod_{n=m+1}^N e^{-2a_n} = \prod_{n=1}^m (1-a_n) \cdot \exp\left(-2\sum_{n=m+1}^N a_n\right) \\ &\ge \prod_{n=1}^m (1-a_n) \cdot \exp\left(-2\sum_{n=m+1}^\infty a_n\right).\end{aligned}$$

So again letting $N \to \infty$ shows,

$$\prod_{n=1}^\infty (1-a_n) \ge \prod_{n=1}^m (1-a_n) \cdot \exp\left(-2\sum_{n=m+1}^\infty a_n\right) > 0.$$

[1] We will eventually prove this standard real analysis fact later in the course.

Lemma 10.41 (Second Borel-Cantelli Lemma). *Suppose that* $\{A_n\}_{n=1}^{\infty}$ *are independent sets. If*

$$\sum_{n=1}^{\infty} P(A_n) = \infty, \tag{10.34}$$

then

$$P(\{A_n \text{ i.o.}\}) = 1. \tag{10.35}$$

Combining this with the first Borel Cantelli Lemma 7.14 gives the ***(Borel) Zero-One law****,*

$$P(A_n \text{ i.o.}) = \begin{cases} 0 & \text{if} \quad \sum_{n=1}^{\infty} P(A_n) < \infty \\ 1 & \text{if} \quad \sum_{n=1}^{\infty} P(A_n) = \infty \end{cases}.$$

Proof. We are going to prove Eq. (10.35) by showing,

$$0 = P(\{A_n \text{ i.o.}\}^c) = P(\{A_n^c \text{ a.a}\}) = P\left(\cup_{n=1}^{\infty} \cap_{k\geq n} A_k^c\right).$$

Since $\cap_{k\geq n} A_k^c \uparrow \cup_{n=1}^{\infty} \cap_{k\geq n} A_k^c$ as $n \to \infty$ and $\cap_{k=n}^{m} A_k^c \downarrow \cap_{n=1}^{\infty} \cup_{k\geq n} A_k$ as $m \to \infty$,

$$P\left(\cup_{n=1}^{\infty} \cap_{k\geq n} A_k^c\right) = \lim_{n\to\infty} P\left(\cap_{k\geq n} A_k^c\right) = \lim_{n\to\infty} \lim_{m\to\infty} P\left(\cap_{m\geq k\geq n} A_k^c\right).$$

Making use of the independence of $\{A_k\}_{k=1}^{\infty}$ and hence the independence of $\{A_k^c\}_{k=1}^{\infty}$, we have

$$P\left(\cap_{m\geq k\geq n} A_k^c\right) = \prod_{m\geq k\geq n} P(A_k^c) = \prod_{m\geq k\geq n} (1 - P(A_k)). \tag{10.36}$$

Using the upper estimate in Eq. (10.33) along with Eq. (10.36) shows

$$P\left(\cap_{m\geq k\geq n} A_k^c\right) \leq \prod_{m\geq k\geq n} e^{-P(A_k)} = \exp\left(-\sum_{k=n}^{m} P(A_k)\right).$$

Using Eq. (10.34), we find from the above inequality that $\lim_{m\to\infty} P\left(\cap_{m\geq k\geq n} A_k^c\right) = 0$ and hence

$$P\left(\cup_{n=1}^{\infty} \cap_{k\geq n} A_k^c\right) = \lim_{n\to\infty} \lim_{m\to\infty} P\left(\cap_{m\geq k\geq n} A_k^c\right) = \lim_{n\to\infty} 0 = 0$$

Note: we could also appeal to Exercise 10.10 above to give a proof of the Borel Zero-One law without appealing to the first Borel Cantelli Lemma. ∎

Example 10.42 (Example 7.15 continued). Suppose that $\{X_n\}$ are now independent Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. Then $P(\lim_{n\to\infty} X_n = 0) = 1$ iff $\sum p_n < \infty$. Indeed, $P(\lim_{n\to\infty} X_n = 0) = 1$ iff $P(X_n = 0 \text{ a.a.}) = 1$ iff $P(X_n = 1 \text{ i.o.}) = 0$ iff $\sum p_n = \sum P(X_n = 1) < \infty$.

Proposition 10.43 (Extremal behaviour of iid random variables). *Suppose that* $\{X_n\}_{n=1}^{\infty}$ *is a sequence of i.i.d. random variables and* c_n *is an increasing sequence of positive real numbers such that for all* $\alpha > 1$ *we have*

$$\sum_{n=1}^{\infty} P\left(X_1 > \alpha^{-1} c_n\right) = \infty \tag{10.37}$$

while

$$\sum_{n=1}^{\infty} P(X_1 > \alpha c_n) < \infty. \tag{10.38}$$

Then

$$\limsup_{n\to\infty} \frac{X_n}{c_n} = 1 \text{ a.s.} \tag{10.39}$$

Proof. By the second Borel-Cantelli Lemma, Eq. (10.37) implies

$$P\left(X_n > \alpha^{-1} c_n \text{ i.o. } n\right) = 1$$

from which it follows that

$$\limsup_{n\to\infty} \frac{X_n}{c_n} \geq \alpha^{-1} \text{ a.s.}.$$

Taking $\alpha = \alpha_k = 1 + 1/k$, we find

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} \geq 1\right) = P\left(\cap_{k=1}^{\infty} \left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \geq \frac{1}{\alpha_k}\right\}\right) = 1.$$

Similarly, by the first Borel-Cantelli lemma, Eq. (10.38) implies

$$P(X_n > \alpha c_n \text{ i.o. } n) = 0$$

or equivalently,

$$P(X_n \leq \alpha c_n \text{ a.a. } n) = 1.$$

That is to say,

$$\limsup_{n\to\infty} \frac{X_n}{c_n} \leq \alpha \text{ a.s.}$$

and hence working as above,

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} \leq 1\right) = P\left(\cap_{k=1}^{\infty} \left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \leq \alpha_k\right\}\right) = 1.$$

Hence,

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} = 1\right) = P\left(\left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \geq 1\right\} \cap \left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \leq 1\right\}\right) = 1.$$

∎

Example 10.44. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. standard normal random variables. Then by Mills' ratio (see Lemma 7.59),

$$P(X_n \geq \alpha c_n) \sim \frac{1}{\sqrt{2\pi}\alpha c_n} e^{-\alpha^2 c_n^2/2}.$$

Now, suppose that we take c_n so that

$$e^{-c_n^2/2} = \frac{1}{n} \implies c_n = \sqrt{2\ln(n)}.$$

It then follows that

$$P(X_n \geq \alpha c_n) \sim \frac{1}{\sqrt{2\pi}\alpha\sqrt{2\ln(n)}} e^{-\alpha^2 \ln(n)} = \frac{1}{2\alpha\sqrt{\pi\ln(n)}} \frac{1}{n^{-\alpha^2}}$$

and therefore

$$\sum_{n=1}^{\infty} P(X_n \geq \alpha c_n) = \infty \text{ if } \alpha < 1$$

and

$$\sum_{n=1}^{\infty} P(X_n \geq \alpha c_n) < \infty \text{ if } \alpha > 1.$$

Hence an application of Proposition 10.43 shows

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2\ln n}} = 1 \text{ a.s..}$$

Example 10.45. Let $\{E_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with exponential distributions determined by

$$P(E_n > x) = e^{-(x\vee 0)} \text{ or } P(E_n \leq x) = 1 - e^{-(x\vee 0)}.$$

(Observe that $P(E_n \leq 0) = 0$) so that $E_n > 0$ a.s.) Then for $c_n > 0$ and $\alpha > 0$, we have

$$\sum_{n=1}^{\infty} P(E_n > \alpha c_n) = \sum_{n=1}^{\infty} e^{-\alpha c_n} = \sum_{n=1}^{\infty} \left(e^{-c_n}\right)^{\alpha}.$$

Hence if we choose $c_n = \ln n$ so that $e^{-c_n} = 1/n$, then we have

$$\sum_{n=1}^{\infty} P(E_n > \alpha \ln n) = \sum_{n=1}^{\infty} \left(\frac{1}{n}\right)^{\alpha}$$

which is convergent iff $\alpha > 1$. So by Proposition 10.43, it follows that

$$\limsup_{n\to\infty} \frac{E_n}{\ln n} = 1 \text{ a.s.}$$

Example 10.46. * Suppose now that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. distributed by the Poisson distribution with intensity, λ, i.e.

$$P(X_1 = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

In this case we have

$$P(X_1 \geq n) = e^{-\lambda} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} \geq \frac{\lambda^n}{n!} e^{-\lambda}$$

and

$$\begin{aligned} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=n}^{\infty} \frac{n!}{k!} \lambda^{k-n} \\ &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{n!}{(k+n)!} \lambda^k \leq \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k = \frac{\lambda^n}{n!}. \end{aligned}$$

Thus we have shown that

$$\frac{\lambda^n}{n!} e^{-\lambda} \leq P(X_1 \geq n) \leq \frac{\lambda^n}{n!}.$$

Thus in terms of convergence issues, we may assume that

$$P(X_1 \geq x) \sim \frac{\lambda^x}{x!} \sim \frac{\lambda^x}{\sqrt{2\pi x} e^{-x} x^x}$$

wherein we have used Stirling's formula,

$$x! \sim \sqrt{2\pi x} e^{-x} x^x.$$

Now suppose that we wish to choose c_n so that

$$P(X_1 \geq c_n) \sim 1/n.$$

This suggests that we need to solve the equation, $x^x = n$. Taking logarithms of this equation implies that

$$x = \frac{\ln n}{\ln x}$$

and upon iteration we find,

$$\begin{aligned} x = \frac{\ln n}{\ln\left(\frac{\ln n}{\ln x}\right)} &= \frac{\ln n}{\ell_2(n) - \ell_2(x)} = \frac{\ln n}{\ell_2(n) - \ell_2\left(\frac{\ln n}{\ln x}\right)} \\ &= \frac{\ln n}{\ell_2(n) - \ell_3(n) + \ell_3(x)}. \end{aligned}$$

where $\ell_k = \overbrace{\ln \circ \ln \circ \dots \circ \ln}^{k\text{ - times}}$. Since, $x \le \ln(n)$, it follows that $\ell_3(x) \le \ell_3(n)$ and hence

$$x = \frac{\ln(n)}{\ell_2(n) + O(\ell_3(n))} = \frac{\ln(n)}{\ell_2(n)}\left(1 + O\left(\frac{\ell_3(n)}{\ell_2(n)}\right)\right).$$

Thus we are lead to take $c_n := \frac{\ln(n)}{\ell_2(n)}$. We then have, for $\alpha \in (0,\infty)$ that

$$\begin{aligned}(\alpha c_n)^{\alpha c_n} &= \exp(\alpha c_n [\ln \alpha + \ln c_n]) \\ &= \exp\left(\alpha \frac{\ln(n)}{\ell_2(n)}[\ln \alpha + \ell_2(n) - \ell_3(n)]\right) \\ &= \exp\left(\alpha\left[\frac{\ln \alpha - \ell_3(n)}{\ell_2(n)} + 1\right]\ln(n)\right) \\ &= n^{\alpha(1+\varepsilon_n(\alpha))}\end{aligned}$$

where

$$\varepsilon_n(\alpha) := \frac{\ln \alpha - \ell_3(n)}{\ell_2(n)}.$$

Hence we have

$$P(X_1 \ge \alpha c_n) \sim \frac{\lambda^{\alpha c_n}}{\sqrt{2\pi\alpha c_n} e^{-\alpha c_n}(\alpha c_n)^{\alpha c_n}} \sim \frac{(\lambda/e)^{\alpha c_n}}{\sqrt{2\pi\alpha c_n}} \frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}}.$$

Since

$$\ln(\lambda/e)^{\alpha c_n} = \alpha c_n \ln(\lambda/e) = \alpha \frac{\ln n}{\ell_2(n)} \ln(\lambda/e) = \ln n^{\alpha \frac{\ln(\lambda/e)}{\ell_2(n)}},$$

it follows that

$$(\lambda/e)^{\alpha c_n} = n^{\alpha \frac{\ln(\lambda/e)}{\ell_2(n)}}.$$

Therefore,

$$P(X_1 \ge \alpha c_n) \sim \frac{n^{\alpha \frac{\ln(\lambda/e)}{\ell_2(n)}}}{\sqrt{\frac{\ln(n)}{\ell_2(n)}}} \frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}} = \sqrt{\frac{\ell_2(n)}{\ln(n)}} \frac{1}{n^{\alpha(1+\delta_n(\alpha))}}$$

where $\delta_n(\alpha) \to 0$ as $n \to \infty$. From this observation, we may show,

$$\sum_{n=1}^{\infty} P(X_1 \ge \alpha c_n) < \infty \text{ if } \alpha > 1 \text{ and}$$
$$\sum_{n=1}^{\infty} P(X_1 \ge \alpha c_n) = \infty \text{ if } \alpha < 1$$

and so by Proposition 10.43 we may conclude that

$$\limsup_{n\to\infty} \frac{X_n}{\ln(n)/\ell_2(n)} = 1 \text{ a.s.}$$

10.8 Kolmogorov and Hewitt-Savage Zero-One Laws

Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables on a measurable space, $(\Omega, \mathcal{B})$. Let $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$, $\mathcal{B}_\infty := \sigma(X_1, X_2, \dots)$, $\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \dots)$, and $\mathcal{T} := \cap_{n=1}^\infty \mathcal{T}_n \subset \mathcal{B}_\infty$. We call $\mathcal{T}$ the **tail σ – field** and events, $A \in \mathcal{T}$, are called **tail events.**

Example 10.47. Let $S_n := X_1 + \dots + X_n$ and $\{b_n\}_{n=1}^\infty \subset (0,\infty)$ such that $b_n \uparrow \infty$. Here are some example of tail events and tail measurable random variables:

1. $\{\sum_{n=1}^\infty X_n \text{ converges}\} \in \mathcal{T}$. Indeed,
$$\left\{\sum_{k=1}^\infty X_k \text{ converges}\right\} = \left\{\sum_{k=n+1}^\infty X_k \text{ converges}\right\} \in \mathcal{T}_n$$
for all $n \in \mathbb{N}$.
2. Both $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are $\mathcal{T}$ – measurable as are $\limsup_{n\to\infty} \frac{S_n}{b_n}$ and $\liminf_{n\to\infty} \frac{S_n}{b_n}$.
3. $\{\lim X_n \text{ exists in } \bar{\mathbb{R}}\} = \left\{\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n\right\} \in \mathcal{T}$ and similarly,
$$\left\{\lim \frac{S_n}{b_n} \text{ exists in } \bar{\mathbb{R}}\right\} = \left\{\limsup_{n\to\infty} \frac{S_n}{b_n} = \liminf_{n\to\infty} \frac{S_n}{b_n}\right\} \in \mathcal{T}$$
and
$$\left\{\lim \frac{S_n}{b_n} \text{ exists in } \mathbb{R}\right\} = \left\{-\infty < \limsup_{n\to\infty} \frac{S_n}{b_n} = \liminf_{n\to\infty} \frac{S_n}{b_n} < \infty\right\} \in \mathcal{T}.$$
4. $\left\{\lim_{n\to\infty} \frac{S_n}{b_n} = 0\right\} \in \mathcal{T}$. Indeed, for any $k \in \mathbb{N}$,
$$\lim_{n\to\infty} \frac{S_n}{b_n} = \lim_{n\to\infty} \frac{(X_{k+1} + \dots + X_n)}{b_n}$$
from which it follows that $\left\{\lim_{n\to\infty} \frac{S_n}{b_n} = 0\right\} \in \mathcal{T}_k$ for all k.

Definition 10.48. *Let $(\Omega, \mathcal{B}, P)$ be a probability space. A σ – field, $\mathcal{F} \subset \mathcal{B}$ is* ***almost trivial*** *iff $P(\mathcal{F}) = \{0,1\}$, i.e. $P(A) \in \{0,1\}$ for all $A \in \mathcal{F}$.*

The following conditions on a sub-σ-algebra, $\mathcal{F} \subset \mathcal{B}$ are equivalent; 1) $\mathcal{F}$ is almost trivial, 2) $P(A) = P(A)^2$ for all $A \in \mathcal{F}$, and 3) $\mathcal{F}$ is independent of itself. For example if $\mathcal{F}$ is independent of itself, then $P(A) = P(A \cap A) = P(A)P(A)$ for all $A \in \mathcal{F}$ which implies $P(A) = 0$ or 1. If $\mathcal{F}$ is almost trivial and $A, B \in \mathcal{F}$, then $P(A \cap B) = 1 = P(A)P(B)$ if $P(A) = P(B) = 1$ and $P(A \cap B) = 0 = P(A)P(B)$ if either $P(A) = 0$ or $P(B) = 0$. Therefore $\mathcal{F}$ is independent of itself.

Lemma 10.49. *Suppose that $X : \Omega \to \bar{\mathbb{R}}$ is a random variable which is $\mathcal{F}$ measurable, where $\mathcal{F} \subset \mathcal{B}$ is almost trivial. Then there exists $c \in \bar{\mathbb{R}}$ such that $X = c$ a.s.*

Proof. Since $\{X = \infty\}$ and $\{X = -\infty\}$ are in $\mathcal{F}$, if $P(X = \infty) > 0$ or $P(X = -\infty) > 0$, then $P(X = \infty) = 1$ or $P(X = -\infty) = 1$ respectively. Hence, it suffices to finish the proof under the added condition that $P(X \in \mathbb{R}) = 1$.

For each $x \in \mathbb{R}$, $\{X \le x\} \in \mathcal{F}$ and therefore, $P(X \le x)$ is either 0 or 1. Since the function, $F(x) := P(X \le x) \in \{0,1\}$ is right continuous, non-decreasing and $F(-\infty) = 0$ and $F(+\infty) = 1$, there is a unique point $c \in \mathbb{R}$ where $F(c) = 1$ and $F(c-) = 0$. At this point, we have $P(X = c) = 1$.

Alternatively if $X : \Omega \to \mathbb{R}$ is an integrable $\mathcal{F}$ measurable random variable, we know that X is independent of itself and therefore X^2 is integrable and $\mathbb{E}X^2 = (\mathbb{E}X)^2 =: c^2$. Thus it follows that $\mathbb{E}\left[(X-c)^2\right] = 0$, i.e. $X = c$ a.s. For general $X : \Omega \to \mathbb{R}$, let $X_M := (M \wedge X) \vee (-M)$, then $X_M = \mathbb{E}X_M$ a.s. For sufficiently large M we know by MCT that $P(|X| < M) > 0$ and since $X = X_M = \mathbb{E}X_M$ a.s. on $\{|X| < M\}$, it follows that $c = \mathbb{E}X_M$ is constant independent of M for M large. Therefore, $X = \lim_{M\to\infty} X_M \overset{\text{a.s.}}{=} \lim_{M\to\infty} c = c$. ■

Proposition 10.50 (Kolmogorov's Zero-One Law). *Suppose that P is a probability measure on $(\Omega, \mathcal{B})$ such that $\{X_n\}_{n=1}^\infty$ are independent random variables. Then $\mathcal{T}$ is almost trivial, i.e. $P(A) \in \{0,1\}$ for all $A \in \mathcal{T}$. In particular the tail events in Example 10.47 have probability either 0 or 1.*

Proof. For each $n \in \mathbb{N}$, $\mathcal{T} \subset \sigma(X_{n+1}, X_{n+2}, \dots)$ which is independent of $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$. Therefore $\mathcal{T}$ is independent of $\cup\mathcal{B}_n$ which is a multiplicative system. Therefore $\mathcal{T}$ and is independent of $\mathcal{B}_\infty = \sigma(\cup\mathcal{B}_n) = \vee_{n=1}^\infty \mathcal{B}_n$. As $\mathcal{T} \subset \mathcal{B}_\infty$ it follows that $\mathcal{T}$ is independent of itself, i.e. $\mathcal{T}$ is almost trivial. ■

Corollary 10.51. *Keeping the assumptions in Proposition 10.50 and let $\{b_n\}_{n=1}^\infty \subset (0,\infty)$ such that $b_n \uparrow \infty$. Then $\limsup_{n\to\infty} X_n$, $\liminf_{n\to\infty} X_n$, $\limsup_{n\to\infty} \frac{S_n}{b_n}$, and $\liminf_{n\to\infty} \frac{S_n}{b_n}$ are all constant almost surely. In particular, either $P\left(\left\{\lim_{n\to\infty} \frac{S_n}{b_n} \text{ exists}\right\}\right) = 0$ or $P\left(\left\{\lim_{n\to\infty} \frac{S_n}{b_n} \text{ exists}\right\}\right) = 1$ and in the latter case $\lim_{n\to\infty} \frac{S_n}{b_n} = c$ a.s for some $c \in \bar{\mathbb{R}}$.*

Example 10.52. Suppose that $\{A_n\}_{n=1}^\infty$ are independent sets and let $X_n := 1_{A_n}$ for all n and $\mathcal{T} = \cap_{n\ge1}\sigma(X_n, X_{n+1}, \dots)$. Then $\{A_n \text{ i.o.}\} \in \mathcal{T}$ and therefore by the Kolmogorov 0-1 law, $P(\{A_n \text{ i.o.}\}) = 0$ or 1. Of course, in this case the Borel zero - one law (Lemma 10.41) tells when $P(\{A_n \text{ i.o.}\})$ is 0 and when it is 1 depending on whether $\sum_{n=1}^\infty P(A_n)$ is finite or infinite respectively.

10.8.1 Hewitt-Savage Zero-One Law

In this subsection, let $\Omega := \mathbb{R}^\infty = \mathbb{R}^\mathbb{N}$ and $X_n(\omega) = \omega_n$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$, and $\mathcal{B} := \sigma(X_1, X_2, \dots)$ be the product σ – algebra on Ω. We say a permutation (i.e. a bijective map on $\mathbb{N}$), $\pi : \mathbb{N} \to \mathbb{N}$ is finite if $\pi(n) = n$ for a.a. n. Define $T_\pi : \Omega \to \Omega$ by $T_\pi(\omega) = (\omega_{\pi 1}, \omega_{\pi 2}, \dots)$. Since $X_i \circ T_\pi(\omega) = \omega_{\pi i} = X_{\pi i}(\omega)$ for all i, it follows that T_π is $\mathcal{B}/\mathcal{B}$ – measurable.

Let us further suppose that μ is a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and let $P = \otimes_{n=1}^\infty \mu$ be the infinite product measure on $\left(\Omega = \mathbb{R}^\mathbb{N}, \mathcal{B}\right)$. Then $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables with $\mathrm{Law}_P(X_n) = \mu$ for all n. If $\pi : \mathbb{N} \to \mathbb{N}$ is a finite permutation and $A_i \in \mathcal{B}_\mathbb{R}$ for all i, then

$$T_\pi^{-1}(A_1 \times A_2 \times A_3 \times \dots) = A_{\pi^{-1}1} \times A_{\pi^{-1}2} \times \dots.$$

Since sets of the form, $A_1 \times A_2 \times A_3 \times \dots$, form a π – system generating $\mathcal{B}$ and

$$\begin{aligned} P \circ T_\pi^{-1}(A_1 \times A_2 \times A_3 \times \dots) &= \prod_{i=1}^\infty \mu(A_{\pi^{-1}i}) \\ &= \prod_{i=1}^\infty \mu(A_i) = P(A_1 \times A_2 \times A_3 \times \dots), \end{aligned}$$

we may conclude that $P \circ T_\pi^{-1} = P$.

Definition 10.53. *The **permutation invariant** σ – field, $\mathcal{S} \subset \mathcal{B}$, is the collection of sets, $A \in \mathcal{B}$ such that $T_\pi^{-1}(A) = A$ for all finite permutations π. (You should check that $\mathcal{S}$ is a σ – field!)*

Proposition 10.54 (Hewitt-Savage Zero-One Law). *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $P = \otimes_{n=1}^\infty \mu$ be the infinite product measure on $\left(\Omega = \mathbb{R}^\mathbb{N}, \mathcal{B}\right)$ so that $\{X_n\}_{n=1}^\infty$ (recall that $X_n(\omega) = \omega_n$) is an i.i.d. sequence with $\mathrm{Law}_P(X_n) = \mu$ for all n. Then $\mathcal{S}$ is P – almost trivial.*

Proof. Let $B \in \mathcal{S}$, $f = 1_B$, and $g = G(X_1, \dots, X_n)$ be a $\sigma(X_1, X_2, \dots, X_n)$ – measurable function such that $\sup_{\omega\in\Omega} |g(\omega)| \le 1$. Further let π be a finite permutation such that $\{\pi 1, \dots, \pi n\} \cap \{1, 2, \dots, n\} = \emptyset$ – for example we could take $\pi(j) = j + n$, $\pi(j+n) = j$ for $j = 1, 2, \dots, n$, and $\pi(j + 2n) = j + 2n$ for all $j \in \mathbb{N}$. Then $g \circ T_\pi = G(X_{\pi 1}, \dots, X_{\pi n})$ is independent of g and therefore,

$$(\mathbb{E}g)^2 = \mathbb{E}g \cdot \mathbb{E}[g \circ T_\pi] = \mathbb{E}[g \cdot g \circ T_\pi].$$

Since $f \circ T_\pi = 1_{T_\pi^{-1}(B)} = 1_B = f$, it follows that $\mathbb{E}f = \mathbb{E}f^2 = \mathbb{E}[f \cdot f \circ T_\pi]$ and therefore,

$$\left|\mathbb{E}f - (\mathbb{E}g)^2\right| = |\mathbb{E}\left[f \cdot f \circ T_\pi - g \cdot g \circ T_\pi\right]|$$
$$\le \mathbb{E}\left|[f-g]\, f \circ T_\pi\right| + \mathbb{E}\left|g\left[f \circ T_\pi - g \circ T_\pi\right]\right|$$
$$\le \mathbb{E}|f-g| + \mathbb{E}|f \circ T_\pi - g \circ T_\pi| = 2\mathbb{E}|f-g|. \tag{10.40}$$

According to Corollary 8.13 (or see Corollary 5.28 or Theorem 5.44 or Exercise 8.5)), we may choose $g = g_k$ as above with $\mathbb{E}|f - g_k| \to 0$ as $n \to \infty$ and so passing to the limit in Eq. (10.40) with $g = g_k$, we may conclude,

$$\left|P(B) - [P(B)]^2\right| = \left|\mathbb{E}f - (\mathbb{E}f)^2\right| \le 0.$$

That is $P(B) \in \{0,1\}$ for all $B \in \mathcal{S}$. ∎

In a nutshell, here is the crux of the above proof. First off we know that for $B \in \mathcal{S} \subset \mathcal{B}$, there exists g which is $\sigma(X_1, \dots, X_n)$ – measurable such that $f := 1_B \cong g$. Since $P \circ T_\pi^{-1} = P$ it also follows that $f = f \circ T_\pi \cong g \circ T_\pi$. For judiciously chosen π, we know that g and $g \circ T_\pi$ are independent. Therefore

$$\mathbb{E}f^2 = \mathbb{E}[f \cdot f \circ T_\pi] \cong \mathbb{E}[g \cdot g \circ T_\pi] = \mathbb{E}[g] \cdot \mathbb{E}[g \circ T_\pi] = (\mathbb{E}g)^2 \cong (\mathbb{E}f)^2.$$

As the approximation f by g may be made as accurate as we please, it follows that $P(B) = \mathbb{E}f^2 = (\mathbb{E}f)^2 = [P(B)]^2$ for all $B \in \mathcal{S}$.

Example 10.55 (Some Random Walk $0-1$ Law Results). Continue the notation in Proposition 10.54.

1. As above, if $S_n = X_1 + \dots + X_n$, then $P(S_n \in B \text{ i.o.}) \in \{0,1\}$ for all $B \in \mathcal{B}_\mathbb{R}$. Indeed, if π is a finite permutation,
$$T_\pi^{-1}(\{S_n \in B \text{ i.o.}\}) = \{S_n \circ T_\pi \in B \text{ i.o.}\} = \{S_n \in B \text{ i.o.}\}.$$
Hence $\{S_n \in B \text{ i.o.}\}$ is in the permutation invariant σ – field, $\mathcal{S}$. The same goes for $\{S_n \in B \text{ a.a.}\}$
2. If $P(X_1 \neq 0) > 0$, then $\limsup_{n\to\infty} S_n = \infty$ a.s. or $\limsup_{n\to\infty} S_n = -\infty$ a.s. Indeed,
$$T_\pi^{-1}\left\{\limsup_{n\to\infty} S_n \le x\right\} = \left\{\limsup_{n\to\infty} S_n \circ T_\pi \le x\right\} = \left\{\limsup_{n\to\infty} S_n \le x\right\}$$
which shows that $\limsup_{n\to\infty} S_n$ is $\mathcal{S}$ – measurable. Therefore, $\limsup_{n\to\infty} S_n = c$ a.s. for some $c \in \bar{\mathbb{R}}$. Since $(X_2, X_3, \dots) \overset{d}{=} (X_1, X_2, \dots)$ it follows (see Corollary 6.47 and Exercise 6.10) that
$$c = \limsup_{n\to\infty} S_n \overset{d}{=} \limsup_{n\to\infty}(X_2 + X_3 + \dots + X_{n+1})$$
$$= \limsup_{n\to\infty}(S_{n+1} - X_1) = \limsup_{n\to\infty} S_{n+1} - X_1 = c - X_1.$$

By Exercise 10.11 below we may now conclude that $c = c - X_1$ a.s. which is possible iff $c \in \{\pm\infty\}$ or $X_1 = 0$ a.s. Since the latter is not allowed, $\limsup_{n\to\infty} S_n = \infty$ or $\limsup_{n\to\infty} S_n = -\infty$ a.s.
3. Now assume that $P(X_1 \neq 0) > 0$ and $X_1 \overset{\mathrm{d}}{=} -X_1$, i.e. $P(X_1 \in A) = P(-X_1 \in A)$ for all $A \in \mathcal{B}_\mathbb{R}$. By 2. we know $\limsup_{n\to\infty} S_n = c$ a.s. with $c \in \{\pm\infty\}$. Since $\{X_n\}_{n=1}^\infty$ and $\{-X_n\}_{n=1}^\infty$ are i.i.d. and $-X_n \overset{\mathrm{d}}{=} X_n$, it follows that $\{X_n\}_{n=1}^\infty \overset{\mathrm{d}}{=} \{-X_n\}_{n=1}^\infty$ .The results of Exercises 6.10 and 10.11 then imply that $c \overset{\mathrm{d}}{=} \limsup_{n\to\infty} S_n \overset{\mathrm{d}}{=} \limsup_{n\to\infty}(-S_n)$ and in particular
$$c \overset{\text{a.s.}}{=} \limsup_{n\to\infty}(-S_n) = -\liminf_{n\to\infty} S_n \ge -\limsup_{n\to\infty} S_n = -c.$$
Since the $c = -\infty$ does not satisfy, $c \ge -c$, we must $c = \infty$. Hence in this symmetric case we have shown,
$$\limsup_{n\to\infty} S_n = \infty \text{ and } \liminf_{n\to\infty} S_n = -\infty \text{ a.s.}$$

Exercise 10.11. Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $Y : \Omega \to \bar{\mathbb{R}}$ is a random variable and $c \in \bar{\mathbb{R}}$ is a constant. Then $Y = c$ a.s. iff $Y \overset{d}{=} c$.

Solution to Exercise (10.11). If $Y = c$ a.s. then $P(Y \in A) = P(c \in A)$ for all $A \in \mathcal{B}_\mathbb{R}$ and therefore $Y \overset{d}{=} c$. Conversely, if $Y \overset{d}{=} c$, then $P(Y = c) = P(c = c) = 1$, i.e. $Y = c$ a.s.

10.9 Another Construction of Independent Random Variables*

This section may be skipped as the results are a special case of those given above. The arguments given here avoid the use of Kolmogorov's existence theorem for product measures.

Example 10.56. Suppose that $\Omega = \Lambda^n$ where Λ is a finite set, $\mathcal{B} = 2^\Omega$, $P(\{\omega\}) = \prod_{j=1}^n q_j(\omega_j)$ where $q_j : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda\in\Lambda} q_j(\lambda) = 1$. Let $\mathcal{C}_i := \left\{\Lambda^{i-1} \times A \times \Lambda^{n-i} : A \subset \Lambda\right\}$. Then $\{\mathcal{C}_i\}_{i=1}^n$ are independent. Indeed, if $B_i := \Lambda^{i-1} \times A_i \times \Lambda^{n-i}$, then

$$\cap B_i = A_1 \times A_2 \times \dots \times A_n$$

and we have

$$P(\cap B_i) = \sum_{\omega \in A_1 \times A_2 \times \dots \times A_n} \prod_{i=1}^n q_i(\omega_i) = \prod_{i=1}^n \sum_{\lambda \in A_i} q_i(\lambda)$$

while

$$P(B_i) = \sum_{\omega \in \Lambda^{i-1} \times A_i \times \Lambda^{n-i}} \prod_{i=1}^n q_i(\omega_i) = \sum_{\lambda \in A_i} q_i(\lambda).$$

Example 10.57. Continue the notation of Example 10.56 and further assume that $\Lambda \subset \mathbb{R}$ and let $X_i : \Omega \to \Lambda$ be defined by, $X_i(\omega) = \omega_i$. Then $\{X_i\}_{i=1}^n$ are independent random variables. Indeed, $\sigma(X_i) = \mathcal{C}_i$ with $\mathcal{C}_i$ as in Example 10.56.

Alternatively, from Exercise 4.10, we know that

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)]$$

for all $f_i : \Lambda \to \mathbb{R}$. Taking $A_i \subset \Lambda$ and $f_i := 1_{A_i}$ in the above identity shows that

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{E}_P\left[\prod_{i=1}^n 1_{A_i}(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[1_{A_i}(X_i)]$$
$$= \prod_{i=1}^n P(X_i \in A_i)$$

as desired.

Theorem 10.58 (Existence of i.i.d simple R.V.'s). *Suppose that* $\{q_i\}_{i=0}^n$ *is a sequence of positive numbers such that* $\sum_{i=0}^n q_i = 1$. *Then there exists a sequence* $\{X_k\}_{k=1}^\infty$ *of simple random variables taking values in* $\Lambda = \{0, 1, 2 \dots, n\}$ *on* $((0,1], \mathcal{B}, m)$ *such that*

$$m(\{X_1 = i_1, \dots, X_k = i_i\}) = q_{i_1} \dots q_{i_k}$$

for all $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$ *and all* $k \in \mathbb{N}$. *(See Example 10.15 above and Theorem 10.62 below for the general case of this theorem.)*

Proof. For $i = 0, 1, \dots, n$, let $\sigma_{-1} = 0$ and $\sigma_j := \sum_{i=0}^j q_i$ and for any interval, $(a, b]$, let

$$T_i((a,b]) := (a + \sigma_{i-1}(b-a), a + \sigma_i(b-a)].$$

Given $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$, let

$$J_{i_1, i_2, \dots, i_k} := T_{i_k}\left(T_{i_{k-1}}(\dots T_{i_1}((0,1]))\right)$$

and define $\{X_k\}_{k=1}^\infty$ on $(0,1]$ by

$$X_k := \sum_{i_1, i_2, \dots, i_k \in \{0,1,2,\dots,n\}} i_k 1_{J_{i_1, i_2, \dots, i_k}},$$

see Figure 10.3. Repeated applications of Corollary 6.27 shows the functions, $X_k : (0,1] \to \mathbb{R}$ are measurable.

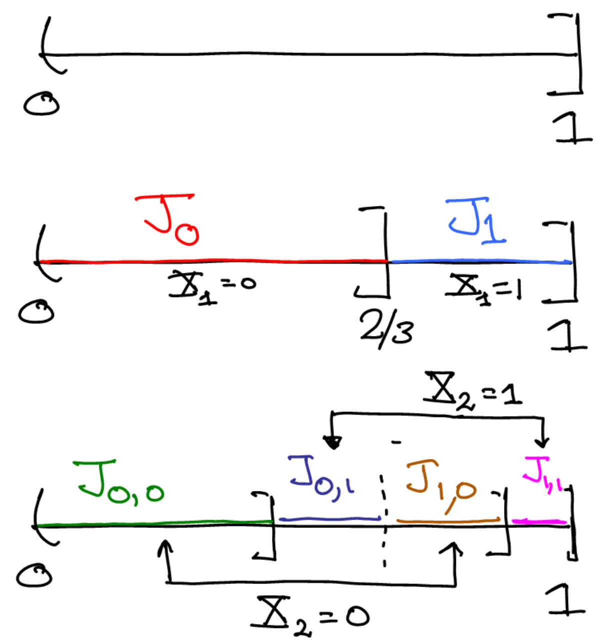


Fig. 10.3. Here we suppose that $p_0 = 2/3$ and $p_1 = 1/3$ and then we construct J_l and $J_{l,k}$ for $l, k \in \{0, 1\}$.

Observe that

$$m(T_i((a,b])) = q_i(b-a) = q_i m((a,b]), \tag{10.41}$$

and so by induction,

$$m(J_{i_1, i_2, \dots, i_k}) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}.$$

The reader should convince herself/himself that

$$\{X_1 = i_1, \dots X_k = i_i\} = J_{i_1, i_2, \dots, i_k}$$

and therefore, we have

$$m(\{X_1 = i_1, \dots, X_k = i_i\}) = m(J_{i_1,i_2,\dots,i_k}) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}$$

as desired. ■

Corollary 10.59 (Independent variables on product spaces). *Suppose $\Lambda = \{0,1,2\dots,n\}$, $q_i > 0$ with $\sum_{i=0}^n q_i = 1$, $\Omega = \Lambda^\infty = \Lambda^{\mathbb{N}}$, and for $i \in \mathbb{N}$, let $Y_i : \Omega \to \mathbb{R}$ be defined by $Y_i(\omega) = \omega_i$ for all $\omega \in \Omega$. Further let $\mathcal{B} := \sigma(Y_1, Y_2, \dots, Y_n, \dots)$. Then there exists a unique probability measure, $P : \mathcal{B} \to [0,1]$ such that*

$$P(\{Y_1 = i_1, \dots, Y_k = i_i\}) = q_{i_1} \dots q_{i_k}.$$

Proof. Let $\{X_i\}_{i=1}^n$ be as in Theorem 10.58 and define $T : (0,1] \to \Omega$ by

$$T(x) = (X_1(x), X_2(x), \dots, X_k(x), \dots).$$

Observe that T is measurable since $Y_i \circ T = X_i$ is measurable for all i. We now define, $P := T_* m$. Then we have

$$\begin{aligned} P(\{Y_1 = i_1, \dots, Y_k = i_i\}) &= m\left(T^{-1}(\{Y_1 = i_1, \dots, Y_k = i_i\})\right) \\ &= m(\{Y_1 \circ T = i_1, \dots, Y_k \circ T = i_i\}) \\ &= m(\{X_1 = i_1, \dots, X_k = i_i\}) = q_{i_1} \dots q_{i_k}. \end{aligned}$$

■

Theorem 10.60. *Given a finite subset, $\Lambda \subset \mathbb{R}$ and a function $q : \Lambda \to [0,1]$ such that $\sum_{\lambda\in\Lambda} q(\lambda) = 1$, there exists a probability space, $(\Omega, \mathcal{B}, P)$ and an independent sequence of random variables, $\{X_n\}_{n=1}^\infty$ such that $P(X_n = \lambda) = q(\lambda)$ for all $\lambda \in \Lambda$.*

Proof. Use Corollary 10.10 to shows that random variables constructed in Example 5.41 or Theorem 10.58 fit the bill. ■

Proposition 10.61. *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables with distribution, $P(X_n = 0) = P(X_n = 1) = \frac{1}{2}$. If we let $U := \sum_{n=1}^\infty 2^{-n} X_n$, then $P(U \le x) = (0 \vee x) \wedge 1$, i.e. U has the uniform distribution on $[0,1]$.*

Proof. Let us recall that $P(X_n = 0 \text{ a.a.}) = 0 = P(X_n = 1 \text{ a.a.})$. Hence we may, by shrinking Ω if necessary, assume that $\{X_n = 0 \text{ a.a.}\} = \emptyset = \{X_n = 1 \text{ a.a.}\}$. With this simplification, we have

$$\begin{aligned} \left\{U < \frac{1}{2}\right\} &= \{X_1 = 0\}, \\ \left\{U < \frac{1}{4}\right\} &= \{X_1 = 0, X_2 = 0\} \text{ and} \\ \left\{\frac{1}{2} \le U < \frac{3}{4}\right\} &= \{X_1 = 1, X_2 = 0\} \end{aligned}$$

and hence that

$$\begin{aligned} \left\{U < \frac{3}{4}\right\} &= \left\{U < \frac{1}{2}\right\} \cup \left\{\frac{1}{2} \le U < \frac{3}{4}\right\} \\ &= \{X_1 = 0\} \cup \{X_1 = 1, X_2 = 0\}. \end{aligned}$$

From these identities, it follows that

$$P(U < 0) = 0,\ P\left(U < \frac{1}{4}\right) = \frac{1}{4},\ P\left(U < \frac{1}{2}\right) = \frac{1}{2}, \text{ and } P\left(U < \frac{3}{4}\right) = \frac{3}{4}.$$

More generally, we claim that if $x = \sum_{j=1}^n \varepsilon_j 2^{-j}$ with $\varepsilon_j \in \{0,1\}$, then

$$P(U < x) = x. \tag{10.42}$$

The proof is by induction on n. Indeed, we have already verified (10.42) when $n = 1, 2$. Suppose we have verified (10.42) up to some $n \in \mathbb{N}$ and let $x = \sum_{j=1}^n \varepsilon_j 2^{-j}$ and consider

$$\begin{aligned} P\left(U < x + 2^{-(n+1)}\right) &= P(U < x) + P\left(x \le U < x + 2^{-(n+1)}\right) \\ &= x + P\left(x \le U < x + 2^{-(n+1)}\right). \end{aligned}$$

Since

$$\left\{x \le U < x + 2^{-(n+1)}\right\} = \left[\cap_{j=1}^n \{X_j = \varepsilon_j\}\right] \cap \{X_{n+1} = 0\}$$

we see that

$$P\left(x \le U < x + 2^{-(n+1)}\right) = 2^{-(n+1)}$$

and hence

$$P\left(U < x + 2^{-(n+1)}\right) = x + 2^{-(n+1)}$$

which completes the induction argument.

Since $x \to P(U < x)$ is left continuous we may now conclude that $P(U < x) = x$ for all $x \in (0,1)$ and since $x \to x$ is continuous we may also deduce that $P(U \le x) = x$ for all $x \in (0,1)$. Hence we may conclude that

$$P(U \le x) = (0 \vee x) \wedge 1.$$

■

We may now show the existence of independent random variables with arbitrary distributions.

Theorem 10.62. *Suppose that* $\{\mu_n\}_{n=1}^{\infty}$ *are a sequence of probability measures on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. *Then there exists a probability space,* $(\Omega, \mathcal{B}, P)$ *and a sequence* $\{Y_n\}_{n=1}^{\infty}$ *independent random variables with* $\mathrm{Law}\,(Y_n) := P \circ Y_n^{-1} = \mu_n$ *for all* n.

Proof. By Theorem 10.60, there exists a sequence of i.i.d. random variables, $\{Z_n\}_{n=1}^{\infty}$, such that $P(Z_n = 1) = P(Z_n = 0) = \frac{1}{2}$. These random variables may be put into a two dimensional array, $\{X_{i,j} : i, j \in \mathbb{N}\}$, see the proof of Lemma 3.8. For each i, let $U_i := \sum_{j=1}^{\infty} 2^{-i} X_{i,j}$ – $\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)$ – measurable random variable. According to Proposition 10.61, U_i is uniformly distributed on $[0,1]$. Moreover by the grouping Lemma 10.16, $\left\{\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)\right\}_{i=1}^{\infty}$ are independent σ – algebras and hence $\{U_i\}_{i=1}^{\infty}$ is a sequence of i.i.d.. random variables with the uniform distribution.

Finally, let $F_i(x) := \mu((-\infty, x])$ for all $x \in \mathbb{R}$ and let $G_i(y) = \inf\{x : F_i(x) \geq y\}$. Then according to Theorem 6.48, $Y_i := G_i(U_i)$ has μ_i as its distribution. Moreover each Y_i is $\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)$ – measurable and therefore the $\{Y_i\}_{i=1}^{\infty}$ are independent random variables. ■

11

The Standard Poisson Process

11.1 Poisson Random Variables

Recall from Exercise 7.5 that a Random variable, X, is Poisson distributed with intensity, a, if

$$P(X=k)=\frac{a^k}{k!}e^{-a} \text{ for all } k\in\mathbb{N}_0.$$

We will abbreviate this in the future by writing $X \overset{d}{=} \mathrm{Poi}(a)$. Let us also recall that

$$\mathbb{E}\left[z^X\right]=\sum_{k=0}^{\infty} z^k \frac{a^k}{k!}e^{-a}=e^{az}e^{-a}=e^{a(z-1)}$$

and as in Exercise 7.5 we have $\mathbb{E}X=a=\mathrm{Var}(X)$.

Lemma 11.1. *If $X=\mathrm{Poi}(a)$ and $Y=\mathrm{Poi}(b)$ and X and Y are independent, then $X+Y=\mathrm{Poi}(a+b)$.*

Proof. For $k\in\mathbb{N}_0$,

$$\begin{aligned}
P(X+Y=k)&=\sum_{l=0}^{k}P(X=l,Y=k-l)=\sum_{l=0}^{k}P(X=l)\,P(Y=k-l)\\
&=\sum_{l=0}^{k}e^{-a}\frac{a^l}{l!}e^{-b}\frac{b^{k-l}}{(k-l)!}=\frac{e^{-(a+b)}}{k!}\sum_{l=0}^{k}\binom{k}{l}a^l b^{k-l}\\
&=\frac{e^{-(a+b)}}{k!}(a+b)^k.
\end{aligned}$$

Alternative Proof. Notice that

$$\mathbb{E}\left[z^{X+Y}\right]=\mathbb{E}\left[z^X\right]\mathbb{E}\left[z^Y\right]=e^{a(z-1)}e^{b(z-1)}=\exp\left((a+b)(z-1)\right).$$

This suffices to complete the proof. ■

Lemma 11.2. *Suppose that $\{N_i\}_{i=1}^{\infty}$ are independent Poisson random variables with parameters, $\{\lambda_i\}_{i=1}^{\infty}$ such that $\sum_{i=1}^{\infty}\lambda_i=\infty$. Then $\sum_{i=1}^{\infty}N_i=\infty$ a.s.*

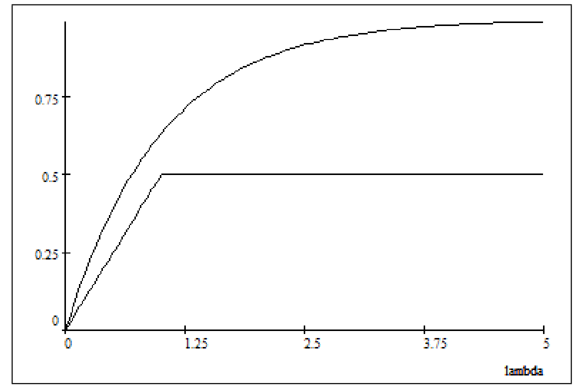


Fig. 11.1. This plot shows, $1-e^{-\lambda}\geq\frac{1}{2}(1\wedge\lambda)$.

Proof. From Figure 11.1 we see that $1-e^{-\lambda}\geq\frac{1}{2}(1\wedge\lambda)$ for all $\lambda\geq 0$. Therefore,

$$\sum_{i=1}^{\infty}P(N_i\geq 1)=\sum_{i=1}^{\infty}(1-P(N_i=0))=\sum_{i=1}^{\infty}\left(1-e^{-\lambda_i}\right)\geq\frac{1}{2}\sum_{i=1}^{\infty}\lambda_i\wedge 1=\infty$$

and so by the second Borel Cantelli Lemma, $P(\{N_i\geq 1 \text{ i.o.}\})=1$. From this it certainly follows that $\sum_{i=1}^{\infty}N_i=\infty$ a.s.

Alternatively, let $\Lambda_n=\lambda_1+\dots+\lambda_n$, then

$$P\left(\sum_{i=1}^{\infty}N_i\geq k\right)\geq P\left(\sum_{i=1}^{n}N_i\geq k\right)=1-e^{-\Lambda_n}\sum_{l=0}^{k-1}\frac{\Lambda_n^l}{l!}\to 1 \text{ as } n\to\infty.$$

Therefore $P\left(\sum_{i=1}^{\infty}N_i\geq k\right)=1$ for all $k\in\mathbb{N}$ and hence,

$$P\left(\sum_{i=1}^{\infty}N_i\geq\infty\right)=P\left(\cap_{k=1}^{\infty}\left\{\sum_{i=1}^{\infty}N_i\geq k\right\}\right)=1.$$

■

11.2 Exponential Random Variables

Recall from Definition 7.55 that $T \overset{d}{=} E(\lambda)$ is an exponential random variable with parameter $\lambda \in [0,\infty)$ provided, $P(T>t) = e^{-\lambda t}$ for all $t \geq 0$. We have seen that

$$\mathbb{E}\left[e^{aT}\right] = \frac{1}{1-a\lambda^{-1}} \text{ for } a<\lambda. \tag{11.1}$$

$\mathbb{E}T = \lambda^{-1}$ and $\operatorname{Var}(T) = \lambda^{-2}$, and (see Theorem 7.56) that T being exponential is characterized by the following memoryless property;

$$P(T>s+t|T>s) = P(T>t) \text{ for all } s,t\geq 0.$$

Theorem 11.3. *Let $\{T_j\}_{j=1}^\infty$ be independent random variables such that $T_j \overset{d}{=} E(\lambda_j)$ with $0<\lambda_j<\infty$ for all j. Then:*

1. *If $\sum_{n=1}^\infty \lambda_n^{-1} < \infty$ then $P(\sum_{n=1}^\infty T_n = \infty) = 0$ (i.e. $P(\sum_{n=1}^\infty T_n < \infty) = 1$).*
2. *If $\sum_{n=1}^\infty \lambda_n^{-1} = \infty$ then $P(\sum_{n=1}^\infty T_n = \infty) = 1$.*

(By Kolmogorov's zero-one law (see Proposition 10.50) it follows that $P(\sum_{n=1}^\infty T_n = \infty)$ is always either 0 or 1. We are showing here that $P(\sum_{n=1}^\infty T_n = \infty) = 1$ iff $\mathbb{E}[\sum_{n=1}^\infty T_n] = \infty$.)

Proof. 1. Since

$$\mathbb{E}\left[\sum_{n=1}^\infty T_n\right] = \sum_{n=1}^\infty \mathbb{E}[T_n] = \sum_{n=1}^\infty \lambda_n^{-1} < \infty$$

it follows that $\sum_{n=1}^\infty T_n < \infty$ a.s., i.e. $P(\sum_{n=1}^\infty T_n = \infty) = 0$.

2. By the DCT, independence, and Eq. (11.1) with $a=-1$,

$$\begin{aligned}\mathbb{E}\left[e^{-\sum_{n=1}^\infty T_n}\right] &= \lim_{N\to\infty} \mathbb{E}\left[e^{-\sum_{n=1}^N T_n}\right] = \lim_{N\to\infty}\prod_{n=1}^N \mathbb{E}\left[e^{-T_n}\right] \\ &= \lim_{N\to\infty}\prod_{n=1}^N\left(\frac{1}{1+\lambda_n^{-1}}\right) = \prod_{n=1}^\infty (1-a_n)\end{aligned}$$

where

$$a_n = 1 - \frac{1}{1+\lambda_n^{-1}} = \frac{1}{1+\lambda_n}.$$

Hence by Exercise 10.10, $\mathbb{E}\left[e^{-\sum_{n=1}^\infty T_n}\right] = 0$ iff $\infty = \sum_{n=1}^\infty a_n$ which happens iff $\sum_{n=1}^\infty \lambda_n^{-1} = \infty$ as you should verify. This completes the proof since $\mathbb{E}\left[e^{-\sum_{n=1}^\infty T_n}\right] = 0$ iff $e^{-\sum_{n=1}^\infty T_n} = 0$ a.s. or equivalently $\sum_{n=1}^\infty T_n = \infty$ a.s. ■

11.2.1 Appendix: More properties of Exponential random Variables*

Theorem 11.4. *Let I be a countable set and let $\{T_k\}_{k\in I}$ be independent random variables such that $T_k \sim E(q_k)$ with $q := \sum_{k\in I} q_k \in (0,\infty)$. Let $T := \inf_k T_k$ and let $K = k$ on the set where $T_j > T_k$ for all $j \neq k$. On the complement of all these sets, define $K = *$ where $*$ is some point not in I. Then $P(K=*) = 0$, K and T are independent, $T \sim E(q)$, and $P(K=k) = q_k/q$.*

Proof. Let $k\in I$ and $t\in\mathbb{R}_+$ and $\Lambda_n \subset_f I$ such that $\Lambda_n \uparrow I\setminus\{k\}$, then

$$\begin{aligned}P(K=k,T>t) &= P(\cap_{j\neq k}\{T_j>T_k\}, T_k>t) = \lim_{n\to\infty} P(\cap_{j\in\Lambda_n}\{T_j>T_k\}, T_k>t) \\ &= \lim_{n\to\infty}\int_{[0,\infty)^{\Lambda_n\cup\{k\}}} \prod_{j\in\Lambda_n} 1_{t_j>t_k}\cdot 1_{t_k>t}\, d\mu_n\left(\{t_j\}_{j\in\Lambda_n}\right) q_k e^{-q_k t_k} dt_k\end{aligned}$$

where μ_n is the joint distribution of $\{T_j\}_{j\in\Lambda_n}$. So by Fubini's theorem,

$$\begin{aligned}P(K=k,T>t) &= \lim_{n\to\infty}\int_t^\infty q_k e^{-q_k t_k} dt_k \int_{[0,\infty)^{\Lambda_n}} \prod_{j\in\Lambda_n} 1_{t_j>t_k}\cdot 1_{t_k>t}\, d\mu_n\left(\{t_j\}_{j\in\Lambda_n}\right) \\ &= \lim_{n\to\infty}\int_t^\infty P(\cap_{j\in\Lambda_n}\{T_j>t_k\}) q_k e^{-q_k t_k} dt_k \\ &= \int_t^\infty P(\cap_{j\neq k}\{T_j>\tau\}) q_k e^{-q_k\tau} d\tau \\ &= \int_t^\infty \prod_{j\neq k} e^{-q_j\tau} q_k e^{-q_k\tau} d\tau = \int_t^\infty \prod_{j\in I} e^{-q_j\tau} q_k d\tau \\ &= \int_t^\infty e^{-\sum_{j=1}^\infty q_j\tau} q_k d\tau = \int_t^\infty e^{-q\tau} q_k d\tau = \frac{q_k}{q} e^{-qt}. \end{aligned} \tag{11.2}$$

Taking $t=0$ shows that $P(K=k) = \frac{q_k}{q}$ and summing this on k shows $P(K\in I) = 1$ so that $P(K=*) = 0$. Moreover summing Eq. (11.2) on k now shows that $P(T>t) = e^{-qt}$ so that T is exponential. Moreover we have shown that

$$P(K=k,T>t) = P(K=k)\,P(T>t)$$

proving the desired independence. ■

Theorem 11.5. *Suppose that $S\sim E(\lambda)$ and $R\sim E(\mu)$ are independent. Then for $t\geq 0$ we have*

$$\mu P(S\leq t<S+R) = \lambda P(R\leq t<R+S).$$

Proof. We have

$$\begin{aligned}\mu P(S\le t<S+R) &= \mu\int_0^t \lambda e^{-\lambda s}P(t<s+R)\,ds\\ &= \mu\lambda\int_0^t e^{-\lambda s}e^{-\mu(t-s)}ds\\ &= \mu\lambda e^{-\mu t}\int_0^t e^{-(\lambda-\mu)s}ds = \mu\lambda e^{-\mu t}\cdot\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu}\\ &= \mu\lambda\cdot\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}\end{aligned}$$

which is symmetric in the interchanged of μ and λ.**Alternatively:**

$$\begin{aligned}P(S\le t<S+R) &= \lambda\mu\int_{\mathbb{R}_+^2} 1_{s\le t<s+r}e^{-\lambda s}e^{-\mu r}dsdr\\ &= \lambda\mu\int_0^t ds\int_{t-s}^\infty dre^{-\lambda s}e^{-\mu r}\\ &= \lambda\int_0^t dse^{-\lambda s}e^{-\mu(t-s)}\\ &= \lambda e^{-\mu t}\int_0^t dse^{-(\lambda-\mu)s}\\ &= \lambda e^{-\mu t}\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu}\\ &= \lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}.\end{aligned}$$

Therefore,

$$\mu P(S\le t<S+R)=\mu\lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}$$

which is symmetric in the interchanged of μ and λ and hence

$$\lambda P(R\le t<S+R)=\mu\lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}.$$

■

Example 11.6. Suppose T is a positive random variable such that $P(T\ge t+s|T\ge s)=P(T\ge t)$ for all $s,t\ge 0$, or equivalently

$$P(T\ge t+s)=P(T\ge t)\,P(T\ge s)\ \text{ for all } s,t\ge 0,$$

then $P(T\ge t)=e^{-at}$ for some $a>0$. (Such exponential random variables are often used to model "waiting times.") The distribution function for T is $F_T(t):=P(T\le t)=1-e^{-a(t\vee 0)}$. Since $F_T(t)$ is piecewise differentiable, the law of T, $\mu:=P\circ T^{-1}$, has a density,

$$d\mu(t)=F_T'(t)\,dt=ae^{-at}1_{t\ge 0}dt.$$

Therefore,

$$\mathbb{E}\left[e^{iaT}\right]=\int_0^\infty ae^{-at}e^{i\lambda t}dt=\frac{a}{a-i\lambda}=\hat{\mu}(\lambda).$$

Since

$$\hat{\mu}'(\lambda)=i\frac{a}{(a-i\lambda)^2}\ \text{and}\ \hat{\mu}''(\lambda)=-2\frac{a}{(a-i\lambda)^3}$$

it follows that

$$\mathbb{E}T=\frac{\hat{\mu}'(0)}{i}=a^{-1}\ \text{and}\ \mathbb{E}T^2=\frac{\hat{\mu}''(0)}{i^2}=\frac{2}{a^2}$$

and hence $\mathrm{Var}(T)=\frac{2}{a^2}-\left(\frac{1}{a}\right)^2=a^{-2}$.

11.3 The Standard Poisson Process

Let $\{T_k\}_{k=1}^\infty$ be an i.i.d. sequence of random exponential times with parameter λ, i.e. $P(T_k\in[t,t+dt])=\lambda e^{-\lambda t}dt$. For each $n\in\mathbb{N}$ let $W_n:=T_1+\cdots+T_n$ be the "**waiting time"** for the n^{th} event to occur. Because of Theorem 11.3 we know that $\lim_{n\to\infty}W_n=\infty$ a.s.

Definition 11.7 (Poisson Process I). *For any subset $A\subset\mathbb{R}_+$ let $N(A):=\sum_{n=1}^\infty 1_A(W_n)$ count the number of waiting times which occurred in A. When $A=(0,t]$ we will write, $N_t:=N((0,t])$ for all $t\ge 0$ and refer to $\{N_t\}_{t\ge 0}$ as the **Poisson Process with intensity** λ. (Observe that $\{N_t=n\}=W_n\le t<W_{n+1}$.)*

The next few results summarize a number of the basic properties of this Poisson process. Many of the proofs will be left as exercises to the reader. We will use the following notation below; for each $n\in\mathbb{N}$ and $T\ge 0$ let

$$\Delta_n(T):=\{(w_1,\dots,w_n)\in\mathbb{R}^n:0<w_1<w_2<\cdots<w_n<T\}$$

and let

$$\Delta_n:=\cup_{T>0}\Delta_n(T)=\{(w_1,\dots,w_n)\in\mathbb{R}^n:0<w_1<w_2<\cdots<w_n<\infty\}.$$

(We equip each of these spaces with their Borel σ − algebras.)

Exercise 11.1. Show $m_n(\Delta_n(T)) = T^n/n!$ where m_n is Lebesgue measure on $\mathcal{B}_{\mathbb{R}^n}$.

Exercise 11.2. If $n \in \mathbb{N}$ and $g : \Delta_n \to \mathbb{R}$ bounded (non-negative) measurable, then

$$\mathbb{E}[g(W_1, \dots, W_n)] = \int_{\Delta_n} g(w_1, w_2, \dots, w_n)\, \lambda^n e^{-\lambda w_n} dw_1 \dots dw_n. \tag{11.3}$$

As a simple corollary we have the following direct proof of Example 10.32.

Corollary 11.8. *If $n \in \mathbb{N}$, then $W_n \overset{d}{=} Gamma(n, \lambda^{-1})$.*

Proof. Taking $g(w_1, w_2, \dots, w_n) = f(w_n)$ in Eq. (11.3) we find with the aid of Exercise 11.1 that

$$\begin{aligned}\mathbb{E}[f(W_n)] &= \int_{\Delta_n} f(w_n)\, \lambda^n e^{-\lambda w_n} dw_1 \dots dw_n \\ &= \int_0^\infty f(w)\, \lambda^n \frac{w^{n-1}}{(n-1)!} e^{-\lambda w} dw\end{aligned}$$

which shows that $W_n \overset{d}{=} \mathrm{Gamma}(n, \lambda^{-1})$. ∎

Corollary 11.9. *If $t \in \mathbb{R}_+$ and $f : \Delta_n(t) \to \mathbb{R}$ is a bounded (or non-negative) measurable function, then*

$$\begin{aligned}&\mathbb{E}[f(W_1, \dots, W_n) : N_t = n] \\ &\quad = \lambda^n e^{-\lambda t} \int_{\Delta_n(t)} f(w_1, w_2, \dots, w_n)\, dw_1 \dots dw_n.\end{aligned} \tag{11.4}$$

Proof. Making use of the observation that $\{N_t = n\} = \{W_n \le t < W_{n+1}\}$, we may apply Eq. (11.3) at level $n+1$ with

$$g(w_1, w_2, \dots, w_{n+1}) = f(w_1, w_2, \dots, w_n)\, 1_{w_n \le t < w_{n+1}}$$

to learn

$$\begin{aligned}&\mathbb{E}[f(W_1, \dots, W_n) : N_t = n] \\ &\quad = \int_{0 < w_1 < \dots < w_n < t < w_{n+1}} f(w_1, w_2, \dots, w_n)\, \lambda^{n+1} e^{-\lambda w_{n+1}} dw_1 \dots dw_n dw_{n+1} \\ &\quad = \int_{\Delta_n(t)} f(w_1, w_2, \dots, w_n)\, \lambda^n e^{-\lambda t} dw_1 \dots dw_n.\end{aligned}$$

∎

Exercise 11.3. Show $N_t \overset{d}{=} \mathrm{Poi}(\lambda t)$ for all $t > 0$.

Definition 11.10 (Order Statistics). *Suppose that $X_1, \dots, X_n$ are non-negative random variables such that $P(X_i = X_j) = 0$ for all $i \ne j$. The **order statistics** of $X_1, \dots, X_n$ are the random variables, $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ defined by*

$$\tilde{X}_k = \max_{\#(\Lambda) = k} \min\{X_i : i \in \Lambda\} \tag{11.5}$$

where Λ always denotes a subset of $\{1, 2, \dots, n\}$ in Eq. (11.5).

The reader should verify that $\tilde{X}_1 \le \tilde{X}_2 \le \dots \le \tilde{X}_n$, $\{X_1, \dots, X_n\} = \left\{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n\right\}$ with repetitions, and that $\tilde{X}_1 < \tilde{X}_2 < \dots < \tilde{X}_n$ if $X_i \ne X_j$ for all $i \ne j$. In particular if $P(X_i = X_j) = 0$ for all $i \ne j$ then $P(\cup_{i \ne j}\{X_i = X_j\}) = 0$ and $\tilde{X}_1 < \tilde{X}_2 < \dots < \tilde{X}_n$ a.s.

Exercise 11.4. Suppose that $X_1, \dots, X_n$ are non-negative[1] random variables such that $P(X_i = X_j) = 0$ for all $i \ne j$. Show;

1. If $f : \Delta_n \to \mathbb{R}$ is bounded (non-negative) measurable, then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \dots, \tilde{X}_n\right)\right] = \sum_{\sigma \in S_n} \mathbb{E}[f(X_{\sigma 1}, \dots, X_{\sigma n}) : X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}], \tag{11.6}$$
where S_n is the permutation group on $\{1, 2, \dots, n\}$.
2. If we further assume that $\{X_1, \dots, X_n\}$ are i.i.d. random variables, then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \dots, \tilde{X}_n\right)\right] = n! \cdot \mathbb{E}[f(X_1, \dots, X_n) : X_1 < X_2 < \dots < X_n]. \tag{11.7}$$
(It is not important that $f\left(\tilde{X}_1, \dots, \tilde{X}_n\right)$ is not defined on the null set, $\cup_{i \ne j}\{X_i = X_j\}$.)
3. $f : \mathbb{R}_+^n \to \mathbb{R}$ is a bounded (non-negative) measurable symmetric function (i.e. $f(w_{\sigma 1}, \dots, w_{\sigma n}) = f(w_1, \dots, w_n)$ for all $\sigma \in S_n$ and $(w_1, \dots, w_n) \in \mathbb{R}_+^n$) then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \dots, \tilde{X}_n\right)\right] = \mathbb{E}[f(X_1, \dots, X_n)].$$
4. Suppose that $Y_1, \dots, Y_n$ is another collection of non-negative random variables such that $P(Y_i = Y_j) = 0$ for all $i \ne j$ such that
$$\mathbb{E}[f(X_1, \dots, X_n)] = \mathbb{E}[f(Y_1, \dots, Y_n)]$$

[1] The non-negativity of the X_i are not really necessary here but this is all we need to consider.

for all bounded (non-negative) measurable symmetric functions from $\mathbb{R}_+^n \to \mathbb{R}$. Show that $\left(\tilde{X}_1, \ldots, \tilde{X}_n\right) \overset{d}{=} \left(\tilde{Y}_1, \ldots, \tilde{Y}_n\right)$.
Hint: if $g : \Delta_n \to \mathbb{R}$ is a bounded measurable function, define $f : \mathbb{R}_+^n \to \mathbb{R}$ by;

$$f(y_1, \ldots, y_n) = \sum_{\sigma \in S_n} 1_{y_{\sigma 1} < y_{\sigma 2} < \cdots < y_{\sigma n}} g(y_{\sigma 1}, y_{\sigma 2}, \ldots, y_{\sigma n})$$

and then show f is symmetric.

Exercise 11.5. Let $t \in \mathbb{R}_+$ and $\{U_i\}_{i=1}^n$ be i.i.d. uniformly distributed random variables on $[0,t]$. Show that the order statistics, $\left(\tilde{U}_1, \ldots, \tilde{U}_n\right)$, of $(U_1, \ldots, U_n)$ has the same distribution as $(W_1, \ldots, W_n)$ given $N_t = n$. (Thus, given $N_t = n$, the collection of points, $\{W_1, \ldots, W_n\}$, has the same distribution as the collection of points, $\{U_1, \ldots, U_n\}$, in $[0,t]$.)

Theorem 11.11 (Joint Distributions). *If $\{A_i\}_{i=1}^k \subset \mathcal{B}_{[0,t]}$ is a partition of $[0,t]$, then $\{N(A_i)\}_{i=1}^k$ are independent random variables and $N(A) \overset{d}{=} \operatorname{Poi}(\lambda m(A))$ for all $A \in \mathcal{B}_{[0,t]}$ with $m(A) < \infty$. In particular, if $0 < t_1 < t_2 < \cdots < t_n$, then $\left\{N_{t_i} - N_{t_{i-1}}\right\}_{i=1}^n$ are independent random variables and $N_t - N_s \overset{d}{=} \operatorname{Poi}(\lambda(t-s))$ for all $0 \le s < t < \infty$. (We say that $\{N_t\}_{t\ge 0}$ is a stochastic process with **independent increments**.)*

Proof. If $z \in \mathbb{C}$ and $A \in \mathcal{B}_{[0,t]}$, then

$$z^{N(A)} = z^{\sum_{i=1}^n 1_A(W_i)} \text{ on } \{N_t = n\}.$$

Let $n \in \mathbb{N}$, $z_i \in \mathbb{C}$, and define

$$f(w_1, \ldots, w_n) = z_1^{\sum_{i=1}^n 1_{A_1}(w_i)} \ldots z_k^{\sum_{i=1}^n 1_{A_k}(w_i)}$$

which is a symmetric function. On $N_t = n$ we have,

$$z_1^{N(A_1)} \ldots z_k^{N(A_k)} = f(W_1, \ldots, W_n)$$

and therefore,

$$\begin{aligned}
\mathbb{E}\left[z_1^{N(A_1)} \ldots z_k^{N(A_k)} | N_t = n\right] &= \mathbb{E}\left[f(W_1, \ldots, W_n) | N_t = n\right] \\
&= \mathbb{E}\left[f(U_1, \ldots, U_n)\right] \\
&= \mathbb{E}\left[z_1^{\sum_{i=1}^n 1_{A_1}(U_i)} \ldots z_k^{\sum_{i=1}^n 1_{A_k}(U_i)}\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[\left(z_1^{1_{A_1}(U_i)} \ldots z_k^{1_{A_k}(U_i)}\right)\right] \\
&= \left(\mathbb{E}\left[\left(z_1^{1_{A_1}(U_1)} \ldots z_k^{1_{A_k}(U_1)}\right)\right]\right)^n \\
&= \left(\frac{1}{t}\sum_{i=1}^k m(A_i) \cdot z_i\right)^n,
\end{aligned}$$

wherein we have made use of the fact that $\{A_i\}_{i=1}^n$ is a partition of $[0,t]$ so that

$$z_1^{1_{A_1}(U_1)} \ldots z_k^{1_{A_k}(U_1)} = \sum_{i=1}^k z_i 1_{A_i}(U_i).$$

Thus it follows that

$$\begin{aligned}
\mathbb{E}\left[z_1^{N(A_1)} \ldots z_k^{N(A_k)}\right] &= \sum_{n=0}^\infty \mathbb{E}\left[z_1^{N(A_1)} \ldots z_k^{N(A_k)} | N_t = n\right] P(N_t = n) \\
&= \sum_{n=0}^\infty \left(\frac{1}{t}\sum_{i=1}^k m(A_i) \cdot z_i\right)^n \frac{(\lambda t)^n}{n!} e^{-\lambda t} \\
&= \sum_{n=0}^\infty \frac{1}{n!}\left(\lambda \sum_{i=1}^k m(A_i) \cdot z_i\right)^n e^{-\lambda t} \\
&= \exp\left(\lambda\left[\sum_{i=1}^k m(A_i) z_i - t\right]\right) \\
&= \exp\left(\lambda\left[\sum_{i=1}^k m(A_i)(z_i - 1)\right]\right).
\end{aligned}$$

From this result it follows that $\{N(A_i)\}_{i=1}^n$ are independent random variables and $N(A) = \operatorname{Poi}(\lambda m(A))$ for all $A \in \mathcal{B}_{\mathbb{R}}$ with $m(A) < \infty$.

Alternatively; suppose that $a_i \in \mathbb{N}_0$ and $n := a_1 + \cdots + a_k$, then

$$\begin{aligned}P[N(A_1)=a_1,\dots,N(A_k)=a_k|N_t=n] &= P\left[\sum_{i=1}^n 1_{A_l}(U_i)=a_l \text{ for } 1\le l\le k\right]\\ &= \frac{n!}{a_1!\dots a_k!}\prod_{l=1}^k\left[\frac{m(A_l)}{t}\right]^{a_l}\\ &= \frac{n!}{t^n}\cdot\prod_{l=1}^k\frac{[m(A_l)]^{a_l}}{a_l!}\end{aligned}$$

and therefore,

$$\begin{aligned}&P[N(A_1)=a_1,\dots,N(A_k)=a_k]\\ &\quad= P[N(A_1)=a_1,\dots,N(A_k)=a_k|N_t=n]\cdot P(N_t=n)\\ &\quad= \frac{n!}{t^n}\cdot\prod_{l=1}^k\frac{[m(A_l)]^{a_l}}{a_l!}\cdot e^{-\lambda t}\frac{(\lambda t)^n}{n!}\\ &\quad= \prod_{l=1}^k\frac{[m(A_l)]^{a_l}}{a_l!}\cdot e^{-\lambda t}\lambda^n\\ &\quad= \prod_{l=1}^k\frac{[m(A_l)\lambda]^{a_l}}{a_l!}e^{-\lambda a_l}\end{aligned}$$

which shows that $\{N(A_l)\}_{l=1}^k$ are independent and that $N(A_l)\overset{d}{=}\mathrm{Poi}(\lambda m(A_l))$ for each l. ∎

Remark 11.12. If $A\in\mathcal{B}_{[0,\infty)}$ with $m(A)=\infty$, then $N(A)=\infty$ a.s. To prove this observe that $N(A)=\uparrow\lim_{n\to\infty}N(A\cap[0,n])$. Therefore for any $k\in\mathbb{N}$, we have

$$\begin{aligned}P(N(A)\ge k)&\ge P(N(A\cap[0,n])\ge k)\\ &= 1-e^{-\lambda m(A\cap[0,n])}\sum_{0\le l<k}\frac{(\lambda m(A\cap[0,n]))^l}{l!}\to 1 \text{ as } n\to\infty.\end{aligned}$$

This shows that $N(A)\ge k$ a.s. for all $k\in\mathbb{N}$, i.e. $N(A)=\infty$ a.s.

Exercise 11.6 (A Generalized Poisson Process I). Suppose that $(S,\mathcal{B}_S,\mu)$ is a finite measure space with $\mu(S)<\infty$. Define $\Omega=\sum_{n=0}^\infty S^n$ where $S^0=\{*\}$, were $*$ is some arbitrary point. Define $\mathcal{B}_\Omega$ to be those sets, $B=\sum_{n=0}^\infty B_n$ where $B_n\in\mathcal{B}_{S^n}:=\mathcal{B}_S^{\otimes n}$ – the product σ – algebra on S^n. Now define a probability measure, P, on $(\Omega,\mathcal{B}_\Omega)$ by

$$P(B):=e^{-\mu(S)}\sum_{n=0}^\infty\frac{1}{n!}\mu^{\otimes n}(B_n)$$

where $\mu^{\otimes 0}(\{*\})=1$ by definition. (We denote P schematically by $P:=e^{-\mu(S)}e^{\mu\otimes}$.) Finally for ever $\omega\in\Omega$, let N_ω, be the point measure on $(S,\mathcal{B}_S)$ defined by; $N_*=0$ and

$$N_\omega=\sum_{i=1}^n\delta_{s_i}\text{ if }\omega=(s_1,\dots,s_n)\in S^n\text{ for }n\ge 1.$$

So for $A\in\mathcal{B}_S$, we have $N_*(A)=0$ and $N_\omega(A)=\sum_{i=1}^n 1_A(s_i)$. Show;

1. For each $A\in\mathcal{B}_S$, $\omega\to N_\omega(A)$ is a Poisson random variable with intensity $\mu(A)$, i.e. $N(A)=\mathrm{Poi}(\mu(A))$.
2. If $\{A_k\}_{k=1}^m\subset\mathcal{B}_S$ are disjoint sets, the $\{\omega\to N_\omega(A_k)\}_{k=1}^m$ are independent random variables.

An integer valued random measure on $(S,\mathcal{B}_S)$ $(\Omega\ni\omega\to N_\omega)$ satisfying properties 1. and 2. of Exercise 11.6 is called a **Poisson process on** $(S,\mathcal{B}_S)$ **with intensity measure** μ. For more motivation as to why Poisson processes are important see Proposition 21.11 below.

Exercise 11.7 (A Generalized Poisson Process II). Let $(S,\mathcal{B}_S,\mu)$ be as in Exercise 11.6, $\{Y_i\}_{i=1}^\infty$ be i.i.d. S – valued Random variables with $\mathrm{Law}_P(Y_i)=\mu(\cdot)/\mu(S)$ and ν be a $\mathrm{Poi}(\mu(S))$ – random variable which is independent of $\{Y_i\}$. Show $N:=\sum_{i=1}^\nu\delta_{Y_i}$ is a Poisson process on $(S,\mathcal{B}_S)$ with intensity measure, μ.

Exercise 11.8 (A Generalized Poisson Process III). Suppose now that $(S,\mathcal{B}_S,\mu)$ is a σ – finite measure space and $S=\sum_{l=1}^\infty S_l$ is a partition of S such that $0<\mu(S_l)<\infty$ for all l. For each $l\in\mathbb{N}$, using either of the construction above we may construct a Poisson point process, N_l, on $(S,\mathcal{B}_S)$ with intensity measure, μ_l where $\mu_l(A):=\mu(A\cap S_l)$ for all $A\in\mathcal{B}_S$. We do this in such a what that $\{N_l\}_{l=1}^\infty$ are all **independent**. Show that $N:=\sum_{l=1}^\infty N_l$ is a Poisson point process on $(S,\mathcal{B}_S)$ with intensity measure, μ. To be more precise observe that N is a random measure on $(S,\mathcal{B}_S)$ which satisfies (as you should show);

1. For each $A\in\mathcal{B}_S$ with $\mu(A)<\infty$, show $N(A)\overset{d}{=}\mathrm{Poi}(\mu(A))$.
2. If $\{A_k\}_{k=1}^m\subset\mathcal{B}_S$ are disjoint sets with $\mu(A_k)<\infty$, show $\{N(A_k)\}_{k=1}^m$ are independent random variables.
3. If $A\in\mathcal{B}_S$ with $\mu(A)=\infty$, show $N(A)=\infty$ a.s.

11.4 Poission Process Extras*

(This subsection still needs work!) In Definition 11.7 we really gave a construction of a Poisson process as defined in Definition 11.13. The goal of this section is to show that the Poisson process, $\{N_t\}_{t\geq 0}$, as defined in Definition 11.13 is uniquely determined and is essentially equivalent to what we have already done above.

Definition 11.13 (Poisson Process II). *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $N_t : \Omega \to \mathbb{N}_0$ be a random variable for each $t \geq 0$. We say that $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ if; 1) $N_0 = 0$, 2) $N_t - N_s \overset{d}{=} \operatorname{Poi}(\lambda(t-s))$ for all $0 \leq s < t < \infty$, 3) $\{N_t\}_{t\geq 0}$ has independent increments, and 4) $t \to N_t(\omega)$ is right continuous and non-decreasing for all $\omega \in \Omega$.*

Let $N_\infty(\omega) :=\uparrow \lim_{t\uparrow\infty} N_t(\omega)$ and observe that $N_\infty = \sum_{k=0}^\infty (N_k - N_{k-1}) = \infty$ a.s. by Lemma 11.2. Therefore, we may and do assume that $N_\infty(\omega) = \infty$ for all $\omega \in \Omega$.

Lemma 11.14. *There is zero probability that $\{N_t\}_{t\geq 0}$ makes a jump greater than or equal to 2.*

Proof. Suppose that $T \in (0, \infty)$ is fixed and $\omega \in \Omega$ is sample point where $t \to N_t(\omega)$ makes a jump of 2 or more for $t \in [0, T]$. Then for all $n \in \mathbb{N}$ we must have $\omega \in \cup_{k=1}^n \left\{N_{\frac{k}{n}T} - N_{\frac{k-1}{n}T} \geq 2\right\}$. Therefore,

$$\begin{aligned} &P^*(\{\omega : [0,T] \ni t \to N_t(\omega) \text{ has jump } \geq 2\}) \\ &\quad\leq \sum_{k=1}^n P\left(N_{\frac{k}{n}T} - N_{\frac{k-1}{n}T} \geq 2\right) = \sum_{k=1}^n O\left(T^2/n^2\right) = O(1/n) \to 0 \end{aligned}$$

as $n \to \infty$. I am leaving open the possibility that the set of ω where a jump size 2 or larger is not measurable. ■

Theorem 11.15. *Suppose that $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ as in Definition 11.13,*

$$W_n := \inf\{t : N_t = n\} \text{ for all } n \in \mathbb{N}_0$$

be the first time N_t reaches n. (The $\{W_n\}_{n=0}^\infty$ are well defined off a set of measure zero and $W_n < W_{n+1}$ for all n by the right continuity of $\{N_t\}_{t\geq 0}$.) Then the $\{T_n := W_n - W_{n-1}\}_{n=1}^\infty$ are i.i.d. $E(\lambda)$ – random variables. Thus the two descriptions of a Poisson process given in Definitions 11.7 and 11.13 are equivalent.

Proof. Suppose that $J_i = (a_i, b_i]$ with $b_i \leq a_{i+1} < \infty$ for all i. We will begin by showing

$$\begin{aligned} P\left(\cap_{i=1}^n \{W_i \in J_i\}\right) &= \lambda^n \prod_{i=1}^{n-1} m(J_i) \cdot \int_{J_n} e^{-\lambda w_n} dw_n & (11.8) \\ &= \lambda^n \int_{J_1\times J_2\times\dots\times J_n} e^{-\lambda w_n} dw_1 \dots dw_n. & (11.9) \end{aligned}$$

To show this let $K_i := (b_{i-1}, a_i]$ where $b_0 = 0$. Then

$$\cap_{i=1}^n \{W_i \in J_i\} = \cap_{i=1}^n \{N(K_i) = 0\} \cap \cap_{i=1}^{n-1} \{N(J_i) = 0\} \cap \{N(J_n) \geq 2\}$$

and therefore,

$$\begin{aligned} P\left(\cap_{i=1}^n \{W_i \in J_i\}\right) &= \prod_{i=1}^n e^{-\lambda m(K_i)} \cdot \prod_{i=1}^{n-1} e^{-\lambda m(J_i)} \lambda m(J_i) \cdot \left(1 - e^{-\lambda m(J_n)}\right) \\ &= \lambda^{n-1} \prod_{i=1}^{n-1} m(J_i) \cdot \left[e^{-\lambda a_n} - e^{-\lambda b_n}\right] \\ &= \lambda^{n-1} \prod_{i=1}^{n-1} m(J_i) \cdot \int_{J_n} \lambda e^{-\lambda w_n} dw_n. \end{aligned}$$

We may now apply a $\pi - \lambda$ – argument, using $\sigma(\{J_1 \times \dots \times J_n\}) = \mathcal{B}_{\Delta_n}$, to show

$$\mathbb{E}[g(W_1, \dots, W_n)] = \int_{\Delta_n} g(w_1, \dots, w_n) \lambda^n e^{-\lambda w_n} dw_1 \dots dw_n$$

holds for all bounded $\mathcal{B}_{\Delta_n}/\mathcal{B}_{\mathbb{R}}$ measurable functions, $g : \Delta_n \to \mathbb{R}$. Undoing the change of variables you made in Exercise 11.2 allows us to conclude that $\{T_n\}_{n=1}^\infty$ are i.i.d. $E(\lambda)$ – distributed random variables. ■

L^p – spaces

Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and for $0 < p < \infty$ and a measurable function $f : \Omega \to \mathbb{C}$ let

$$\|f\|_p := \left(\int_\Omega |f|^p \, d\mu \right)^{1/p} \tag{12.1}$$

and when $p = \infty$, let

$$\|f\|_\infty = \inf \{a \geq 0 : \mu(|f| > a) = 0\} \tag{12.2}$$

For $0 < p \leq \infty$, let

$$L^p(\Omega, \mathcal{B}, \mu) = \{f : \Omega \to \mathbb{C} : f \text{ is measurable and } \|f\|_p < \infty\} / \sim$$

where $f \sim g$ iff $f = g$ a.e. Notice that $\|f - g\|_p = 0$ iff $f \sim g$ and if $f \sim g$ then $\|f\|_p = \|g\|_p$. In general we will (by abuse of notation) use f to denote both the function f and the equivalence class containing f.

Remark 12.1. Suppose that $\|f\|_\infty \leq M$, then for all $a > M$, $\mu(|f| > a) = 0$ and therefore $\mu(|f| > M) = \lim_{n\to\infty} \mu(|f| > M + 1/n) = 0$, i.e. $|f(\omega)| \leq M$ for μ - a.e. ω. Conversely, if $|f| \leq M$ a.e. and $a > M$ then $\mu(|f| > a) = 0$ and hence $\|f\|_\infty \leq M$. This leads to the identity:

$$\|f\|_\infty = \inf \{a \geq 0 : |f(\omega)| \leq a \text{ for } \mu - \text{a.e. } \omega\}.$$

12.1 Modes of Convergence

Let $\{f_n\}_{n=1}^\infty \cup \{f\}$ be a collection of complex valued measurable functions on Ω. We have the following notions of convergence and Cauchy sequences.

Definition 12.2. *1. $f_n \to f$ a.e. if there is a set $E \in \mathcal{B}$ such that $\mu(E) = 0$ and $\lim_{n\to\infty} 1_{E^c} f_n = 1_{E^c} f$.*
2. $f_n \to f$ in μ – measure if $\lim_{n\to\infty} \mu(|f_n - f| > \varepsilon) = 0$ for all $\varepsilon > 0$. We will abbreviate this by saying $f_n \to f$ in L^0 or by $f_n \xrightarrow{\mu} f$.
3. $f_n \to f$ in L^p iff $f \in L^p$ and $f_n \in L^p$ for all n, and $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Definition 12.3. *1. $\{f_n\}$ is a.e. Cauchy if there is a set $E \in \mathcal{B}$ such that $\mu(E) = 0$ and$\{1_{E^c} f_n\}$ is a pointwise Cauchy sequences.*
2. $\{f_n\}$ is Cauchy in μ – measure (or L^0 – Cauchy) if $\lim_{m,n\to\infty} \mu(|f_n - f_m| > \varepsilon) = 0$ for all $\varepsilon > 0$.
3. $\{f_n\}$ is Cauchy in L^p if $\lim_{m,n\to\infty} \|f_n - f_m\|_p = 0$.

When μ is a probability measure, we describe, $f_n \xrightarrow{\mu} f$ as f_n **converging to f in probability.** If a sequence $\{f_n\}_{n=1}^\infty$ is L^p – convergent, then it is L^p – Cauchy. For example, when $p \in [1, \infty]$ and $f_n \to f$ in L^p, we have (using Minikowski's inequality of Theorem 12.22 below)

$$\|f_n - f_m\|_p \leq \|f_n - f\|_p + \|f - f_m\|_p \to 0 \text{ as } m, n \to \infty.$$

The case where $p = 0$ will be handled in Theorem 12.8 below.

Lemma 12.4 (L^p – convergence implies convergence in probability). *Let $p \in [1, \infty)$. If $\{f_n\} \subset L^p$ is L^p – convergent (Cauchy) then $\{f_n\}$ is also convergent (Cauchy) in measure.*

Proof. By Chebyshev's inequality (7.2),

$$\mu(|f| \geq \varepsilon) = \mu(|f|^p \geq \varepsilon^p) \leq \frac{1}{\varepsilon^p} \int_\Omega |f|^p \, d\mu = \frac{1}{\varepsilon^p} \|f\|_p^p$$

and therefore if $\{f_n\}$ is L^p – Cauchy, then

$$\mu(|f_n - f_m| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \|f_n - f_m\|_p^p \to 0 \text{ as } m, n \to \infty$$

showing $\{f_n\}$ is L^0 – Cauchy. A similar argument holds for the L^p – convergent case. ■

Example 12.5. Let us consider a number of examples here to get a feeling for these different notions of convergence. In each of these examples we will work in the measure space, $\left(\mathbb{R}_+, \mathcal{B} = \mathcal{B}_{\mathbb{R}_+}, m\right)$.

1. Let $f_n = \frac{1}{n} 1_{[0,n]}$ as in Figure 12.1. In this case $f_n \nrightarrow 0$ in L^1 but $f_n \to 0$ a.e.,$f_n \to 0$ in L^p for all $p > 1$ and $f_n \xrightarrow{m} 0$.

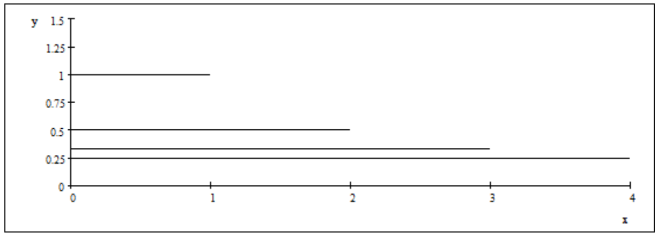
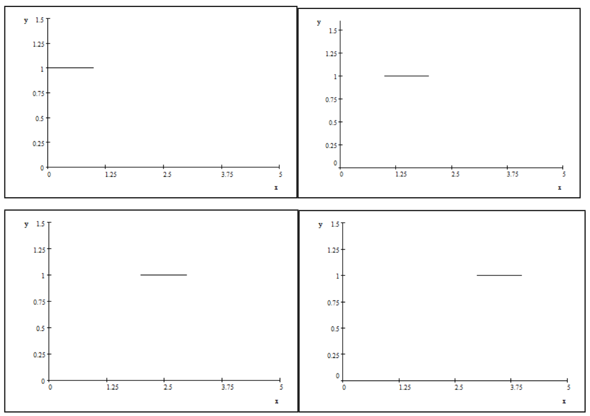


Fig. 12.1. Graphs of $f_n = \frac{1}{n} 1_{[0,n]}$ for $n = 1, 2, 3, 4$.

2. Let $f_n = 1_{[n-1,n]}$ as in the figure below. Then $f_n \to 0$ a.e., yet $f_n \nrightarrow 0$ in any L^p –space or in measure.



3. Now suppose that $f_n = n \cdot 1_{[0,1/n]}$ as in Figure 12.2. In this case $f_n \to 0$ a.e., $f_n \overset{m}{\to} 0$ but $f_n \nrightarrow 0$ in L^1 or in any L^p for $p \geq 1$. Observe that $\|f_n\|_p = n^{1-1/p}$ for all $p \geq 1$.

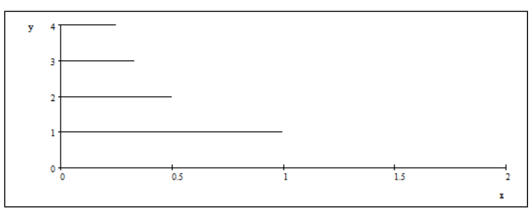
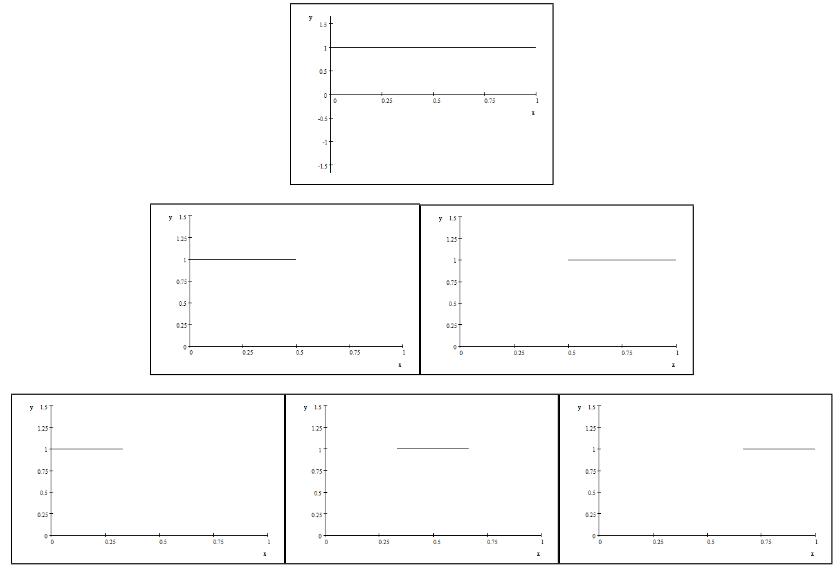


Fig. 12.2. Graphs of $f_n = n \cdot 1_{[0,n]}$ for $n = 1, 2, 3, 4$.

4. For $n \in \mathbb{N}$ and $1 \leq k \leq n$, let $g_{n,k} := 1_{(\frac{k-1}{n}, \frac{k}{n}]}$. Then define $\{f_n\}$ as

$$(f_1, f_2, f_3, \dots) = (g_{1,1}, g_{2,1}, g_{2,2}, g_{3,1}, g_{3,2}, g_{3,3}, g_{4,1}, g_{4,2}, g_{4,3}, g_{4,4}, \dots)$$

as depicted in the figures below.



For this sequence of functions we have $f_n \to 0$ in L^p for all $1 \leq p < \infty$ and $f_n \overset{m}{\to} 0$ but $f_n \nrightarrow 0$ a.e. and $f_n \nrightarrow 0$ in L^∞. In this case, $\|g_{n,k}\|_p = \left(\frac{1}{n}\right)^{1/p}$ for $1 \leq p < \infty$ while $\|g_{n,k}\|_\infty = 1$ for all n, k.

12.2 Almost Everywhere and Measure Convergence

Theorem 12.6 (Egorov: *a.s.* $\implies$ **convergence in probability).** *Suppose $\mu(\Omega) = 1$ and $f_n \to f$ a.s. Then for all $\varepsilon > 0$ there exists $E = E_\varepsilon \in \mathcal{B}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. In particular $f_n \xrightarrow{\mu} f$ as $n \to \infty$.*

Proof. Let $f_n \to f$ a.e. Then for all $\varepsilon > 0$,

$$\begin{aligned} 0 &= \mu(\{|f_n - f| > \varepsilon \text{ i.o. } n\}) \\ &= \lim_{N\to\infty} \mu\left(\bigcup_{n \geq N} \{|f_n - f| > \varepsilon\}\right) \quad (12.3) \\ &\geq \limsup_{N\to\infty} \mu\left(\{|f_N - f| > \varepsilon\}\right) \end{aligned}$$

from which it follows that $f_n \xrightarrow{\mu} f$ as $n \to \infty$.

We now prove that the convergence is uniform off a small exceptional set. By Eq. (12.3), there exists an increasing sequence $\{N_k\}_{k=1}^{\infty}$, such that $\mu(E_k) < \varepsilon 2^{-k}$, where

$$E_k := \bigcup_{n \geq N_k} \left\{ |f_n - f| > \frac{1}{k} \right\}.$$

If we now set $E := \cup_{k=1}^{\infty} E_k$, then $\mu(E) < \sum_k \varepsilon 2^{-k} = \varepsilon$ and for $\omega \notin E$ we have $|f_n(\omega) - f(\omega)| \leq \frac{1}{k}$ for all $n \geq N_k$ and $k \in \mathbb{N}$. That is $f_n \to f$ uniformly on E^c. ■

Lemma 12.7. *Suppose $a_n \in \mathbb{C}$ and $|a_{n+1} - a_n| \leq \varepsilon_n$ and $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Then $\lim_{n\to\infty} a_n = a \in \mathbb{C}$ exists and $|a - a_n| \leq \delta_n := \sum_{k=n}^{\infty} \varepsilon_k$.*

Proof. Let $m > n$ then

$$|a_m - a_n| = \left| \sum_{k=n}^{m-1} (a_{k+1} - a_k) \right| \leq \sum_{k=n}^{m-1} |a_{k+1} - a_k| \leq \sum_{k=n}^{\infty} \varepsilon_k := \delta_n. \tag{12.4}$$

So $|a_m - a_n| \leq \delta_{\min(m,n)} \to 0$ as $, m, n \to \infty$, i.e. $\{a_n\}$ is Cauchy. Let $m \to \infty$ in (12.4) to find $|a - a_n| \leq \delta_n$. ■

Theorem 12.8. *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions on Ω.*

1. *If f and g are measurable functions and $f_n \xrightarrow{\mu} f$ and $f_n \xrightarrow{\mu} g$ then $f = g$ a.e.*
2. *If $f_n \xrightarrow{\mu} f$ and $g_n \xrightarrow{\mu} g$ then $\lambda f_n \to \lambda f$ for all $\lambda \in \mathbb{C}$ and $f_n + g_n \xrightarrow{\mu} f + g$.*
3. *If $f_n \xrightarrow{\mu} f$ then $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure.*
4. *If $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure, there exists a measurable function, f, and a subsequence $g_j = f_{n_j}$ of $\{f_n\}$ such that $\lim_{j\to\infty} g_j := f$ exists a.e.*
5. *(**Completeness of convergence in measure.**) If $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure and f is as in item 4. then $f_n \xrightarrow{\mu} f$.*

Proof. One of the basic tricks here is to observe that if $\varepsilon > 0$ and $a, b \geq 0$ such that $a + b \geq \varepsilon$, then either $a \geq \varepsilon/2$ or $b \geq \varepsilon/2$.

1. Suppose that f and g are measurable functions such that $f_n \xrightarrow{\mu} g$ and $f_n \xrightarrow{\mu} f$ as $n \to \infty$ and $\varepsilon > 0$ is given. Since
$$|f - g| \leq |f - f_n| + |f_n - g|,$$
if $\varepsilon > 0$ and $|f - g| \geq \varepsilon$, then either $|f - f_n| \geq \varepsilon/2$ or $|f_n - g| \geq \varepsilon/2$. Thus it follows
$$\{|f - g| > \varepsilon\} \subset \{|f - f_n| > \varepsilon/2\} \cup \{|g - f_n| > \varepsilon/2\},$$
and therefore,
$$\mu(|f - g| > \varepsilon) \leq \mu(|f - f_n| > \varepsilon/2) + \mu(|g - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty.$$
Hence
$$\mu(|f - g| > 0) = \mu\left(\cup_{n=1}^{\infty} \left\{ |f - g| > \frac{1}{n} \right\} \right) \leq \sum_{n=1}^{\infty} \mu\left(|f - g| > \frac{1}{n} \right) = 0,$$
i.e. $f = g$ a.e.
2. The first claim is easy and the second follows similarly to the proof of the first item.
3. Suppose $f_n \xrightarrow{\mu} f$, $\varepsilon > 0$ and $m, n \in \mathbb{N}$, then $|f_n - f_m| \leq |f - f_n| + |f_m - f|$. So by the basic trick,
$$\mu(|f_n - f_m| > \varepsilon) \leq \mu(|f_n - f| > \varepsilon/2) + \mu(|f_m - f| > \varepsilon/2) \to 0 \text{ as } m, n \to \infty.$$
4. Suppose $\{f_n\}$ is $L^0(\mu)$ – Cauchy and let $\varepsilon_n > 0$ such that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$ ($\varepsilon_n = 2^{-n}$ would do) and set $\delta_n = \sum_{k=n}^{\infty} \varepsilon_k$. Choose $g_j = f_{n_j}$ where $\{n_j\}$ is a subsequence of $\mathbb{N}$ such that
$$\mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \leq \varepsilon_j.$$
Let
$$F_N := \cup_{j \geq N} \{|g_{j+1} - g_j| > \varepsilon_j\} \text{ and}$$
$$E := \cap_{N=1}^{\infty} F_N = \{|g_{j+1} - g_j| > \varepsilon_j \text{ i.o.}\}.$$
Since
$$\mu(F_N) \leq \delta_N < \infty$$
and $F_N \downarrow E$ it follows[1] that $0 = \mu(E) = \lim_{N\to\infty} \mu(F_N)$. For $\omega \notin E$, $|g_{j+1}(\omega) - g_j(\omega)| \leq \varepsilon_j$ for a.a. j and so by Lemma 12.7, $f(\omega) := \lim_{j\to\infty} g_j(\omega)$ exists. For $\omega \in E$ we may define $f(\omega) \equiv 0$.
5. Next we will show $g_N \xrightarrow{\mu} f$ as $N \to \infty$ where f and g_N are as above. If
$$\omega \in F_N^c = \cap_{j \geq N} \{|g_{j+1} - g_j| \leq \varepsilon_j\},$$
then

[1] Alternatively, $\mu(E) = 0$ by the first Borel Cantelli lemma and the fact that $\sum_{j=1}^{\infty} \mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \leq \sum_{j=1}^{\infty} \varepsilon_j < \infty$.

$$|g_{j+1}(\omega) - g_j(\omega)| \le \varepsilon_j \text{ for all } j \ge N.$$

Another application of Lemma 12.7 shows $|f(\omega) - g_j(\omega)| \le \delta_j$ for all $j \ge N$, i.e.

$$F_N^c \subset \cap_{j\ge N} \{|f - g_j| \le \delta_j\} \subset \{|f - g_N| \le \delta_N\}.$$

Therefore, by taking complements of this equation, $\{|f - g_N| > \delta_N\} \subset F_N$ and hence

$$\mu(|f - g_N| > \delta_N) \le \mu(F_N) \le \delta_N \to 0 \text{ as } N \to \infty$$

and in particular, $g_N \overset{\mu}{\to} f$ as $N \to \infty$.

With this in hand, it is straightforward to show $f_n \overset{\mu}{\to} f$. Indeed, by the usual trick, for all $j \in \mathbb{N}$,

$$\mu(\{|f_n - f| > \varepsilon\}) \le \mu(\{|f - g_j| > \varepsilon/2\}) + \mu(|g_j - f_n| > \varepsilon/2).$$

Therefore, letting $j \to \infty$ in this inequality gives,

$$\mu(\{|f_n - f| > \varepsilon\}) \le \limsup_{j\to\infty} \mu(|g_j - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty,$$

wherein we have used $\{f_n\}_{n=1}^\infty$ is Cauchy in measure and $g_j \overset{\mu}{\to} f$. ■

Corollary 12.9 (Dominated Convergence Theorem). *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. Suppose $\{f_n\}$, $\{g_n\}$, and g are in L^1 and $f \in L^0$ are functions such that*

$$|f_n| \le g_n \text{ a.e., } f_n \xrightarrow{\mu} f, \; g_n \xrightarrow{\mu} g, \text{ and } \int g_n \to \int g \text{ as } n \to \infty.$$

Then $f \in L^1$ and $\lim_{n\to\infty} \|f - f_n\|_1 = 0$, i.e. $f_n \to f$ in L^1. In particular $\lim_{n\to\infty} \int f_n = \int f$.

Proof. First notice that $|f| \le g$ a.e. and hence $f \in L^1$ since $g \in L^1$. To see that $|f| \le g$, use item 4. of Theorem 12.8 to find subsequences $\{f_{n_k}\}$ and $\{g_{n_k}\}$ of $\{f_n\}$ and $\{g_n\}$ respectively which are almost everywhere convergent. Then

$$|f| = \lim_{k\to\infty} |f_{n_k}| \le \lim_{k\to\infty} g_{n_k} = g \text{ a.e.}$$

If (for sake of contradiction) $\lim_{n\to\infty} \|f - f_n\|_1 \ne 0$ there exists $\varepsilon > 0$ and a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that

$$\int |f - f_{n_k}| \ge \varepsilon \text{ for all } k. \tag{12.5}$$

Using item 4. of Theorem 12.8 again, we may assume (by passing to a further subsequences if necessary) that $f_{n_k} \to f$ and $g_{n_k} \to g$ almost everywhere. Noting, $|f - f_{n_k}| \le g + g_{n_k} \to 2g$ and $\int (g + g_{n_k}) \to \int 2g$, an application of the dominated convergence Theorem 7.27 implies $\lim_{k\to\infty} \int |f - f_{n_k}| = 0$ which contradicts Eq. (12.5). ■

Exercise 12.1 (Fatou's Lemma). Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. If $f_n \ge 0$ and $f_n \to f$ in measure, then $\int_\Omega f d\mu \le \liminf_{n\to\infty} \int_\Omega f_n d\mu$.

Lemma 12.10. *Suppose $1 \le p < \infty$, $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$, and $f_n \overset{\mu}{\to} f$, then $\|f\|_p \le \liminf_{n\to\infty} \|f_n\|_p$. Moreover if $\{f_n\}_{n=1}^\infty \cup \{f\} \subset L^p(\mu)$, then $\|f - f_n\|_p \to 0$ as $n \to \infty$ iff $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p < \infty$ and $f_n \xrightarrow{\mu} f$.*

Proof. Choose a subsequence, $g_k = f_{n_k}$, such that $\liminf_{n\to\infty} \|f_n\|_p = \lim_{k\to\infty} \|g_k\|_p$. By passing to a further subsequence if necessary, we may further assume that $g_k \to f$ a.e. Therefore, by Fatou's lemma,

$$\|f\|_p^p = \int_\Omega |f|^p d\mu = \int_\Omega \lim_{k\to\infty} |g_k|^p d\mu \le \liminf_{k\to\infty} \int_\Omega |g_k|^p d\mu = \liminf_{n\to\infty} \|f_n\|_p^p$$

which proves the first assertion.

If $\|f - f_n\|_p \to 0$ as $n \to \infty$, then by the triangle inequality, $\left|\|f\|_p - \|f_n\|_p\right| \le \|f - f_n\|_p$ which shows $\int |f_n|^p \to \int |f|^p$ if $f_n \to f$ in L^p. Chebyschev's inequality implies $f_n \xrightarrow{\mu} f$ if $f_n \to f$ in L^p.

Conversely if $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p < \infty$ and $f_n \xrightarrow{\mu} f$, let $F_n := |f - f_n|^p$ and $G_n := 2^{p-1}\left[|f|^p + |f_n|^p\right]$. Then $F_n \xrightarrow{\mu} 0$[2], $F_n \le G_n \in L^1$, and $\int G_n \to \int G$ where $G := 2^p |f|^p \in L^1$. Therefore, by Corollary 12.9, $\int |f - f_n|^p = \int F_n \to \int 0 = 0$. ■

Exercise 12.2. Let $(\Omega, \mathcal{B}, \mu)$ be a measure space, $p \in [1, \infty)$, and suppose that $0 \le f \in L^1(\mu)$, $0 \le f_n \in L^1(\mu)$ for all n, $f_n \xrightarrow{\mu} f$, and $\int f_n d\mu \to \int f d\mu$. Then $f_n \to f$ in $L^1(\mu)$. In particular if $f, f_n \in L^p(\mu)$ and $f_n \to f$ in $L^p(\mu)$, then $|f_n|^p \to |f|^p$ in $L^1(\mu)$.

Solution to Exercise (12.2). Let $F_n := |f - f_n| \le f + f_n := g_n$ and $g := 2f$. Then $F_n \xrightarrow{\mu} 0$, $g_n \xrightarrow{\mu} g$, and $\int g_n d\mu \to \int g d\mu$. So by Corollary 12.9, $\int |f - f_n| d\mu = \int F_n d\mu \to 0$ as $n \to \infty$.

Proposition 12.11. *Suppose $(\Omega, \mathcal{B}, \mu)$ is a probability space and $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions on Ω. Then $\{f_n\}_{n=1}^\infty$ converges to f in probability iff every subsequence, $\{f'_n\}_{n=1}^\infty$ of $\{f_n\}_{n=1}^\infty$ has a further subsequence, $\{f''_n\}_{n=1}^\infty$, which is almost surely convergent to f.*

[2] This is becuase $|F_n| \ge \varepsilon$ iff $|f - f_n| \ge \varepsilon^{1/p}$.

Proof. If $\{f_n\}_{n=1}^\infty$ is convergent and hence Cauchy in probability then any subsequence, $\{f'_n\}_{n=1}^\infty$ is also Cauchy in probability. Hence by item 4. of Theorem 12.8 there is a further subsequence, $\{f''_n\}_{n=1}^\infty$ of $\{f'_n\}_{n=1}^\infty$ which is convergent almost surely.

Conversely if $\{f_n\}_{n=1}^\infty$ does not converge to f in probability, then there exists an $\varepsilon > 0$ and a subsequence, $\{n_k\}$ such that $\inf_k \mu(|f - f_{n_k}| \geq \varepsilon) > 0$. Any subsequence of $\{f_{n_k}\}$ would have the same property and hence can not be almost surely convergent because of Egorov's Theorem 12.6. ■

Corollary 12.12. *Suppose $(\Omega, \mathcal{B}, \mu)$ is a probability space, $f_n \xrightarrow{\mu} f$ and $g_n \xrightarrow{\mu} g$ and $\varphi : \mathbb{R} \to \mathbb{R}$ and $\psi : \mathbb{R}^2 \to \mathbb{R}$ are continuous functions. Then*

1. $\varphi(f_n) \xrightarrow{\mu} \varphi(f)$,
2. $\psi(f_n, g_n) \xrightarrow{\mu} \psi(f, g)$, *and*
3. $f_n \cdot g_n \xrightarrow{\mu} f \cdot g$.

Proof. Item 1. and 3. follow from item 2. by taking $\psi(x, y) = \varphi(x)$ and $\psi(x, y) = x \cdot y$ respectively. So it suffices to prove item 2. To do this we will make repeated use of Theorem 12.8.

Given any subsequence, $\{n_k\}$, of $\mathbb{N}$ there is a subsequence, $\{n'_k\}$ of $\{n_k\}$ such that $f_{n'_k} \to f$ a.s. and yet a further subsequence $\{n''_k\}$ of $\{n'_k\}$ such that $g_{n''_k} \to g$ a.s. Hence, by the continuity of ψ, it now follows that

$$\lim_{k\to\infty} \psi\left(f_{n''_k}, g_{n''_k}\right) = \psi(f, g) \text{ a.s.}$$

which completes the proof. ■

Example 12.13. It is not possible to drop the assumption that $\mu(\Omega) < \infty$ in Corollary 12.12. For example, let $\Omega = \mathbb{R}$, $\mathcal{B} = \mathcal{B}_\mathbb{R}$, $\mu = m$ be Lebesgue measure, $f_n(x) = \frac{1}{n}$ and $g_n(x) = x^2 = g(x)$. Then $f_n \xrightarrow{\mu} 0$, $g_n \xrightarrow{\mu} g$ while $f_n g_n$ does not converge to $0 = 0 \cdot g$ in measure. Also if we let $\varphi(y) = y^2$, $f_n(x) = x + 1/n$ and $f(x) = x$ for all $x \in \mathbb{R}$, then $f_n \xrightarrow{\mu} f$ while

$$[\varphi(f_n) - \varphi(f)](x) = (x + 1/n)^2 - x^2 = \frac{2}{n}x + \frac{1}{n^2}$$

does not go to 0 in measure as $n \to \infty$.

12.3 Jensen's, Hölder's and Minikowski's Inequalities

Theorem 12.14 (Jensen's Inequality). *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a probability space, i.e. μ is a positive measure and $\mu(\Omega) = 1$. Also suppose that $f \in L^1(\mu)$, $f : \Omega \to (a, b)$, and $\varphi : (a, b) \to \mathbb{R}$ is a convex function, (i.e. $\varphi''(x) \geq 0$ on (a, b).) Then*

$$\varphi\left(\int_\Omega f d\mu\right) \leq \int_\Omega \varphi(f) d\mu$$

where if $\varphi \circ f \notin L^1(\mu)$, then $\varphi \circ f$ is integrable in the extended sense and $\int_\Omega \varphi(f) d\mu = \infty$.

Proof. Let $t = \int_\Omega f d\mu \in (a, b)$ and let $\beta \in \mathbb{R}$ ($\beta = \dot\varphi(t)$ when $\dot\varphi(t)$ exists), be such that $\varphi(s) - \varphi(t) \geq \beta(s - t)$ for all $s \in (a, b)$. (See Lemma 12.52) and Figure 12.5 when φ is C^1 and Theorem 12.55 below for the existence of such a β in the general case.) Then integrating the inequality, $\varphi(f) - \varphi(t) \geq \beta(f - t)$, implies that

$$0 \leq \int_\Omega \varphi(f) d\mu - \varphi(t) = \int_\Omega \varphi(f) d\mu - \varphi(\int_\Omega f d\mu).$$

Moreover, if $\varphi(f)$ is not integrable, then $\varphi(f) \geq \varphi(t) + \beta(f - t)$ which shows that negative part of $\varphi(f)$ is integrable. Therefore, $\int_\Omega \varphi(f) d\mu = \infty$ in this case. ■

Example 12.15. Since e^x for $x \in \mathbb{R}$, $-\ln x$ for $x > 0$, and x^p for $x \geq 0$ and $p \geq 1$ are all convex functions, we have the following inequalities

$$\exp\left(\int_\Omega f d\mu\right) \leq \int_\Omega e^f d\mu, \tag{12.6}$$

$$\int_\Omega \log(|f|) d\mu \leq \log\left(\int_\Omega |f| d\mu\right)$$

and for $p \geq 1$,

$$\left|\int_\Omega f d\mu\right|^p \leq \left(\int_\Omega |f| d\mu\right)^p \leq \int_\Omega |f|^p d\mu.$$

Example 12.16. As a special case of Eq. (12.6), if $p_i, s_i > 0$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \ldots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i^{p_i}} \leq \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i^{p_i}} = \sum_{i=1}^n \frac{s_i^{p_i}}{p_i}. \tag{12.7}$$

Indeed, we have applied Eq. (12.6) with $\Omega = \{1, 2, \ldots, n\}$, $\mu = \sum_{i=1}^n \frac{1}{p_i} \delta_i$ and $f(i) := \ln s_i^{p_i}$. As a special case of Eq. (12.7), suppose that $s, t, p, q \in (1, \infty)$ with $q = \frac{p}{p-1}$ (i.e. $\frac{1}{p} + \frac{1}{q} = 1$) then

$$st \leq \frac{1}{p} s^p + \frac{1}{q} t^q. \tag{12.8}$$

(When $p = q = 1/2$, the inequality in Eq. (12.8) follows from the inequality, $0 \leq (s-t)^2$.)

As another special case of Eq. (12.7), take $p_i = n$ and $s_i = a_i^{1/n}$ with $a_i > 0$, then we get the arithmetic geometric mean inequality,

$$\sqrt[n]{a_1 \dots a_n} \leq \frac{1}{n} \sum_{i=1}^{n} a_i. \tag{12.9}$$

Example 12.17. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space, $0 < p < q < \infty$, and $f : \Omega \to \mathbb{C}$ be a measurable function. Then by Jensen's inequality,

$$\left(\int_\Omega |f|^p \, d\mu\right)^{q/p} \leq \int_\Omega (|f|^p)^{q/p} \, d\mu = \int_\Omega |f|^q \, d\mu$$

from which it follows that $\|f\|_p \leq \|f\|_q$. In particular, $L^p(\mu) \subset L^q(\mu)$ for all $0 < p < q < \infty$. See Corollary 12.31 for an alternative proof.

Theorem 12.18 (Hölder's inequality). *Suppose that $1 \leq p \leq \infty$ and $q := \frac{p}{p-1}$, or equivalently $p^{-1} + q^{-1} = 1$. If f and g are measurable functions then*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q. \tag{12.10}$$

Assuming $p \in (1, \infty)$ and $\|f\|_p \cdot \|g\|_q < \infty$, equality holds in Eq. (12.10) iff $|f|^p$ and $|g|^q$ are linearly dependent as elements of L^1 which happens iff

$$|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p \quad a.e. \tag{12.11}$$

Proof. The cases $p = 1$ and $q = \infty$ or $p = \infty$ and $q = 1$ are easy to deal with and will be left to the reader. So we now assume that $p, q \in (1, \infty)$. If $\|f\|_q = 0$ or ∞ or $\|g\|_p = 0$ or ∞, Eq. (12.10) is again easily verified. So we will now assume that $0 < \|f\|_q, \|g\|_p < \infty$. Taking $s = |f| / \|f\|_p$ and $t = |g| / \|g\|_q$ in Eq. (12.8) gives,

$$\frac{|fg|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f|^p}{\|f\|_p} + \frac{1}{q} \frac{|g|^q}{\|g\|^q} \tag{12.12}$$

with equality iff $|g/\|g\|_q| = |f|^{p-1} / \|f\|_p^{(p-1)} = |f|^{p/q} / \|f\|_p^{p/q}$, i.e. $|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p$. Integrating Eq. (12.12) implies

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1$$

with equality iff Eq. (12.11) holds. The proof is finished since it is easily checked that equality holds in Eq. (12.10) when $|f|^p = c\,|g|^q$ of $|g|^q = c\,|f|^p$ for some constant c. ∎

Example 12.19. Suppose that $a_k \in \mathbb{C}$ for $k = 1, 2, \dots, n$ and $p \in [1, \infty)$, then

$$\left|\sum_{k=1}^{n} a_k\right|^p \leq n^{p-1} \sum_{k=1}^{n} |a_k|^p. \tag{12.13}$$

Indeed, by Hölder's inequality applied using the measure space, $\{1, 2, \dots, n\}$ equipped with counting measure, we have

$$\left|\sum_{k=1}^{n} a_k\right| = \left|\sum_{k=1}^{n} a_k \cdot 1\right| \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} 1^q\right)^{1/q} = n^{1/q} \left(\sum_{k=1}^{n} |a_k|^p\right)^{1/p}$$

where $q = \frac{p}{p-1}$. Taking the p^{th} – power of this inequality then gives, Eq. (12.14).

Theorem 12.20 (Generalized Hölder's inequality). *Suppose that $f_i : \Omega \to \mathbb{C}$ are measurable functions for $i = 1, \dots, n$ and $p_1, \dots, p_n$ and r are positive numbers such that $\sum_{i=1}^n p_i^{-1} = r^{-1}$, then*

$$\left\|\prod_{i=1}^{n} f_i\right\|_r \leq \prod_{i=1}^{n} \|f_i\|_{p_i}. \tag{12.14}$$

Proof. One may prove this theorem by induction based on Hölder's Theorem 12.18 above. Alternatively we may give a proof along the lines of the proof of Theorem 12.18 which is what we will do here.

Since Eq. (12.14) is easily seen to hold if $\|f_i\|_{p_i} = 0$ for some i, we will assume that $\|f_i\|_{p_i} > 0$ for all i. By assumption, $\sum_{i=1}^n \frac{r_i}{p_i} = 1$, hence we may replace s_i by s_i^r and p_i by p_i/r for each i in Eq. (12.7) to find

$$s_1^r \dots s_n^r \leq \sum_{i=1}^{n} \frac{(s_i^r)^{p_i/r}}{p_i/r} = r \sum_{i=1}^{n} \frac{s_i^{p_i}}{p_i}.$$

Now replace s_i by $|f_i| / \|f_i\|_{p_i}$ in the previous inequality and integrate the result to find

$$\frac{1}{\prod_{i=1}^n \|f_i\|_{p_i}} \left\|\prod_{i=1}^{n} f_i\right\|_r^r \leq r \sum_{i=1}^{n} \frac{1}{p_i} \frac{1}{\|f_i\|_{p_i}^{p_i}} \int_\Omega |f_i|^{p_i} \, d\mu = \sum_{i=1}^{n} \frac{r}{p_i} = 1.$$

∎

Definition 12.21. *A **norm** on a vector space Z is a function $\|\cdot\| : Z \to [0, \infty)$ such that*

1. *(Homogeneity) $\|\lambda f\| = |\lambda| \, \|f\|$ for all $\lambda \in \mathbb{F}$ and $f \in Z$.*
2. *(Triangle inequality) $\|f + g\| \leq \|f\| + \|g\|$ for all $f, g \in Z$.*

3. (Positive definite) $\|f\| = 0$ implies $f = 0$.

A pair $(Z, \|\cdot\|)$ where Z is a vector space and $\|\cdot\|$ is a norm on Z is called a ***normed vector space.***

Theorem 12.22 (Minkowski's Inequality). *If $1 \le p \le \infty$ and $f, g \in L^p(\mu)$ then*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p. \tag{12.15}$$

In particular, $\left(L^p(\mu), \|\cdot\|_p\right)$ is a normed vector space for all $1 \le p \le \infty$.

Proof. When $p = \infty$, $|f| \le \|f\|_\infty$ a.e. and $|g| \le \|g\|_\infty$ a.e. so that $|f+g| \le |f| + |g| \le \|f\|_\infty + \|g\|_\infty$ a.e. and therefore

$$\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty .$$

When $p < \infty$,

$$|f+g|^p \le (2\max(|f|, |g|))^p = 2^p \max(|f|^p, |g|^p) \le 2^p (|f|^p + |g|^p),$$

which implies[3] $f + g \in L^p$ since

$$\|f+g\|_p^p \le 2^p \left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

Furthermore, when $p = 1$ we have

$$\|f+g\|_1 = \int_\Omega |f+g| d\mu \le \int_\Omega |f|\, d\mu + \int_\Omega |g| d\mu = \|f\|_1 + \|g\|_1.$$

We now consider $p \in (1, \infty)$. We may assume $\|f+g\|_p$, $\|f\|_p$ and $\|g\|_p$ are all positive since otherwise the theorem is easily verified. Integrating

$$|f+g|^p = |f+g||f+g|^{p-1} \le (|f|+|g|)|f+g|^{p-1}$$

and then applying Holder's inequality with $q = p/(p-1)$ gives

$$\begin{aligned}\int_\Omega |f+g|^p d\mu &\le \int_\Omega |f|\ |f+g|^{p-1} d\mu + \int_\Omega |g|\ |f+g|^{p-1} d\mu \\ &\le (\|f\|_p + \|g\|_p)\, \|\, |f+g|^{p-1}\, \|_q, \end{aligned} \tag{12.16}$$

where

$$\||f+g|^{p-1}\|_q^q = \int_\Omega (|f+g|^{p-1})^q d\mu = \int_\Omega |f+g|^p d\mu = \|f+g\|_p^p. \tag{12.17}$$

Combining Eqs. (12.16) and (12.17) implies

$$\|f+g\|_p^p \le \|f\|_p \|f+g\|_p^{p/q} + \|g\|_p \|f+g\|_p^{p/q} \tag{12.18}$$

Solving this inequality for $\|f+g\|_p$ gives Eq. (12.15). ∎

[3] In light of Example 12.19, the last 2^p in the above inequality may be replaced by 2^{p-1}.

12.4 Completeness of L^p – spaces

Definition 12.23 (Banach space). *A normed vector space $(Z, \|\cdot\|)$ is a* ***Banach space*** *if is is complete, i.e. all Cauchy sequences are convergent. To be more precise we are assuming that if $\{x_n\}_{n=1}^\infty \subset Z$ satisfies, $\lim_{m,n\to\infty} \|x_n - x_m\| = 0$, then there exists an $x \in Z$ such that $\lim_{n\to\infty} \|x - x_n\| = 0$.*

Theorem 12.24. *Let $\|\cdot\|_\infty$ be as defined in Eq. (12.2), then $(L^\infty(\Omega, \mathcal{B}, \mu), \|\cdot\|_\infty)$ is a Banach space. A sequence $\{f_n\}_{n=1}^\infty \subset L^\infty$ converges to $f \in L^\infty$ iff there exists $E \in \mathcal{B}$ such that $\mu(E) = 0$ and $f_n \to f$ uniformly on E^c. Moreover, bounded simple functions are dense in L^∞.*

Proof. By Minkowski's Theorem 12.22, $\|\cdot\|_\infty$ satisfies the triangle inequality. The reader may easily check the remaining conditions that ensure $\|\cdot\|_\infty$ is a norm. Suppose that $\{f_n\}_{n=1}^\infty \subset L^\infty$ is a sequence such $f_n \to f \in L^\infty$, i.e. $\|f - f_n\|_\infty \to 0$ as $n \to \infty$. Then for all $k \in \mathbb{N}$, there exists $N_k < \infty$ such that

$$\mu\left(|f - f_n| > k^{-1}\right) = 0 \text{ for all } n \ge N_k.$$

Let

$$E = \cup_{k=1}^\infty \cup_{n \ge N_k} \left\{|f - f_n| > k^{-1}\right\}.$$

Then $\mu(E) = 0$ and for $x \in E^c$, $|f(x) - f_n(x)| \le k^{-1}$ for all $n \ge N_k$. This shows that $f_n \to f$ uniformly on E^c. Conversely, if there exists $E \in \mathcal{B}$ such that $\mu(E) = 0$ and $f_n \to f$ uniformly on E^c, then for any $\varepsilon > 0$,

$$\mu(|f - f_n| \ge \varepsilon) = \mu(\{|f - f_n| \ge \varepsilon\} \cap E^c) = 0$$

for all n sufficiently large. That is to say $\limsup_{j\to\infty} \|f - f_n\|_\infty \le \varepsilon$ for all $\varepsilon > 0$. The density of simple functions follows from the approximation Theorem 6.39. So the last item to prove is the completeness of L^∞.

Suppose $\varepsilon_{m,n} := \|f_m - f_n\|_\infty \to 0$ as $m, n \to \infty$. Let $E_{m,n} = \{|f_n - f_m| > \varepsilon_{m,n}\}$ and $E := \cup E_{m,n}$, then $\mu(E) = 0$ and

$$\sup_{x \in E^c} |f_m(x) - f_n(x)| \le \varepsilon_{m,n} \to 0 \text{ as } m, n \to \infty.$$

Therefore, $f := \lim_{n\to\infty} f_n$ exists on E^c and the limit is uniform on E^c. Letting $f = \lim_{n\to\infty} 1_{E^c} f_n$, it then follows that $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$. ∎

Theorem 12.25 (Completeness of $L^p(\mu)$). *For $1 \le p \le \infty$, $L^p(\mu)$ equipped with the L^p – norm, $\|\cdot\|_p$ (see Eq. (12.1)), is a Banach space.*

Proof. By Minkowski's Theorem 12.22, $\|\cdot\|_p$ satisfies the triangle inequality. As above the reader may easily check the remaining conditions that ensure $\|\cdot\|_p$ is a norm. So we are left to prove the completeness of $L^p(\mu)$ for $1 \le p < \infty$, the case $p = \infty$ being done in Theorem 12.24.

Let $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev's inequality (Lemma 12.4), $\{f_n\}$ is L^0-Cauchy (i.e. Cauchy in measure) and by Theorem 12.8 there exists a subsequence $\{g_j\}$ of $\{f_n\}$ such that $g_j \to f$ a.e. By Fatou's Lemma,

$$\|g_j - f\|_p^p = \int \lim_{k\to\infty} \inf |g_j - g_k|^p d\mu \le \lim_{k\to\infty} \inf \int |g_j - g_k|^p d\mu$$
$$= \lim_{k\to\infty} \inf \|g_j - g_k\|_p^p \to 0 \text{ as } j \to \infty.$$

In particular, $\|f\|_p \le \|g_j - f\|_p + \|g_j\|_p < \infty$ so the $f \in L^p$ and $g_j \xrightarrow{L^p} f$. The proof is finished because,

$$\|f_n - f\|_p \le \|f_n - g_j\|_p + \|g_j - f\|_p \to 0 \text{ as } j, n \to \infty.$$

■

See Definition 14.2 for a very important example of where completeness is used. To end this section we are going to record a few results we will need later regarding subspace of $L^p(\mu)$ which are induced by sub – σ – algebras, $\mathcal{B}_0 \subset \mathcal{B}$.

Lemma 12.26. *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathcal{B}_0$ be a sub – σ – algebra of $\mathcal{B}$. Then for $1 \le p < \infty$, the map $i : L^p(\Omega, \mathcal{B}_0, \mu) \to L^p(\Omega, \mathcal{B}, \mu)$ defined by $i([f]_0) = [f]$ is a well defined linear isometry. Here we are writing,*

$$[f]_0 = \{g \in \mathrm{L}^p(\Omega, \mathcal{B}_0, \mu) : g = f \text{ a.e.}\} \text{ and}$$
$$[f] = \{g \in \mathrm{L}^p(\Omega, \mathcal{B}, \mu) : g = f \text{ a.e.}\}.$$

Moreover the image of i, $i(L^p(\Omega, \mathcal{B}_0, \mu))$, is a closed subspace of $L^p(\Omega, \mathcal{B}, \mu)$.

Proof. This is proof is routine and most of it will be left to the reader. Let us just check that $i(L^p(\Omega, \mathcal{B}_0, \mu))$, is a closed subspace of $L^p(\Omega, \mathcal{B}, \mu)$. To this end, suppose that $i([f_n]_0) = [f_n]$ is a convergent sequence in $\mathrm{L}^p(\Omega, \mathcal{B}, \mu)$. Because, i, is an isometry it follows that $\{[f_n]_0\}_{n=1}^\infty$ is a Cauchy and hence convergent sequence in $L^p(\Omega, \mathcal{B}_0, \mu)$. Letting $f \in \mathrm{L}^p(\Omega, \mathcal{B}_0, \mu)$ such that $\|f - f_n\|_{L^p(\mu)} \to 0$, we will have, since i is isometric, that $[f_n] \to [f] = i([f]_0) \in i(L^p(\Omega, \mathcal{B}_0, \mu))$ as desired. ■

Exercise 12.3. Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathcal{B}_0$ be a sub – σ – algebra of $\mathcal{B}$. Further suppose that to every $B \in \mathcal{B}$ there exists $A \in \mathcal{B}_0$ such that $\mu(B \Delta A) = 0$. Show for all $1 \le p < \infty$ that $i(L^p(\Omega, \mathcal{B}_0, \mu)) = L^p(\Omega, \mathcal{B}, \mu)$, i.e. to each $f \in \mathrm{L}^p(\Omega, \mathcal{B}, \mu)$ there exists a $g \in \mathrm{L}^p(\Omega, \mathcal{B}_0, \mu)$ such that $f = g$ a.e. **Hints:** 1. verify the last assertion for simple functions in $\mathrm{L}^p(\Omega, \mathcal{B}_0, \mu)$. 2. then make use of Theorem 6.39 and Exercise 6.4.

Exercise 12.4. Suppose that $1 \le p < \infty$, $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space and $\mathcal{B}_0$ is a sub – σ – algebra of $\mathcal{B}$. Show that $i(L^p(\Omega, \mathcal{B}_0, \mu)) = L^p(\Omega, \mathcal{B}, \mu)$ implies; to every $B \in \mathcal{B}$ there exists $A \in \mathcal{B}_0$ such that $\mu(B \Delta A) = 0$.

Solution to Exercise (12.4). Let $B \in \mathcal{B}$ with $\mu(B) < \infty$. Then $1_B \in \mathrm{L}^p(\Omega, \mathcal{B}, \mu)$ and hence by assumption there exists $g \in \mathrm{L}^p(\Omega, \mathcal{B}_0, \mu)$ such that $g = 1_B$ a.e. Let $A := \{g = 1\} \in \mathcal{B}_0$ and observe that $A \Delta B \subset \{g \ne 1_B\}$. Therefore $\mu(A \Delta B) = \mu(g \ne 1_B) = 0$. For general the case we use the fact that $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space to conclude that each $B \in \mathcal{B}$ may be written as a disjoint union, $B = \sum_{n=1}^\infty B_n$, with $B_n \in \mathcal{B}$ and $\mu(B_n) < \infty$. By what we have just proved we may find $A_n \in \mathcal{B}_0$ such that $\mu(B_n \Delta A_n) = 0$. I now claim that $A := \cup_{n=1}^\infty A_n \in \mathcal{B}_0$ satisfies $\mu(A \Delta B) = 0$. Indeed, notice that

$$A \setminus B = \cup_{n=1}^\infty A_n \setminus B \subset \cup_{n=1}^\infty A_n \setminus B_n,$$

similarly $B \setminus A \subset \cup_{n=1}^\infty B_n \setminus A_n$, and therefore $A \Delta B \subset \cup_{n=1}^\infty A_n \Delta B_n$. Therefore by sub-additivity of μ, $\mu(A \Delta B) \le \sum_{n=1}^\infty \mu(A_n \Delta B_n) = 0$.

Convention: From now on we will drop the cumbersome notation and simply identify $[f]$ with f and $L^p(\Omega, \mathcal{B}_0, \mu)$ with its image, $i(L^p(\Omega, \mathcal{B}_0, \mu))$, in $L^p(\Omega, \mathcal{B}, \mu)$.

12.5 Density Results

Theorem 12.27 (Density Theorem). *Let $p \in [1, \infty)$, $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that*

1. *$\mathbb{M} \subset L^p(\mu, \mathbb{R})$ and $\sigma(\mathbb{M}) = \mathcal{B}$.*
2. *There exists $\psi_k \in \mathbb{M}$ such that $\psi_k \to 1$ boundedly.*

Then to every function $f \in L^p(\mu, \mathbb{R})$, there exist $\varphi_n \in \mathbb{M}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(\mu)} = 0$, i.e. $\mathbb{M}$ is dense in $L^p(\mu, \mathbb{R})$.

Proof. Fix $k \in \mathbb{N}$ for the moment and let $\mathbb{H}$ denote those bounded $\mathcal{B}$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{M}$ such that $\lim_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded measurable $\mathbb{R}$ – valued functions on Ω, $1 \in \mathbb{H}$, $\mathbb{M} \subset \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated convergence theorem, $\lim_{n\to\infty} \|\psi_k(f - f_n)\|_{L^p(\mu)} = 0$.[4] (Take the dominating function to be $g =$

[4] It is at this point that the proof would break down if $p = \infty$.

$[2C|\psi_k|]^p$ where C is a constant bounding all of the $\{|f_n|\}_{n=1}^{\infty}$.) We may now choose $\varphi_n \in \mathbb{M}$ such that $\|\varphi_n - \psi_k f_n\|_{L^p(\mu)} \le \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} \le \limsup_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} + \limsup_{n\to\infty} \|\psi_k f_n - \varphi_n\|_{L^p(\mu)} = 0 \quad (12.19)$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem 8.16, now shows $\mathbb{H}$ contains all bounded measurable functions on Ω. Let $f \in L^p(\mu)$ be given. The dominated convergence theorem implies $\lim_{k\to\infty} \left\|\psi_k 1_{\{|f|\le k\}} f - f\right\|_{L^p(\mu)} = 0$. (Take the dominating function to be $g = [2C|f|]^p$ where C is a bound on all of the $|\psi_k|$.) Using this and what we have just proved, there exists $\varphi_k \in \mathbb{M}$ such that

$$\left\|\psi_k 1_{\{|f|\le k\}} f - \varphi_k\right\|_{L^p(\mu)} \le \frac{1}{k}.$$

The same line of reasoning used in Eq. (12.19) now implies $\lim_{k\to\infty} \|f - \varphi_k\|_{L^p(\mu)} = 0$. ■

Example 12.28. Let μ be a measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu([-M,M]) < \infty$ for all $M < \infty$. Then, $C_c(\mathbb{R},\mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ with compact support) is dense in $L^p(\mu)$ for all $1 \le p < \infty$. To see this, apply Theorem 12.27 with $\mathbb{M} = C_c(\mathbb{R},\mathbb{R})$ and $\psi_k := 1_{[-k,k]}$.

Theorem 12.29. *Suppose $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{B} \subset 2^{\Omega}$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and μ is σ – finite on $\mathcal{A}$. Let $\mathbb{S}(\mathcal{A},\mu)$ denote the measurable simple functions, $\varphi : \Omega \to \mathbb{R}$ such $\{\varphi = y\} \in \mathcal{A}$ for all $y \in \mathbb{R}$ and $\mu(\{\varphi \ne 0\}) < \infty$. Then $\mathbb{S}(\mathcal{A},\mu)$ is dense subspace of $L^p(\mu)$.*

Proof. Let $\mathbb{M} := \mathbb{S}(\mathcal{A},\mu)$. By assumption there exists $\Omega_k \in \mathcal{A}$ such that $\mu(\Omega_k) < \infty$ and $\Omega_k \uparrow \Omega$ as $k \to \infty$. If $A \in \mathcal{A}$, then $\Omega_k \cap A \in \mathcal{A}$ and $\mu(\Omega_k \cap A) < \infty$ so that $1_{\Omega_k \cap A} \in \mathbb{M}$. Therefore $1_A = \lim_{k\to\infty} 1_{\Omega_k\cap A}$ is $\sigma(\mathbb{M})$ – measurable for every $A \in \mathcal{A}$. So we have shown that $\mathcal{A} \subset \sigma(\mathbb{M}) \subset \mathcal{B}$ and therefore $\mathcal{B} = \sigma(\mathcal{A}) \subset \sigma(\mathbb{M}) \subset \mathcal{B}$, i.e. $\sigma(\mathbb{M}) = \mathcal{B}$. The theorem now follows from Theorem 12.27 after observing $\psi_k := 1_{\Omega_k} \in \mathbb{M}$ and $\psi_k \to 1$ boundedly. ■

Theorem 12.30 (Separability of L^p – Spaces). *Suppose, $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{B}$ is a countable algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and μ is σ – finite on $\mathcal{A}$. Then $L^p(\mu)$ is separable and*

$$\mathbb{D} = \{\sum a_j 1_{A_j} : a_j \in \mathbb{Q} + i\mathbb{Q},\ A_j \in \mathcal{A} \text{ with } \mu(A_j) < \infty\}$$

is a countable dense subset.

Proof. It is left to reader to check $\mathbb{D}$ is dense in $\mathbb{S}(\mathcal{A},\mu)$ relative to the $L^p(\mu)$ – norm. Once this is done, the proof is then complete since $\mathbb{S}(\mathcal{A},\mu)$ is a dense subspace of $L^p(\mu)$ by Theorem 12.29. ■

12.6 Relationships between different L^p – spaces

The $L^p(\mu)$ – norm controls two types of behaviors of f, namely the "behavior at infinity" and the behavior of "local singularities." So in particular, if f blows up at a point $x_0 \in \Omega$, then locally near x_0 it is harder for f to be in $L^p(\mu)$ as p increases. On the other hand a function $f \in L^p(\mu)$ is allowed to decay at "infinity" slower and slower as p increases. With these insights in mind, we should not in general expect $L^p(\mu) \subset L^q(\mu)$ or $L^q(\mu) \subset L^p(\mu)$. However, there are two notable exceptions. (1) If $\mu(\Omega) < \infty$, then there is no behavior at infinity to worry about and $L^q(\mu) \subset L^p(\mu)$ for all $q \ge p$ as is shown in Corollary 12.31 below. (2) If μ is counting measure, i.e. $\mu(A) = \#(A)$, then all functions in $L^p(\mu)$ for any p can not blow up on a set of positive measure, so there are no local singularities. In this case $L^p(\mu) \subset L^q(\mu)$ for all $q \ge p$, see Corollary 12.36 below.

Corollary 12.31 (Example 12.17 revisited). *If $\mu(\Omega) < \infty$ and $0 < p \le q \le \infty$, then $L^q(\mu) \subset L^p(\mu)$, the inclusion map is bounded and in fact*

$$\|f\|_p \le [\mu(\Omega)]^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q.$$

Proof. Take $a \in [1,\infty]$ such that

$$\frac{1}{p} = \frac{1}{a} + \frac{1}{q}, \text{ i.e. } a = \frac{pq}{q-p}.$$

Then by Theorem 12.20,

$$\|f\|_p = \|f \cdot 1\|_p \le \|f\|_q \cdot \|1\|_a = \mu(\Omega)^{1/a} \|f\|_q = \mu(\Omega)^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q.$$

The reader may easily check this final formula is correct even when $q = \infty$ provided we interpret $1/p - 1/\infty$ to be $1/p$. ■

The rest of this section may be skipped.

Example 12.32 (Power Inequalities). Let $a := (a_1, \ldots, a_n)$ with $a_i > 0$ for $i = 1, 2, \ldots, n$ and for $p \in \mathbb{R} \setminus \{0\}$, let

$$\|a\|_p := \left(\frac{1}{n}\sum_{i=1}^{n} a_i^p\right)^{1/p}.$$

Then by Corollary 12.31, $p \to \|a\|_p$ is increasing in p for $p > 0$. For $p = -q < 0$, we have

$$\|a\|_p := \left(\frac{1}{n}\sum_{i=1}^n a_i^{-q}\right)^{-1/q} = \left(\frac{1}{\frac{1}{n}\sum_{i=1}^n \left(\frac{1}{a_i}\right)^q}\right)^{1/q} = \left\|\frac{1}{a}\right\|_q^{-1}$$

where $\frac{1}{a} := (1/a_1, \dots, 1/a_n)$. So for $p < 0$, as p increases, $q = -p$ decreases, so that $\left\|\frac{1}{a}\right\|_q$ is decreasing and hence $\left\|\frac{1}{a}\right\|_q^{-1}$ is increasing. Hence we have shown that $p \to \|a\|_p$ is increasing for $p \in \mathbb{R} \setminus \{0\}$.

We now claim that $\lim_{p\to 0}\|a\|_p = \sqrt[n]{a_1 \dots a_n}$. To prove this, write $a_i^p = e^{p\ln a_i} = 1 + p\ln a_i + O(p^2)$ for p near zero. Therefore,

$$\frac{1}{n}\sum_{i=1}^n a_i^p = 1 + p\frac{1}{n}\sum_{i=1}^n \ln a_i + O(p^2).$$

Hence it follows that

$$\lim_{p\to 0}\|a\|_p = \lim_{p\to 0}\left(\frac{1}{n}\sum_{i=1}^n a_i^p\right)^{1/p} = \lim_{p\to 0}\left(1 + p\frac{1}{n}\sum_{i=1}^n \ln a_i + O(p^2)\right)^{1/p}$$
$$= e^{\frac{1}{n}\sum_{i=1}^n \ln a_i} = \sqrt[n]{a_1 \dots a_n}.$$

So if we now define $\|a\|_0 := \sqrt[n]{a_1 \dots a_n}$, the map $p \in \mathbb{R} \to \|a\|_p \in (0,\infty)$ is continuous and increasing in p.

We will now show that $\lim_{p\to\infty}\|a\|_p = \max_i a_i =: M$ and $\lim_{p\to-\infty}\|a\|_p = \min_i a_i =: m$. Indeed, for $p > 0$,

$$\frac{1}{n}M^p \le \frac{1}{n}\sum_{i=1}^n a_i^p \le M^p$$

and therefore,

$$\left(\frac{1}{n}\right)^{1/p} M \le \|a\|_p \le M.$$

Since $\left(\frac{1}{n}\right)^{1/p} \to 1$ as $p \to \infty$, it follows that $\lim_{p\to\infty}\|a\|_p = M$. For $p = -q < 0$, we have

$$\lim_{p\to-\infty}\|a\|_p = \lim_{q\to\infty}\left(\frac{1}{\left\|\frac{1}{a}\right\|_q}\right) = \frac{1}{\max_i(1/a_i)} = \frac{1}{1/m} = m = \min_i a_i.$$

Conclusion. If we extend the definition of $\|a\|_p$ to $p = \infty$ and $p = -\infty$ by $\|a\|_\infty = \max_i a_i$ and $\|a\|_{-\infty} = \min_i a_i$, then $\bar{\mathbb{R}} \ni p \to \|a\|_p \in (0,\infty)$ is a continuous non-decreasing function of p.

Proposition 12.33. *Suppose that $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ be defined by*

$$\frac{1}{p_\lambda} = \frac{1-\lambda}{p_0} + \frac{\lambda}{p_1} \tag{12.20}$$

with the interpretation that $\lambda/p_1 = 0$ if $p_1 = \infty$.[5] *Then $L^{p_\lambda} \subset L^{p_0} + L^{p_1}$, i.e. every function $f \in L^{p_\lambda}$ may be written as $f = g + h$ with $g \in L^{p_0}$ and $h \in L^{p_1}$. For $1 \le p_0 < p_1 \le \infty$ and $f \in L^{p_0} + L^{p_1}$ let*

$$\|f\| := \inf\left\{\|g\|_{p_0} + \|h\|_{p_1} : f = g + h\right\}.$$

Then $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space and the inclusion map from L^{p_λ} to $L^{p_0} + L^{p_1}$ is bounded; in fact $\|f\| \le 2\|f\|_{p_\lambda}$ for all $f \in L^{p_\lambda}$.

Proof. Let $M > 0$, then the local singularities of f are contained in the set $E := \{|f| > M\}$ and the behavior of f at "infinity" is solely determined by f on E^c. Hence let $g = f1_E$ and $h = f1_{E^c}$ so that $f = g + h$. By our earlier discussion we expect that $g \in L^{p_0}$ and $h \in L^{p_1}$ and this is the case since,

$$\|g\|_{p_0}^{p_0} = \int |f|^{p_0} 1_{|f|>M} = M^{p_0}\int\left|\frac{f}{M}\right|^{p_0} 1_{|f|>M}$$
$$\le M^{p_0}\int\left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|>M} \le M^{p_0-p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty$$

and

$$\|h\|_{p_1}^{p_1} = \left\|f1_{|f|\le M}\right\|_{p_1}^{p_1} = \int |f|^{p_1} 1_{|f|\le M} = M^{p_1}\int\left|\frac{f}{M}\right|^{p_1} 1_{|f|\le M}$$
$$\le M^{p_1}\int\left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|\le M} \le M^{p_1-p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty.$$

Moreover this shows

$$\|f\| \le M^{1-p_\lambda/p_0}\|f\|_{p_\lambda}^{p_\lambda/p_0} + M^{1-p_\lambda/p_1}\|f\|_{p_\lambda}^{p_\lambda/p_1}.$$

Taking $M = \lambda\|f\|_{p_\lambda}$ then gives

$$\|f\| \le \left(\lambda^{1-p_\lambda/p_0} + \lambda^{1-p_\lambda/p_1}\right)\|f\|_{p_\lambda}$$

and then taking $\lambda = 1$ shows $\|f\| \le 2\|f\|_{p_\lambda}$. The proof that $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space is left as Exercise 12.11 to the reader. ∎

[5] A little algebra shows that λ may be computed in terms of p_0, p_λ and p_1 by

$$\lambda = \frac{p_0}{p_\lambda}\cdot\frac{p_1 - p_\lambda}{p_1 - p_0}.$$

Corollary 12.34 (Interpolation of L^p – norms). *Suppose that $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ be defined as in Eq. (12.20), then $L^{p_0} \cap L^{p_1} \subset L^{p_\lambda}$ and*

$$\|f\|_{p_\lambda} \le \|f\|_{p_0}^{\lambda} \|f\|_{p_1}^{1-\lambda}. \tag{12.21}$$

Further assume $1 \le p_0 < p_\lambda < p_1 \le \infty$, and for $f \in L^{p_0} \cap L^{p_1}$ let

$$\|f\| := \|f\|_{p_0} + \|f\|_{p_1}.$$

Then $(L^{p_0} \cap L^{p_1}, \|\cdot\|)$ is a Banach space and the inclusion map of $L^{p_0} \cap L^{p_1}$ into L^{p_λ} is bounded, in fact

$$\|f\|_{p_\lambda} \le \max\left(\lambda^{-1}, (1-\lambda)^{-1}\right)\left(\|f\|_{p_0} + \|f\|_{p_1}\right). \tag{12.22}$$

The heuristic explanation of this corollary is that if $f \in L^{p_0} \cap L^{p_1}$, then f has local singularities no worse than an L^{p_1} function and behavior at infinity no worse than an L^{p_0} function. Hence $f \in L^{p_\lambda}$ for any p_λ between p_0 and p_1.

Proof. Let λ be determined as above, $a = p_0/\lambda$ and $b = p_1/(1-\lambda)$, then by Theorem 12.20,

$$\|f\|_{p_\lambda} = \left\| |f|^\lambda |f|^{1-\lambda} \right\|_{p_\lambda} \le \left\| |f|^\lambda \right\|_a \left\| |f|^{1-\lambda} \right\|_b = \|f\|_{p_0}^{\lambda} \|f\|_{p_1}^{1-\lambda}.$$

It is easily checked that $\|\cdot\|$ is a norm on $L^{p_0} \cap L^{p_1}$. To show this space is complete, suppose that $\{f_n\} \subset L^{p_0} \cap L^{p_1}$ is a $\|\cdot\|$ – Cauchy sequence. Then $\{f_n\}$ is both L^{p_0} and L^{p_1} – Cauchy. Hence there exist $f \in L^{p_0}$ and $g \in L^{p_1}$ such that $\lim_{n\to\infty} \|f - f_n\|_{p_0} = 0$ and $\lim_{n\to\infty} \|g - f_n\|_{p_\lambda} = 0$. By Chebyshev's inequality (Lemma 12.4) $f_n \to f$ and $f_n \to g$ in measure and therefore by Theorem 12.8, $f = g$ a.e. It now is clear that $\lim_{n\to\infty} \|f - f_n\| = 0$. The estimate in Eq. (12.22) is left as Exercise 12.10 to the reader. ■

Remark 12.35. Combining Proposition 12.33 and Corollary 12.34 gives

$$L^{p_0} \cap L^{p_1} \subset L^{p_\lambda} \subset L^{p_0} + L^{p_1}$$

for $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ as in Eq. (12.20).

Corollary 12.36. *Suppose now that μ is counting measure on Ω. Then $L^p(\mu) \subset L^q(\mu)$ for all $0 < p < q \le \infty$ and $\|f\|_q \le \|f\|_p$.*

Proof. Suppose that $0 < p < q = \infty$, then

$$\|f\|_\infty^p = \sup\{|f(x)|^p : x \in \Omega\} \le \sum_{x\in\Omega} |f(x)|^p = \|f\|_p^p,$$

i.e. $\|f\|_\infty \le \|f\|_p$ for all $0 < p < \infty$. For $0 < p \le q \le \infty$, apply Corollary 12.34 with $p_0 = p$ and $p_1 = \infty$ to find

$$\|f\|_q \le \|f\|_p^{p/q} \|f\|_\infty^{1-p/q} \le \|f\|_p^{p/q} \|f\|_p^{1-p/q} = \|f\|_p.$$

■

12.6.1 Summary:

1. $L^{p_0} \cap L^{p_1} \subset L^q \subset L^{p_0} + L^{p_1}$ for any $q \in (p_0, p_1)$.
2. If $p \le q$, then $\ell^p \subset \ell^q$ and $\|f\|_q \le \|f\|_p$.
3. Since $\mu(|f| > \varepsilon) \le \varepsilon^{-p} \|f\|_p^p$, L^p – convergence implies L^0 – convergence.
4. L^0 – convergence implies almost everywhere convergence for some subsequence.
5. If $\mu(\Omega) < \infty$ then almost everywhere convergence implies uniform convergence off certain sets of small measure and in particular we have L^0 – convergence.
6. If $\mu(\Omega) < \infty$, then $L^q \subset L^p$ for all $p \le q$ and L^q – convergence implies L^p – convergence.

12.7 Uniform Integrability

This section will address the question as to what extra conditions are needed in order that an L^0 – convergent sequence is L^p – convergent. This will lead us to the notion of uniform integrability. To simplify matters a bit here, it will be assumed that $(\Omega, \mathcal{B}, \mu)$ is a finite measure space for this section.

Notation 12.37 *For $f \in L^1(\mu)$ and $E \in \mathcal{B}$, let*

$$\mu(f : E) := \int_E f d\mu.$$

and more generally if $A, B \in \mathcal{B}$ let

$$\mu(f : A, B) := \int_{A\cap B} f d\mu.$$

When μ is a probability measure, we will often write $\mathbb{E}[f : E]$ for $\mu(f : E)$ and $\mathbb{E}[f : A, B]$ for $\mu(f : A, B)$.

Definition 12.38. *A collection of functions, $\Lambda \subset L^1(\mu)$ is said to be* ***uniformly integrable*** *if,*

$$\lim_{a\to\infty} \sup_{f\in\Lambda} \mu(|f| : |f| \ge a) = 0. \tag{12.23}$$

In words, $\Lambda \subset L^1(\mu)$ is uniformly integrable if "tail expectations" can be made uniformly small.

The condition in Eq. (12.23) implies $\sup_{f\in\Lambda}\|f\|_1<\infty$.[6] Indeed, choose a sufficiently large so that $\sup_{f\in\Lambda}\mu(|f|:|f|\geq a)\leq 1$, then for $f\in\Lambda$

$$\|f\|_1=\mu(|f|:|f|\geq a)+\mu(|f|:|f|<a)\leq 1+a\mu(\Omega).$$

Example 12.39. If $\Lambda=\{f\}$ with $f\in L^1(\mu)$, then Λ is uniformly integrable. Indeed, $\lim_{a\to\infty}\mu(|f|:|f|\geq a)=0$ by the dominated convergence theorem.

Exercise 12.5. Suppose A is an index set, $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are two collections of random variables. If $\{g_\alpha\}_{\alpha\in A}$ is uniformly integrable and $|f_\alpha|\leq|g_\alpha|$ for all $\alpha\in A$, show $\{f_\alpha\}_{\alpha\in A}$ is uniformly integrable as well.

Solution to Exercise (12.5). For $a>0$ we have

$$\mathbb{E}[|f_\alpha|:|f_\alpha|\geq a]\leq\mathbb{E}[|g_\alpha|:|f_\alpha|\geq a]\leq\mathbb{E}[|g_\alpha|:|g_\alpha|\geq a].$$

Therefore,

$$\lim_{a\to\infty}\sup_\alpha\mathbb{E}[|f_\alpha|:|f_\alpha|\geq a]\leq\lim_{a\to\infty}\sup_\alpha\mathbb{E}[|g_\alpha|:|g_\alpha|\geq a]=0.$$

Definition 12.40. *A collection of functions, $\Lambda\subset L^1(\mu)$ is said to be* ***uniformly absolutely continuous*** *if for all $\varepsilon>0$ there exists $\delta>0$ such that*

$$\sup_{f\in\Lambda}\mu(|f|:E)<\varepsilon \text{ whenever } \mu(E)<\delta. \tag{12.24}$$

Equivalently put,

$$\lim_{\delta\downarrow 0}\sup\{\mu(|f|:E):f\in\Lambda \text{ and } \mu(E)<\delta\}=0. \tag{12.25}$$

Remark 12.41. It is not in general true that if $\{f_n\}\subset L^1(\mu)$ is uniformly absolutely continuous implies $\sup_n\|f_n\|_1<\infty$. For example take $\Omega=\{*\}$ and $\mu(\{*\})=1$. Let $f_n(*)=n$. Since for $\delta<1$ a set $E\subset\Omega$ such that $\mu(E)<\delta$ is in fact the empty set and hence $\{f_n\}_{n=1}^\infty$ is uniformly absolutely continuous. However, for finite measure spaces without "atoms", for every $\delta>0$ we may find a finite partition of Ω by sets $\{E_\ell\}_{\ell=1}^k$ with $\mu(E_\ell)<\delta$. If Eq. (12.24) holds with $\varepsilon=1$, then

$$\mu(|f_n|)=\sum_{\ell=1}^k\mu(|f_n|:E_\ell)\leq k$$

showing that $\mu(|f_n|)\leq k$ for all n.

[6] This is not necessarily the case if $\mu(\Omega)=\infty$. Indeed, if $\Omega=\mathbb{R}$ and $\mu=m$ is Lebesgue measure, the sequences of functions, $\{f_n:=1_{[-n,n]}\}_{n=1}^\infty$ are uniformly integrable but not bounded in $L^1(m)$.

Proposition 12.42. *A subset $\Lambda\subset L^1(\mu)$ is uniformly integrable iff $\Lambda\subset L^1(\mu)$ is bounded and uniformly absolutely continuous.*

Proof. ($\Longrightarrow$) We have already seen that uniformly integrable subsets, Λ, are bounded in $L^1(\mu)$. Moreover, for $f\in\Lambda$, and $E\in\mathcal{B}$,

$$\begin{aligned}\mu(|f|:E)&=\mu(|f|:|f|\geq M,E)+\mu(|f|:|f|<M,E)\\&\leq\mu(|f|:|f|\geq M)+M\mu(E).\end{aligned}$$

Therefore,

$$\lim_{\delta\downarrow 0}\sup\{\mu(|f|:E):f\in\Lambda\text{ and }\mu(E)<\delta\}\leq\sup_{f\in\Lambda}\mu(|f|:|f|\geq M)\to 0\text{ as }M\to\infty$$

which verifies that Λ is uniformly absolutely continuous.

($\Longleftarrow$) Let $K:=\sup_{f\in\Lambda}\|f\|_1<\infty$. Then for $f\in\Lambda$, we have

$$\mu(|f|\geq a)\leq\|f\|_1/a\leq K/a\text{ for all }a>0.$$

Hence given $\varepsilon>0$ and $\delta>0$ as in the definition of uniform absolute continuity, we may choose $a=K/\delta$ in which case

$$\sup_{f\in\Lambda}\mu(|f|:|f|\geq a)<\varepsilon.$$

Since $\varepsilon>0$ was arbitrary, it follows that $\lim_{a\to\infty}\sup_{f\in\Lambda}\mu(|f|:|f|\geq a)=0$ as desired. ∎

Corollary 12.43. *Suppose $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are two uniformly integrable collections of functions, then $\{f_\alpha+g_\alpha\}_{\alpha\in A}$ is also uniformly integrable.*

Proof. By Proposition 12.42, $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are both bounded in $L^1(\mu)$ and are both uniformly absolutely continuous. Since $\|f_\alpha+g_\alpha\|_1\leq\|f_\alpha\|_1+\|g_\alpha\|_1$ it follows that $\{f_\alpha+g_\alpha\}_{\alpha\in A}$ is bounded in $L^1(\mu)$ as well. Moreover, for $\varepsilon>0$ we may choose $\delta>0$ such that $\mu(|f_\alpha|:E)<\varepsilon$ and $\mu(|g_\alpha|:E)<\varepsilon$ whenever $\mu(E)<\delta$. For this choice of ε and δ, we then have

$$\mu(|f_\alpha+g_\alpha|:E)\leq\mu(|f_\alpha|+|g_\alpha|:E)<2\varepsilon\text{ whenever }\mu(E)<\delta,$$

showing $\{f_\alpha+g_\alpha\}_{\alpha\in A}$ uniformly absolutely continuous. Another application of Proposition 12.42 completes the proof. ∎

Exercise 12.6 (Problem 5 on p. 196 of Resnick.). Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of integrable and i.i.d random variables. Then $\left\{\frac{S_n}{n}\right\}_{n=1}^\infty$ is uniformly integrable.

Theorem 12.44 (Vitali Convergence Theorem). *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space, $\Lambda := \{f_n\}_{n=1}^\infty$ be a sequence of functions in $L^1(\mu)$, and $f : \Omega \to \mathbb{C}$ be a measurable function. Then $f \in L^1(\mu)$ and $\|f - f_n\|_1 \to 0$ as $n \to \infty$ iff $f_n \to f$ in μ measure and Λ is uniformly integrable.*

Proof. ($\Longrightarrow$) If $f_n \to f$ in $L^1(\mu)$, then by Chebyschev's inequality it follows that $f_n \to f$ in $\mu-$measure. Given $\varepsilon > 0$ we may choose $N = N_\varepsilon \in \mathbb{N}$ such that $\|f - f_n\|_1 \le \varepsilon/2$ for $n \ge N_\varepsilon$. Since convergent sequences are bounded, we have $K := \sup_n \|f_n\|_1 < \infty$ and $\mu(|f_n| \ge a) \le K/a$ for all $a > 0$. Applying Proposition 12.42 with $\Lambda = \{f\}$, for any a sufficiently large we will have $\sup_n \mu(|f| : |f_n| \ge a) \le \varepsilon/2$. Thus for a sufficiently large and $n \ge N$, it follows that

$$\begin{aligned}\mu(|f_n| : |f_n| \ge a) &\le \mu(|f - f_n| : |f_n| \ge a) + \mu(|f| : |f_n| \ge a) \\ &\le \|f - f_n\|_1 + \mu(|f| : |f_n| \ge a) \le \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

By Example 12.39 we also know that $\limsup_{a\to\infty} \max_{n<N} \mu(|f_n| : |f_n| \ge a) = 0$ for any finite N. Therefore we have shown,

$$\limsup_{a\to\infty} \sup_n \mu(|f_n| : |f_n| \ge a) \le \varepsilon$$

and as $\varepsilon > 0$ was arbitrary it follows that $\{f_n\}_{n=1}^\infty$ is uniformly integrable.

($\Longleftarrow$) If $f_n \to f$ in μ measure and $\Lambda = \{f_n\}_{n=1}^\infty$ is uniformly integrable then we know $M := \sup_n \|f_n\|_1 < \infty$. Hence and application of Fatou's lemma, see Exercise 12.1,

$$\int_\Omega |f| \, d\mu \le \liminf_{n\to\infty} \int_\Omega |f_n| \, d\mu \le M < \infty,$$

i.e. $f \in L^1(\mu)$. It then follows by Example 12.39 and Corollary 12.43 that $\Lambda_0 := \{f - f_n\}_{n=1}^\infty$ is uniformly integrable.

Therefore,

$$\begin{aligned}\|f - f_n\|_1 &= \mu(|f - f_n| : |f - f_n| \ge a) + \mu(|f - f_n| : |f - f_n| < a) \\ &\le \varepsilon(a) + \int_\Omega 1_{|f-f_n|<a} |f - f_n| \, d\mu \end{aligned} \tag{12.26}$$

where

$$\varepsilon(a) := \sup_m \mu(|f - f_m| : |f - f_m| \ge a) \to 0 \text{ as } a \to \infty.$$

Since $1_{|f-f_n|<a} |f - f_n| \le a \in L^1(\mu)$ and $1_{|f-f_n|<a} |f - f_n| \xrightarrow{\mu} 0$ becuase

$$\mu\left(1_{|f-f_n|<a} |f - f_n| > \varepsilon\right) \le \mu(|f - f_n| > \varepsilon) \to 0 \text{ as } n \to \infty,$$

we may pass to the limit in Eq. (12.26), with the aid of the dominated convergence theorem (see Corollary 12.9), to find

$$\limsup_{n\to\infty} \|f - f_n\|_1 \le \varepsilon(a) \to 0 \text{ as } a \to \infty.$$

■

Example 12.45. Let $\Omega = [0,1]$, $\mathcal{B} = \mathcal{B}_{[0,1]}$ and $P = m$ be Lebesgue measure on $\mathcal{B}$. Then the collection of functions, $f_\varepsilon := \frac{1}{\varepsilon} 1_{[0,\varepsilon]}$ for $\varepsilon \in (0,1)$ is bounded in $L^1(P)$, $f_\varepsilon \to 0$ a.e. as $\varepsilon \downarrow 0$ but

$$0 = \int_\Omega \lim_{\varepsilon\downarrow 0} f_\varepsilon dP \neq \lim_{\varepsilon\downarrow 0} \int_\Omega f_\varepsilon dP = 1.$$

This is a typical example of a bounded and pointwise convergent sequence in L^1 which is **not** uniformly integrable. This is easy to check directly as well since,

$$\sup_{\varepsilon\in(0,1)} m(|f_\varepsilon| : |f_\varepsilon| \ge a) = 1 \text{ for all } a > 0.$$

Example 12.46. Let $\Omega = [0,1]$, P be Lebesgue measure on $\mathcal{B} = \mathcal{B}_{[0,1]}$, and for $\varepsilon \in (0,1)$ let $a_\varepsilon > 0$ with $\lim_{\varepsilon\downarrow 0} a_\varepsilon = \infty$ and let $f_\varepsilon := a_\varepsilon 1_{[0,\varepsilon]}$. Then $\mathbb{E} f_\varepsilon = \varepsilon a_\varepsilon$ and so $\sup_{\varepsilon>0} \|f_\varepsilon\|_1 =: K < \infty$ iff $\varepsilon a_\varepsilon \le K$ for all ε. Since

$$\sup_\varepsilon \mathbb{E}[f_\varepsilon : f_\varepsilon \ge M] = \sup_\varepsilon [\varepsilon a_\varepsilon \cdot 1_{a_\varepsilon \ge M}],$$

if $\{f_\varepsilon\}$ is uniformly integrable and $\delta > 0$ is given, for large M we have $\varepsilon a_\varepsilon \le \delta$ for ε small enough so that $a_\varepsilon \ge M$. From this we conclude that $\limsup_{\varepsilon\downarrow 0} (\varepsilon a_\varepsilon) \le \delta$ and since $\delta > 0$ was arbitrary, $\lim_{\varepsilon\downarrow 0} \varepsilon a_\varepsilon = 0$ if $\{f_\varepsilon\}$ is uniformly integrable. By reversing these steps one sees the converse is also true.

Alternatively. No matter how $a_\varepsilon > 0$ is chosen, $\lim_{\varepsilon\downarrow 0} f_\varepsilon = 0$ a.s.. So from Theorem 12.44, if $\{f_\varepsilon\}$ is uniformly integrable we would have to have

$$\lim_{\varepsilon\downarrow 0} (\varepsilon a_\varepsilon) = \lim_{\varepsilon\downarrow 0} \mathbb{E} f_\varepsilon = \mathbb{E}0 = 0.$$

Corollary 12.47. *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space, $p \in [1,\infty)$, $\{f_n\}_{n=1}^\infty$ be a sequence of functions in $L^p(\mu)$, and $f : \Omega \to \mathbb{C}$ be a measurable function. Then $f \in L^p(\mu)$ and $\|f - f_n\|_p \to 0$ as $n \to \infty$ iff $f_n \to f$ in μ measure and $\Lambda := \{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable.*

Proof. ($\Longleftarrow$) Suppose that $f_n \to f$ in μ measure and $\Lambda := \{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable. By Corollary 12.12, $|f_n|^p \xrightarrow{\mu} |f|^p$ in $\mu-$measure, and $h_n := |f - f_n|^p \xrightarrow{\mu} 0$, and by Theorem 12.44, $|f|^p \in L^1(\mu)$ and $|f_n|^p \to |f|^p$ in $L^1(\mu)$. It now follows by an application of Lemma 12.10 that $\|f - f_n\|_p \to 0$ as $n \to \infty$.

($\Longrightarrow$) Suppose $f \in L^p$ and $f_n \to f$ in L^p. Again $f_n \to f$ in $\mu-$measure by Lemma 12.4. Let

$$h_n := ||f_n|^p - |f|^p| \le |f_n|^p + |f|^p =: g_n \in L^1$$

and $g := 2|f|^p \in L^1$. Then $g_n \xrightarrow{\mu} g$, $h_n \xrightarrow{\mu} 0$ and $\int g_n d\mu \to \int g d\mu$. Therefore by the dominated convergence theorem in Corollary 12.9, $\lim_{n\to\infty} \int h_n \, d\mu = 0$, i.e. $|f_n|^p \to |f|^p$ in $L^1(\mu)$.[7] Hence it follows from Theorem 12.44 that Λ is uniformly integrable. ■

The following Lemma gives a concrete necessary and sufficient conditions for verifying a sequence of functions is uniformly integrable.

Lemma 12.48. *Suppose that $\mu(\Omega) < \infty$, and $\Lambda \subset L^0(\Omega)$ is a collection of functions.*

1. *If there exists a measurable function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{x\to\infty} \varphi(x)/x = \infty$ and*
$$K := \sup_{f\in\Lambda} \mu(\varphi(|f|)) < \infty, \tag{12.27}$$
then Λ is uniformly integrable. (A typical example for φ in item 1. is $\varphi(x) = x^p$ for some $p > 1$.)
2. **(Skip this if you like.) Conversely if Λ is uniformly integrable, there exists a non-decreasing continuous function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi(0) = 0$, $\lim_{x\to\infty} \varphi(x)/x = \infty$ and Eq. (12.27) is valid.*

Proof. 1. Let φ be as in item 1. above and set $\varepsilon_a := \sup_{x\ge a} \frac{x}{\varphi(x)} \to 0$ as $a \to \infty$ by assumption. Then for $f \in \Lambda$

$$\begin{aligned}\mu(|f| : |f| \ge a) &= \mu\left(\frac{|f|}{\varphi(|f|)}\varphi(|f|) : |f| \ge a\right) \le \mu(\varphi(|f|) : |f| \ge a)\varepsilon_a \\ &\le \mu(\varphi(|f|))\varepsilon_a \le K\varepsilon_a\end{aligned}$$

and hence

$$\lim_{a\to\infty} \sup_{f\in\Lambda} \mu\left(|f| 1_{|f|\ge a}\right) \le \lim_{a\to\infty} K\varepsilon_a = 0.$$

2. *(Skip this if you like.) By assumption, $\varepsilon_a := \sup_{f\in\Lambda} \mu\left(|f| 1_{|f|\ge a}\right) \to 0$ as $a \to \infty$. Therefore we may choose $a_n \uparrow \infty$ such that

$$\sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n} < \infty$$

where by convention $a_0 := 0$. Now define φ so that $\varphi(0) = 0$ and

$$\varphi'(x) = \sum_{n=0}^{\infty} (n+1)\, 1_{(a_n, a_{n+1}]}(x),$$

i.e.

$$\varphi(x) = \int_0^x \varphi'(y)dy = \sum_{n=0}^{\infty} (n+1)\,(x \wedge a_{n+1} - x \wedge a_n)\,.$$

By construction φ is continuous, $\varphi(0) = 0$, $\varphi'(x)$ is increasing (so φ is convex) and $\varphi'(x) \ge (n+1)$ for $x \ge a_n$. In particular

$$\frac{\varphi(x)}{x} \ge \frac{\varphi(a_n) + (n+1)x}{x} \ge n+1 \text{ for } x \ge a_n$$

from which we conclude $\lim_{x\to\infty} \varphi(x)/x = \infty$. We also have $\varphi'(x) \le (n+1)$ on $[0, a_{n+1}]$ and therefore

$$\varphi(x) \le (n+1)x \text{ for } x \le a_{n+1}.$$

So for $f \in \Lambda$,

$$\begin{aligned}\mu(\varphi(|f|)) &= \sum_{n=0}^{\infty} \mu\left(\varphi(|f|)1_{(a_n,a_{n+1}]}(|f|)\right) \\ &\le \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\, 1_{(a_n,a_{n+1}]}(|f|)\right) \\ &\le \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\, 1_{|f|\ge a_n}\right) \le \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n}\end{aligned}$$

and hence

$$\sup_{f\in\Lambda} \mu(\varphi(|f|)) \le \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n} < \infty.$$

■

[7] Here is an alternative proof. By the mean value theorem,

$$||f|^p - |f_n|^p| \le p(\max(|f|, |f_n|))^{p-1} \,||f| - |f_n|| \le p(|f| + |f_n|)^{p-1} \,||f| - |f_n||$$

and therefore by Hölder's inequality,

$$\begin{aligned}\int ||f|^p - |f_n|^p|\, d\mu &\le p \int (|f| + |f_n|)^{p-1} \,||f| - |f_n||\, d\mu \le p \int (|f| + |f_n|)^{p-1} \,|f - f_n|\, d\mu \\ &\le p\|f - f_n\|_p \|(|f| + |f_n|)^{p-1}\|_q = p\|\,|f| + |f_n|\|_p^{p/q} \|f - f_n\|_p \\ &\le p(\|f\|_p + \|f_n\|_p)^{p/q} \|f - f_n\|_p\end{aligned}$$

where $q := p/(p-1)$. This shows that $\int ||f|^p - |f_n|^p|\, d\mu \to 0$ as $n \to \infty$.

Exercise 12.7. Show directly that if $\mu(\Omega) < \infty$, φ is as in Lemma 12.48, and $\{f_n\} \subset L^1(\Omega)$ such that $f_n \xrightarrow{\mu} f$ and $K := \sup_n \mathbb{E}[\varphi(|f_n|)] < \infty$, then $\|f - f_n\|_1 \to 0$ as $n \to \infty$.

Solution to Exercise (12.7). Letting $\varepsilon_a := \sup_{x \geq a} \frac{x}{\varphi(x)}$ as above we have $x \leq \varepsilon_a \varphi(x)$ for $x \geq a$. Therefore,

$$\begin{aligned}\mathbb{E}|f_n| &= \mathbb{E}[|f_n| : |f_n| \geq a] + \mathbb{E}[|f_n| : |f_n| < a] \\ &\leq \varepsilon_a \mathbb{E}[\varphi(|f_n|) : |f_n| \geq a] + a \\ &\leq \varepsilon_a \mathbb{E}[\varphi(|f_n|)] + a = a + \varepsilon_a K\end{aligned}$$

from which it follows $\sup_n \mathbb{E}|f_n| < \infty$. Hence by Fatou's lemma,

$$\mathbb{E}|f| \leq \liminf_{n\to\infty} \mathbb{E}|f_n| \leq \sup_n \mathbb{E}|f_n| < \infty.$$

Similarly,

$$\begin{aligned}\mathbb{E}[|f - f_n| : |f_n| \geq a] &\leq \mathbb{E}[|f| : |f_n| \geq a] + \mathbb{E}[|f_n| : |f_n| \geq a] \\ &\leq \mathbb{E}[|f| : |f_n| \geq a] + \varepsilon_a K\end{aligned}$$

and for $b > 0$

$$\begin{aligned}\mathbb{E}[|f| : |f_n| \geq a] &= \mathbb{E}[|f| : |f| \geq b, |f_n| \geq a] + \mathbb{E}[|f| : |f| < b, |f_n| \geq a] \\ &\leq \mathbb{E}[|f| : |f| \geq b] + bP(|f_n| \geq a) \\ &\leq \mathbb{E}[|f| : |f| \geq b] + b\frac{1}{a}\sup_n \mathbb{E}|f_n|.\end{aligned}$$

Therefore,

$$\begin{aligned}\limsup_{a\to\infty} \mathbb{E}[|f - f_n| : |f_n| \geq a] &\leq \limsup_{a\to\infty} (\mathbb{E}[|f| : |f_n| \geq a] + \varepsilon_a K) \\ &\leq \mathbb{E}[|f| : |f| \geq b].\end{aligned}$$

Now by the DCT, $\lim_{n\to\infty} \mathbb{E}[|f - f_n| : |f_n| < a] = 0$ and hence

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{E}[|f - f_n|] &\leq \limsup_{n\to\infty} \mathbb{E}[|f - f_n| : |f_n| \geq a] + \limsup_{n\to\infty} \mathbb{E}[|f - f_n| : |f_n| < a] \\ &\leq \mathbb{E}[|f| : |f| \geq b].\end{aligned}$$

Another application of the DCT now show $\mathbb{E}[|f| : |f| \geq b] \to 0$ as $b \to \infty$ which completes the proof.

12.8 Exercises

Exercise 12.8. Let $f \in L^p \cap L^\infty$ for some $p < \infty$. Show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$. If we further assume $\mu(X) < \infty$, show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$ for all measurable functions $f : X \to \mathbb{C}$. In particular, $f \in L^\infty$ iff $\lim_{q\to\infty} \|f\|_q < \infty$. **Hints:** Use Corollary 12.34 to show $\limsup_{q\to\infty} \|f\|_q \leq \|f\|_\infty$ and to show $\liminf_{q\to\infty} \|f\|_q \geq \|f\|_\infty$, let $M < \|f\|_\infty$ and make use of Chebyshev's inequality.

Exercise 12.9. Let $\infty > a, b > 1$ with $a^{-1} + b^{-1} = 1$. Give a calculus proof of the inequality

$$st \leq \frac{s^a}{a} + \frac{t^b}{b} \text{ for all } s, t \geq 0.$$

Hint: by taking $s = xt^{b/a}$, show that it suffices to prove

$$x \leq \frac{x^a}{a} + \frac{1}{b} \text{ for all } x \geq 0.$$

and then maximize the function $f(x) = x - x^a/a$ for $x \in [0, \infty)$.

Exercise 12.10. Prove Eq. (12.22) in Corollary 12.34. (Part of Folland 6.3 on p. 186.) **Hint:** Use the inequality, with $a, b \geq 1$ with $a^{-1} + b^{-1} = 1$ chosen appropriately,

$$st \leq \frac{s^a}{a} + \frac{t^b}{b}$$

applied to the right side of Eq. (12.21).

Exercise 12.11. Complete the proof of Proposition 12.33 by showing $(L^p + L^r, \|\cdot\|)$ is a Banach space.

Exercise 12.12. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space. Show directly that for any $g \in L^1(\mu)$, $\Lambda = \{g\}$ is uniformly absolutely continuous. (We already know this is true by combining Example 12.39 with Proposition 12.42.)

Solution to Exercise (12.12). First Proof. If the statement is false, there would exist $\varepsilon > 0$ and sets E_n such that $\mu(E_n) \to 0$ while $\mu(|g| : E_n) \geq \varepsilon$ for all n. Since $|1_{E_n} g| \leq |g| \in L^1$ and for any $\delta > 0$, $\mu(1_{E_n} \cdot |g| > \delta) \leq \mu(E_n) \to 0$ as $n \to \infty$ so that $1_{E_n} \cdot |g| \xrightarrow{\mu} 0$, the dominated convergence theorem of Corollary 12.9 implies $\lim_{n\to\infty} \mu(|g| : E_n) = 0$. This contradicts $\mu(|g| : E_n) \geq \varepsilon$ for all n and the proof is complete.

Second Proof. Let $\varphi = \sum_{i=1}^n c_i 1_{B_i}$ be a simple function such that $\|g - \varphi\|_1 < \varepsilon/2$. Then

$$\begin{aligned}\mu(|g|:E) &\le \mu(|\varphi|:E)+\mu(|g-\varphi|:E)\\ &\le \sum_{i=1}^n |c_i|\,\mu(E\cap B_i)+\|g-\varphi\|_1 \le \left(\sum_{i=1}^n |c_i|\right)\mu(E)+\varepsilon/2.\end{aligned}$$

This shows $\mu(|g|:E)<\varepsilon$ provided that $\mu(E)<\varepsilon\left(2\sum_{i=1}^n |c_i|\right)^{-1}$.

Exercise 12.13. Suppose that $(\Omega,\mathcal{B},P)$ is a probability space and $\{X_n\}_{n=1}^\infty$ is a sequence of uncorrelated (i.e. $\mathrm{Cov}(X_n,X_m)=0$ if $m\neq n$) square integrable random variables such that $\mu=\mathbb{E}X_n$ and $\sigma^2=\mathrm{Var}(X_n)$ for all n. Let $S_n:=X_1+\dots+X_n$. Show $\left\|\frac{S_n}{n}-\mu\right\|_2^2=\frac{\sigma^2}{n}\to 0$ as $n\to\infty$.

Solution to Exercise (12.13). To say that the $\{X_n\}_{n=1}^\infty$ are uncorrelated is equivalent to saying that $\{X_n-\mu\}_{n=1}^\infty$ is an orthogonal set. Thus by Pythagorean's theorem,

$$\begin{aligned}\left\|\frac{S_n}{n}-\mu\right\|_2^2 &= \frac{1}{n^2}\|S_n-n\mu\|_2^2\\ &=\frac{1}{n^2}\|(X_1-\mu)+\dots+(X_n-\mu)\|_2^2\\ &=\frac{1}{n^2}\sum_{i=1}^n\|X_i-\mu\|_2^2=\frac{n\sigma^2}{n^2}\to 0 \text{ as } n\to\infty.\end{aligned}$$

Exercise 12.14. Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. integrable random variables and $S_n:=X_1+\dots+X_n$ and $\mu:=\mathbb{E}X_n$. Show, $\frac{S_n}{n}\to\mu$ in $L^1(P)$ as $n\to\infty$. (Incidentally, this shows that $\left\{\frac{S_n}{n}\right\}_{n=1}^\infty$ is U.I. **Hint:** for $M\in(0,\infty)$, let $X_i^M:=X_i\cdot 1_{|X_i|\le M}$ and $S_n^M:=X_1^M+\dots+X_n^M$ and use Exercise 12.13 to see that

$$\frac{S_n^M}{n}\to\mathbb{E}X_1^M \text{ in } L^2(P)\subset L^1(P) \text{ for all } M.$$

Using this to show $\lim_{n\to\infty}\left\|\frac{S_n}{n}-\mathbb{E}X_1\right\|_1=0$ by getting good control on $\left\|\frac{S_n}{n}-\frac{S_n^M}{n}\right\|_1$ and $\left|\mathbb{E}X_n-\mathbb{E}X_n^M\right|$.

Exercise 12.15. Suppose $1\le p<\infty$, $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables such that $\mathbb{E}|X_n|^p<\infty$, $S_n:=X_1+\dots+X_n$ and $\mu:=\mathbb{E}X_n$. Show, $\frac{S_n}{n}\to\mu$ in $L^p(P)$ as $n\to\infty$. **Hint:** show $\left\{\left|\frac{S_n}{n}\right|^p\right\}_{n=1}^\infty$ is U.I. – this is not meant to be hard!

12.9 Appendix: Convex Functions

Reference; see the appendix (page 500) of Revuz and Yor.

Definition 12.49. *Given any function, $\varphi:(a,b)\to\mathbb{R}$, we say that φ is **convex** if for all $a<x_0\le x_1<b$ and $t\in[0,1]$,*

$$\varphi(x_t)\le h_t:=(1-t)\varphi(x_0)+t\varphi(x_1) \text{ for all } t\in[0,1], \tag{12.28}$$

where

$$x_t:=x_0+t(x_1-x_0)=(1-t)x_0+tx_1, \tag{12.29}$$

see Figure 12.3 below.

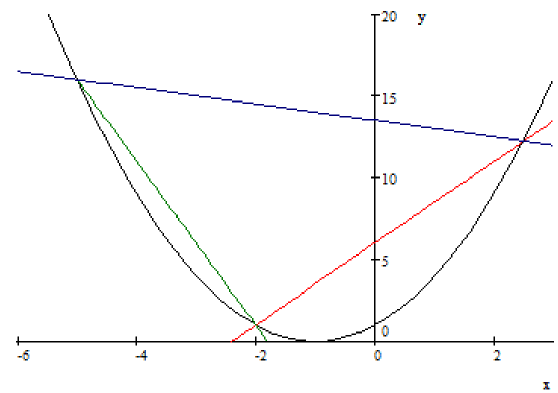


Fig. 12.3. A convex function along with three cords corresponding to $x_0=-5$ and $x_1=-2$ $x_0=-2$ and $x_1=5/2$, and $x_0=-5$ and $x_1=5/2$ with slopes, $m_1=-15/3$, $m_2=15/6$ and $m_3=-1/2$ respectively. Notice that $m_1\le m_3\le m_2$.

Lemma 12.50. *Let $\varphi:(a,b)\to\mathbb{R}$ be a function and*

$$F(x_0,x_1):=\frac{\varphi(x_1)-\varphi(x_0)}{x_1-x_0} \text{ for } a<x_0<x_1<b.$$

Then the following are equivalent;

1. *φ is convex,*
2. *$F(x_0,x_1)$ is non-decreasing in x_0 for all $a<x_0<x_1<b$, and*
3. *$F(x_0,x_1)$ is non-decreasing in x_1 for all $a<x_0<x_1<b$.*

Proof. Let x_t and h_t be as in Eq. (12.28), then (x_t,h_t) is on the line segment joining $(x_0,\varphi(x_0))$ to $(x_1,\varphi(x_1))$ and the statement that φ is convex is then equivalent to the assertion that $\varphi(x_t)\le h_t$ for all $0\le t\le 1$. Since (x_t,h_t) lies on a straight line we always have the following three slopes are equal;

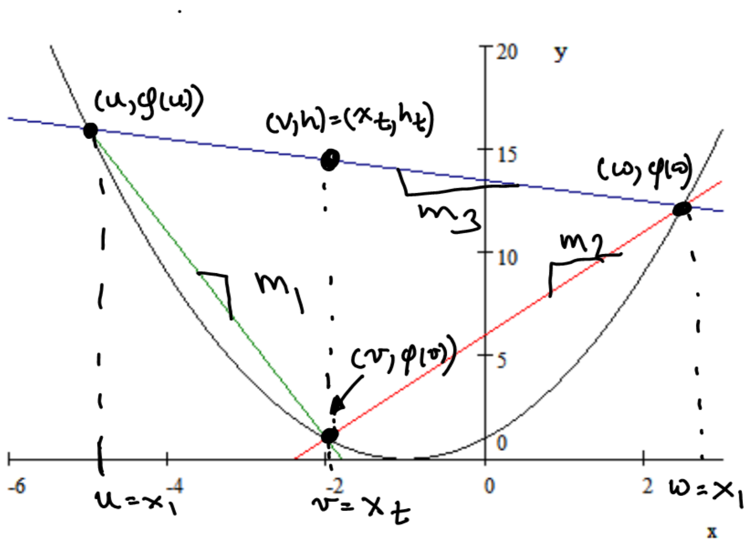


Fig. 12.4. A convex function with three cords. Notice the slope relationships; $m_1 \leq m_3 \leq m_2$.

$$\frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t}.$$

In light of this identity, it is now clear that the convexity of φ is equivalent to either,

$$F(x_0, x_t) = \frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} \leq \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = F(x_0, x_1)$$

or

$$F(x_0, x_1) = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \leq \frac{\varphi(x_1) - \varphi(x_t)}{x_1 - x_t} = F(x_t, x_1)$$

holding for all $x_0 < x_t < x_1$. ■

Lemma 12.51 (A generalized FTC). *If $\varphi \in PC^1((a,b) \to \mathbb{R})$*[8]*, then for all $a < x < y < b$,*

[8] PC^1 denotes the space of piecewise C^1 – functions, i.e. $\varphi \in PC^1((a,b) \to \mathbb{R})$ means the φ is continuous and there are a finite number of points,

$$\{a = a_0 < a_1 < a_2 < \dots < a_{n-1} < a_n = b\},$$

such that $\varphi|_{[a_{j-1}, a_j] \cap (a,b)}$ is C^1 for all $j = 1, 2, \dots, n$.

$$\varphi(y) - \varphi(x) = \int_x^y \varphi'(t)\, dt.$$

Proof. Let $b_1, \dots, b_{l-1}$ be the points of non-differentiability of φ in (x,y) and set $b_0 = x$ and $b_l = y$. Then

$$\varphi(y) - \varphi(x) = \sum_{k=1}^{l} [\varphi(b_k) - \varphi(b_{k-1})]$$
$$= \sum_{k=1}^{l} \int_{b_{k-1}}^{b_k} \varphi'(t)\, dt = \int_x^y \varphi'(t)\, dt.$$

■

Figure 12.5 below serves as motivation for the following elementary lemma on convex functions.

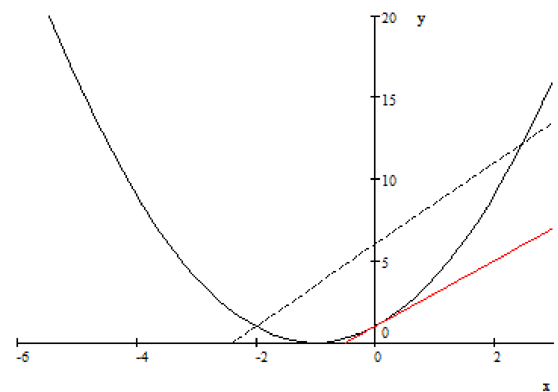


Fig. 12.5. A convex function, φ, along with a cord and a tangent line. Notice that the tangent line is always below φ and the cord lies above φ between the points of intersection of the cord with the graph of φ.

Lemma 12.52 (Convex Functions). *Let $\varphi \in PC^1((a,b) \to \mathbb{R})$ and for $x \in (a,b)$, let*

$$\varphi'(x+) := \lim_{h \downarrow 0} \frac{\varphi(x+h) - \varphi(x)}{h} \quad \text{and}$$
$$\varphi'(x-) := \lim_{h \uparrow 0} \frac{\varphi(x+h) - \varphi(x)}{h}.$$

(Of course, $\varphi'(x\pm) = \varphi'(x)$ at points $x \in (a,b)$ where φ is differentiable.)

1. *If $\varphi'(x) \le \varphi'(y)$ for all $a < x < y < b$ with x and y be points where φ is differentiable, then for any $x_0 \in (a,b)$, we have $\varphi'(x_0-) \le \varphi'(x_0+)$ and for $m \in (\varphi'(x_0-), \varphi'(x_0+))$ we have,*

$$\varphi(x_0) + m(x - x_0) \le \varphi(x) \ \forall \ x_0, x \in (a,b). \tag{12.30}$$

2. *If $\varphi \in PC^2((a,b) \to \mathbb{R})$[9] with $\varphi''(x) \ge 0$ for almost all $x \in (a,b)$, then Eq. (12.30) holds with $m = \varphi'(x_0)$.*
3. *If either of the hypothesis in items 1. and 2. above hold then φ is convex.*

(This lemma applies to the functions, $e^{\lambda x}$ for all $\lambda \in \mathbb{R}$, $|x|^\alpha$ for $\alpha > 1$, and $-\ln x$ to name a few examples. See Appendix 12.9 below for much more on convex functions.)

Proof. 1. If x_0 is a point where φ is not differentiable and $h > 0$ is small, by the mean value theorem, for all $h > 0$ small, there exists $c_+(h) \in (x_0, x_0 + h)$ and $c_-(h) \in (x_0 - h, x_0)$ such that

$$\frac{\varphi(x_0 - h) - \varphi(x_0)}{-h} = \varphi'(c_-(h)) \le \varphi'(c_+(h)) = \frac{\varphi(x_0 + h) - \varphi(x_0)}{h}.$$

Letting $h \downarrow 0$ in this equation shows $\varphi'(x_0-) \le \varphi'(x_0+)$. Furthermore if $x < x_0 < y$ with x and y being points of differentiability of φ, the for small $h > 0$,

$$\varphi'(x) \le \varphi'(c_-(h)) \le \varphi'(c_+(h)) \le \varphi'(y).$$

Letting $h \downarrow 0$ in these inequalities shows,

$$\varphi'(x) \le \varphi'(x_0-) \le \varphi'(x_0+) \le \varphi'(y). \tag{12.31}$$

Now let $m \in (\varphi'(x_0-), \varphi'(x_0+))$. By the fundamental theorem of calculus in Lemma 12.51 and making use of Eq. (12.31), if $x > x_0$ then

$$\varphi(x) - \varphi(x_0) = \int_{x_0}^x \varphi'(t)\,dt \ge \int_{x_0}^x m\ dt = m(x - x_0)$$

and if $x < x_0$, then

$$\varphi(x_0) - \varphi(x) = \int_x^{x_0} \varphi'(t)\,dt \le \int_x^{x_0} m\ dt = m(x_0 - x).$$

[9] PC^2 denotes the space of piecewise C^2 – functions, i.e. $\varphi \in PC^2((a,b) \to \mathbb{R})$ means the φ is C^1 and there are a finite number of points,

$$\{a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b\},$$

such that $\varphi|_{[a_{j-1},a_j]\cap(a,b)}$ is C^2 for all $j = 1, 2, \ldots, n$.

These two equations implies Eq. (12.30) holds.

2. Notice that $\varphi' \in PC^1((a,b))$ and therefore,

$$\varphi'(y) - \varphi'(x) = \int_x^y \varphi''(t)\,dt \ge 0 \text{ for all } a < x \le y < b$$

which shows that item 1. may be used.

Alternatively; by Taylor's theorem with integral remainder (see Eq. (7.55) with $F = \varphi$, $a = x_0$, and $b = x$) implies

$$\begin{aligned}\varphi(x) &= \varphi(x_0) + \varphi'(x_0)(x - x_0) + (x - x_0)^2 \int_0^1 \varphi''(x_0 + \tau(x - x_0))(1 - \tau)\,d\tau \\ &\ge \varphi(x_0) + \varphi'(x_0)(x - x_0).\end{aligned}$$

3. For any $\xi \in (a,b)$, let $h_\xi(x) := \varphi(x_0) + \varphi'(x_0)(x - x_0)$. By Eq. (12.30) we know that $h_\xi(x) \le \varphi(x)$ for all $\xi, x \in (a,b)$ with equality when $\xi = x$ and therefore,

$$\varphi(x) = \sup_{\xi \in (a,b)} h_\xi(x).$$

Since h_ξ is an affine function for each $\xi \in (a,b)$, it follows that

$$h_\xi(x_t) = (1 - t) h_\xi(x_0) + t h_\xi(x_1) \le (1 - t)\varphi(x_0) + t\varphi(x_1)$$

for all $t \in [0,1]$. Thus we may conclude that

$$\varphi(x_t) = \sup_{\xi \in (a,b)} h_\xi(x_t) \le (1 - t)\varphi(x_0) + t\varphi(x_1)$$

as desired.

*For fun, here are three more proofs of Eq. (12.28) under the hypothesis of item 2. Clearly these proofs may be omitted.

3a. By Lemma 12.50 below it suffices to show either

$$\frac{d}{dx}\frac{\varphi(y) - \varphi(x)}{y - x} \ge 0 \text{ or } \frac{d}{dy}\frac{\varphi(y) - \varphi(x)}{y - x} \ge 0 \text{ for } a < x < y < b.$$

For the first case,

$$\begin{aligned}\frac{d}{dx}\frac{\varphi(y) - \varphi(x)}{y - x} &= \frac{\varphi(y) - \varphi(x) - \varphi'(x)(y - x)}{(y - x)^2} \\ &= \int_0^1 \varphi''(x + t(y - x))(1 - t)\,dt \ge 0.\end{aligned}$$

Similarly,

$$\frac{d}{dy}\frac{\varphi(y)-\varphi(x)}{y-x}=\frac{\varphi'(y)(y-x)-[\varphi(y)-\varphi(x)]}{(y-x)^2}$$

where we now use,

$$\varphi(x)-\varphi(y)=\varphi'(y)(x-y)+(x-y)^2\int_0^1\varphi''(y+t(x-y))(1-t)\,dt$$

so that

$$\frac{\varphi'(y)(y-x)-[\varphi(y)-\varphi(x)]}{(y-x)^2}=(x-y)^2\int_0^1\varphi''(y+t(x-y))(1-t)\,dt\geq 0$$

again.

3b. Let

$$f(t):=\varphi(u)+t(\varphi(v)-\varphi(u))-\varphi(u+t(v-u)).$$

Then $f(0)=f(1)=0$ with $\ddot{f}(t)=-(v-u)^2\varphi''(u+t(v-u))\leq 0$ for almost all t. By the mean value theorem, there exists, $t_0\in(0,1)$ such that $\dot{f}(t_0)=0$ and then by the fundamental theorem of calculus it follows that

$$\dot{f}(t)=\int_{t_0}^t\ddot{f}(\tau)\,dt.$$

In particular, $\dot{f}(t)\leq 0$ for $t>t_0$ and $\dot{f}(t)\geq 0$ for $t<t_0$ and hence $f(t)\geq f(1)=0$ for $t\geq t_0$ and $f(t)\geq f(0)=0$ for $t\leq t_0$, i.e. $f(t)\geq 0$.

3c. Let $h:[0,1]\to\mathbb{R}$ be a piecewise C^2 – function. Then by the fundamental theorem of calculus and integration by parts,

$$h(t)=h(0)+\int_0^t h(\tau)\,d\tau=h(0)+th(t)-\int_0^t h(\tau)\,\tau d\tau$$

and

$$h(1)=h(t)+\int_t^1 h(\tau)\,d(\tau-1)=h(t)-(t-1)h(t)-\int_t^1 h(\tau)(\tau-1)\,d\tau.$$

Thus we have shown,

$$h(t)=h(0)+th(t)-\int_0^t h(\tau)\,\tau d\tau \text{ and}$$
$$h(t)=h(1)+(t-1)h(t)+\int_t^1 h(\tau)(\tau-1)\,d\tau.$$

So if we multiply the first equation by $(1-t)$ and add to it the second equation multiplied by t shows,

$$h(t)=(1-t)h(0)+th(1)-\int_0^1 G(t,\tau)\,\ddot{h}(\tau)\,d\tau, \tag{12.32}$$

where

$$G(t,\tau):=\begin{cases}\tau(1-t) \text{ if } \tau\leq t\\ t(1-\tau) \text{ if } \tau\geq t\end{cases}.$$

(The function $G(t,\tau)$ is the "Green's function" for the operator $-d^2/dt^2$ on $[0,1]$ with Dirichlet boundary conditions. The formula in Eq. (12.32) is a standard representation formula for $h(t)$ which appears naturally in the study of harmonic functions.)

We now take $h(t):=\varphi(x_0+t(x_1-x_0))$ in Eq. (12.32) to learn

$$\begin{aligned}\varphi(x_0+t(x_1-x_0))&=(1-t)\varphi(x_0)+t\varphi(x_1)\\&\quad-(x_1-x_0)^2\int_0^1 G(t,\tau)\,\ddot{\varphi}(x_0+\tau(x_1-x_0))\,d\tau\\&\leq(1-t)\varphi(x_0)+t\varphi(x_1),\end{aligned}$$

because $\ddot{\varphi}\geq 0$ and $G(t,\tau)\geq 0$. ■

Example 12.53. The functions $\exp(x)$ and $-\log(x)$ are convex and $|x|^p$ is convex iff $p\geq 1$ as follows from Lemma 12.52.

Example 12.54 (Proof of Lemma 10.40). Taking $\varphi(x)=e^{-x}$ in Lemma 12.52, Eq. (12.30) with $x_0=0$ implies (see Figure 10.1),

$$1-x\leq\varphi(x)=e^{-x}\text{ for all }x\in\mathbb{R}.$$

Taking $\varphi(x)=e^{-2x}$ in Lemma 12.52, Eq. (12.28) with $x_0=0$ and $x_1=1$ implies, for all $t\in[0,1]$,

$$\begin{aligned}e^{-t}&\leq\varphi\left((1-t)0+t\frac{1}{2}\right)\\&\leq(1-t)\varphi(0)+t\varphi\left(\frac{1}{2}\right)=1-t+te^{-1}\leq 1-\frac{1}{2}t,\end{aligned}$$

wherein the last equality we used $e^{-1}<\frac{1}{2}$. Taking $t=2x$ in this equation then gives (see Figure 10.2)

$$e^{-2x}\leq 1-x\text{ for }0\leq x\leq\frac{1}{2}. \tag{12.33}$$

Theorem 12.55. *Suppose that $\varphi:(a,b)\to\mathbb{R}$ is convex and for $x,y\in(a,b)$ with $x<y$, let*[10]

[10] The same formula would define $F(x,y)$ for $x\neq y$. However, since $F(x,y)=F(y,x)$, we would gain no new information by this extension.

$$F(x,y) := \frac{\varphi(y) - \varphi(x)}{y - x}.$$

Then;

1. *$F(x,y)$ is increasing in each of its arguments.*
2. *The following limits exist,*

$$\varphi'_+(x) := F(x, x+) := \lim_{y \downarrow x} F(x,y) < \infty \text{ and} \tag{12.34}$$

$$\varphi'_-(y) := F(y-, y) := \lim_{x \uparrow y} F(x,y) > -\infty. \tag{12.35}$$

3. *The functions, $\varphi'_\pm$ are both increasing functions and further satisfy,*

$$-\infty < \varphi'_-(x) \le \varphi'_+(x) \le \varphi'_-(y) < \infty \ \forall \ a < x < y < b. \tag{12.36}$$

4. *For any $t \in \left[\varphi'_-(x), \varphi'_+(x)\right]$,*

$$\varphi(y) \ge \varphi(x) + t(y - x) \text{ for all } x, y \in (a,b). \tag{12.37}$$

5. *For $a < \alpha < \beta < b$, let $K := \max\left\{\left|\varphi'_+(\alpha)\right|, \left|\varphi'_-(\beta)\right|\right\}$. Then*

$$|\varphi(y) - \varphi(x)| \le K|y - x| \text{ for all } x, y \in [\alpha, \beta].$$

   That is φ is Lipschitz continuous on $[\alpha, \beta]$.
6. *The function φ'_+ is right continuous and φ'_- is left continuous.*
7. *The set of discontinuity points for φ'_+ and for φ'_- are the same as the set of points of non-differentiability of φ. Moreover this set is at most countable.*

Proof. BRUCE: The first two items are a repetition of Lemma 12.50.

1. and 2. If we let $h_t = t\varphi(x_1) + (1-t)\varphi(x_0)$, then (x_t, h_t) is on the line segment joining $(x_0, \varphi(x_0))$ to $(x_1, \varphi(x_1))$ and the statement that φ is convex is then equivalent of $\varphi(x_t) \le h_t$ for all $0 \le t \le 1$. Since

$$\frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t},$$

the convexity of φ is equivalent to

$$\frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} \le \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} \text{ for all } x_0 \le x_t \le x_1$$

and to

$$\frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \le \frac{\varphi(x_1) - \varphi(x_t)}{x_1 - x_t} \text{ for all } x_0 \le x_t \le x_1.$$

Convexity also implies

$$\frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} = \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \le \frac{\varphi(x_1) - \varphi(x_t)}{x_1 - x_t}.$$

These inequalities may be written more compactly as,

$$\frac{\varphi(v) - \varphi(u)}{v - u} \le \frac{\varphi(w) - \varphi(u)}{w - u} \le \frac{\varphi(w) - \varphi(v)}{w - v}, \tag{12.38}$$

valid for all $a < u < v < w < b$, again see Figure 12.4. The first (second) inequality in Eq. (12.38) shows $F(x,y)$ is increasing y (x). This then implies the limits in item 2. are monotone and hence exist as claimed.

3. Let $a < x < y < b$. Using the increasing nature of F,

$$-\infty < \varphi'_-(x) = F(x-, x) \le F(x, x+) = \varphi'_+(x) < \infty$$

and

$$\varphi'_+(x) = F(x, x+) \le F(y-, y) = \varphi'_-(y)$$

as desired.

4. Let $t \in \left[\varphi'_-(x), \varphi'_+(x)\right]$. Then

$$t \le \varphi'_+(x) = F(x, x+) \le F(x,y) = \frac{\varphi(y) - \varphi(x)}{y - x}$$

or equivalently,

$$\varphi(y) \ge \varphi(x) + t(y - x) \text{ for } y \ge x.$$

Therefore Eq. (12.37) holds for $y \ge x$. Similarly, for $y < x$,

$$t \ge \varphi'_-(x) = F(x-, x) \ge F(y, x) = \frac{\varphi(x) - \varphi(y)}{x - y}$$

or equivalently,

$$\varphi(y) \ge \varphi(x) - t(x - y) = \varphi(x) + t(y - x) \text{ for } y \le x.$$

Hence we have proved Eq. (12.37) for all $x, y \in (a,b)$.

5. For $a < \alpha \le x < y \le \beta < b$, we have

$$\varphi'_+(\alpha) \le \varphi'_+(x) = F(x, x+) \le F(x,y) \le F(y-, y) = \varphi'_-(y) \le \varphi'_-(\beta) \tag{12.39}$$

and in particular,

$$-K \le \varphi'_+(\alpha) \le \frac{\varphi(y) - \varphi(x)}{y - x} \le \varphi'_-(\beta) \le K.$$

This last inequality implies, $|\varphi(y) - \varphi(x)| \le K(y - x)$ which is the desired Lipschitz bound.

6. For $a < c < x < y < b$, we have $\varphi'_+(x) = F(x, x+) \le F(x, y)$ and letting $x \downarrow c$ (using the continuity of F) we learn $\varphi'_+(c+) \le F(c, y)$. We may now let $y \downarrow c$ to conclude $\varphi'_+(c+) \le \varphi'_+(c)$. Since $\varphi'_+(c) \le \varphi'_+(c+)$, it follows that $\varphi'_+(c) = \varphi'_+(c+)$ and hence that φ'_+ is right continuous.

Similarly, for $a < x < y < c < b$, we have $\varphi'_-(y) \ge F(x, y)$ and letting $y \uparrow c$ (using the continuity of F) we learn $\varphi'_-(c-) \ge F(x, c)$. Now let $x \uparrow c$ to conclude $\varphi'_-(c-) \ge \varphi'_-(c)$. Since $\varphi'_-(c) \ge \varphi'_-(c-)$, it follows that $\varphi'_-(c) = \varphi'_-(c-)$, i.e. φ'_- is left continuous.

7. Since $\varphi_\pm$ are increasing functions, they have at most countably many points of discontinuity. Letting $x \uparrow y$ in Eq. (12.36), using the left continuity of φ'_-, shows $\varphi'_-(y) = \varphi'_+(y-)$. Hence if φ'_- is continuous at y, $\varphi'_-(y) = \varphi'_-(y+) = \varphi'_+(y)$ and φ is differentiable at y. Conversely if φ is differentiable at y, then

$$\varphi'_+(y-) = \varphi'_-(y) = \varphi'(y) = \varphi'_+(y)$$

which shows φ'_+ is continuous at y. Thus we have shown that set of discontinuity points of φ'_+ is the same as the set of points of non-differentiability of φ. That the discontinuity set of φ'_- is the same as the non-differentiability set of φ is proved similarly. ∎

Corollary 12.56. *If $\varphi : (a, b) \to \mathbb{R}$ is a convex function and $D \subset (a, b)$ is a dense set, then*

$$\varphi(y) = \sup_{x \in D} \left[\varphi(x) + \varphi'_\pm(x)(y - x)\right] \text{ for all } x, y \in (a, b).$$

Proof. Let $\psi_\pm(y) := \sup_{x \in D} [\varphi(x) + \varphi_\pm(x)(y - x)]$. According to Eq. (12.37) above, we know that $\varphi(y) \ge \psi_\pm(y)$ for all $y \in (a, b)$. Now suppose that $x \in (a, b)$ and $x_n \in \Lambda$ with $x_n \uparrow x$. Then passing to the limit in the estimate, $\psi_-(y) \ge \varphi(x_n) + \varphi'_-(x_n)(y - x_n)$, shows $\psi_-(y) \ge \varphi(x) + \varphi'_-(x)(y - x)$. Since $x \in (a, b)$ is arbitrary we may take $x = y$ to discover $\psi_-(y) \ge \varphi(y)$ and hence $\varphi(y) = \psi_-(y)$. The proof that $\varphi(y) = \psi_+(y)$ is similar. ∎

Lemma 12.57. *Suppose that $\varphi : (a, b) \to \mathbb{R}$ is a non-decreasing function such that*

$$\varphi\left(\frac{1}{2}(x + y)\right) \le \frac{1}{2}[\varphi(x) + \varphi(y)] \text{ for all } x, y \in (a, b), \tag{12.40}$$

then φ is convex. The result remains true if φ is assumed to be continuous rather than non-decreasing.

Proof. Let $x_0, x_1 \in (a, b)$ and $x_t := x_0 + t(x_1 - x_0)$ as above. For $n \in \mathbb{N}$ let $\mathbb{D}_n = \left\{\frac{k}{2^n} : 1 \le k < 2^n\right\}$. We are going to being by showing Eq. (12.40) implies

$$\varphi(x_t) \le (1 - t)\varphi(x_0) + t\varphi(x_1) \text{ for all } t \in \mathbb{D} := \cup_n \mathbb{D}_n. \tag{12.41}$$

We will do this by induction on n. For $n = 1$, this follows directly from Eq. (12.40). So now suppose that Eq. (12.41) holds for all $t \in \mathbb{D}_n$ and now suppose that $t = \frac{2k+1}{2^n} \in \mathbb{D}_{n+1}$. Observing that

$$x_t = \frac{1}{2}\left(x_{\frac{k}{2^{n-1}}} + x_{\frac{k+1}{2^n}}\right)$$

we may again use Eq. (12.40) to conclude,

$$\varphi(x_t) \le \frac{1}{2}\left(\varphi\left(x_{\frac{k}{2^{n-1}}}\right) + \varphi\left(x_{\frac{k+1}{2^{n-1}}}\right)\right).$$

Then use the induction hypothesis to conclude,

$$\begin{aligned}\varphi(x_t) &\le \frac{1}{2}\begin{pmatrix} \left(1 - \frac{k}{2^{n-1}}\right)\varphi(x_0) + \frac{k}{2^{n-1}}\varphi(x_1) \\ + \left(1 - \frac{k+1}{2^{n-1}}\right)\varphi(x_0) + \frac{k+1}{2^{n-1}}\varphi(x_1)\end{pmatrix} \\ &= (1 - t)\varphi(x_0) + t\varphi(x_1)\end{aligned}$$

as desired.

For general $t \in (0, 1)$, let $\tau \in \mathbb{D}$ such that $\tau > t$. Since φ is increasing and by Eq. (12.41) we conclude,

$$\varphi(x_t) \le \varphi(x_\tau) \le (1 - \tau)\varphi(x_0) + \tau\varphi(x_1).$$

We may now let $\tau \downarrow t$ to complete the proof. This same technique clearly also works if we were to assume that φ is continuous rather than monotonic. ∎

13

Hilbert Space Basics

Definition 13.1. *Let H be a complex vector space. An inner product on H is a function, $\langle\cdot|\cdot\rangle : H\times H \to \mathbb{C}$, such that*

1. $\langle ax+by|z\rangle = a\langle x|z\rangle + b\langle y|z\rangle$ *i.e.* $x\to\langle x|z\rangle$ *is linear.*
2. $\overline{\langle x|y\rangle} = \langle y|x\rangle$.
3. $\|x\|^2 := \langle x|x\rangle \geq 0$ *with* $\|x\|^2 = 0$ *iff* $x=0$.

Notice that combining properties (1) and (2) that $x\to\langle z|x\rangle$ is conjugate linear for fixed $z\in H$, i.e.

$$\langle z|ax+by\rangle = \bar{a}\langle z|x\rangle + \bar{b}\langle z|y\rangle.$$

The following identity will be used frequently in the sequel without further mention,

$$\begin{aligned}\|x+y\|^2 &= \langle x+y|x+y\rangle = \|x\|^2+\|y\|^2+\langle x|y\rangle+\langle y|x\rangle\\ &= \|x\|^2+\|y\|^2+2\mathrm{Re}\langle x|y\rangle. \end{aligned}\tag{13.1}$$

Theorem 13.2 (Schwarz Inequality). *Let $(H,\langle\cdot|\cdot\rangle)$ be an inner product space, then for all $x,y\in H$*

$$|\langle x|y\rangle| \leq \|x\|\|y\|$$

and equality holds iff x and y are linearly dependent.

Proof. If $y=0$, the result holds trivially. So assume that $y\neq 0$ and observe; if $x=\alpha y$ for some $\alpha\in\mathbb{C}$, then $\langle x|y\rangle = \bar{\alpha}\,\|y\|^2$ and hence

$$|\langle x|y\rangle| = |\alpha|\,\|y\|^2 = \|x\|\|y\|.$$

Now suppose that $x\in H$ is arbitrary, let $z := x-\|y\|^{-2}\langle x|y\rangle y$. (So $\|y\|^{-2}\langle x|y\rangle y$ is the "orthogonal projection" of x along y, see Figure 13.1.) Then

$$\begin{aligned}0\leq\|z\|^2 &= \left\|x-\frac{\langle x|y\rangle}{\|y\|^2}y\right\|^2 = \|x\|^2+\frac{|\langle x|y\rangle|^2}{\|y\|^4}\|y\|^2 - 2\mathrm{Re}\langle x|\frac{\langle x|y\rangle}{\|y\|^2}y\rangle\\ &= \|x\|^2-\frac{|\langle x|y\rangle|^2}{\|y\|^2}\end{aligned}$$

from which it follows that $0\leq\|y\|^2\|x\|^2-|\langle x|y\rangle|^2$ with equality iff $z=0$ or equivalently iff $x=\|y\|^{-2}\langle x|y\rangle y$. ∎

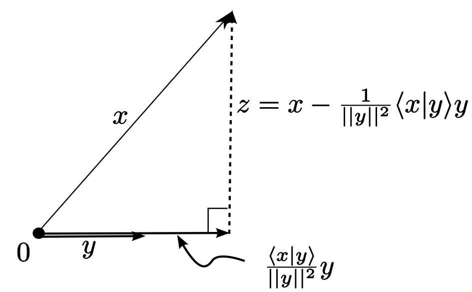


Fig. 13.1. The picture behind the proof of the Schwarz inequality.

Corollary 13.3. *Let $(H,\langle\cdot|\cdot\rangle)$ be an inner product space and $\|x\| := \sqrt{\langle x|x\rangle}$. Then the* ***Hilbertian norm***, $\|\cdot\|$, *is a norm on H. Moreover $\langle\cdot|\cdot\rangle$ is continuous on $H\times H$, where H is viewed as the normed space $(H,\|\cdot\|)$.*

Proof. If $x,y\in H$, then, using Schwarz's inequality,

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2+\|y\|^2+2\mathrm{Re}\langle x|y\rangle\\ &\leq \|x\|^2+\|y\|^2+2\|x\|\|y\| = (\|x\|+\|y\|)^2.\end{aligned}$$

Taking the square root of this inequality shows $\|\cdot\|$ satisfies the triangle inequality.

Checking that $\|\cdot\|$ satisfies the remaining axioms of a norm is now routine and will be left to the reader. If $x,y,\Delta x,\Delta y\in H$, then

$$\begin{aligned}|\langle x+\Delta x|y+\Delta y\rangle - \langle x|y\rangle| &= |\langle x|\Delta y\rangle + \langle\Delta x|y\rangle + \langle\Delta x|\Delta y\rangle|\\ &\leq \|x\|\|\Delta y\| + \|y\|\|\Delta x\| + \|\Delta x\|\|\Delta y\|\\ &\to 0 \text{ as } \Delta x,\Delta y\to 0,\end{aligned}$$

from which it follows that $\langle\cdot|\cdot\rangle$ is continuous. ∎

Definition 13.4. *Let $(H,\langle\cdot|\cdot\rangle)$ be an inner product space, we say $x,y\in H$ are* ***orthogonal*** *and write $x\perp y$ iff $\langle x|y\rangle = 0$. More generally if $A\subset H$ is a set, $x\in H$ is* ***orthogonal to*** *A (write $x\perp A$) iff $\langle x|y\rangle = 0$ for all $y\in A$. Let $A^\perp = \{x\in H : x\perp A\}$ be the set of vectors orthogonal to A. A subset $S\subset H$*

is an ***orthogonal set*** *if $x \perp y$ for all distinct elements $x, y \in S$. If S further satisfies, $\|x\| = 1$ for all $x \in S$, then S is said to be an* ***orthonormal set.***

Proposition 13.5. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space then*

1. *(**Parallelogram Law**)*
$$\|a+b\|^2 + \|a-b\|^2 = 2\|a\|^2 + 2\|b\|^2 \tag{13.2}$$
for all $a, b \in H$.
2. *(**Pythagorean Theorem**) If $S \subset\subset H$ is a finite orthogonal set, then*
$$\left\|\sum_{x\in S} x\right\|^2 = \sum_{x\in S} \|x\|^2. \tag{13.3}$$
3. *If $A \subset H$ is a set, then $A^\perp$ is a* ***closed*** *linear subspace of H.*

Proof. I will assume that H is a complex Hilbert space, the real case being easier. Items 1. and 2. are proved by the following elementary computations;

$$\begin{aligned}\|a+b\|^2 &+ \|a-b\|^2 \\ &= \|a\|^2 + \|b\|^2 + 2\mathrm{Re}\langle a|b\rangle + \|a\|^2 + \|b\|^2 - 2\mathrm{Re}\langle a|b\rangle \\ &= 2\|a\|^2 + 2\|b\|^2,\end{aligned}$$

and

$$\begin{aligned}\left\|\sum_{x\in S} x\right\|^2 &= \langle \sum_{x\in S} x | \sum_{y\in S} y\rangle = \sum_{x,y\in S} \langle x|y\rangle \\ &= \sum_{x\in S} \langle x|x\rangle = \sum_{x\in S} \|x\|^2.\end{aligned}$$

Item 3. is a consequence of the continuity of $\langle\cdot|\cdot\rangle$ and the fact that

$$A^\perp = \cap_{x\in A} \mathrm{Nul}(\langle\cdot|x\rangle)$$

where $\mathrm{Nul}(\langle\cdot|x\rangle) = \{y \in H : \langle y|x\rangle = 0\}$ – a closed subspace of H. Alternatively, if $x_n \in A^\perp$ and $x_n \to x$ in H, then

$$0 = \lim_{n\to\infty} 0 = \lim_{n\to\infty} \langle x_n|a\rangle = \left\langle \lim_{n\to\infty} x_n|a\right\rangle = \langle x|a\rangle \ \forall\, a \in A$$

which shows that $x \in A^\perp$. ∎

Definition 13.6. *A* ***Hilbert space*** *is an inner product space $(H, \langle\cdot|\cdot\rangle)$ such that the induced Hilbertian norm is complete.*

Example 13.7. For any measure space, $(\Omega, \mathcal{B}, \mu)$, $H := L^2(\mu)$ with inner product,

$$\langle f|g\rangle = \int_\Omega f(\omega)\,\bar{g}(\omega)\,d\mu(\omega)$$

is a Hilbert space – see Theorem 12.25 for the completeness assertion.

Definition 13.8. *A subset C of a vector space X is said to be convex if for all $x, y \in C$ the line segment $[x, y] := \{tx + (1-t)y : 0 \le t \le 1\}$ joining x to y is contained in C as well. (Notice that any vector subspace of X is convex.)*

Theorem 13.9 (Best Approximation Theorem). *Suppose that H is a Hilbert space and $M \subset H$ is a closed convex subset of H. Then for any $x \in H$ there exists a unique $y \in M$ such that*

$$\|x - y\| = d(x, M) = \inf_{z\in M} \|x - z\|.$$

Moreover, if M is a vector subspace of H, then the point y may also be characterized as the unique point in M such that $(x - y) \perp M$.

Proof. Let $x \in H$, $\delta := d(x, M)$, $y, z \in M$, and, referring to Figure 13.2, let $w = z + (y - x)$ and $c = (z + y)/2 \in M$. It then follows by the parallelogram law (Eq. (13.2) with $a = (y - x)$ and $b = (z - x)$) and the fact that $c \in M$ that

$$\begin{aligned}2\|y-x\|^2 + 2\|z-x\|^2 &= \|w-x\|^2 + \|y-z\|^2 \\ &= \|z+y-2x\|^2 + \|y-z\|^2 \\ &= 4\|x-c\|^2 + \|y-z\|^2 \\ &\ge 4\delta^2 + \|y-z\|^2.\end{aligned}$$

Thus we have shown for all $y, z \in M$ that,

$$\|y-z\|^2 \le 2\|y-x\|^2 + 2\|z-x\|^2 - 4\delta^2. \tag{13.4}$$

Uniqueness. If $y, z \in M$ minimize the distance to x, then $\|y - x\| = \delta = \|z - x\|$ and it follows from Eq. (13.4) that $y = z$.

Existence. Let $y_n \in M$ be chosen such that $\|y_n - x\| = \delta_n \to \delta = d(x, M)$. Taking $y = y_m$ and $z = y_n$ in Eq. (13.4) shows

$$\|y_n - y_m\|^2 \le 2\delta_m^2 + 2\delta_n^2 - 4\delta^2 \to 0 \text{ as } m, n \to \infty.$$

Therefore, by completeness of H, $\{y_n\}_{n=1}^\infty$ is convergent. Because M is closed, $y := \lim_{n\to\infty} y_n \in M$ and because the norm is continuous,

$$\|y - x\| = \lim_{n\to\infty} \|y_n - x\| = \delta = d(x, M).$$

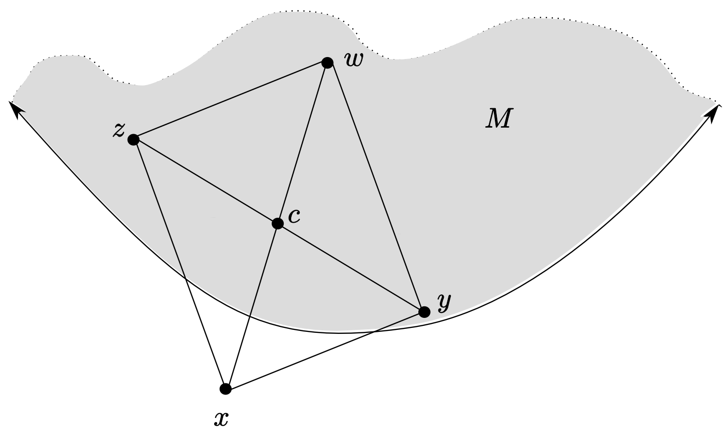


Fig. 13.2. In this figure $y, z \in M$ and by convexity, $c = (z + y)/2 \in M$.

So y is the desired point in M which is closest to x.

Orthogonality property. Now suppose M is a closed subspace of H and $x \in H$. Let $y \in M$ be the closest point in M to x. Then for $w \in M$, the function

$$g(t) := \|x - (y + tw)\|^2 = \|x - y\|^2 - 2t\mathrm{Re}\langle x - y|w\rangle + t^2\|w\|^2$$

has a minimum at $t = 0$ and therefore $0 = g'(0) = -2\mathrm{Re}\langle x - y|w\rangle$. Since $w \in M$ is arbitrary, this implies that $(x - y) \perp M$, see Figure 13.3.

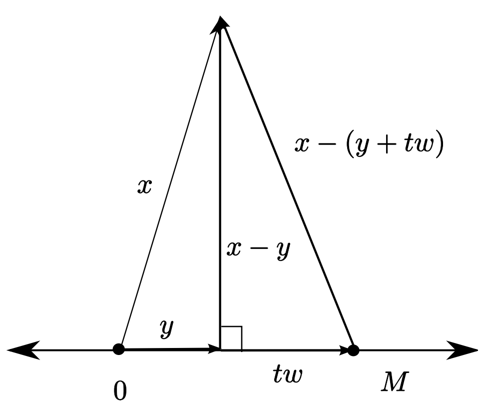


Fig. 13.3. The orthogonality relationships of closest points.

Finally suppose $y \in M$ is any point such that $(x - y) \perp M$. Then for $z \in M$, by Pythagorean's theorem,

$$\|x - z\|^2 = \|x - y + y - z\|^2 = \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2$$

which shows $d(x, M)^2 \geq \|x - y\|^2$. That is to say y is the point in M closest to x. ■

Notation 13.10 *If $A : X \to Y$ is a linear operator between two normed spaces, we let*

$$\|A\| := \sup_{x \in X\setminus\{0\}} \frac{\|Ax\|_Y}{\|x\|_X} = \sup_{\|x\|_X = 1} \|Ax\|_Y .$$

We refer to $\|A\|$ as the operator norm of A and call A a bounded operator if $\|A\| < \infty$. We further let $L(X, Y)$ be the set of bounded operators from X to Y.

Exercise 13.1. Show that a linear operator, $A : X \to Y$, is a bounded iff it is continuous.

Solution to Exercise (13.1). If A is continuous at $x = 0$, then (as $A0 = 0$) there exists $\delta > 0$ such that $\|Ax\|_Y \leq 1$ for $\|x\|_X \leq \delta$. Thus if $x \neq 0$, we have $\left\|\frac{\delta}{\|x\|_X} x\right\|_X = \delta$ and therefore,

$$\frac{\delta}{\|x\|_X} \|Ax\|_Y = \left\|A\frac{\delta}{\|x\|_X} x\right\|_Y \leq 1$$

from which it follows that $\|Ax\|_Y \leq \delta^{-1} \|x\|_X$ which shows that $\|A\| \leq \delta^{-1} < \infty$. Conversely if $\|A\| < \infty$, then

$$\|Ax - Ax'\|_Y = \|A(x - x')\|_Y \leq \|A\| \|x - x'\|_X$$

from which it follows that A is continuous.

Definition 13.11. *Suppose that $A : H \to H$ is a bounded operator. The **adjoint** of A, denoted A^*, is the unique operator $A^* : H \to H$ such that $\langle Ax|y\rangle = \langle x|A^*y\rangle$. (The proof that A^* exists and is unique will be given in Proposition 13.16 below.) A bounded operator $A : H \to H$ is **self - adjoint** or **Hermitian** if $A = A^*$.*

Definition 13.12. *H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection of H onto M is the function $P_M : H \to H$ such that for $x \in H$, $P_M(x)$ is the unique element in M such that $(x - P_M(x)) \perp M$, i.e. $P_M(x)$ is the unique element in M such that*

$$\langle x|m\rangle = \langle P_M(x)|m\rangle \text{ for all } m \in M. \tag{13.5}$$

Given a linear transformation A, we will let $\mathrm{Ran}\,(A)$ and $\mathrm{Nul}\,(A)$ denote the **range** and the **null-space** of A respectively.

Theorem 13.13 (Projection Theorem). *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection P_M satisfies:*

1. *P_M is linear and hence we will write $P_M x$ rather than $P_M(x)$.*
2. *$P_M^2 = P_M$ (P_M is a projection).*
3. *$P_M^* = P_M$ (P_M is self-adjoint).*
4. *$\mathrm{Ran}(P_M) = M$ and $\mathrm{Nul}(P_M) = M^\perp$.*
5. *If $N \subset M \subset H$ is another closed subspace, the $P_N P_M = P_M P_N = P_N$.*

Proof.

1. Let $x_1, x_2 \in H$ and $\alpha \in \mathbb{C}$, then $P_M x_1 + \alpha P_M x_2 \in M$ and
$$P_M x_1 + \alpha P_M x_2 - (x_1 + \alpha x_2) = [P_M x_1 - x_1 + \alpha(P_M x_2 - x_2)] \in M^\perp$$
showing $P_M x_1 + \alpha P_M x_2 = P_M(x_1 + \alpha x_2)$, i.e. P_M is linear.
2. Obviously $\mathrm{Ran}(P_M) = M$ and $P_M x = x$ for all $x \in M$. Therefore $P_M^2 = P_M$.
3. Let $x, y \in H$, then since $(x - P_M x)$ and $(y - P_M y)$ are in $M^\perp$,
$$\begin{aligned}\langle P_M x | y\rangle &= \langle P_M x | P_M y + y - P_M y\rangle = \langle P_M x | P_M y\rangle \\ &= \langle P_M x + (x - P_M x) | P_M y\rangle = \langle x | P_M y\rangle.\end{aligned}$$
4. We have already seen, $\mathrm{Ran}(P_M) = M$ and $P_M x = 0$ iff $x = x - 0 \in M^\perp$, i.e. $\mathrm{Nul}(P_M) = M^\perp$.
5. If $N \subset M \subset H$ it is clear that $P_M P_N = P_N$ since $P_M = Id$ on $N = \mathrm{Ran}(P_N) \subset M$. Taking adjoints gives the other identity, namely that $P_N P_M = P_N$.
Alternative proof 1 of $P_N P_M = P_N$. If $x \in H$, then $(x - P_M x) \perp M$ and therefore $(x - P_M x) \perp N$. We also have $(P_M x - P_N P_M x) \perp N$ and therefore,
$$x - P_N P_M x = (x - P_M x) + (P_M x - P_N P_M x) \in N^\perp$$
which shows $P_N P_M x = P_N x$.
Alternative proof 2 of $P_N P_M = P_N$. If $x \in H$ and $n \in N$, we have
$$\langle P_N P_M x | n\rangle = \langle P_M x | P_N n\rangle = \langle P_M x | n\rangle = \langle x | P_M n\rangle = \langle x | n\rangle\,.$$
Since this holds for all n we may conclude that $P_N P_M x = P_N x$.

■

Corollary 13.14. *If $M \subset H$ is a proper closed subspace of a Hilbert space H, then $H = M \oplus M^\perp$.*

Proof. Given $x \in H$, let $y = P_M x$ so that $x - y \in M^\perp$. Then $x = y + (x - y) \in M + M^\perp$. If $x \in M \cap M^\perp$, then $x \perp x$, i.e. $\|x\|^2 = \langle x | x\rangle = 0$. So $M \cap M^\perp = \{0\}$. ■

Exercise 13.2. Suppose M is a subset of H, then $M^{\perp\perp} = \overline{\mathrm{span}(M)}$ where (as usual), $\mathrm{span}\,(M)$ denotes all finite linear combinations of elements from M.

Theorem 13.15 (Riesz Theorem). *Let H^* be the dual space of H, i.e. $f \in H^*$ iff $f : H \to \mathbb{F}$ is linear and continuous. The map*
$$z \in H \xrightarrow{j} \langle \cdot | z\rangle \in H^* \tag{13.6}$$
is a conjugate linear[1] isometric isomorphism, where for $f \in H^*$ we let,
$$\|f\|_{H^*} := \sup_{x \in H \setminus \{0\}} \frac{|f(x)|}{\|x\|} = \sup_{\|x\|=1} |f(x)|\,.$$

Proof. Let $f \in H^*$ and $M = \mathrm{Nul}(f)$ – a closed proper subspace of H since f is continuous. If $f = 0$, then clearly $f(\cdot) = \langle \cdot | 0\rangle$. If $f \neq 0$ there exists $y \in H \setminus M$. Then for any $\alpha \in \mathbb{C}$ we have $e := \alpha(y - P_M y) \in M^\perp$. We now choose α so that $f(e) = 1$. Hence if $x \in H$,
$$f(x - f(x)\,e) = f(x) - f(x)\,f(e) = f(x) - f(x) = 0,$$
which shows $x - f(x)\,e \in M$. As $e \in M^\perp$ it follows that
$$0 = \langle x - f(x)\,e | e\rangle = \langle x | e\rangle - f(x)\,\|e\|^2$$
which shows $f(\cdot) = \langle \cdot | z\rangle = jz$ where $z := e / \|e\|^2$ and thus j is surjective.

The map j is conjugate linear by the axioms of the inner products. Moreover, for $x, z \in H$,
$$|\langle x | z\rangle| \le \|x\|\,\|z\| \text{ for all } x \in H$$
with equality when $x = z$. This implies that $\|jz\|_{H^*} = \|\langle \cdot | z\rangle\|_{H^*} = \|z\|$. Therefore j is isometric and this implies j is injective.

■

Proposition 13.16 (Adjoints). *Let H and K be Hilbert spaces and $A : H \to K$ be a bounded operator. Then there exists a unique bounded operator $A^* : K \to H$ such that*
$$\langle Ax | y\rangle_K = \langle x | A^* y\rangle_H \text{ for all } x \in H \text{ and } y \in K. \tag{13.7}$$
Moreover, for all $A, B \in L(H, K)$ and $\lambda \in \mathbb{C}$,

[1] Recall that j is conjugate linear if
$$j(z_1 + \alpha z_2) = jz_1 + \bar{\alpha} jz_2$$
for all $z_1, z_2 \in H$ and $\alpha \in \mathbb{C}$.

1. $(A+\lambda B)^* = A^* + \bar{\lambda} B^*$,
2. $A^{**} := (A^*)^* = A$,
3. $\|A^*\| = \|A\|$ *and*
4. $\|A^*A\| = \|A\|^2$.
5. *If* $K = H$, *then* $(AB)^* = B^*A^*$. *In particular* $A \in L(H)$ *has a bounded inverse iff* A^* *has a bounded inverse and* $(A^*)^{-1} = (A^{-1})^*$.

Proof. For each $y \in K$, the map $x \to \langle Ax|y\rangle_K$ is in H^* and therefore there exists, by Theorem 13.15, a unique vector $z \in H$ (we will denote this z by $A^*(y)$) such that
$$\langle Ax|y\rangle_K = \langle x|z\rangle_H \text{ for all } x \in H.$$
This shows there is a unique map $A^* : K \to H$ such that $\langle Ax|y\rangle_K = \langle x|A^*(y)\rangle_H$ for all $x \in H$ and $y \in K$.

To see A^* is linear, let $y_1, y_2 \in K$ and $\lambda \in \mathbb{C}$, then for any $x \in H$,
$$\begin{aligned}\langle Ax|y_1 + \lambda y_2\rangle_K &= \langle Ax|y_1\rangle_K + \bar{\lambda}\langle Ax|y_2\rangle_K \\ &= \langle x|A^*(y_1)\rangle_K + \bar{\lambda}\langle x|A^*(y_2)\rangle_H \\ &= \langle x|A^*(y_1) + \lambda A^*(y_2)\rangle_H\end{aligned}$$
and by the uniqueness of $A^*(y_1 + \lambda y_2)$ we find
$$A^*(y_1 + \lambda y_2) = A^*(y_1) + \lambda A^*(y_2).$$
This shows A^* is linear and so we will now write A^*y instead of $A^*(y)$.

Since
$$\langle A^*y|x\rangle_H = \overline{\langle x|A^*y\rangle_H} = \overline{\langle Ax|y\rangle_K} = \langle y|Ax\rangle_K$$
it follows that $A^{**} = A$. The assertion that $(A+\lambda B)^* = A^* + \bar{\lambda}B^*$ is Exercise 13.3.

Items 3. and 4. Making use of Schwarz's inequality (Theorem 13.2), we have
$$\begin{aligned}\|A^*\| &= \sup_{k \in K:\|k\|=1} \|A^*k\| \\ &= \sup_{k \in K:\|k\|=1} \sup_{h \in H:\|h\|=1} |\langle A^*k|h\rangle| \\ &= \sup_{h \in H:\|h\|=1} \sup_{k \in K:\|k\|=1} |\langle k|Ah\rangle| = \sup_{h \in H:\|h\|=1} \|Ah\| = \|A\|\end{aligned}$$
so that $\|A^*\| = \|A\|$. Since
$$\|A^*A\| \le \|A^*\| \, \|A\| = \|A\|^2$$
and
$$\begin{aligned}\|A\|^2 &= \sup_{h \in H:\|h\|=1} \|Ah\|^2 = \sup_{h \in H:\|h\|=1} |\langle Ah|Ah\rangle| \\ &= \sup_{h \in H:\|h\|=1} |\langle h|A^*Ah\rangle| \le \sup_{h \in H:\|h\|=1} \|A^*Ah\| = \|A^*A\| \end{aligned} \tag{13.8}$$
we also have $\|A^*A\| \le \|A\|^2 \le \|A^*A\|$ which shows $\|A\|^2 = \|A^*A\|$.

Alternatively, from Eq. (13.8),
$$\|A\|^2 \le \|A^*A\| \le \|A\| \, \|A^*\| \tag{13.9}$$
which then implies $\|A\| \le \|A^*\|$. Replacing A by A^* in this last inequality shows $\|A^*\| \le \|A\|$ and hence that $\|A^*\| = \|A\|$. Using this identity back in Eq. (13.9) proves $\|A\|^2 = \|A^*A\|$.

Now suppose that $K = H$. Then
$$\langle ABh|k\rangle = \langle Bh|A^*k\rangle = \langle h|B^*A^*k\rangle$$
which shows $(AB)^* = B^*A^*$. If A^{-1} exists then
$$\begin{aligned}(A^{-1})^* A^* &= (AA^{-1})^* = I^* = I \text{ and} \\ A^* (A^{-1})^* &= (A^{-1}A)^* = I^* = I.\end{aligned}$$
This shows that A^* is invertible and $(A^*)^{-1} = (A^{-1})^*$. Similarly if A^* is invertible then so is $A = A^{**}$. ■

Exercise 13.3. Let H, K, M be Hilbert spaces, $A, B \in L(H,K)$, $C \in L(K,M)$ and $\lambda \in \mathbb{C}$. Show $(A+\lambda B)^* = A^* + \bar{\lambda}B^*$ and $(CA)^* = A^*C^* \in L(M,H)$.

Exercise 13.4. Let $H = \mathbb{C}^n$ and $K = \mathbb{C}^m$ equipped with the usual inner products, i.e. $\langle z|w\rangle_H = z \cdot \bar{w}$ for $z, w \in H$. Let A be an $m \times n$ matrix thought of as a linear operator from H to K. Show the matrix associated to $A^* : K \to H$ is the conjugate transpose of A.

Lemma 13.17. *Suppose* $A : H \to K$ *is a bounded operator, then:*

1. $\mathrm{Nul}(A^*) = \mathrm{Ran}(A)^\perp$.
2. $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$.
3. *if* $K = H$ *and* $V \subset H$ *is an* A *– invariant subspace (i.e.* $A(V) \subset V$*), then* $V^\perp$ *is* A^* *– invariant.*

Proof. An element $y \in K$ is in $\mathrm{Nul}(A^*)$ iff $0 = \langle A^*y|x\rangle = \langle y|Ax\rangle$ for all $x \in H$ which happens iff $y \in \mathrm{Ran}(A)^\perp$. Because, by Exercise 13.2, $\overline{\mathrm{Ran}(A)} = \mathrm{Ran}(A)^{\perp\perp}$, and so by the first item, $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$. Now suppose $A(V) \subset V$ and $y \in V^\perp$, then
$$\langle A^*y|x\rangle = \langle y|Ax\rangle = 0 \text{ for all } x \in V$$

which shows $A^*y \in V^\perp$. ■

The next elementary theorem (referred to as the bounded linear transformation theorem, or B.L.T. theorem for short) is often useful.

Theorem 13.18 (B. L. T. Theorem). *Suppose that Z is a normed space, X is a Banach space, and $\mathcal{S} \subset Z$ is a dense linear subspace of Z. If $T : \mathcal{S} \to X$ is a bounded linear transformation (i.e. there exists $C < \infty$ such that $\|Tz\| \le C\|z\|$ for all $z \in \mathcal{S}$), then T has a unique extension to an element $\bar{T} \in L(Z, X)$ and this extension still satisfies*

$$\left\|\bar{T}z\right\| \le C\|z\| \text{ for all } z \in \bar{\mathcal{S}}.$$

Proof. Let $z \in Z$ and choose $z_n \in \mathcal{S}$ such that $z_n \to z$. Since

$$\|Tz_m - Tz_n\| \le C\|z_m - z_n\| \to 0 \text{ as } m, n \to \infty,$$

it follows by the completeness of X that $\lim_{n\to\infty} Tz_n =: \bar{T}z$ exists. Moreover, if $w_n \in \mathcal{S}$ is another sequence converging to z, then

$$\|Tz_n - Tw_n\| \le C\|z_n - w_n\| \to C\|z - z\| = 0$$

and therefore $\bar{T}z$ is well defined. It is now a simple matter to check that $\bar{T} : Z \to X$ is still linear and that

$$\left\|\bar{T}z\right\| = \lim_{n\to\infty}\|Tz_n\| \le \lim_{n\to\infty} C\|z_n\| = C\|z\| \text{ for all } x \in Z.$$

Thus $\bar{T}$ is an extension of T to all of the Z. The uniqueness of this extension is easy to prove and will be left to the reader. ■

13.1 Compactness Results for L^p – Spaces*

In this section we are going to identify the sequentially "weak" compact subsets of $L^p(\Omega, \mathcal{B}, P)$ for $1 \le p < \infty$, where $(\Omega, \mathcal{B}, P)$ is a probability space. The key to our proofs will be the following Hilbert space compactness result.

Theorem 13.19. *Suppose $\{x_n\}_{n=1}^\infty$ is a bounded sequence in a Hilbert space H (i.e. $C := \sup_n \|x_n\| < \infty$), then there exists a sub-sequence, $y_k := x_{n_k}$ and an $x \in H$ such that $\lim_{k\to\infty}\langle y_k|h\rangle = \langle x|h\rangle$ for all $h \in H$. We say that y_k converges to x weakly in this case and denote this by $y_k \xrightarrow{w} x$.*

Proof. Let $H_0 := \overline{\operatorname{span}(x_k : k \in \mathbb{N})}$. Then H_0 is a closed separable Hilbert subspace of H and $\{x_k\}_{k=1}^\infty \subset H_0$. Let $\{h_n\}_{n=1}^\infty$ be a countable dense subset of H_0. Since $|\langle x_k|h_n\rangle| \le \|x_k\|\|h_n\| \le C\|h_n\| < \infty$, the sequence, $\{\langle x_k|h_n\rangle\}_{k=1}^\infty \subset \mathbb{C}$, is bounded and hence has a convergent sub-sequence for all $n \in \mathbb{N}$. By the Cantor's diagonalization argument we can find a a sub-sequence, $y_k := x_{n_k}$, of $\{x_n\}$ such that $\lim_{k\to\infty}\langle y_k|h_n\rangle$ exists for all $n \in \mathbb{N}$.

We now show $\varphi(z) := \lim_{k\to\infty}\langle y_k|z\rangle$ exists for all $z \in H_0$. Indeed, for any $k, l, n \in \mathbb{N}$, we have

$$\begin{aligned}|\langle y_k|z\rangle - \langle y_l|z\rangle| &= |\langle y_k - y_l|z\rangle| \le |\langle y_k - y_l|h_n\rangle| + |\langle y_k - y_l|z - h_n\rangle| \\ &\le |\langle y_k - y_l|h_n\rangle| + 2C\|z - h_n\|.\end{aligned}$$

Letting $k, l \to \infty$ in this estimate then shows

$$\limsup_{k,l\to\infty}|\langle y_k|z\rangle - \langle y_l|z\rangle| \le 2C\|z - h_n\|.$$

Since we may choose $n \in \mathbb{N}$ such that $\|z - h_n\|$ is as small as we please, we may conclude that $\limsup_{k,l\to\infty}|\langle y_k|z\rangle - \langle y_l|z\rangle|$, i.e. $\varphi(z) := \lim_{k\to\infty}\langle y_k|z\rangle$ exists.

The function, $\bar{\varphi}(z) = \lim_{k\to\infty}\langle z|y_k\rangle$ is a bounded linear functional on H because

$$|\bar{\varphi}(z)| = \liminf_{k\to\infty}|\langle z|y_k\rangle| \le C\|z\|.$$

Therefore by the Riesz Theorem 13.15, there exists $x \in H_0$ such that $\bar{\varphi}(z) = \langle z|x\rangle$ for all $z \in H_0$. Thus, for this $x \in H_0$ we have shown

$$\lim_{k\to\infty}\langle y_k|z\rangle = \langle x|z\rangle \text{ for all } z \in H_0. \tag{13.10}$$

To finish the proof we need only observe that Eq. (13.10) is valid for all $z \in H$. Indeed if $z \in H$, then $z = z_0 + z_1$ where $z_0 = P_{H_0}z \in H_0$ and $z_1 = z - P_{H_0}z \in H_0^\perp$. Since $y_k, x \in H_0$, we have

$$\lim_{k\to\infty}\langle y_k|z\rangle = \lim_{k\to\infty}\langle y_k|z_0\rangle = \langle x|z_0\rangle = \langle x|z\rangle \text{ for all } z \in H.$$

■

Since unbounded subsets of H are clearly not sequentially weakly compact, Theorem 13.19 states that a set is sequentially precompact in H iff it is bounded. Let us now use Theorem 13.19 to identify the sequentially compact subsets of $L^p(\Omega, \mathcal{B}, P)$ for all $1 \le p < \infty$. We begin with the case $p = 1$.

Theorem 13.20. *If $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, \mathcal{B}, P)$, there exists a subsequence $Y_k := X_{n_k}$ of $\{X_n\}_{n=1}^\infty$ and $X \in L^1(\Omega, \mathcal{B}, P)$ such that*

$$\lim_{k\to\infty}\mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h \in \mathcal{B}_b. \tag{13.11}$$

Proof. For each $m \in \mathbb{N}$ let $X_n^m := X_n 1_{|X_n| \le m}$. The truncated sequence $\{X_n^m\}_{n=1}^\infty$ is a bounded subset of the Hilbert space, $L^2(\Omega, \mathcal{B}, P)$, for all $m \in \mathbb{N}$. Therefore by Theorem 13.19, $\{X_n^m\}_{n=1}^\infty$ has a weakly convergent sub-sequence for all $m \in \mathbb{N}$. By Cantor's diagonalization argument, we can find $Y_k^m := X_{n_k}^m$ and $X^m \in L^2(\Omega, \mathcal{B}, P)$ such that $Y_k^m \xrightarrow{w} X^m$ as $m \to \infty$ and in particular

$$\lim_{k\to\infty} \mathbb{E}[Y_k^m h] = \mathbb{E}[X^m h] \text{ for all } h \in \mathcal{B}_b.$$

Our next goal is to show $X^m \to X$ in $L^1(\Omega, \mathcal{B}, P)$. To this end, for $m < M$ and $h \in \mathcal{B}_b$ we have

$$\begin{aligned}\left|\mathbb{E}\left[\left(X^M - X^m\right)h\right]\right| &= \lim_{k\to\infty}\left|\mathbb{E}\left[\left(Y_k^M - Y_k^m\right)h\right]\right| \le \liminf_{k\to\infty}\mathbb{E}\left[\left|Y_k^M - Y_k^m\right||h|\right] \\ &\le \|h\|_\infty \cdot \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : M \ge |Y_k| > m\right] \\ &\le \|h\|_\infty \cdot \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : |Y_k| > m\right].\end{aligned}$$

Taking $h = \overline{\operatorname{sgn}(X^M - X^m)}$ in this inequality shows

$$\mathbb{E}\left[\left|X^M - X^m\right|\right] \le \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : |Y_k| > m\right]$$

with the right member of this inequality going to zero as $m, M \to \infty$ with $M \ge m$ by the assumed uniform integrability of the $\{X_n\}$. Therefore there exists $X \in L^1(\Omega, \mathcal{B}, P)$ such that $\lim_{m\to\infty}\mathbb{E}|X - X^m| = 0$.

We are now ready to verify Eq. (13.11) is valid. For $h \in \mathcal{B}_b$,

$$\begin{aligned}|\mathbb{E}[(X - Y_k)h]| &\le |\mathbb{E}[(X^m - Y_k^m)h]| + |\mathbb{E}[(X - X^m)h]| + |\mathbb{E}[(Y_k - Y_k^m)h]| \\ &\le |\mathbb{E}[(X^m - Y_k^m)h]| + \|h\|_\infty \cdot (\mathbb{E}[|X - X^m|] + \mathbb{E}[|Y_k| : |Y_k| > m]) \\ &\le |\mathbb{E}[(X^m - Y_k^m)h]| + \|h\|_\infty \cdot \left(\mathbb{E}[|X - X^m|] + \sup_l \mathbb{E}[|Y_l| : |Y_l| > m]\right).\end{aligned}$$

Passing to the limit as $k \to \infty$ in the above inequality shows

$$\limsup_{k\to\infty}|\mathbb{E}[(X - Y_k)h]| \le \|h\|_\infty \cdot \left(\mathbb{E}[|X - X^m|] + \sup_l \mathbb{E}[|Y_l| : |Y_l| > m]\right).$$

Since $X^m \to X$ in L^1 and $\sup_l \mathbb{E}[|Y_l| : |Y_l| > m] \to 0$ by uniform integrability, it follows that, $\limsup_{k\to\infty}|\mathbb{E}[(X - Y_k)h]| = 0$. ■

Example 13.21. Let $(\Omega, \mathcal{B}, P) = \left((0,1), \mathcal{B}_{(0,1)}, m\right)$ where m is Lebesgue measure and let $X_n(\omega) = 2^n 1_{0<\omega<2^{-n}}$. Then $\mathbb{E}X_n = 1$ for all n and hence $\{X_n\}_{n=1}^\infty$ is bounded in $L^1(\Omega, \mathcal{B}, P)$ (but is not uniformly integrable). Suppose for sake of contradiction that there existed $X \in L^1(\Omega, \mathcal{B}, P)$ and subsequence, $Y_k := X_{n_k}$ such that $Y_k \xrightarrow{w} X$. Then for $h \in \mathcal{B}_b$ and any $\varepsilon > 0$ we would have

$$\mathbb{E}\left[Xh1_{(\varepsilon,1)}\right] = \lim_{k\to\infty}\mathbb{E}\left[Y_k h 1_{(\varepsilon,1)}\right] = 0.$$

Then by DCT it would follow that $\mathbb{E}[Xh] = 0$ for all $h \in \mathcal{B}_b$ and hence that $X \equiv 0$. On the other hand we would also have

$$0 = \mathbb{E}[X \cdot 1] = \lim_{k\to\infty}\mathbb{E}[Y_k \cdot 1] = 1$$

and we have reached the desired contradiction. Hence we must conclude that bounded subset of $L^1(\Omega, \mathcal{B}, P)$ need not be weakly compact and thus we can not drop the uniform integrability assumption made in Theorem 13.20.

When $1 < p < \infty$, the situation is simpler.

Theorem 13.22. *Let $p \in (1,\infty)$ and $q = p(p-1)^{-1} \in (1,\infty)$ be its conjugate exponent. If $\{X_n\}_{n=1}^\infty$ is a bounded sequence in $L^p(\Omega, \mathcal{B}, P)$, there exists $X \in L^p(\Omega, \mathcal{B}, P)$ and a subsequence $Y_k := X_{n_k}$ of $\{X_n\}_{n=1}^\infty$ such that*

$$\lim_{k\to\infty}\mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h \in L^q(\Omega, \mathcal{B}, P). \tag{13.12}$$

Proof. Let $C := \sup_{n\in\mathbb{N}}\|X_n\|_p < \infty$ and recall that Lemma 12.48 guarantees that $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, \mathcal{B}, P)$. Therefore by Theorem 13.20, there exists $X \in L^1(\Omega, \mathcal{B}, P)$ and a subsequence, $Y_k := X_{n_k}$, such that Eq. (13.11) holds. We will complete the proof by showing; a) $X \in L^p(\Omega, \mathcal{B}, P)$ and b) and Eq. (13.12) is valid.

a) For $h \in \mathcal{B}_b$ we have

$$|\mathbb{E}[Xh]| \le \liminf_{k\to\infty}\mathbb{E}[|Y_k h|] \le \liminf_{k\to\infty}\|Y_k\|_p \cdot \|h\|_q \le C\|h\|_q.$$

For $M < \infty$, taking $h = \overline{\operatorname{sgn}(X)}|X|^{p-1}1_{|X|\le M}$ in the previous inequality shows

$$\begin{aligned}\mathbb{E}\left[|X|^p 1_{|X|\le M}\right] &\le C\left\|\overline{\operatorname{sgn}(X)}|X|^{p-1}1_{|X|\le M}\right\|_q \\ &= C\left(\mathbb{E}\left[|X|^{(p-1)q}1_{|X|\le M}\right]\right)^{1/q} \le C\left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1/q}\end{aligned}$$

from which it follows that

$$\left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1/p} \le \left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1-1/q} \le C.$$

Using the monotone convergence theorem, we may let $M \to \infty$ in this equation to find $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p} \le C < \infty$.

b) Now that we know $X \in L^p(\Omega, \mathcal{B}, P)$, in make sense to consider $\mathbb{E}[(X - Y_k)h]$ for all $h \in L^p(\Omega, \mathcal{B}, P)$. For $M < \infty$, let $h^M := h1_{|h|\le M}$, then

$$\begin{aligned}|\mathbb{E}[(X - Y_k)h]| &\le \left|\mathbb{E}\left[(X - Y_k)h^M\right]\right| + \left|\mathbb{E}\left[(X - Y_k)h1_{|h|>M}\right]\right| \\ &\le \left|\mathbb{E}\left[(X - Y_k)h^M\right]\right| + \|X - Y_k\|_p\left\|h1_{|h|>M}\right\|_q \\ &\le \left|\mathbb{E}\left[(X - Y_k)h^M\right]\right| + 2C\left\|h1_{|h|>M}\right\|_q.\end{aligned}$$

Since $h^M \in \mathcal{B}_b$, we may pass to the limit $k \to \infty$ in the previous inequality to find,

$$\limsup_{k\to\infty} |\mathbb{E}[(X - Y_k)h]| \le 2C \left\|h1_{|h|>M}\right\|_q .$$

This completes the proof, since $\left\|h1_{|h|>M}\right\|_q \to 0$ as $M \to \infty$ by DCT. ■

13.2 Exercises

Exercise 13.5. Suppose that $\{M_n\}_{n=1}^\infty$ is an increasing sequence of closed subspaces of a Hilbert space, H. Let M be the closure of $M_0 := \cup_{n=1}^\infty M_n$. Show $\lim_{n\to\infty} P_{M_n}x = P_M x$ for all $x \in H$. **Hint:** first prove this for $x \in M_0$ and then for $x \in M$. Also consider the case where $x \in M^\perp$.

Solution to Exercise (13.5). Let $P_n := P_{M_n}$ and $P = P_M$. If $y \in M_0$, then $P_n y = y = Py$ for all n sufficiently large. and therefore, $\lim_{n\to\infty} P_n y = Py$. Now suppose that $x \in M$ and $y \in M_0$. Then

$$\begin{aligned}\|Px - P_n x\| &\le \|Px - Py\| + \|Py - P_n y\| + \|P_n y - P_n x\| \\ &\le 2\|x - y\| + \|Py - P_n y\|\end{aligned}$$

and passing to the limit as $n \to \infty$ then shows

$$\limsup_{n\to\infty} \|Px - P_n x\| \le 2\|x - y\| .$$

The left hand side may be made as small as we like by choosing $y \in M_0$ arbitrarily close to $x \in M = \bar{M}_0$.

For the general case, if $x \in H$, then $x = Px + y$ where $y = x - Px \in M^\perp \subset M_n^\perp$ for all n. Therefore,

$$P_n x = P_n P x \to Px \text{ as } n \to \infty$$

by what we have just proved.

Exercise 13.6 (A "Martingale" Convergence Theorem). Suppose that $\{M_n\}_{n=1}^\infty$ is an increasing sequence of closed subspaces of a Hilbert space, H, $P_n := P_{M_n}$, and $\{x_n\}_{n=1}^\infty$ is a sequence of elements from H such that $x_n = P_n x_{n+1}$ for all $n \in \mathbb{N}$. Show;

1. $P_m x_n = x_m$ for all $1 \le m \le n < \infty$,
2. $(x_n - x_m) \perp M_m$ for all $n \ge m$,
3. $\|x_n\|$ is increasing as n increases,
4. if $\sup_n \|x_n\| = \lim_{n\to\infty} \|x_n\| < \infty$, then $x := \lim_{n\to\infty} x_n$ exists in M and that $x_n = P_n x$ for all $n \in \mathbb{N}$. (**Hint:** show $\{x_n\}_{n=1}^\infty$ is a Cauchy sequence.)

Remark 13.23. Let $H = \ell^2 := L^2(\mathbb{N}, \text{counting measure})$,

$$M_n = \{(a(1), \dots, a(n), 0, 0, \dots) : a(i) \in \mathbb{C} \text{ for } 1 \le i \le n\},$$

and $x_n(i) = 1_{i \le n}$, then $x_m = P_m x_n$ for all $n \ge m$ while $\|x_n\|^2 = n \uparrow \infty$ as $n \to \infty$. Thus, we can not drop the assumption that $\sup_n \|x_n\| < \infty$ in Exercise 13.6.

The rest of this section may be safely skipped.

Exercise 13.7. *Suppose that $(X, \|\cdot\|)$ is a normed space such that parallelogram law, Eq. (13.2), holds for all $x, y \in X$, then there exists a unique inner product on $\langle\cdot|\cdot\rangle$ such that $\|x\| := \sqrt{\langle x|x\rangle}$ for all $x \in X$. In this case we say that $\|\cdot\|$ is a Hilbertian norm.

Solution to Exercise (13.7). If $\|\cdot\|$ is going to come from an inner product $\langle\cdot|\cdot\rangle$, it follows from Eq. (13.1) that

$$2\mathrm{Re}\langle x|y\rangle = \|x + y\|^2 - \|x\|^2 - \|y\|^2$$

and

$$-2\mathrm{Re}\langle x|y\rangle = \|x - y\|^2 - \|x\|^2 - \|y\|^2.$$

Subtracting these two equations gives the "polarization identity,"

$$4\mathrm{Re}\langle x|y\rangle = \|x + y\|^2 - \|x - y\|^2. \tag{13.13}$$

Replacing y by iy in this equation then implies that

$$4\mathrm{Im}\langle x|y\rangle = \|x + iy\|^2 - \|x - iy\|^2$$

from which we find

$$\langle x|y\rangle = \frac{1}{4}\sum_{\varepsilon\in G} \varepsilon\|x + \varepsilon y\|^2 \tag{13.14}$$

where $G = \{\pm 1, \pm i\}$ – a cyclic subgroup of $S^1 \subset \mathbb{C}$. Hence, if $\langle\cdot|\cdot\rangle$ is going to exist we must define it by Eq. (13.14) and the uniqueness has been proved.

For existence, define $\langle x|y\rangle$ by Eq. (13.14) in which case,

$$\begin{aligned}\langle x|x\rangle &= \frac{1}{4}\sum_{\varepsilon\in G} \varepsilon\|x + \varepsilon x\|^2 = \frac{1}{4}\left[\|2x\|^2 + i\|x + ix\|^2 - i\|x - ix\|^2\right] \\ &= \|x\|^2 + \frac{i}{4}\left|1 + i|^2\right|\|x\|^2 - \frac{i}{4}\left|1 - i|^2\right|\|x\|^2 = \|x\|^2 .\end{aligned}$$

So to finish the proof, it only remains to show that $\langle x|y\rangle$ defined by Eq. (13.14) is an inner product.

Since

$$\begin{aligned}4\langle y|x\rangle &= \sum_{\varepsilon\in G}\varepsilon\|y+\varepsilon x\|^2 = \sum_{\varepsilon\in G}\varepsilon\|\varepsilon\left(y+\varepsilon x\right)\|^2\\ &= \sum_{\varepsilon\in G}\varepsilon\|\varepsilon y+\varepsilon^2 x\|^2\\ &= \|y+x\|^2 - \|-y+x\|^2 + i\|iy-x\|^2 - i\|-iy-x\|^2\\ &= \|x+y\|^2 - \|x-y\|^2 + i\|x-iy\|^2 - i\|x+iy\|^2\\ &= 4\overline{\langle x|y\rangle}\end{aligned}$$

it suffices to show $x\to\langle x|y\rangle$ is linear for all $y\in H$. For this we will need to derive an identity from Eq. (13.2),. To do this we make use of Eq. (13.2), three times to find

$$\begin{aligned}\|x+y+z\|^2 &= -\|x+y-z\|^2 + 2\|x+y\|^2 + 2\|z\|^2\\ &= \|x-y-z\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2\\ &= \|y+z-x\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2\\ &= -\|y+z+x\|^2 + 2\|y+z\|^2 + 2\|x\|^2\\ &\quad - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2.\end{aligned}$$

Solving this equation for $\|x+y+z\|^2$ gives

$$\|x+y+z\|^2 = \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2. \quad (13.15)$$

Using Eq. (13.15), for $x,y,z\in H$,

$$\begin{aligned}4\,\mathrm{Re}\langle x+z|y\rangle &= \|x+z+y\|^2 - \|x+z-y\|^2\\ &= \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2\\ &\quad - \left(\|z-y\|^2 + \|x-y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2\right)\\ &= \|z+y\|^2 - \|z-y\|^2 + \|x+y\|^2 - \|x-y\|^2\\ &= 4\,\mathrm{Re}\langle x|y\rangle + 4\,\mathrm{Re}\langle z|y\rangle. \quad (13.16)\end{aligned}$$

Now suppose that $\delta\in G$, then since $|\delta|=1$,

$$\begin{aligned}4\langle\delta x|y\rangle &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|\delta x+\varepsilon y\|^2 = \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\delta^{-1}\varepsilon y\|^2\\ &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\delta\|x+\delta\varepsilon y\|^2 = 4\delta\langle x|y\rangle \quad (13.17)\end{aligned}$$

where in the third inequality, the substitution $\varepsilon\to\varepsilon\delta$ was made in the sum. So Eq. (13.17) says $\langle\pm ix|y\rangle = \pm i\langle x|y\rangle$ and $\langle -x|y\rangle = -\langle x|y\rangle$. Therefore

$$\mathrm{Im}\langle x|y\rangle = \mathrm{Re}\left(-i\langle x|y\rangle\right) = \mathrm{Re}\langle -ix|y\rangle$$

which combined with Eq. (13.16) shows

$$\begin{aligned}\mathrm{Im}\langle x+z|y\rangle &= \mathrm{Re}\langle -ix-iz|y\rangle = \mathrm{Re}\langle -ix|y\rangle + \mathrm{Re}\langle -iz|y\rangle\\ &= \mathrm{Im}\langle x|y\rangle + \mathrm{Im}\langle z|y\rangle\end{aligned}$$

and therefore (again in combination with Eq. (13.16)),

$$\langle x+z|y\rangle = \langle x|y\rangle + \langle z|y\rangle \text{ for all } x,y\in H.$$

Because of this equation and Eq. (13.17) to finish the proof that $x\to\langle x|y\rangle$ is linear, it suffices to show $\langle\lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda>0$. Now if $\lambda = m\in\mathbb{N}$, then

$$\langle mx|y\rangle = \langle x+(m-1)x|y\rangle = \langle x|y\rangle + \langle (m-1)x|y\rangle$$

so that by induction $\langle mx|y\rangle = m\langle x|y\rangle$. Replacing x by x/m then shows that $\langle x|y\rangle = m\langle m^{-1}x|y\rangle$ so that $\langle m^{-1}x|y\rangle = m^{-1}\langle x|y\rangle$ and so if $m,n\in\mathbb{N}$, we find

$$\langle\frac{n}{m}x|y\rangle = n\langle\frac{1}{m}x|y\rangle = \frac{n}{m}\langle x|y\rangle$$

so that $\langle\lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda>0$ and $\lambda\in\mathbb{Q}$. By continuity, it now follows that $\langle\lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda>0$.

An alternate ending: In the case where X is real, the latter parts of the proof are easier to digest as we can use Eq. (13.13) for the formula for the inner product. For example, we have

$$\begin{aligned}4\left\langle x|2z\right\rangle &= \|x+2z\|^2 - \|x-2z\|^2\\ &= \|x+z+z\|^2 + \|x+z-z\|^2 - \|x-z+z\|^2 - \|x-z-z\|^2\\ &= \frac{1}{2}\left[\|x+z\|^2 + \|z\|^2\right] - \frac{1}{2}\left[\|x-z\|^2 + \|z\|^2\right]\\ &= \frac{1}{2}\left[\|x+z\|^2 - \|x-z\|^2\right] = 2\left\langle x|z\right\rangle\end{aligned}$$

from which it follows that $\langle x|2z\rangle = 2\left\langle x|z\right\rangle$. Similarly,

$$\begin{aligned}4\left[\langle x|z\rangle + \langle y|z\rangle\right] &= \|x+z\|^2 - \|x-z\|^2 + \|y+z\|^2 - \|y-z\|^2\\ &= \|x+z\|^2 + \|y+z\|^2 - \|x-z\|^2 - \|y-z\|^2\\ &= \frac{1}{2}\left(\|x+y+2z\|^2 + \|x-y\|^2\right) - \frac{1}{2}\left(\|x+y-2z\|^2 + \|x-y\|^2\right)\\ &= 2\left\langle x+y|2z\right\rangle = 4\left\langle x+y|z\right\rangle\end{aligned}$$

from which it follows that and $\langle x+y|z\rangle = \langle x|z\rangle + \langle y|z\rangle$. From this identity one shows as above that $\langle\cdot|\cdot\rangle$ is a real inner product on X.

Now suppose that X is complex and now let

$$Q(x,y) = \frac{1}{4}\left[\|x+z\|^2 - \|x-z\|^2\right].$$

We should expect that $Q(\cdot,\cdot) = \operatorname{Re}\langle\cdot|\cdot\rangle$ and therefore we should define

$$\langle x|y\rangle := Q(x,y) - iQ(ix,y).$$

Since

$$\begin{aligned}4Q(ix,y) &= \|ix+y\|^2 - \|ix-y\|^2 = \|-i(ix+y)\|^2 - \|-i(ix-y)\|^2 \\ &= \|x-iy\|^2 - \|x+iy\|^2 = -4Q(x,iy),\end{aligned}$$

it follows that $Q(ix,x) = 0$ so that $\langle x|x\rangle = \|x\|^2$ and that

$$\langle y|x\rangle = Q(y,x) - iQ(iy,x) = Q(y,x) + iQ(y,ix) = \overline{\langle x|y\rangle}.$$

Since $x \to \langle x|y\rangle$ is real linear, we now need only show that $\langle ix|y\rangle = i\langle x|y\rangle$. However,

$$\begin{aligned}\langle ix|y\rangle &= Q(ix,y) - iQ(i(ix),y) \\ &= Q(ix,y) + iQ(x,y) = i\langle x|y\rangle\end{aligned}$$

as desired.

Conditional Expectation

In this section let $(\Omega, \mathcal{B}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{B}$ be a sub – sigma algebra of $\mathcal{B}$. We will write $f \in \mathcal{G}_b$ iff $f : \Omega \to \mathbb{C}$ is bounded and f is $(\mathcal{G}, \mathcal{B}_{\mathbb{C}})$ – measurable. If $A \in \mathcal{B}$ and $P(A) > 0$, we will let

$$\mathbb{E}[X|A] := \frac{\mathbb{E}[X : A]}{P(A)} \text{ and } P(B|A) := \mathbb{E}[1_B|A] := \frac{P(A \cap B)}{P(A)}$$

for all integrable random variables, X, and $B \in \mathcal{B}$. We will often use the factorization Lemma 6.40 in this section. Because of this let us repeat it here.

Lemma 14.1. *Suppose that $(\mathbb{Y}, \mathcal{F})$ is a measurable space and $Y : \Omega \to \mathbb{Y}$ is a map. Then to every $(\sigma(Y), \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $H : \Omega \to \bar{\mathbb{R}}$, there is a $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function $h : \mathbb{Y} \to \bar{\mathbb{R}}$ such that $H = h \circ Y$.*

Proof. First suppose that $H = 1_A$ where $A \in \sigma(Y) = Y^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = Y^{-1}(B)$ then $1_A = 1_{Y^{-1}(B)} = 1_B \circ Y$ and hence the lemma is valid in this case with $h = 1_B$. More generally if $H = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ Y$ and hence $H = h \circ Y$ with $h := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$.

For a general $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, H, from $\Omega \to \bar{\mathbb{R}}$, choose simple functions H_n converging to H. Let $h_n : \mathbb{Y} \to \bar{\mathbb{R}}$ be simple functions such that $H_n = h_n \circ Y$. Then it follows that

$$H = \lim_{n\to\infty} H_n = \limsup_{n\to\infty} H_n = \limsup_{n\to\infty} h_n \circ Y = h \circ Y$$

where $h := \limsup_{n\to\infty} h_n$ – a measurable function from $\mathbb{Y}$ to $\bar{\mathbb{R}}$. ∎

Definition 14.2 (Conditional Expectation). *Let $\mathbb{E}_{\mathcal{G}} : L^2(\Omega, \mathcal{B}, P) \to L^2(\Omega, \mathcal{G}, P)$ denote orthogonal projection of $L^2(\Omega, \mathcal{B}, P)$ onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$. For $f \in L^2(\Omega, \mathcal{B}, P)$, we say that $\mathbb{E}_{\mathcal{G}} f \in L^2(\Omega, \mathcal{G}, P)$ is the* ***conditional expectation of f given $\mathcal{G}$.***

Remark 14.3 (Basic Properties of $\mathbb{E}_{\mathcal{G}}$). Let $f \in L^2(\Omega, \mathcal{B}, P)$. By the orthogonal projection Theorem 13.13 we know that $F \in L^2(\Omega, \mathcal{G}, P)$ is $\mathbb{E}_{\mathcal{G}} f$ a.s. iff either of the following two conditions hold;

1. $\|f - F\|_2 \le \|f - g\|_2$ for all $g \in L^2(\Omega, \mathcal{G}, P)$ or
2. $\mathbb{E}[fh] = \mathbb{E}[Fh]$ for all $h \in L^2(\Omega, \mathcal{G}, P)$.

Moreover if $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$ then $L^2(\Omega, \mathcal{G}_0, P) \subset L^2(\Omega, \mathcal{G}_1, P) \subset L^2(\Omega, \mathcal{B}, P)$ and therefore,

$$\mathbb{E}_{\mathcal{G}_0} \mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^2(\Omega, \mathcal{B}, P). \tag{14.1}$$

It is also useful to observe that condition 2. above may expressed as

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G} \tag{14.2}$$

or

$$\mathbb{E}[fh] = \mathbb{E}[Fh] \text{ for all } h \in \mathcal{G}_b. \tag{14.3}$$

Indeed, if Eq. (14.2) holds, then by linearity we have $\mathbb{E}[fh] = \mathbb{E}[Fh]$ for all $\mathcal{G}$ – measurable simple functions, h and hence by the approximation Theorem 6.39 and the DCT for all $h \in \mathcal{G}_b$. Therefore Eq. (14.2) implies Eq. (14.3). If Eq. (14.3) holds and $h \in L^2(\Omega, \mathcal{G}, P)$, we may use DCT to show

$$\mathbb{E}[fh] \overset{\text{DCT}}{=} \lim_{n\to\infty} \mathbb{E}\left[fh 1_{|h|\le n}\right] \overset{(14.3)}{=} \lim_{n\to\infty} \mathbb{E}\left[Fh 1_{|h|\le n}\right] \overset{\text{DCT}}{=} \mathbb{E}[Fh],$$

which is condition 2. in Remark 14.3. Taking $h = 1_A$ with $A \in \mathcal{G}$ in condition 2. or Remark 14.3, we learn that Eq. (14.2) is satisfied as well.

Theorem 14.4. *Let $(\Omega, \mathcal{B}, P)$ and $\mathcal{G} \subset \mathcal{B}$ be as above and let $f, g \in L^1(\Omega, \mathcal{B}, P)$. The operator $\mathbb{E}_{\mathcal{G}} : L^2(\Omega, \mathcal{B}, P) \to L^2(\Omega, \mathcal{G}, P)$ extends uniquely to a linear contraction from $L^1(\Omega, \mathcal{B}, P)$ to $L^1(\Omega, \mathcal{G}, P)$. This extension enjoys the following properties;*

1. *If $f \ge 0$, P – a.e. then $\mathbb{E}_{\mathcal{G}} f \ge 0$, P – a.e.*
2. ***Monotonicity.*** *If $f \ge g$, P – a.e. there $\mathbb{E}_{\mathcal{G}} f \ge \mathbb{E}_{\mathcal{G}} g$, P – a.e.*
3. *L^∞* ***– contraction property.*** *$|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, P – a.e.*
4. ***Averaging Property.*** *If $f \in L^1(\Omega, \mathcal{B}, P)$ then $F = \mathbb{E}_{\mathcal{G}} f$ iff $F \in L^1(\Omega, \mathcal{G}, P)$ and*
$$\mathbb{E}(Fh) = \mathbb{E}(fh) \text{ for all } h \in \mathcal{G}_b. \tag{14.4}$$
5. ***Pull out property or product rule.*** *If $g \in \mathcal{G}_b$ and $f \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}}(gf) = g \cdot \mathbb{E}_{\mathcal{G}} f$, P – a.e.*

6. ***Tower or smoothing property.*** *If $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$. Then*

$$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^1(\Omega, \mathcal{B}, P). \tag{14.5}$$

Proof. By the definition of orthogonal projection, $f \in L^2(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$,

$$\mathbb{E}(fh) = \mathbb{E}(f \cdot \mathbb{E}_{\mathcal{G}} h) = \mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h). \tag{14.6}$$

Taking

$$h = \overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}} f)} := \frac{\overline{\mathbb{E}_{\mathcal{G}} f}}{\mathbb{E}_{\mathcal{G}} f} 1_{|\mathbb{E}_{\mathcal{G}} f|>0} \tag{14.7}$$

in Eq. (14.6) shows

$$\mathbb{E}(|\mathbb{E}_{\mathcal{G}} f|) = \mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h) = \mathbb{E}(fh) \le \mathbb{E}(|fh|) \le \mathbb{E}(|f|). \tag{14.8}$$

It follows from this equation and the BLT (Theorem 13.18) that $\mathbb{E}_{\mathcal{G}}$ extends uniquely to a contraction form $L^1(\Omega, \mathcal{B}, P)$ to $L^1(\Omega, \mathcal{G}, P)$. Moreover, by a simple limiting argument, Eq. (14.6) remains valid for all $f \in L^1(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$. Indeed, (without reference to Theorem 13.18) if $f_n := f 1_{|f| \le n} \in L^2(\Omega, \mathcal{B}, P)$, then $f_n \to f$ in $L^1(\Omega, \mathcal{B}, P)$ and hence

$$\mathbb{E}[|\mathbb{E}_{\mathcal{G}} f_n - \mathbb{E}_{\mathcal{G}} f_m|] = \mathbb{E}[|\mathbb{E}_{\mathcal{G}}(f_n - f_m)|] \le \mathbb{E}[|f_n - f_m|] \to 0 \text{ as } m, n \to \infty.$$

By the completeness of $L^1(\Omega, \mathcal{G}, P)$, $F := L^1(\Omega, \mathcal{G}, P)\text{-}\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n$ exists. Moreover the function F satisfies,

$$\mathbb{E}(F \cdot h) = \mathbb{E}(\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n \cdot h) = \lim_{n\to\infty} \mathbb{E}(f_n \cdot h) = \mathbb{E}(f \cdot h) \tag{14.9}$$

for all $h \in \mathcal{G}_b$ and by Proposition 7.22 there is at most one, $F \in L^1(\Omega, \mathcal{G}, P)$, which satisfies Eq. (14.9). We will again denote F by $\mathbb{E}_{\mathcal{G}} f$. This proves the existence and uniqueness of F satisfying the defining relation in Eq. (14.4) of item 4. The same argument used in Eq. (14.8) again shows $\mathbb{E}|F| \le \mathbb{E}|f|$ and therefore that $\mathbb{E}_{\mathcal{G}} : L^1(\Omega, \mathcal{B}, P) \to L^1(\Omega, \mathcal{G}, P)$ is a contraction.

Items 1 and 2. If $f \in L^1(\Omega, \mathcal{B}, P)$ with $f \ge 0$, then

$$\mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h) = \mathbb{E}(fh) \ge 0 \;\forall\, h \in \mathcal{G}_b \text{ with } h \ge 0. \tag{14.10}$$

An application of Lemma 7.23 then shows that $\mathbb{E}_{\mathcal{G}} f \ge 0$ a.s.[1] The proof of item 2. follows by applying item 1. with f replaced by $f - g \ge 0$.

Item 3. If f is real, $\pm f \le |f|$ and so by Item 2., $\pm\mathbb{E}_{\mathcal{G}} f \le \mathbb{E}_{\mathcal{G}} |f|$, i.e. $|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, P – a.e. For complex f, let $h \ge 0$ be a bounded and $\mathcal{G}$ – measurable function. Then

$$\mathbb{E}[|\mathbb{E}_{\mathcal{G}} f|\, h] = \mathbb{E}\left[\mathbb{E}_{\mathcal{G}} f \cdot \overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}} f)} h\right] = \mathbb{E}\left[f \cdot \overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}} f)} h\right]$$
$$\le \mathbb{E}[|f|\, h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}} |f| \cdot h].$$

Since $h \ge 0$ is an arbitrary $\mathcal{G}$ – measurable function, it follows, by Lemma 7.23, that $|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, P – a.s. Recall the item 4. has already been proved.

Item 5. If $h, g \in \mathcal{G}_b$ and $f \in L^1(\Omega, \mathcal{B}, P)$, then

$$\mathbb{E}[(g\mathbb{E}_{\mathcal{G}} f)\, h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}} f \cdot hg] = \mathbb{E}[f \cdot hg] = \mathbb{E}[gf \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}}(gf) \cdot h].$$

Thus $\mathbb{E}_{\mathcal{G}}(gf) = g \cdot \mathbb{E}_{\mathcal{G}} f$, P – a.e.

Item 6., by the item 5. of the projection Theorem 13.13, Eq. (14.5) holds on $L^2(\Omega, \mathcal{B}, P)$. By continuity of conditional expectation on $L^1(\Omega, \mathcal{B}, P)$ and the density of L^1 probability spaces in L^2 – probability spaces shows that Eq. (14.5) continues to hold on $L^1(\Omega, \mathcal{B}, P)$.

Second Proof. For $h \in (\mathcal{G}_0)_b$, we have

$$\mathbb{E}[\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}_1} f \cdot h] = \mathbb{E}[f \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}_0} f \cdot h]$$

which shows $\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_0} f$ a.s. By the product rule in item 5., it also follows that

$$\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f] = \mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f \cdot 1] = \mathbb{E}_{\mathcal{G}_0} f \cdot \mathbb{E}_{\mathcal{G}_1}[1] = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s.}$$

Notice that $\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f]$ need only be $\mathcal{G}_1$ – measurable. What the statement says there are representatives of $\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f]$ which is $\mathcal{G}_0$ – measurable and any such representative is also a representative of $\mathbb{E}_{\mathcal{G}_0} f$. ■

Remark 14.5. There is another standard construction of $\mathbb{E}_{\mathcal{G}} f$ based on the characterization in Eq. (14.4) and the Radon Nikodym Theorem 15.8 below. It goes as follows, for $0 \le f \in L^1(P)$, let Q be the measure defined by $dQ := f dP$. Then $Q|_{\mathcal{G}} \ll P|_{\mathcal{G}}$ and hence there exists $0 \le g \in L^1(\Omega, \mathcal{G}, P)$ such that $dQ|_{\mathcal{G}} = g dP|_{\mathcal{G}}$. This then implies that

$$\int_A f dP = Q(A) = \int_A g dP \text{ for all } A \in \mathcal{G},$$

i.e. $g = \mathbb{E}_{\mathcal{G}} f$. For general real valued, $f \in L^1(P)$, define $\mathbb{E}_{\mathcal{G}} f = \mathbb{E}_{\mathcal{G}} f_+ - \mathbb{E}_{\mathcal{G}} f_-$ and then for complex $f \in L^1(P)$ let $\mathbb{E}_{\mathcal{G}} f = \mathbb{E}_{\mathcal{G}} \operatorname{Re} f + i\mathbb{E}_{\mathcal{G}} \operatorname{Im} f$.

Notation 14.6 *In the future, we will often write $\mathbb{E}_{\mathcal{G}} f$ as $\mathbb{E}[f|\mathcal{G}]$. Moreover, if $(\mathbb{X}, \mathcal{M})$ is a measurable space and $X : \Omega \to \mathbb{X}$ is a measurable map. We will often simply denote $\mathbb{E}[f|\sigma(X)]$ simply by $\mathbb{E}[f|X]$. We will further let $P(A|\mathcal{G}) := \mathbb{E}[1_A|\mathcal{G}]$ be the* ***conditional probability of*** *A* ***given*** *$\mathcal{G}$, and $P(A|X) := P(A|\sigma(X))$ be* ***conditional probability of*** *A* ***given*** *X.*

Exercise 14.1. Suppose $f \in L^1(\Omega, \mathcal{B}, P)$ and $f > 0$ a.s. Show $\mathbb{E}[f|\mathcal{G}] > 0$ a.s. (i.e. show $g > 0$ a.s. for any version, g, of $\mathbb{E}[f|\mathcal{G}]$.) Use this result to conclude if $f \in (a, b)$ a.s. for some a, b such that $-\infty \le a < b \le \infty$, then $\mathbb{E}[f|\mathcal{G}] \in (a, b)$ a.s. More precisely you are to show that any version, g, of $\mathbb{E}[f|\mathcal{G}]$ satisfies, $g \in (a, b)$ a.s.

[1] This can also easily be proved directly here by taking $h = 1_{\mathbb{E}_{\mathcal{G}} f < 0}$ in Eq. (14.10).

14.1 Examples

Example 14.7. Suppose $\mathcal{G}$ is the trivial σ – algebra, i.e. $\mathcal{G} = \{\emptyset, \Omega\}$. In this case $\mathbb{E}_{\mathcal{G}} f = \mathbb{E} f$ a.s.

Example 14.8. On the opposite extreme, if $\mathcal{G} = \mathcal{B}$, then $\mathbb{E}_{\mathcal{G}} f = f$ a.s.

Exercise 14.2 (Exercise 4.15 revisited.). Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\mathcal{P} := \{A_i\}_{i=1}^{\infty} \subset \mathcal{B}$ is a partition of Ω. (Recall this means $\Omega = \sum_{i=1}^{\infty} A_i$.) Let $\mathcal{G}$ be the σ – algebra generated by $\mathcal{P}$. Show:

1. $B \in \mathcal{G}$ iff $B = \cup_{i \in \Lambda} A_i$ for some $\Lambda \subset \mathbb{N}$.
2. $g : \Omega \to \mathbb{R}$ is $\mathcal{G}$ – measurable iff $g = \sum_{i=1}^{\infty} \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$.
3. For $f \in L^1(\Omega, \mathcal{B}, P)$, let $\mathbb{E}[f|A_i] := \mathbb{E}[1_{A_i} f] / P(A_i)$ if $P(A_i) \neq 0$ and $\mathbb{E}[f|A_i] = 0$ otherwise. Show

$$\mathbb{E}_{\mathcal{G}} f = \sum_{i=1}^{\infty} \mathbb{E}[f|A_i] 1_{A_i} \text{ a.s.} \tag{14.11}$$

Solution to Exercise (14.2). We will only prove part 3. here. To do this, suppose that $\mathbb{E}_{\mathcal{G}} f = \sum_{i=1}^{\infty} \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$. Then

$$\mathbb{E}[f : A_j] = \mathbb{E}[\mathbb{E}_{\mathcal{G}} f : A_j] = \mathbb{E}\left[\sum_{i=1}^{\infty} \lambda_i 1_{A_i} : A_j\right] = \lambda_j P(A_j)$$

which holds automatically if $P(A_j) = 0$ no matter how λ_j is chosen. Therefore, we must take

$$\lambda_j = \frac{\mathbb{E}[f : A_j]}{P(A_j)} = \mathbb{E}[f|A_j]$$

which verifies Eq. (14.11).

Example 14.9. If S is a countable or finite set equipped with the σ – algebra, 2^S, and $X : \Omega \to S$ is a measurable map. Then

$$\mathbb{E}[Z|X] = \sum_{s \in S} \mathbb{E}[Z|X = s] 1_{X=s} \text{ a.s.}$$

where by convention we set $\mathbb{E}[Z|X = s] = 0$ if $P(X = s) = 1$. This is an immediate consequence of Exercise 14.2 with $\mathcal{G} = \sigma(X)$ which is generated by the partition, $\{X = s\}$ for $s \in S$. Thus if we define $F(s) := \mathbb{E}[Z|X = s]$, we will have $\mathbb{E}[Z|X] = F(X)$ a.s.

Lemma 14.10. *Suppose $(\mathbb{X}, \mathcal{M})$ is a measurable space, $X : \Omega \to \mathbb{X}$ is a measurable function, and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{B}$. If X is independent of $\mathcal{G}$ and $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $f(X) \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}}[f(X)] = \mathbb{E}[f(X)]$ a.s.. Conversely if $\mathbb{E}_{\mathcal{G}}[f(X)] = \mathbb{E}[f(X)]$ a.s. for all bounded measurable functions, $f : \mathbb{X} \to \mathbb{R}$, then X is independent of $\mathcal{G}$.*

Proof. Suppose that X is independent of $\mathcal{G}$, $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $f(X) \in L^1(\Omega, \mathcal{B}, P)$, $\mu := \mathbb{E}[f(X)]$, and $A \in \mathcal{G}$. Then, by independence,

$$\mathbb{E}[f(X) : A] = \mathbb{E}[f(X) 1_A] = \mathbb{E}[f(X)] \mathbb{E}[1_A] = \mathbb{E}[\mu 1_A] = \mathbb{E}[\mu : A].$$

Therefore $\mathbb{E}_{\mathcal{G}}[f(X)] = \mu = \mathbb{E}[f(X)]$ a.s.

Conversely if $\mathbb{E}_{\mathcal{G}}[f(X)] = \mathbb{E}[f(X)] = \mu$ and $A \in \mathcal{G}$, then

$$\mathbb{E}[f(X) 1_A] = \mathbb{E}[f(X) : A] = \mathbb{E}[\mu : A] = \mu \mathbb{E}[1_A] = \mathbb{E}[f(X)] \mathbb{E}[1_A].$$

Since this last equation is assumed to hold true for all $A \in \mathcal{G}$ and all bounded measurable functions, $f : \mathbb{X} \to \mathbb{R}$, X is independent of $\mathcal{G}$. ∎

The following remark is often useful in computing conditional expectations. The following Exercise should help you gain some more intuition about conditional expectations.

*Remark 14.11 (**Note well.**).* According to Lemma 14.1, $\mathbb{E}(f|X) = \tilde{f}(X)$ a.s. for some measurable function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$. So computing $\mathbb{E}(f|X) = \tilde{f}(X)$ is equivalent to finding a function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$, such that

$$\mathbb{E}[f \cdot h(X)] = \mathbb{E}\left[\tilde{f}(X) h(X)\right] \tag{14.12}$$

for all bounded and measurable functions, $h : \mathbb{X} \to \mathbb{R}$. "The" function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$, is often denoted by writing $\tilde{f}(x) = \mathbb{E}(f|X = x)$. If $P(X = x) > 0$, then $\mathbb{E}(f|X = x) = \mathbb{E}(f : X = x) / P(X = x)$ consistent with our previous definitions – compare with Example 14.9. If $P(X = x)$, $\mathbb{E}(f|X = x)$ is not given a value but is **just** a convenient notational way to denote a function $\tilde{f} : \mathbb{X} \to \mathbb{R}$ such that Eq. (14.12) holds. (Roughly speaking, you should think that $\mathbb{E}(f|X = x) = \mathbb{E}[f \cdot \delta_x(X)] / \mathbb{E}[\delta_x(X)]$ where δ_x is the "Dirac delta function" at x. If this last comment is confusing to you, please ignore it!)

Example 14.12. Suppose that X is a random variable, $t \in \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $f(X) \in L^1(P)$. We wish to compute $\mathbb{E}[f(X) | X \wedge t] = h(X \wedge t)$. So we are looking for a function, $h : (-\infty, t] \to \mathbb{R}$ such that

$$\mathbb{E}[f(X) u(X \wedge t)] = \mathbb{E}[h(X \wedge t) u(X \wedge t)] \tag{14.13}$$

for all bounded measurable functions, $u : (-\infty, t] \to \mathbb{R}$. Taking $u = 1_{\{t\}}$ in Eq. (14.13) implies,

$$\mathbb{E}[f(X) : X \geq t] = h(t) P(X \geq t)$$

and therefore we should take,

$$h(t) = \mathbb{E}[f(X) | X \geq t]$$

which by convention we set to be (say) zero if $P(X \geq t) = 0$. Now suppose that $u(t) = 0$, then Eq. (8.6) becomes,

$$\mathbb{E}[f(X)u(X) : X < t] = \mathbb{E}[h(X)u(X) : X < t]$$

from which it follows that $f(X)1_{X<t} = h(X)1_{X<t}$ a.s. Thus we can take

$$h(x) := \begin{cases} f(x) & \text{if } x < t \\ \mathbb{E}[f(X)|X \geq t] & \text{if } x = t \end{cases}$$

and we have shown,

$$\begin{aligned}\mathbb{E}[f(X)|X \wedge t] &= 1_{X<t}f(X) + 1_{X\geq t}\mathbb{E}[f(X)|X \geq t] \\ &= 1_{X\wedge t<t}f(X) + 1_{X\wedge t=t}\mathbb{E}[f(X)|X \geq t].\end{aligned}$$

Proposition 14.13. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $(\mathbb{X}, \mathcal{M}, \mu)$ and $(\mathbb{Y}, \mathcal{N}, \nu)$ are two σ – finite measure spaces, $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are measurable functions, and there exists $0 \leq \rho \in L^1(\Omega, \mathcal{B}, \mu \otimes \nu)$ such that $P((X,Y) \in U) = \int_U \rho(x,y)\,d\mu(x)\,d\nu(y)$ for all $U \in \mathcal{M} \otimes \mathcal{N}$. Let*

$$\bar{\rho}(x) := \int_{\mathbb{Y}} \rho(x,y)\,d\nu(y) \tag{14.14}$$

and $x \in \mathbb{X}$ and $B \in \mathcal{N}$, let

$$Q(x,B) := \begin{cases} \frac{1}{\bar{\rho}(x)} \int_B \rho(x,y)\,d\nu(y) & \text{if } \bar{\rho}(x) \in (0,\infty) \\ \delta_{y_0}(B) & \text{if } \bar{\rho}(x) \in \{0,\infty\} \end{cases} \tag{14.15}$$

where y_0 is some arbitrary but fixed point in Y. Then for any bounded (or non-negative) measurable function, $f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$, we have

$$\mathbb{E}[f(X,Y)|X] = Q(X, f(X,\cdot)) =: \int_{\mathbb{Y}} f(X,y)\,Q(X,dy) = g(X) \quad a.s. \tag{14.16}$$

where,

$$g(x) := \int_{\mathbb{Y}} f(x,y)\,Q(x,dy) = Q(x, f(x,\cdot)).$$

As usual we use the notation,

$$Q(x,v) := \int_{\mathbb{Y}} v(y)\,Q(x,dy) = \begin{cases} \frac{1}{\bar{\rho}(x)} \int_{\mathbb{Y}} v(y)\rho(x,y)\,d\nu(y) & \text{if } \bar{\rho}(x) \in (0,\infty) \\ \delta_{y_0}(v) = v(y_0) & \text{if } \bar{\rho}(x) \in \{0,\infty\}. \end{cases}$$

for all bounded measurable functions, $v : \mathbb{Y} \to \mathbb{R}$.

Proof. Our goal is to compute $\mathbb{E}[f(X,Y)|X]$. According to Remark 14.11, we are searching for a bounded measurable function, $g : \mathbb{X} \to \mathbb{R}$, such that

$$\mathbb{E}[f(X,Y)h(X)] = \mathbb{E}[g(X)h(X)] \text{ for all } h \in \mathcal{M}_b. \tag{14.17}$$

(Throughout this argument we are going to repeatedly use the Tonelli - Fubini theorems.) We now explicitly write out both sides of Eq. (14.17);

$$\begin{aligned}\mathbb{E}[f(X,Y)h(X)] &= \int_{\mathbb{X}\times\mathbb{Y}} h(x)f(x,y)\rho(x,y)\,d\mu(x)\,d\nu(y) \\ &= \int_{\mathbb{X}} h(x)\left[\int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y)\right]d\mu(x) \end{aligned} \tag{14.18}$$

$$\begin{aligned}\mathbb{E}[g(X)h(X)] &= \int_{\mathbb{X}\times\mathbb{Y}} h(x)g(x)\rho(x,y)\,d\mu(x)\,d\nu(y) \\ &= \int_{\mathbb{X}} h(x)g(x)\bar{\rho}(x)\,d\mu(x). \end{aligned} \tag{14.19}$$

Since the right sides of Eqs. (14.18) and (14.19) must be equal for all $h \in \mathcal{M}_b$, we must demand (see Lemma 7.23 and 7.24) that

$$\int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y) = g(x)\bar{\rho}(x) \text{ for } \mu - \text{a.e. } x. \tag{14.20}$$

There are two possible problems in solving this equation for $g(x)$ at a particular point x; the first is when $\bar{\rho}(x) = 0$ and the second is when $\bar{\rho}(x) = \infty$. Since

$$\int_{\mathbb{X}} \bar{\rho}(x)\,d\mu(x) = \int_{\mathbb{X}}\left[\int_{\mathbb{Y}} \rho(x,y)\,d\nu(y)\right]d\mu(x) = 1,$$

we know that $\bar{\rho}(x) < \infty$ for μ – a.e. x and therefore it does not matter how g is defined on $\{\bar{\rho} = \infty\}$ as long as it is measurable. If

$$0 = \bar{\rho}(x) = \int_{\mathbb{Y}} \rho(x,y)\,d\nu(y),$$

then $\rho(x,y) = 0$ for ν – a.e. y and therefore,

$$\int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y) = 0. \tag{14.21}$$

Hence Eq. (14.20) will be valid no matter how we choose $g(x)$ for $x \in \{\bar{\rho} = 0\}$. So a valid solution of Eq. (14.20) is

$$g(x) := \begin{cases} \frac{1}{\bar{\rho}(x)} \int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y) & \text{if } \bar{\rho}(x) \in (0,\infty) \\ f(x,y_0) = \delta_{y_0}(f(x,\cdot)) & \text{if } \bar{\rho}(x) \in \{0,\infty\} \end{cases}$$

and with this choice we will have $\mathbb{E}[f(X,Y)|X] = g(X) = Q(X,f)$ a.s. as desired. (Observe here that when $\bar{\rho}(x) < \infty$, $\rho(x,\cdot) \in L^1(\nu)$ and hence $\int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y)$ is a well defined integral.) ■

It is comforting to observe that

$$P(X \in \{\bar{\rho} = 0\}) = P(\bar{\rho}(X) = 0) = \int_{\mathbb{X}} 1_{\bar{\rho}=0}\bar{\rho}\,d\mu = 0$$

and similarly

$$P(X \in \{\bar{\rho} = \infty\}) = \int_{\mathbb{X}} 1_{\bar{\rho}=\infty}\bar{\rho}\,d\mu = 0.$$

Thus it follows that $P(X \in \{x \in \mathbb{X} : \bar{\rho}(x) = 0 \text{ of } \infty\}) = 0$ while the set $\{x \in \mathbb{X} : \bar{\rho}(x) = 0 \text{ of } \infty\}$ is precisely where there is ambiguity in defining $g(x)$. Just for added security, let us check directly that $g(X) = \mathbb{E}[f(X,Y)|X]$ a.s. According to Eq. (14.19) we have

$$\begin{aligned}
\mathbb{E}[g(X)h(X)] &= \int_{\mathbb{X}} h(x)g(x)\bar{\rho}(x)\,d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)g(x)\bar{\rho}(x)\,d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)\bar{\rho}(x)\left(\frac{1}{\bar{\rho}(x)}\int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y)\right)d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)\left(\int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y)\right)d\mu(x) \\
&= \int_{\mathbb{X}} h(x)\left(\int_{\mathbb{Y}} f(x,y)\rho(x,y)\,d\nu(y)\right)d\mu(x) \\
&= \mathbb{E}[f(X,Y)h(X)] \qquad \text{(by Eq. (14.18))},
\end{aligned}$$

wherein we have repeatedly used $\mu(\bar{\rho} = \infty) = 0$ and Eq. (14.21) holds when $\bar{\rho}(x) = 0$. This completes the verification that $g(X) = \mathbb{E}[f(X,Y)|X]$ a.s..

Proposition 14.13 shows that conditional expectation is a generalization of the notion of performing integration over a partial subset of the variables in the integrand. Whereas to compute the expectation, one should integrate over all of the variables. Proposition 14.13 also gives an example of regular conditional probabilities which we now define.

Definition 14.14. *Let $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces. A function, $Q : \mathbb{X}\times\mathcal{N} \to [0,1]$ is a* ***probability kernel on*** *$\mathbb{X}\times\mathbb{Y}$ if*

1. *$Q(x,\cdot) : \mathcal{N} \to [0,1]$ is a probability measure on $(\mathbb{Y},\mathcal{N})$ for each $x \in \mathbb{X}$ and*
2. *$Q(\cdot,B) : \mathbb{X} \to [0,1]$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $B \in \mathcal{N}$.*

If Q is a probability kernel on $\mathbb{X}\times\mathbb{Y}$ and $f : \mathbb{Y} \to \mathbb{R}$ is a bounded measurable function or a positive measurable function, then $x \to Q(x,f) := \int_{\mathbb{Y}} f(y)Q(x,dy)$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable. This is clear for simple functions and then for general functions via simple limiting arguments.

Definition 14.15. *Let $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces and $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ be measurable functions. A probability kernel, Q, on $\mathbb{X}\times\mathbb{Y}$ is said to be a* ***regular conditional distribution of*** *Y* ***given*** *X iff $Q(X,B)$ is a version of $P(Y \in B|X)$ for each $B \in \mathcal{N}$. Equivalently, we should have $Q(X,f) = \mathbb{E}[f(Y)|X]$ a.s. for all $f \in \mathcal{N}_b$.*

The probability kernel, Q, defined in Eq. (14.15) is an example of a regular conditional distribution of Y given X.

Remark 14.16. Unfortunately, regular conditional distributions do not always exists, see Doob [12, p. 624]. However, if we require $\mathbb{Y}$ to be a "standard Borel space," (i.e. $\mathbb{Y}$ is isomorphic to a Borel subset of $\mathbb{R}$), then a conditional distribution of Y given X will always exists. See Theorem 14.32 in the appendix to this chapter. Moreover, it is known that "reasonable" measure spaces are standard Borel spaces, see Section 9.10 above for more details. So in most instances of interest a regular conditional distribution of Y given X **will** exist.

Exercise 14.3. Suppose that $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ are measurable spaces, $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are measurable functions, and there exists a regular conditional distribution, Q, of Y given X. Show:

1. For all bounded measurable functions, $f : (\mathbb{X}\times\mathbb{Y}, \mathcal{M}\otimes\mathcal{N}) \to \mathbb{R}$, the function $\mathbb{X} \ni x \to Q(x, f(x,\cdot))$ is measurable and
$$Q(X, f(X,\cdot)) = \mathbb{E}[f(X,Y)|X] \text{ a.s.} \tag{14.22}$$
Hint: let $\mathbb{H}$ denote the set of bounded measurable functions, f, on $\mathbb{X}\times\mathbb{Y}$ such that the two assertions are valid.
2. If $A \in \mathcal{M}\otimes\mathcal{N}$ and $\mu := P\circ X^{-1}$ be the law of X, then
$$P((X,Y) \in A) = \int_{\mathbb{X}} Q(x, 1_A(x,\cdot))\,d\mu(x) = \int_{\mathbb{X}} d\mu(x)\int_{\mathbb{Y}} 1_A(x,y)Q(x,dy). \tag{14.23}$$

Exercise 14.4. Keeping the same notation as in Exercise 14.3 and further assume that X and Y are independent. Find a regular conditional distribution of Y given X and prove

$$\mathbb{E}[f(X,Y)|X] = h_f(X) \text{ a.s. } \forall \text{ bounded measurable } f : \mathbb{X}\times\mathbb{Y} \to \mathbb{R},$$

where

$$h_f(x) := \mathbb{E}[f(x,Y)] \text{ for all } x \in X,$$

i.e.

$$\mathbb{E}[f(X,Y)|X] = \mathbb{E}[f(x,Y)]|_{x=X} \text{ a.s.}$$

Exercise 14.5 (Exercise 14.4 strengthened). Let $(\mathbb{X},\mathcal{M}), (\mathbb{Y},\mathcal{N})$ be measurable spaces, $(\Omega,\mathcal{F},P)$ a probability space, and $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ be measurable functions. Let $\mathcal{G} \subset \mathcal{F}$ be a σ–field such that X is $\mathcal{G}/\mathcal{M}$ – measurable and Y is independent of $\mathcal{G}$. Then for any bounded measurable $f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ we have

$$\mathbb{E}[f(X,Y)|\mathcal{G}] = h_f(X) \text{ where } h_f(x) := \mathbb{E}[f(x,Y)]. \tag{14.24}$$

Exercise 14.6. Suppose $(\Omega,\mathcal{B},P)$ and $(\Omega',\mathcal{B}',P')$ are two probability spaces, $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ are measurable spaces, $X : \Omega \to \mathbb{X}$, $X' : \Omega' \to \mathbb{X}$, $Y : \Omega \to \mathbb{Y}$,and $Y' : \Omega \to \mathbb{Y}$ are measurable functions such that $P \circ (X,Y)^{-1} = P' \circ (X',Y')$, i.e. $(X,Y) \stackrel{d}{=} (X',Y')$. If $f : (\mathbb{X} \times \mathbb{Y}, \mathcal{M} \otimes \mathcal{N}) \to \mathbb{R}$ is a bounded measurable function and $\tilde{f} : (\mathbb{X},\mathcal{M}) \to \mathbb{R}$ is a measurable function such that $\tilde{f}(X) = \mathbb{E}[f(X,Y)|X]$ P - a.s. then

$$\mathbb{E}'[f(X',Y')|X'] = \tilde{f}(X') \;\; P' \text{ a.s.}$$

Let now suppose that $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{B}$ and let $P_{\mathcal{G}} : \mathcal{B} \to L^1(\Omega,\mathcal{G},P)$ be defined by, $P_{\mathcal{G}}(B) = P(B|\mathcal{G}) := \mathbb{E}_{\mathcal{G}} 1_B \in L^1(\Omega,\mathcal{B},P)$ for all $B \in \mathcal{B}$. If $B = \sum_{n=1}^\infty B_n$ with $B_n \in \mathcal{B}$, then $1_B = \sum_{n=1}^\infty 1_{B_n}$ and this sum converges in $L^1(P)$ (in fact in all $L^p(P)$) by the DCT. Since $\mathbb{E}_{\mathcal{G}} : L^1(\Omega,\mathcal{B},P) \to L^1(\Omega,\mathcal{G},P)$ is a contraction and therefore continuous it follows that

$$P_{\mathcal{G}}(B) = \mathbb{E}_{\mathcal{G}} 1_B = \mathbb{E}_{\mathcal{G}} \sum_{n=1}^\infty 1_{B_n} = \sum_{n=1}^\infty \mathbb{E}_{\mathcal{G}} 1_{B_n} = \sum_{n=1}^\infty P_{\mathcal{G}}(B_n) \tag{14.25}$$

where all equalities are in $L^1(\Omega,\mathcal{G},P)$. Now suppose that we have chosen a representative, $\bar{P}_{\mathcal{G}}(B) : \Omega \to [0,1]$, of $P_{\mathcal{G}}(B)$ for each $B \in \mathcal{B}$. From Eq. (14.25) it follows that

$$\bar{P}_{\mathcal{G}}(B)(\omega) = \sum_{n=1}^\infty \bar{P}_{\mathcal{G}}(B_n)(\omega) \text{ for } P\text{-a.e. } \omega. \tag{14.26}$$

However, **note well**, the exceptional set of ω's depends on the sets $B, B_n \in \mathcal{B}$. The goal of regular conditioning is to carefully choose the representative, $\bar{P}_{\mathcal{G}}(B) : \Omega \to [0,1]$, such that Eq. (14.26) holds for all $\omega \in \Omega$ and all $B, B_n \in \mathcal{B}$ with $B = \sum_{n=1}^\infty B_n$.

Definition 14.17. *If $\mathcal{G}$ is a sub-σ – algebra of $\mathcal{B}$, a* ***regular conditional distribution given*** *$\mathcal{G}$ is a probability kernel on $Q : (\Omega,\mathcal{G}) \times (\Omega,\mathcal{B}) \to [0,1]$ such that*

$$Q(\cdot,B) = P(B|\mathcal{G})(\cdot) \;\; a.s. \text{ for every } B \in \mathcal{B}. \tag{14.27}$$

This corresponds to the Q in Definition 14.15 provided, $(\mathbb{X},\mathcal{M}) = (\Omega,\mathcal{G})$, $(\mathbb{Y},\mathcal{N}) = (\Omega,\mathcal{B})$, and $X(\omega) = Y(\omega) = \omega$ for all $\omega \in \Omega$.

14.2 Additional Properties of Conditional Expectations

The next theorem is devoted to extending the notion of conditional expectations to all non-negative functions and to proving conditional versions of the MCT, DCT, and Fatou's lemma.

Theorem 14.18 (Extending $\mathbb{E}_{\mathcal{G}}$). *If $f : \Omega \to [0,\infty]$ is $\mathcal{B}$ – measurable, there is a $\mathcal{G}$ – measurable function, $F : \Omega \to [0,\infty]$, satisfying*

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}. \tag{14.28}$$

By Lemma 7.24, the function F is uniquely determined up to sets of measure zero and hence we denote any such version of F by $\mathbb{E}_{\mathcal{G}} f$.

1. *Properties 2., 5. (with $0 \le g \in \mathcal{G}_b$), and 6. of Theorem 14.4 still hold for any $\mathcal{B}$ – measurable functions such that $0 \le f \le g$. Namely;*
   a) ***Order Preserving.*** *$\mathbb{E}_{\mathcal{G}} f \le \mathbb{E}_{\mathcal{G}} g$ a.s. when $0 \le f \le g$,*
   b) ***Pull out Property.*** *$\mathbb{E}_{\mathcal{G}}[hf] = h\mathbb{E}_{\mathcal{G}}[f]$ a.s. for all $h \ge 0$ and $\mathcal{G}$ – measurable.*
   c) ***Tower or smoothing property.*** *If $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$. Then*
   $$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \;\; a.s.$$
2. ***Conditional Monotone Convergence (cMCT).*** *Suppose that, almost surely, $0 \le f_n \le f_{n+1}$ for all n, then $\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n = \mathbb{E}_{\mathcal{G}}[\lim_{n\to\infty} f_n]$ a.s.*
3. ***Conditional Fatou's Lemma (cFatou).*** *Suppose again that $0 \le f_n \in L^1(\Omega,\mathcal{B},P)$ a.s., then*
   $$\mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty} f_n\right] \le \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f_n] \;\; a.s. \tag{14.29}$$

Proof. Since $f \wedge n \in L^1(\Omega,\mathcal{B},P)$ and $f \wedge n$ is increasing, it follows that $F :=\uparrow \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n]$ exists a.s. Moreover, by two applications of the standard MCT, we have for any $A \in \mathcal{G}$, that

$$\mathbb{E}[F : A] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}_{\mathcal{G}}[f \wedge n] : A] = \lim_{n\to\infty} \mathbb{E}[f \wedge n : A] = \lim_{n\to\infty} \mathbb{E}[f : A].$$

Thus Eq. (14.28) holds and this uniquely determines F follows from Lemma 7.24.

Item 1. a) If $0 \le f \le g$, then

$$\mathbb{E}_{\mathcal{G}} f = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n] \le \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[g \wedge n] = \mathbb{E}_{\mathcal{G}} g \text{ a.s.}$$

and so $\mathbb{E}_{\mathcal{G}}$ still preserves order. We will prove items 1b and 1c at the end of this proof.

Item 2. Suppose that, almost surely, $0 \le f_n \le f_{n+1}$ for all n, then $\mathbb{E}_{\mathcal{G}} f_n$ is a.s. increasing in n. Hence, again by two applications of the MCT, for any $A \in \mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}\left[\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n : A\right] &= \lim_{n\to\infty} \mathbb{E}[\mathbb{E}_{\mathcal{G}} f_n : A] = \lim_{n\to\infty} \mathbb{E}[f_n : A] \\ &= \mathbb{E}\left[\lim_{n\to\infty} f_n : A\right] = \mathbb{E}\left[\mathbb{E}_{\mathcal{G}}\left[\lim_{n\to\infty} f_n\right] : A\right]\end{aligned}$$

which combined with Lemma 7.24 implies that $\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n = \mathbb{E}_{\mathcal{G}}[\lim_{n\to\infty} f_n]$ a.s.

Item 3. For $0 \le f_n$, let $g_k := \inf_{n \ge k} f_n$. Then $g_k \le f_k$ for all k and $g_k \uparrow \liminf_{n\to\infty} f_n$ and hence by cMCT and item 1.,

$$\mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty} f_n\right] = \lim_{k\to\infty} \mathbb{E}_{\mathcal{G}} g_k \le \liminf_{k\to\infty} \mathbb{E}_{\mathcal{G}} f_k \text{ a.s.}$$

Item 1. b) If $h \ge 0$ is a $\mathcal{G}$ – measurable function and $f \ge 0$, then by cMCT,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}[hf] &\overset{\text{cMCT}}{=} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[(h \wedge n)(f \wedge n)] \\ &= \lim_{n\to\infty} (h \wedge n) \mathbb{E}_{\mathcal{G}}[(f \wedge n)] \overset{\text{cMCT}}{=} h \mathbb{E}_{\mathcal{G}} f \text{ a.s.}\end{aligned}$$

Item 1. c) Similarly by multiple uses of cMCT,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f &= \mathbb{E}_{\mathcal{G}_0} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_1}(f \wedge n) = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1}(f \wedge n) \\ &= \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \mathbb{E}_{\mathcal{G}_0} f\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f &= \mathbb{E}_{\mathcal{G}_1} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0}[f \wedge n] \\ &= \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \mathbb{E}_{\mathcal{G}_0} f.\end{aligned}$$

■

Theorem 14.19 (Conditional Dominated Convergence (cDCT)). *If $f_n \overset{a.s}{\to} f$, and $|f_n| \le g \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ a.s.*

Proof. From Corollary 12.9 we know that $f_n \to f$ in $L^1(P)$ and therefore $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ in $L^1(P)$ as conditional expectation is a contraction on $L^1(P)$. So we need only prove the almost sure convergence. As usual it suffices to consider the real case.

Following the proof of the Dominated convergence theorem, we start with the fact that $0 \le g \pm f_n$ a.s. for all n. Hence by cFatou,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}(g \pm f) &= \mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty}(g \pm f_n)\right] \\ &\le \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(g \pm f_n) = \mathbb{E}_{\mathcal{G}} g + \begin{cases} \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) & \text{in } + \text{ case} \\ -\limsup_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) & \text{in } - \text{ case,}\end{cases}\end{aligned}$$

where the above equations hold a.s. Cancelling $\mathbb{E}_{\mathcal{G}} g$ from both sides of the equation then implies

$$\limsup_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) \le \mathbb{E}_{\mathcal{G}} f \le \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) \text{ a.s.}$$

■

Remark 14.20. Suppose that $f_n \overset{P}{\to} f$, $|f_n| \le g_n \in L^1(\Omega, \mathcal{B}, P)$, $g_n \overset{P}{\to} g \in L^1(\Omega, \mathcal{B}, P)$ and $\mathbb{E} g_n \to \mathbb{E} g$. Then by the DCT in Corollary 12.9, we know that $f_n \to f$ in $L^1(\Omega, \mathcal{B}, P)$. Since $\mathbb{E}_{\mathcal{G}}$ is a contraction, it follows that $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ in $L^1(\Omega, \mathcal{B}, P)$ and hence $\mathbb{E}_{\mathcal{G}} f_n \overset{P}{\to} \mathbb{E}_{\mathcal{G}} f$.

Exercise 14.7. Suppose that $(\mathbb{X}, \mathcal{M})$ and $(\mathbb{Y}, \mathcal{N})$ are measurable spaces, $(\mathbb{Y}, \mathcal{N})$ $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are measurable functions. Further assume that $\mathcal{G}$ is a sub – σ – algebra of $\mathcal{B}$, X is $\mathcal{G}/\mathcal{M}$ – measurable and Y is independent of $\mathcal{G}$. Show for all bounded measurable functions, $f : (\mathbb{X} \times \mathbb{Y}, \mathcal{M} \otimes \mathcal{N}) \to \mathbb{R}$, that

$$\mathbb{E}[f(X, Y) | \mathcal{G}] = h_f(X) = \mathbb{E}[f(x, Y)]|_{x=X} \text{ a.s.}$$

where if $\mu := \mathrm{Law}_P(Y)$,

$$h_f(x) := \mathbb{E}[f(x, Y)] = \int_{\mathbb{Y}} f(x, y)\, d\mu(y). \tag{14.30}$$

Solution to Exercise (14.7). Notice by Fubini's theorem, $h_f(x)$ is $\mathcal{N}/\mathcal{B}_{\mathbb{R}}$ – measurable and therefore $h_f(X)$ is $\mathcal{G}$ – measurable. If $f(x, y) = u(x)v(y)$ where $u : (\mathbb{X}, \mathcal{M}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $v : (\mathbb{Y}, \mathcal{N}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ are measurable functions, then by Lemma 14.10

$$\begin{aligned}\mathbb{E}[f(X, Y) | \mathcal{G}] &= \mathbb{E}[u(X) v(Y) | \mathcal{G}] = u(X) \mathbb{E}[v(Y) | \mathcal{G}] \\ &= u(X) \mathbb{E}[v(Y)] = u(X) \mu(v) = h_f(X) \text{ a.s.}\end{aligned}$$

The proof may now be completed using the multiplicative systems Theorem 8.2. In more detail, let

$$\mathbb{H} := \{f \in [\mathcal{M} \otimes \mathcal{N}]_b : \mathbb{E}[f(X,Y)|\mathcal{G}] = h_f(X) \text{ a.s.}\}.$$

Using the linearity of conditional expectations and of expectations along with the DCT and the cDCT (Theorem 14.19), it is easily seen that $\mathbb{H}$ is a linear subspace which is closed under bounded convergence. Moreover we have just seen that $\mathbb{H}$ contains the multiplicative system of product functions of the form $f(x,y) = u(x)\,v(y)$. Since such functions generate $\mathcal{M} \otimes \mathcal{N}$, it follows that $\mathbb{H}$ consists of all bounded measurable functions on $\mathbb{X} \times \mathbb{Y}$.

The next result in Lemma 14.23 shows how to localize conditional expectations. In order to state and prove the lemma we need a little ground work first.

Definition 14.21. *Suppose that $\mathcal{F}$ and $\mathcal{G}$ are sub-σ-fields of $\mathcal{B}$ and $A \in \mathcal{B}$. We say that $\mathcal{F} = \mathcal{G}$ on A iff $\mathcal{F}_A = \mathcal{G}_A$. Recall that $\mathcal{F}_A = \{B \cap A : B \in \mathcal{F}\}$.*

Lemma 14.22. *If $A \in \mathcal{F} \cap \mathcal{G}$, then $\mathcal{F}_A \cap \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A$ and $\mathcal{F}_A = \mathcal{G}_A$ implies*

$$\mathcal{F}_A = \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A. \tag{14.31}$$

Proof. If $A \in \mathcal{F}$ we have $B \in \mathcal{F}_A$ iff there exists $B' \in \mathcal{F}$ such that $B = A \cap B'$. As $A \in \mathcal{F}$ it follows that $B \in \mathcal{F}$ and therefore we have

$$\mathcal{F}_A = \{B \subset A : B \in \mathcal{F}\}.$$

Thus if $A \in \mathcal{F} \cap \mathcal{G}$ it follows that $\mathcal{F}_A = \{B \subset A : B \in \mathcal{F}\}$ and $\mathcal{G}_A = \{B \subset A : B \in \mathcal{G}\}$ and therefore

$$\mathcal{F}_A \cap \mathcal{G}_A = \{B \subset A : B \in \mathcal{F} \cap \mathcal{G}\} = [\mathcal{F} \cap \mathcal{G}]_A.$$

Equation (14.31) now clearly follows from this identity when $\mathcal{F}_A = \mathcal{G}_A$. ■

Lemma 14.23 (Localizing Conditional Expectations). *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $\mathcal{F}$ and $\mathcal{G}$ be sub-sigma-fileds of $\mathcal{B}$, $X, Y \in L^1(\Omega, \mathcal{B}, P)$ or $X, Y : (\Omega, \mathcal{B}) \to [0, \infty]$ are measurable, and $A \in \mathcal{F} \cap \mathcal{G}$. If $\mathcal{F} = \mathcal{G}$ on A and $X = Y$ a.s. on A, then*

$$\mathbb{E}_\mathcal{F} X = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} X = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} Y = \mathbb{E}_\mathcal{G} Y \text{ a.s. on } A. \tag{14.32}$$

Proof. It suffices to prove, $\mathbb{E}_\mathcal{F} X = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} Y$ a.s. on A and this is equivalent to $1_A \mathbb{E}_\mathcal{F} X = 1_A \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} Y$ a.s. As both sides of this equation are $\mathcal{F}$ – measurable, by the comparison Lemma 7.24 we need to show

$$\mathbb{E}[1_A \mathbb{E}_\mathcal{F} X : B] = \mathbb{E}[1_A \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} Y : B] \text{ for all } B \in \mathcal{F}.$$

So let $B \in \mathcal{F}$ in which case $A \cap B \in \mathcal{F}_A = \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A \subset \mathcal{F} \cap \mathcal{G}$ (see Lemma 14.22). Therefore, using the basic properties of conditional expectations, we have

$$\mathbb{E}[1_A \mathbb{E}_\mathcal{F} X : B] = \mathbb{E}[\mathbb{E}_\mathcal{F} X : A \cap B] = \mathbb{E}[X : A \cap B]$$

and similarly

$$\mathbb{E}[1_A \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} Y : B] = \mathbb{E}[\mathbb{E}_{\mathcal{F}\cap\mathcal{G}} Y : A \cap B] = \mathbb{E}[Y : A \cap B].$$

This completes the proof as $\mathbb{E}[X : A \cap B] = \mathbb{E}[Y : A \cap B]$ because of the assumption that $X = Y$ a.s. on A. ■

Example 14.24. Let us use Lemma 14.23 to show $\mathbb{E}[f(X)|X \wedge t] = f(X) = f(X \wedge t)$ on $\{X < t\}$ – a fact we have already seen to be true in Example 14.12. Let us begin by observing that $\{X < t\} = \{X \wedge t < t\} \in \sigma(X) \cap \sigma(X \wedge t)$. Moreover, using $\sigma(X)_A = \sigma(X|_A)$ for all $A \in \mathcal{B}$,[2] we see that

$$\sigma(X)_{\{X<t\}} = \sigma\left(X|_{\{X<t\}}\right) = \sigma\left((X \wedge t)|_{\{X<t\}}\right) = \sigma(X \wedge t)_{\{X<t\}}.$$

Therefore it follows that

$$\mathbb{E}[f(X)|X \wedge t] = \mathbb{E}[f(X)|\sigma(X \wedge t)] = \mathbb{E}[f(X)|\sigma(X)] = f(X) \text{ a.s. on } \{X < t\}.$$

What goes wrong with the above argument if you replace $\{X < t\}$ by $\{X \le t\}$ everywhere? (Notice that the same argument shows; if $X = Y$ on $A \in \sigma(X) \cap \sigma(Y)$ then $\mathbb{E}[f(X)|Y] = f(Y) = f(X)$ a.s. on A.)

Theorem 14.25 (Conditional Jensen's inequality). *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $-\infty \le a < b \le \infty$, and $\varphi : (a,b) \to \mathbb{R}$ be a convex function. Assume $f \in L^1(\Omega, \mathcal{B}, P; \mathbb{R})$ is a random variable satisfying, $f \in (a,b)$ a.s. and $\varphi(f) \in L^1(\Omega, \mathcal{B}, P; \mathbb{R})$. Then $\varphi(\mathbb{E}_\mathcal{G} f) \in L^1(\Omega, \mathcal{G}, P)$,*

$$\varphi(\mathbb{E}_\mathcal{G} f) \le \mathbb{E}_\mathcal{G}[\varphi(f)] \text{ a.s.} \tag{14.33}$$

and

$$\mathbb{E}[\varphi(\mathbb{E}_\mathcal{G} f)] \le \mathbb{E}[\varphi(f)] \tag{14.34}$$

Proof. Let $\Lambda := \mathbb{Q} \cap (a,b)$ – a countable dense subset of (a,b). By Theorem 12.55 (also see Lemma 12.52) and Figure 12.5 when φ is C^1)

$$\varphi(y) \ge \varphi(x) + \varphi'_-(x)(y - x) \text{ for all for all } x, y \in (a,b), \tag{14.35}$$

[2] Here is the verification that $\sigma(X)_A = \sigma(X|_A)$. Let $i_A : A \to \Omega$ be the inclusion map. Since $\sigma(X) = X^{-1}(\mathcal{B}_\mathbb{R})$ and $\sigma(X)_A = i_A^{-1}\sigma(X)$ it follows that

$$\begin{aligned}\sigma(X)_A &= i_A^{-1}\left(X^{-1}(\mathcal{B}_\mathbb{R})\right) = (X \circ i_A)^{-1}(\mathcal{B}_\mathbb{R}) \\ &= \sigma(X \circ i_A) = \sigma(X|_A).\end{aligned}$$

where $\varphi'_-(x)$ is the left hand derivative of φ at x. Taking $y = f$ and then taking conditional expectations imply,

$$\mathbb{E}_{\mathcal{G}}[\varphi(f)] \ge \mathbb{E}_{\mathcal{G}}\left[\varphi(x) + \varphi'_-(x)(f - x)\right] = \varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x) \text{ a.s.} \quad (14.36)$$

Since this is true for all $x \in (a, b)$ (and hence all x in the countable set, Λ) we may conclude that

$$\mathbb{E}_{\mathcal{G}}[\varphi(f)] \ge \sup_{x \in \Lambda}\left[\varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x)\right] \text{ a.s.}$$

By Exercise 14.1, $\mathbb{E}_{\mathcal{G}} f \in (a, b)$, and hence it follows from Corollary 12.56 that

$$\sup_{x \in \Lambda}\left[\varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x)\right] = \varphi(\mathbb{E}_{\mathcal{G}} f) \text{ a.s.}$$

Combining the last two estimates proves Eq. (14.33).

From Eq. (14.33) and Eq. (14.35) with $y = \mathbb{E}_{\mathcal{G}} f$ and $x \in (a, b)$ fixed we find,

$$\varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x) \le \varphi(\mathbb{E}_{\mathcal{G}} f) \le \mathbb{E}_{\mathcal{G}}[\varphi(f)]. \quad (14.37)$$

Therefore

$$|\varphi(\mathbb{E}_{\mathcal{G}} f)| \le |\mathbb{E}_{\mathcal{G}}[\varphi(f)]| \vee \left|\varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x)\right| \in L^1(\Omega, \mathcal{G}, P) \quad (14.38)$$

which implies that $\varphi(\mathbb{E}_{\mathcal{G}} f) \in L^1(\Omega, \mathcal{G}, P)$. Taking expectations of Eq. (14.33) is now allowed and immediately gives Eq. (14.34). ■

Remark 14.26 (On Theorem 14.25 and its proof.). *From Eq. (14.35),

$$\varphi(f) \ge \varphi(\mathbb{E}_{\mathcal{G}} f) + \varphi'_-(\mathbb{E}_{\mathcal{G}} f)(f - \mathbb{E}_{\mathcal{G}} f). \quad (14.39)$$

Therefore taking $\mathbb{E}_{\mathcal{G}}$ of this equation "implies" that

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}[\varphi(f)] \ge& \varphi(\mathbb{E}_{\mathcal{G}} f) + \mathbb{E}_{\mathcal{G}}\left[\varphi'_-(\mathbb{E}_{\mathcal{G}} f)(f - \mathbb{E}_{\mathcal{G}} f)\right] \\ &= \varphi(\mathbb{E}_{\mathcal{G}} f) + \varphi'_-(\mathbb{E}_{\mathcal{G}} f)\mathbb{E}_{\mathcal{G}}[(f - \mathbb{E}_{\mathcal{G}} f)] = \varphi(\mathbb{E}_{\mathcal{G}} f). \quad (14.40)\end{aligned}$$

The technical problem with this argument is the justification that $\mathbb{E}_{\mathcal{G}}\left[\varphi'_-(\mathbb{E}_{\mathcal{G}} f)(f - \mathbb{E}_{\mathcal{G}} f)\right] = \varphi'_-(\mathbb{E}_{\mathcal{G}} f)\mathbb{E}_{\mathcal{G}}[(f - \mathbb{E}_{\mathcal{G}} f)]$ since there is no reason for φ'_- to be a bounded function. The proof we give in Theorem 14.25 circumvents this technical detail.

On the other hand let us now suppose that φ is $C^1(\mathbb{R})$ is convex and for the moment that $|f| \le M < \infty$ a.s. Then $\mathbb{E}_{\mathcal{G}} f \in [-M, M]$ a.s. and hence $\varphi'_-(\mathbb{E}_{\mathcal{G}} f) = \varphi'(\mathbb{E}_{\mathcal{G}} f)$ is bounded and Eq. (14.40) is now valid. Moreover, taking $x = 0$ in Eq. (14.37) shows

$$\varphi(0) + \varphi'(0)\mathbb{E}_{\mathcal{G}} f \le \varphi(\mathbb{E}_{\mathcal{G}} f) \le \mathbb{E}_{\mathcal{G}}[\varphi(f)].$$

If f is unbounded we may apply the above inequality with f replaced by $f_M := f \cdot 1_{|f| \le M}$ in order to conclude,

$$\varphi(0) + \varphi'(0)\mathbb{E}_{\mathcal{G}} f_M \le \varphi(\mathbb{E}_{\mathcal{G}} f_M) \le \mathbb{E}_{\mathcal{G}}[\varphi(f_M)].$$

If we further assume that $\varphi(f_M) \ge 0$ is increasing as M increase (for example this is the case if $\varphi(x) = |x|^p$ for some $p > 1$) we may conclude by passing to the limit along a nicely chosen subsequence that

$$\varphi(0) + \varphi'(0)\mathbb{E}_{\mathcal{G}} f \le \varphi(\mathbb{E}_{\mathcal{G}} f) \le \mathbb{E}_{\mathcal{G}}[\varphi(f)]$$

where we used $\mathbb{E}_{\mathcal{G}}[\varphi(f_M)] \to \mathbb{E}_{\mathcal{G}}[\varphi(f)]$ by cMCT.

Corollary 14.27. *The conditional expectation operator, $\mathbb{E}_{\mathcal{G}}$ maps $L^p(\Omega, \mathcal{B}, P)$ into $L^p(\Omega, \mathcal{B}, P)$ and the map remains a contraction for all $1 \le p \le \infty$.*

Proof. The case $p = \infty$ and $p = 1$ have already been covered in Theorem 14.4. So now suppose, $1 < p < \infty$, and apply Jensen's inequality with $\varphi(x) = |x|^p$ to find $|\mathbb{E}_{\mathcal{G}} f|^p \le \mathbb{E}_{\mathcal{G}}|f|^p$ a.s. Taking expectations of this inequality gives the desired result. ■

Exercise 14.8 (Martingale Convergence Theorem for $p = 1$ and 2."). Let $(\Omega, \mathcal{B}, P)$ be a probability space and $\{\mathcal{B}_n\}_{n=1}^\infty$ be an increasing sequence of sub-σ-algebras of $\mathcal{B}$. Show;

1. The closure, M, of $\cup_{n=1}^\infty L^2(\Omega, \mathcal{B}_n, P)$ is $L^2(\Omega, \mathcal{B}_\infty, P)$ where $\mathcal{B}_\infty = \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$. **Hint:** make use of Theorem 12.29.
2. For every $X \in L^2(\Omega, \mathcal{B}, P)$, $X_n := \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^2(P)$. **Hint:** see Exercise 13.5.
3. For every $X \in L^1(\Omega, \mathcal{B}, P)$, $X_n := \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^1(P)$. **Hint:** make use of item 2. by a truncation argument using the contractive properties of conditional expectations.

(Eventually we will show that $X_n = \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ a.s. as well.)

Exercise 14.9 (Martingale Convergence Theorem for general p). Let $1 \le p < \infty$, $(\Omega, \mathcal{B}, P)$ be a probability space, and $\{\mathcal{B}_n\}_{n=1}^\infty$ be an increasing sequence of sub-σ-algebras of $\mathcal{B}$. Show for all $X \in L^p(\Omega, \mathcal{B}, P)$, $X_n := \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^p(P)$. (**Hint:** show that $\{|\mathbb{E}[X|\mathcal{B}_n]|^p\}_{n=1}^\infty$ is uniformly integrable and $\mathbb{E}[X|\mathcal{B}_n] \xrightarrow{P} \mathbb{E}[X|\mathcal{B}_\infty]$ with the aid of item 3. of Exercise 14.8.)

14.3 Construction of Regular Conditional Distributions*

Lemma 14.28. *Suppose that $h : \mathbb{Q} \to [0,1]$ is an increasing (i.e. non-decreasing) function and $H(t) := \inf\{h(s) : t < s \in \mathbb{Q}\}$ for all $t \in \mathbb{R}$. Then $H : \mathbb{R} \to [0,1]$ is an increasing right continuous function.*

Proof. If $t_1 < t_2$, then

$$\{h(s) : t_1 < s \in \mathbb{Q}\} \subset \{h(s) : t_2 < s \in \mathbb{Q}\}$$

and therefore $H(t_1) \leq H(t_2)$. Let $H(t+) := \lim_{\tau \downarrow t} H(\tau)$. Then for any $s \in \mathbb{Q}$ with $s > t$ we have $H(t) \leq H(t+) \leq h(s)$ and then taking the infimum over such s we learn that $H(t) \leq H(t+) \leq H(t)$, i.e. $H(t+) = H(t)$. ■

Lemma 14.29. *Suppose that $(\mathbb{X}, \mathcal{M})$ is a measurable space and $F : \mathbb{X} \times \mathbb{R} \to \mathbb{R}$ is a function such that; 1) $F(\cdot, t) : \mathbb{X} \to \mathbb{R}$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $t \in \mathbb{R}$, and 2) $F(x, \cdot) : \mathbb{R} \to \mathbb{R}$ is right continuous for all $x \in \mathbb{X}$. Then F is $\mathcal{M} \otimes \mathcal{B}_{\mathbb{R}}/\mathcal{B}_{\mathbb{R}}$ – measurable.*

Proof. For $n \in \mathbb{N}$, the function,

$$F_n(x,t) := \sum_{k=-\infty}^{\infty} F\left(x, (k+1) 2^{-n}\right) 1_{(k2^{-n},(k+1)2^{-n}]}(t),$$

is $\mathcal{M} \otimes \mathcal{B}_{\mathbb{R}}/\mathcal{B}_{\mathbb{R}}$ – measurable. Using the right continuity assumption, it follows that $F(x,t) = \lim_{n\to\infty} F_n(x,t)$ for all $(x,t) \in \mathbb{X} \times \mathbb{R}$ and therefore F is also $\mathcal{M} \otimes \mathcal{B}_{\mathbb{R}}/\mathcal{B}_{\mathbb{R}}$ – measurable. ■

Proposition 14.30. *Let $\mathcal{B}_{\mathbb{R}}$ be the Borel σ – algebra on $\mathbb{R}$. Then $\mathcal{B}_{\mathbb{R}}$ contains a countable sub-algebra, $\mathcal{A}_{\mathbb{R}} \subset \mathcal{B}_{\mathbb{R}}$, which generates $\mathcal{B}_{\mathbb{R}}$ and has the amazing property that every finitely additive probability measure on $\mathcal{A}_{\mathbb{R}}$ extends uniquely to a countably additive probability measure on $\mathcal{B}_{\mathbb{R}}$.*

Proof. By the results in Appendix 9.10, we know that $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measure theoretically isomorphic to $\left(\{0,1\}^{\mathbb{N}}, \mathcal{F}\right)$ where $\mathcal{F}$ is the product σ – algebra. As we saw in Section 5.5, $\mathcal{F}$ is generated by the countable algebra, $\mathcal{A} := \cup_{n=1}^{\infty} \mathcal{A}_n$ where

$$\mathcal{A}_n := \{B \times \Omega : B \subset \{0,1\}^n\} \text{ for all } n \in \mathbb{N}.$$

According to the baby Kolmogorov Theorem 5.40, any finitely additive probability measure on $\mathcal{A}$ has a unique extension to a probability measure on $\mathcal{F}$. The algebra $\mathcal{A}$ may now be transferred by the measure theoretic isomorphism to the desired sub-algebra, $\mathcal{A}_{\mathbb{R}}$, of $\mathcal{B}_{\mathbb{R}}$. ■

Theorem 14.31. *Suppose that $(\mathbb{X}, \mathcal{M})$ is a measurable space, $X : \Omega \to \mathbb{X}$ is a measurable function and $Y : \Omega \to \mathbb{R}$ is a random variable. Then there exists a probability kernel, Q, on $\mathbb{X} \times \mathbb{R}$ such that $\mathbb{E}[f(Y)|X] = Q(X, f)$, P – a.s., for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$.*

Proof. First proof. For each $r \in \mathbb{Q}$, let $q_r : \mathbb{X} \to [0,1]$ be a measurable function such that

$$\mathbb{E}[1_{Y \leq r}|X] = q_r(X) \text{ a.s.}$$

Let $\nu := P \circ X^{-1}$ be the law of X. Then using the basic properties of conditional expectation, $q_r \leq q_s$ ν – a.s. for all $r \leq s$, $\lim_{r\uparrow\infty} q_r = 1$ and $\lim_{r\downarrow\infty} q_r = 0$, ν – a.s. Hence the set, $\mathbb{X}_0 \subset \mathbb{X}$ where $q_r(x) \leq q_s(x)$ for all $r \leq s$, $\lim_{r\uparrow\infty} q_r(x) = 1$, and $\lim_{r\downarrow\infty} q_r(x) = 0$ satisfies, $\nu(\mathbb{X}_0) = P(X \in \mathbb{X}_0) = 1$. For $t \in \mathbb{R}$, let

$$F(x,t) := 1_{\mathbb{X}_0}(x) \cdot \inf\{q_r(x) : r > t\} + 1_{\mathbb{X}\setminus\mathbb{X}_0}(x) \cdot 1_{t\geq 0}.$$

Then $F(\cdot, t) : \mathbb{X} \to \mathbb{R}$ is measurable for each $t \in \mathbb{R}$ and by Lemma 14.28, $F(x, \cdot)$ is a distribution function on $\mathbb{R}$ for each $x \in \mathbb{X}$. Hence an application of Lemma 14.29 shows $F : \mathbb{X} \times \mathbb{R} \to [0,1]$ is measurable.

For each $x \in \mathbb{X}$ and $B \in \mathcal{B}_{\mathbb{R}}$, let $Q(x, B) = \mu_{F(x,\cdot)}(B)$ where μ_F denotes the probability measure on $\mathbb{R}$ determined by a distribution function, $F : \mathbb{R} \to [0,1]$.

We will now show that Q is the desired probability kernel. To prove this, let $\mathbb{H}$ be the collection of bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$, such that $\mathbb{X} \ni x \to Q(x,f) \in \mathbb{R}$ is measurable and $\mathbb{E}[f(Y)|X] = Q(X,f)$, P – a.s. It is easily seen that $\mathbb{H}$ is a linear subspace which is closed under bounded convergence. We will finish the proof by showing that $\mathbb{H}$ contains the multiplicative class, $\mathbb{M} = \left\{1_{(-\infty,t]} : t \in \mathbb{R}\right\}$ so that multiplicative systems Theorem 8.2 may be applied.

Notice that $Q\left(x, 1_{(-\infty,t]}\right) = F(x,t)$ is measurable. Now let $r \in \mathbb{Q}$ and $g : \mathbb{X} \to \mathbb{R}$ be a bounded measurable function, then

$$\begin{aligned}\mathbb{E}[1_{Y\leq r} \cdot g(X)] &= \mathbb{E}[\mathbb{E}[1_{Y\leq r}|X]\, g(X)] = \mathbb{E}[q_r(X)\, g(X)] \\ &= \mathbb{E}[q_r(X)\, 1_{\mathbb{X}_0}(X)\, g(X)].\end{aligned}$$

For $t \in \mathbb{R}$, we may let $r \downarrow t$ in the above equality (use DCT) to learn,

$$\mathbb{E}[1_{Y\leq t} \cdot g(X)] = \mathbb{E}[F(X,t)\, 1_{\mathbb{X}_0}(X)\, g(X)] = \mathbb{E}[F(X,t)\, g(X)].$$

Since g was arbitrary, we may conclude that

$$Q\left(X, 1_{(-\infty,t]}\right) = F(X,t) = \mathbb{E}[1_{Y\leq t}|X] \text{ a.s.}$$

This completes the proof.

Second proof. Let $\mathcal{A} := \mathcal{A}_{\mathbb{R}}$ be the algebra described in Proposition 14.30. For each $A \in \mathcal{A}$, let $\mu_A : \mathbb{X} \to \mathbb{R}$ be a measurable function such that $\mu_A(X) = P(Y \in A|X)$ a.s. If $A = A_1 \cup A_2$ with $A_i \in \mathcal{A}$ and $A_1 \cap A_2 = \emptyset$, then

$$\mu_{A_1}(X) + \mu_{A_2}(X) = P(Y \in A_1|X) + P(Y \in A_2|X)$$
$$= P(Y \in A_1 \cup A_2|X) = \mu_{A_1+A_2}(X) \text{ a.s.}$$

Thus if $\nu := \mathrm{Law}_P(X)$, we have $\mu_{A_1}(x) + \mu_{A_2}(x) = \mu_{A_1+A_2}(x)$ for ν – a.e. x. Since

$$\mu_{\mathbb{R}}(X) = P(Y \in \mathbb{R}|X) = 1 \text{ a.s.}$$

we know that $\mu_{\mathbb{R}}(x) = 1$ for ν – a.e. x.

Thus if we let X_0 denote those $x \in X$ such that $\mu_{\mathbb{R}}(x) = 1$ and $\mu_{A_1}(x) + \mu_{A_2}(x) = \mu_{A_1+A_2}(x)$ for all disjoint pairs, $(A_1, A_2) \in \mathcal{A}^2$, we have $\nu(X_0) = 1$ and $\mathcal{A} \ni A \to Q_0(x, A) := \mu_A(x)$ is a finitely additive probability measure on $\mathcal{A}$. According to Proposition 14.30, $Q_0(x, \cdot)$ extends to a probability measure, $Q(x, \cdot)$ on $\mathcal{B}_{\mathbb{R}}$ for all $x \in X_0$. For $x \notin X_0$ we let $Q_0(x, \cdot) = \delta_0$ where $\delta_0(B) = 1_B(0)$ for all $B \in \mathcal{B}_{\mathbb{R}}$.

We will now show that Q is the desired probability kernel. To prove this, let $\mathbb{H}$ be the collection of bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$, such that $\mathbb{X} \ni x \to Q(x, f) \in \mathbb{R}$ is measurable and $\mathbb{E}[f(Y)|X] = Q(X, f)$, P – a.s. By construction, $\mathbb{H}$ contains the multiplicative system, $\{1_A : A \in \mathcal{A}\}$. Moreover it is easily seen that $\mathbb{H}$ is a linear subspace which is closed under bounded convergence. Therefore by the multiplicative systems Theorem 8.2, $\mathbb{H}$ consists of all bounded measurable functions on $\mathbb{R}$. ∎

This result leads fairly immediately to the following far reaching generalization.

Theorem 14.32. *Suppose that $(\mathbb{X}, \mathcal{M})$ is a measurable space and $(\mathbb{Y}, \mathcal{N})$ is a standard Borel space*[3]*, see Appendix 9.10. Suppose that $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are measurable functions. Then there exists a probability kernel, Q, on $\mathbb{X} \times \mathbb{Y}$ such that $\mathbb{E}[f(Y)|X] = Q(X, f)$, P – a.s., for all bounded measurable functions, $f : \mathbb{Y} \to \mathbb{R}$.*

Proof. By definition of a standard Borel space, we may assume that $\mathbb{Y} \in \mathcal{B}_{\mathbb{R}}$ and $\mathcal{N} = \mathcal{B}_{\mathbb{Y}}$. In this case Y may also be viewed to be a measurable map form $\Omega \to \mathbb{R}$ such that $Y(\Omega) \subset \mathbb{Y}$. By Theorem 14.31, we may find a probability kernel, Q_0, on $\mathbb{X} \times \mathbb{R}$ such that

$$\mathbb{E}[f(Y)|X] = Q_0(X, f), \ P - a.s., \tag{14.41}$$

for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$.

Taking $f = 1_{\mathbb{Y}}$ in Eq. (14.41) shows

$$1 = \mathbb{E}[1_{\mathbb{Y}}(Y)|X] = Q_0(X, \mathbb{Y}) \text{ a.s..}$$

[3] According to the counter example in Doob [12, p. 624], it is not sufficient to assume that $\mathcal{N}$ is countably generated!

Thus if we let $\mathbb{X}_0 := \{x \in \mathbb{X} : Q_0(x, \mathbb{Y}) = 1\}$, we know that $P(X \in \mathbb{X}_0) = 1$. Let us now define

$$Q(x, B) := 1_{\mathbb{X}_0}(x) Q_0(x, B) + 1_{\mathbb{X} \setminus \mathbb{X}_0}(x) \delta_y(B) \text{ for } (x, B) \in \mathbb{X} \times \mathcal{B}_{\mathbb{Y}},$$

where y is an arbitrary but fixed point in Y. Then and hence Q is a probability kernel on $\mathbb{X} \times \mathbb{Y}$. Moreover if $B \in \mathcal{B}_{\mathbb{Y}} \subset \mathcal{B}_{\mathbb{R}}$, then

$$Q(X, B) = 1_{\mathbb{X}_0}(X) Q_0(X, B) = 1_{\mathbb{X}_0}(X) \mathbb{E}[1_B(Y)|X] = \mathbb{E}[1_B(Y)|X] \text{ a.s.}$$

This shows that Q is the desired regular conditional probability. ∎

Corollary 14.33. *Suppose $\mathcal{G}$ is a sub-σ – algebra of $\mathcal{B}$, $(\mathbb{Y}, \mathcal{N})$ is a standard Borel space, and $Y : \Omega \to \mathbb{Y}$ is a measurable function. Then there exists a probability kernel, Q, on $(\Omega, \mathcal{G}) \times (\mathbb{Y}, \mathcal{N})$ such that $\mathbb{E}[f(Y)|\mathcal{G}] = Q(\cdot, f)$, P - a.s. for all bounded measurable functions, $f : \mathbb{Y} \to \mathbb{R}$.*

Proof. This is a special case of Theorem 14.32 applied with $(\mathbb{X}, \mathcal{M}) = (\Omega, \mathcal{G})$ and $X : \Omega \to \Omega$ being the identity map which is $\mathcal{B}/\mathcal{G}$ – measurable. ∎

Corollary 14.34. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space such that $(\Omega, \mathcal{B})$ is a standard Borel space and $\mathcal{G}$ is a sub-σ – algebra $\mathcal{B}$. Then there exists a probability kernel, Q on $(\Omega, \mathcal{G}) \times (\Omega, \mathcal{B})$ such that $\mathbb{E}[Z|\mathcal{G}] = Q(\cdot, Z)$, P - a.s. for all bounded $\mathcal{B}$ – measurable random variables, $Z : \Omega \to \mathbb{R}$.*

Proof. This is a special case of Corollary 14.33 with $(\mathbb{Y}, \mathcal{N}) = (\Omega, \mathcal{B})$ and $Y : \Omega \to \Omega$ being the identity map which is $\mathcal{B}/\mathcal{B}$ – measurable. ∎

Remark 14.35. It turns out that every standard Borel space $(\mathbb{X}, \mathcal{M})$ possess a countable sub-algebra $\mathcal{A}$ generating $\mathcal{M}$ with the property that every finitely additive probability measure on $\mathcal{A}$ extends to a probability measure on $\mathcal{M}$, see [6]. With this in hand, the second proof of Theorem 14.31 extends easily to give another proof of Theorem 14.32 all in one go. As the next example shows it is a bit tricky to produce the algebra $\mathcal{A}$.

Example 14.36. Let $\Omega := \{0,1\}^{\mathbb{N}}$, $\pi_i : \Omega \to \{0,1\}$ be projection onto the i^{th} component and $\mathcal{B} := \sigma(\pi_1, \pi_2, \dots)$ be the product σ – algebra on Ω. Further let $\mathcal{A} := \cup_{n=1}^{\infty} \mathcal{A}_n$ where

$$\mathcal{A}_n := \{B \times \Omega : B \subset \{0,1\}^n\} \text{ for all } n \in \mathbb{N}.$$

Suppose that $X = \{e_n\}_{n=1}^{\infty} \subset \Omega$ where $e_n(i) = \delta_{in}$ for $i, n \in \mathbb{N}$. I now claim that

$$\mathcal{A}_X = \{A \subset X : \#(A) < \infty \text{ or } \#(A^c) < \infty\} =: \mathcal{C}$$

is the so called cofinite σ – algebra. To see this observe that $\mathcal{A}$ is generated by sets of the form $\{\pi_i = 1\}$ for $i \in \mathbb{N}$. Therefore $\mathcal{A}_X$ is generated by sets of the form $\{\pi_i = 1\}_X = \{e_i\}$. But these one point sets are easily seen to generate $\mathcal{C}$.

Now suppose that $\lambda : X \to [0,1]$ is a function such that $Z := \sum_{n\in\mathbb{N}} \lambda(e_n) \in (0,1)$ and let $\mu(B) := \sum_{a\in B} \lambda(a)$ for all $B \subset X$. Then μ is a measure on 2^X with $\mu(X) = Z < 1$.

Using this measure μ, we may define $P_0 : \mathcal{A}_X = \mathcal{C} \to [0,1]$ by,

$$P_0(A) := \begin{cases} \mu(A) & \text{if } \#(A) < \infty \\ 1 - \mu(A^c) & \text{if } \#(A^c) < \infty \end{cases}.$$

I claim that P_0 is a finitely additive probability measure on $\mathcal{A}_X = \mathcal{C}$ which has no -extension to a probability measure on 2^X. To see that P_0 is finitely additive, let $A, B \in \mathcal{C}$ be disjoint sets. If both A and B are finite sets, then

$$P_0(A \cup B) = \mu(A \cup B) = \mu(A) + \mu(B) = P_0(A) + P_0(B).$$

If one of the sets is an infinite set, say B, then $\#(B^c) < \infty$ and $\#(A) < \infty$ for otherwise $A \cap B \neq \emptyset$. As $A \cap B = \emptyset$ we know that $A \subset B^c$ and therefore,

$$\begin{aligned} P_0(A \cup B) &= 1 - \mu([A \cup B]^c) = 1 - \mu(A^c \cap B^c) \\ &= 1 - \mu(B^c \setminus A) = 1 - (\mu(B^c) - \mu(A)) \\ &= 1 - \mu(B^c) + \mu(A) = P_0(B) + P_0(A). \end{aligned}$$

Thus we have shown that $P_0 : \mathcal{A}_X \to [0,1]$ is a finitely additive probability measure. If P were a countably additive extension of P_0, we would have to have,

$$\begin{aligned} 1 = P_0(X) = P(X) &= \sum_{n=1}^{\infty} P(\{e_n\}) \\ = \sum_{n=1}^{\infty} P_0(\{e_n\}) &= \sum_{n=1}^{\infty} \mu(\{e_n\}) = Z < 1 \end{aligned}$$

which is clearly a contradiction.

There is however a way to fix this example as shown in [6]. It is to replace $\mathcal{A}_X$ in this example by the algebra, $\mathcal{A}$, generated by $\mathcal{E} := \{\{n\} : n \geq 2\}$. This algebra may be described as those $A \subset \mathbb{N}$ such that either $A \subset\subset \{2, 3, \dots\}$ for $1 \in A$ and $\#(A^c) < \infty$. Thus if $A_k \in \mathcal{A}$ with $A_k \downarrow \emptyset$ we must have that $1 \notin A_k$ for k large and therefore $\#(A_k) < \infty$ for k large. Moreover $\#(A_k)$ is decreasing in k. If $\lim_{k\to\infty} \#(A_k) = m > 0$, we must have that $A_k = A_l$ for all k, l large and therefore $\cap A_k \neq \emptyset$. Thus we must conclude that $A_k = \emptyset$ for large k. We therefore may conclude that any finitely additive probability measure, P_0, on $\mathcal{A}$ has a unique extension to a probability measure on $\sigma(\mathcal{A}) = 2^{\mathbb{N}}$.

15

The Radon-Nikodym Theorem

Theorem 15.1 (A Baby Radon-Nikodym Theorem). *Suppose $(X,\mathcal{M})$ is a measurable space, λ and ν are two finite positive measures on $\mathcal{M}$ such that $\nu(A)\le\lambda(A)$ for all $A\in\mathcal{M}$. Then there exists a measurable function, $\rho:X\to[0,1]$ such that $d\nu=\rho d\lambda$.*

Proof. If f is a non-negative simple function, then

$$\nu(f)=\sum_{a\ge 0}a\nu(f=a)\le\sum_{a\ge 0}a\lambda(f=a)=\lambda(f).$$

In light of Theorem 6.39 and the MCT, this inequality continues to hold for all non-negative measurable functions. Furthermore if $f\in L^1(\lambda)$, then $\nu(|f|)\le\lambda(|f|)<\infty$ and hence $f\in L^1(\nu)$ and

$$|\nu(f)|\le\nu(|f|)\le\lambda(|f|)\le\lambda(X)^{1/2}\cdot\|f\|_{L^2(\lambda)}.$$

Therefore, $L^2(\lambda)\ni f\to\nu(f)\in\mathbb{C}$ is a continuous linear functional on $L^2(\lambda)$. By the Riesz representation Theorem 13.15, there exists a unique $\rho\in L^2(\lambda)$ such that

$$\nu(f)=\int_X f\rho d\lambda \text{ for all } f\in L^2(\lambda).$$

In particular this equation holds for all bounded measurable functions, $f:X\to\mathbb{R}$ and for such a function we have

$$\nu(f)=\operatorname{Re}\nu(f)=\operatorname{Re}\int_X f\rho d\lambda=\int_X f\operatorname{Re}\rho\, d\lambda. \tag{15.1}$$

Thus by replacing ρ by $\operatorname{Re}\rho$ if necessary we may assume ρ is real.

Taking $f=1_{\rho<0}$ in Eq. (15.1) shows

$$0\le\nu(\rho<0)=\int_X 1_{\rho<0}\rho\, d\lambda\le 0,$$

from which we conclude that $1_{\rho<0}\rho=0$, λ – a.e., i.e. $\lambda(\rho<0)=0$. Therefore $\rho\ge 0$, λ – a.e. Similarly for $\alpha>1$,

$$\lambda(\rho>\alpha)\ge\nu(\rho>\alpha)=\int_X 1_{\rho>\alpha}\rho\, d\lambda\ge\alpha\lambda(\rho>\alpha)$$

which is possible iff $\lambda(\rho>\alpha)=0$. Letting $\alpha\downarrow 1$, it follows that $\lambda(\rho>1)=0$ and hence $0\le\rho\le 1$, λ - a.e. ■

Definition 15.2. *Let μ and ν be two positive measure on a measurable space, $(X,\mathcal{M})$. Then:*

1. *μ and ν are **mutually singular** (written as $\mu\perp\nu$) if there exists $A\in\mathcal{M}$ such that $\nu(A)=0$ and $\mu(A^c)=0$. We say that ν lives on A and μ lives on A^c.*
2. *The measure ν **is absolutely continuous relative to** μ (written as $\nu\ll\mu$) provided $\nu(A)=0$ whenever $\mu(A)=0$.*

As an example, suppose that μ is a positive measure and $\rho\ge 0$ is a measurable function. Then the measure, $\nu:=\rho\mu$ is absolutely continuous relative to μ. Indeed, if $\mu(A)=0$ then

$$\nu(A)=\int_A\rho d\mu=0.$$

We will eventually show that if μ and ν are $\sigma-$finite and $\nu\ll\mu$, then $d\nu=\rho d\mu$ for some measurable function, $\rho\ge 0$.

Definition 15.3 (Lebesgue Decomposition). *Let μ and ν be two positive measure on a measurable space, $(X,\mathcal{M})$. Two positive measures ν_a and ν_s form a **Lebesgue decomposition** of ν relative to μ if $\nu=\nu_a+\nu_s$, $\nu_a\ll\mu$, and $\nu_s\perp\mu$.*

Lemma 15.4. *If μ_1, μ_2 and ν are positive measures on $(X,\mathcal{M})$ such that $\mu_1\perp\nu$ and $\mu_2\perp\nu$, then $(\mu_1+\mu_2)\perp\nu$. More generally if $\{\mu_i\}_{i=1}^\infty$ is a sequence of positive measures such that $\mu_i\perp\nu$ for all i then $\mu=\sum_{i=1}^\infty\mu_i$ is singular relative to ν.*

Proof. It suffices to prove the second assertion since we can then take $\mu_j\equiv 0$ for all $j\ge 3$. Choose $A_i\in\mathcal{M}$ such that $\nu(A_i)=0$ and $\mu_i(A_i^c)=0$ for all i. Letting $A:=\cup_i A_i$ we have $\nu(A)=0$. Moreover, since $A^c=\cap_i A_i^c\subset A_m^c$ for all m, we have $\mu_i(A^c)=0$ for all i and therefore, $\mu(A^c)=0$. This shows that $\mu\perp\nu$. ■

Lemma 15.5. *Let ν and μ be positive measures on $(X,\mathcal{M})$. If there exists a Lebesgue decomposition, $\nu=\nu_s+\nu_a$, of the measure ν relative to μ then this decomposition is unique. Moreover: if ν is a σ – finite measure then so are ν_s and ν_a.*

Proof. Since $\nu_s \perp \mu$, there exists $A \in \mathcal{M}$ such that $\mu(A) = 0$ and $\nu_s(A^c) = 0$ and because $\nu_a \ll \mu$, we also know that $\nu_a(A) = 0$. So for $C \in \mathcal{M}$,

$$\nu(C \cap A) = \nu_s(C \cap A) + \nu_a(C \cap A) = \nu_s(C \cap A) = \nu_s(C) \tag{15.2}$$

and

$$\nu(C \cap A^c) = \nu_s(C \cap A^c) + \nu_a(C \cap A^c) = \nu_a(C \cap A^c) = \nu_a(C). \tag{15.3}$$

Now suppose we have another Lebesgue decomposition, $\nu = \tilde{\nu}_a + \tilde{\nu}_s$ with $\tilde{\nu}_s \perp \mu$ and $\tilde{\nu}_a \ll \mu$. Working as above, we may choose $\tilde{A} \in \mathcal{M}$ such that $\mu(\tilde{A}) = 0$ and $\tilde{A}^c$ is $\tilde{\nu}_s$ – null. Then $B = A \cup \tilde{A}$ is still a μ – null set and and $B^c = A^c \cap \tilde{A}^c$ is a null set for both ν_s and $\tilde{\nu}_s$. Therefore we may use Eqs. (15.2) and (15.3) with A being replaced by B to conclude,

$$\nu_s(C) = \nu(C \cap B) = \tilde{\nu}_s(C) \text{ and}$$
$$\nu_a(C) = \nu(C \cap B^c) = \tilde{\nu}_a(C) \text{ for all } C \in \mathcal{M}.$$

Lastly if ν is a σ – finite measure then there exists $X_n \in \mathcal{M}$ such that $X = \sum_{n=1}^\infty X_n$ and $\nu(X_n) < \infty$ for all n. Since $\infty > \nu(X_n) = \nu_a(X_n) + \nu_s(X_n)$, we must have $\nu_a(X_n) < \infty$ and $\nu_s(X_n) < \infty$, showing ν_a and ν_s are σ – finite as well. ∎

Lemma 15.6. *Suppose μ is a positive measure on $(X, \mathcal{M})$ and $f, g : X \to [0, \infty]$ are functions such that the measures, $f d\mu$ and $g d\mu$ are σ – finite and further satisfy,*

$$\int_A f d\mu = \int_A g d\mu \text{ for all } A \in \mathcal{M}. \tag{15.4}$$

Then $f(x) = g(x)$ for μ – a.e. x. (BRUCE: this lemma is very closely related to Lemma 7.24 above.)

Proof. By assumption there exists $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\int_{X_n} f d\mu < \infty$ and $\int_{X_n} g d\mu < \infty$ for all n. Replacing A by $A \cap X_n$ in Eq. (15.4) implies

$$\int_A 1_{X_n} f d\mu = \int_{A \cap X_n} f d\mu = \int_{A \cap X_n} g d\mu = \int_A 1_{X_n} g d\mu$$

for all $A \in \mathcal{M}$. Since $1_{X_n} f$ and $1_{X_n} g$ are in $L^1(\mu)$ for all n, this equation implies $1_{X_n} f = 1_{X_n} g$, μ – a.e. Letting $n \to \infty$ then shows that $f = g$, μ – a.e. ∎

Remark 15.7. Lemma 15.6 is in general false without the σ – finiteness assumption. A trivial counterexample is to take $\mathcal{M} = 2^X$, $\mu(A) = \infty$ for all non-empty $A \in \mathcal{M}$, $f = 1_X$ and $g = 2 \cdot 1_X$. Then Eq. (15.4) holds yet $f \neq g$.

Theorem 15.8 (Radon Nikodym Theorem for Positive Measures). *Suppose that μ and ν are σ – finite positive measures on $(X, \mathcal{M})$. Then ν has a unique Lebesgue decomposition $\nu = \nu_a + \nu_s$ relative to μ and there exists a unique (modulo sets of μ – measure 0) function $\rho : X \to [0, \infty)$ such that $d\nu_a = \rho d\mu$. Moreover, $\nu_s = 0$ iff $\nu \ll \mu$.*

Proof. The uniqueness assertions follow directly from Lemmas 15.5 and 15.6.

Existence when μ and ν are both finite measures. (Von-Neumann's Proof. See Remark 15.9 for the motivation for this proof.) First suppose that μ and ν are **finite** measures and let $\lambda = \mu + \nu$. By Theorem 15.1, $d\nu = h d\lambda$ with $0 \le h \le 1$ and this implies, for all non-negative measurable functions f, that

$$\nu(f) = \lambda(fh) = \mu(fh) + \nu(fh) \tag{15.5}$$

or equivalently

$$\nu(f(1-h)) = \mu(fh). \tag{15.6}$$

Taking $f = 1_{\{h=1\}}$ in Eq. (15.6) shows that

$$\mu(\{h = 1\}) = \nu(1_{\{h=1\}}(1-h)) = 0,$$

i.e. $0 \le h(x) < 1$ for μ - a.e. x. Let

$$\rho := 1_{\{h<1\}} \frac{h}{1-h}$$

and then take $f = g 1_{\{h<1\}}(1-h)^{-1}$ with $g \ge 0$ in Eq. (15.6) to learn

$$\nu(g 1_{\{h<1\}}) = \mu(g 1_{\{h<1\}}(1-h)^{-1} h) = \mu(\rho g).$$

Hence if we define

$$\nu_a := 1_{\{h<1\}} \nu \text{ and } \nu_s := 1_{\{h=1\}} \nu,$$

we then have $\nu_s \perp \mu$ (since ν_s "lives" on $\{h = 1\}$ while $\mu(h = 1) = 0$) and $\nu_a = \rho\mu$ and in particular $\nu_a \ll \mu$. Hence $\nu = \nu_a + \nu_s$ is the desired Lebesgue decomposition of ν. If we further assume that $\nu \ll \mu$, then $\mu(h = 1) = 0$ implies $\nu(h = 1) = 0$ and hence that $\nu_s = 0$ and we conclude that $\nu = \nu_a = \rho\mu$.

Existence when μ and ν are σ-finite measures. Write $X = \sum_{n=1}^\infty X_n$ where $X_n \in \mathcal{M}$ are chosen so that $\mu(X_n) < \infty$ and $\nu(X_n) < \infty$ for all n. Let $d\mu_n = 1_{X_n} d\mu$ and $d\nu_n = 1_{X_n} d\nu$. Then by what we have just proved there exists $\rho_n \in L^1(X, \mu_n) \subset L^1(X, \mu)$ and measure ν_n^s such that $d\nu_n = \rho_n d\mu_n + d\nu_n^s$ with $\nu_n^s \perp \mu_n$. Since μ_n and ν_n^s "live" on X_n there exists $A_n \in \mathcal{M}_{X_n}$ such that $\mu(A_n) = \mu_n(A_n) = 0$ and

$$\nu_n^s(X \setminus A_n) = \nu_n^s(X_n \setminus A_n) = 0.$$

This shows that $\nu_n^s \perp \mu$ for all n and so by Lemma 15.4, $\nu_s := \sum_{n=1}^\infty \nu_n^s$ is singular relative to μ. Since

$$\nu = \sum_{n=1}^\infty \nu_n = \sum_{n=1}^\infty (\rho_n \mu_n + \nu_n^s) = \sum_{n=1}^\infty (\rho_n 1_{X_n} \mu + \nu_n^s) = \rho\mu + \nu_s, \tag{15.7}$$

where $\rho := \sum_{n=1}^\infty 1_{X_n} \rho_n$, it follows that $\nu = \nu_a + \nu_s$ with $\nu_a = \rho\mu$. Hence this is the desired Lebesgue decomposition of ν relative to μ. ■

Remark 15.9. Here is the motivation for the above construction. Suppose that $d\nu = d\nu_s + \rho d\mu$ is the Radon-Nikodym decomposition and $X = A \sum B$ such that $\nu_s(B) = 0$ and $\mu(A) = 0$. Then we find

$$\nu_s(f) + \mu(\rho f) = \nu(f) = \lambda(hf) = \nu(hf) + \mu(hf).$$

Letting $f \to 1_A f$ then implies that

$$\nu(1_A f) = \nu_s(1_A f) = \nu(1_A h f)$$

which show that $h = 1$, ν –a.e. on A. Also letting $f \to 1_B f$ implies that

$$\mu(\rho 1_B f) = \nu(h 1_B f) + \mu(h 1_B f) = \mu(\rho h 1_B f) + \mu(h 1_B f)$$

which implies, $\rho = \rho h + h$, μ – a.e. on B, i.e.

$$\rho(1 - h) = h, \ \mu\text{– a.e. on } B.$$

In particular it follows that $h < 1$, $\mu = \nu$ – a.e. on B and that $\rho = \frac{h}{1-h} 1_{h<1}$, μ – a.e. So up to sets of ν – measure zero, $A = \{h = 1\}$ and $B = \{h < 1\}$ and therefore,

$$d\nu = 1_{\{h=1\}} d\nu + 1_{\{h<1\}} d\nu = 1_{\{h=1\}} d\nu + \frac{h}{1-h} 1_{h<1} d\mu.$$

Some Ergodic Theory

The goal of this chapter is to show (in certain circumstances) that "time averages" are the same as "spatial averages." We start with a "simple" Hilbert space version of the type of theorem that we are after. For more on the following mean Ergodic theorem, see [24] and [13].

Theorem 16.1 (Von-Neumann's Mean Ergodic Theorem). *Let $U : H \to H$ be an isometry on a Hilbert space H, $M = \mathrm{Nul}(U-I)$, $P = P_M$ be orthogonal projection onto M, and $S_n = \sum_{k=0}^{n-1} U^k$. Show $\frac{S_n}{n} \to P_M$* ***strongly*** *by which we mean $\lim_{n\to\infty} \frac{S_n}{n} x = P_M x$ for all $x \in H$.*

Proof. Since U is an isometry we have $(Ux, Uy) = (x, y)$ for all $x, y \in H$ and therefore that $U^*U = I$. In general it is not true that $UU^* = I$ but instead, $UU^* = P_{\mathrm{Ran}(U)}$. Thus $U^*U = I$ iff U is surjective, i.e. U is unitary.

Before starting the proof in earnest we need to prove

$$\mathrm{Nul}(U^* - I) = \mathrm{Nul}(U - I).$$

If $x \in \mathrm{Nul}\,(U - I)$ then $x = Ux$ and therefore $U^*x = U^*Ux = x$, i.e. $x \in \mathrm{Nul}(U^* - I)$. Conversely if $x \in \mathrm{Nul}(U^* - I)$ then $U^*x = x$ and we have

$$\begin{aligned}\|Ux - x\|^2 &= 2\,\|x\|^2 - 2\,\mathrm{Re}\,(Ux, x)\\ &= 2\,\|x\|^2 - 2\,\mathrm{Re}\,(x, U^*x) = 2\,\|x\|^2 - 2\,\mathrm{Re}\,(x, x) = 0\end{aligned}$$

which shows that $Ux = x$, i.e. $x \in \mathrm{Nul}\,(U - I)$. With this remark in hand we can easily complete the proof.

Let us first observe that

$$\frac{S_n}{n}(U - I) = \frac{1}{n}[U^n - I] \to 0 \text{ as } n \to \infty.$$

Thus if $x = (U - I)y \in \mathrm{Ran}(U - I)$, we have

$$\frac{S_n}{n}x = \frac{1}{n}(U^n y - y) \to 0 \text{ as } n \to \infty.$$

More generally if $x \in \overline{\mathrm{Ran}(U - I)}$ and $x' \in \mathrm{Ran}(U-I)$, we have, since $\left\|\frac{S_n}{n}\right\| \le 1$, that

$$\left\|\frac{S_n}{n}x - \frac{S_n}{n}x'\right\| \le \|x - x'\|$$

and hence

$$\limsup_{n\to\infty}\left\|\frac{S_n}{n}x\right\| = \limsup_{n\to\infty}\left\|\frac{S_n}{n}x - \frac{S_n}{n}x'\right\| \le \|x - x'\|.$$

Letting $x' \in \mathrm{Ran}\,(U - I)$ tend to $x \in \overline{\mathrm{Ran}(U - I)}$ allows us to conclude that $\limsup_{n\to\infty}\left\|\frac{S_n}{n}x\right\| = 0$.

For

$$x \in \overline{\mathrm{Ran}(U - I)}^{\perp} = \mathrm{Ran}\,(U - I)^{\perp} = \mathrm{Nul}\,(U^* - I) = \mathrm{Nul}\,(U - I) = M$$

we have $\frac{S_n}{n}x = x$. So for general $x \in H$, we have $x = P_M x + y$ with $y \in M^\perp = \overline{\mathrm{Ran}(U - I)}$ and therefore,

$$\frac{S_n}{n}x = \frac{S_n}{n}P_M x + \frac{S_n}{n}y = P_M x + \frac{S_n}{n}y \to P_M x \text{ as } n \to \infty.$$

■

For the rest of this section, suppose that $(\Omega, \mathcal{B}, \mu)$ is a σ − finite measure space and $\theta : \Omega \to \Omega$ is a measurable map such that $\theta_*\mu = \mu$. After Theorem 16.6 we will further assume that $\mu = P$ is a probability measure. For more results along the lines of this chapter, the reader is referred to Kallenberg [30, Chapter 10]. The reader may also benefit from Norris's notes in [39].

Definition 16.2. *Let*

$$\begin{aligned}\mathcal{B}_\theta &:= \left\{A \in \mathcal{B} : \theta^{-1}(A) = A\right\} \text{ and}\\ \mathcal{B}'_\theta &:= \left\{A \in \mathcal{B} : \mu\left(\theta^{-1}(A)\,\Delta A\right) = 0\right\}\end{aligned}$$

be the ***invariant*** σ − ***field*** *and* ***almost invariant*** σ *–fields respectively.*

In what follows we will make use of the following easily proved set identities. Let $\{A_n\}_{n=1}^\infty$, $\{B_n\}_{n=1}^\infty$, and A, B, C be a collection of subsets of Ω, then;

1. $A\Delta C \subset [A\Delta B] \cup [B\Delta C]$,
2. $[\cup_{n=1}^\infty A_n]\,\Delta\,[\cup_{n=1}^\infty B_n] \subset \cup_{n=1}^\infty A_n \Delta B_n$,
3. $[\cap_{n=1}^\infty A_n]\,\Delta\,[\cap_{n=1}^\infty B_n] \subset \cup_{n=1}^\infty A_n \Delta B_n$,
4. $B\Delta\{A_n \text{ i.o.}\} \subset \cup_{n=1}^\infty [B\Delta A_n]$.

Lemma 16.3. *The elements of $\mathcal{B}'_\theta$ are the same as the elements in $\mathcal{B}_\theta$ modulo null sets, i.e.*

$$\mathcal{B}'_\theta = \{B \in \mathcal{B} : \exists\ A \in \mathcal{B}_\theta \ \ni\ \mu(A\Delta B) = 0\}.$$

Moreover if $B \in \mathcal{B}'_\theta$, then

$$A := \left\{\omega \in \Omega : \theta^k(\omega) \in B \text{ i.o. } k\right\} \in \mathcal{B}_\theta \tag{16.1}$$

and $\mu(A\Delta B) = 0$. (We could have just as well taken A to be equal to $\left\{\omega \in \Omega : \theta^k(\omega) \in B \text{ a.a.}\right\}$.)

Proof. If $A \in \mathcal{B}_\theta$ and $B \in \mathcal{B}$ such that $\mu(A\Delta B) = 0$, then

$$\mu\left(A\Delta\theta^{-1}(B)\right) = \mu\left(\theta^{-1}(A)\,\Delta\theta^{-1}(B)\right) = \mu\theta^{-1}(A\Delta B) = \mu(A\Delta B) = 0$$

and therefore it follows that

$$\mu\left(B\Delta\theta^{-1}(B)\right) \le \mu(B\Delta A) + \mu\left(A\Delta\theta^{-1}(B)\right) = 0.$$

This shows that $B \in \mathcal{B}'_\theta$.

Conversely if $B \in \mathcal{B}'_\theta$ then by the invariance of μ under θ it follows that $\mu\left(\theta^{-l}(B)\,\Delta\theta^{-(l+1)}(B)\right) = 0$ for all $k = 0, 1, 2, 3 \dots$. In particular we learn that

$$\begin{aligned}\mu\left(\theta^{-k}(B)\,\Delta B\right) &= \mu\left(\left|1_{\theta^{-k}(B)} - 1_B\right|\right)\\ &\le \sum_{l=0}^{k-1} \mu\left(\left|1_{\theta^{-l}(B)} - 1_{\theta^{-(l+1)}(B)}\right|\right)\\ &= \sum_{l=0}^{k-1} \mu\left(\theta^{-l}(B)\,\Delta\theta^{-(l+1)}(B)\right) = 0.\end{aligned}$$

Thus if $A = \left\{\theta^{-k}(B) \text{ i.o. } k\right\}$ as in Eq. (16.1) we have,

$$\mu(B\Delta A) \le \sum_{k=1}^{\infty} \mu\left(B\Delta\theta^{-k}(B)\right) = 0.$$

This completes the proof since

$$\theta^{-1}(A) = \left\{\omega \in \Omega : \theta^{k+1}(\omega) \in B \text{ i.o. } k\right\} = A$$

and thus $A \in \mathcal{B}_\theta$. ■

Definition 16.4. *A $\mathcal{B}$ – measurable function, $f : \Omega \to \mathbb{R}$ is (almost) invariant iff $f \circ \theta = f$ ($f \circ \theta = f$ a.s.).*

Lemma 16.5. *A $\mathcal{B}$ – measurable function, $f : \Omega \to \mathbb{R}$ is (almost) invariant iff f is $\mathcal{B}_\theta$ ($\mathcal{B}'_\theta$) measurable. Moreover, if f is almost invariant, then there exists and invariant function, $g : \Omega \to \mathbb{R}$, such that $f = g$, μ – a.e. (This latter assertion has already been explained in Exercises 12.3 and 12.4.)*

Proof. If f is invariant, $f \circ \theta = f$, then $\theta^{-1}(\{f \le x\}) = \{f \circ \theta \le x\} = \{f \le x\}$ which shows that $\{f \le x\} \in \mathcal{B}_\theta$ for all $x \in \mathbb{R}$ and therefore f is $\mathcal{B}_\theta$ – measurable. Similarly if f is almost invariant so that $f \circ \theta = f$ (μ – a.e.), then

$$\begin{aligned}\mu\left(\left|1_{\theta^{-1}(\{f\le x\})} - 1_{\{f\le x\}}\right|\right) &= \mu\left(\left|1_{\{f\circ\theta\le x\}} - 1_{\{f\le x\}}\right|\right)\\ &= \mu\left(\left|1_{(-\infty,x]} \circ f \circ \theta - 1_{(-\infty,x]} \circ f\right|\right) = 0\end{aligned}$$

from which it follows that $\{f \le x\} \in \mathcal{B}'_\theta$ for all $x \in \mathbb{R}$, that is f is $\mathcal{B}'_\theta$ – measurable.

Conversely if $f : \Omega \to \mathbb{R}$ is ($\mathcal{B}'_\theta$) $\mathcal{B}_\theta$ -measurable, then for all $-\infty < a < b < \infty$, ($\{a < f \le b\} \in \mathcal{B}'_\theta$) $\{a < f \le b\} \in \mathcal{B}_\theta$ from which it follows that $1_{\{a<f\le b\}}$ is (almost) invariant. Thus for every $N \in \mathbb{N}$ the function defined by;

$$f_N := \sum_{n=-N^2}^{N^2} \frac{n}{N} 1_{\left\{\frac{n-1}{N} < f \le \frac{n}{N}\right\}},$$

is (almost) invariant. As $f = \lim_{N\to\infty} f_N$, it follows that f is (almost) invariant as well.

In the case where f is almost invariant, we can choose $D_N(n) \in \mathcal{B}_\theta$ such that $\mu\left(D_N(n)\,\Delta\left\{\frac{n-1}{N} < f \le \frac{n}{N}\right\}\right) = 0$ for all n and N and then set

$$g_N := \sum_{n=-N^2}^{N^2} \frac{n}{N} 1_{D_N(n)}.$$

We then have $g_N = f_N$ a.e. and g_N is $\mathcal{B}_\theta$ – measurable. We may thus conclude that $\tilde{g} := \limsup_{N\to\infty} g_N$ is $\mathcal{B}_\theta$ – measurable. It now follows that $g := \tilde{g}1_{|\tilde{g}|<\infty}$ is $\mathcal{B}_\theta$ – measurable function such that $g = f$ a.e. ■

Theorem 16.6. *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space and $\theta : \Omega \to \Omega$ is a measurable map such that $\theta_*\mu = \mu$. Then;*

1. *$U : L^2(\mu) \to L^2(\mu)$ defined by $Uf := f \circ \theta$ is an isometry. The isometry U is unitary if θ^{-1} exists as a measurable map.*
2. *The map,*
$$L^2(\Omega, \mathcal{B}_\theta, \mu) \ni f \to f \in \mathrm{Nul}(U - I)$$
is unitary. In other words, $Uf = f$ iff there exists $g \in L^2(\Omega, \mathcal{B}_\theta, \mu)$ such that $f = g$ a.e.

3. *For every $f \in L^2(\mu)$ we have,*

$$L^2(\mu) - \lim_{n\to\infty} \frac{f + f\circ\theta + \dots + f\circ\theta^{n-1}}{n} = \mathbb{E}_{\mathcal{B}_\theta}[f]$$

where $\mathbb{E}_{\mathcal{B}_\theta}$ denotes orthogonal projection from $L^2(\Omega,\mathcal{B},\mu)$ onto $L^2(\Omega,\mathcal{B}_\theta,\mu)$, i.e. $\mathbb{E}_{\mathcal{B}_\theta}$ is conditional expectation.

Proof. 1. To see that U is an isometry observe that

$$\|Uf\|^2 = \int_\Omega |f\circ\theta|^2 d\mu = \int_\Omega |f|^2 d(\theta_*\mu) = \int_\Omega |f|^2 d\mu = \|f\|^2$$

for all $f \in L^2(\mu)$.

2. $f \in \mathrm{Nul}(U-I)$ iff $f\circ\theta = Uf = f$ a.e., i.e. iff f is almost invariant. According to Lemma 16.5 this happen iff there exists a $\mathcal{B}_\theta$ – measurable function, g, such that $f = g$ a.e. Necessarily, $g \in L^2(\mu)$ so that $g \in L^2(\Omega,\mathcal{B}_\theta,\mu)$ as required.

3. The last assertion now follows from items 1. and 2. and the mean ergodic Theorem 16.1. ■

Assumption 1 *From now on we will assume that $\mu = P$ is a probability measure such that $P\theta^{-1} = \theta$.*

Exercise 16.1. For every $Z \in L^1(P)$, show that $\mathbb{E}[Z\circ\theta|\mathcal{B}_\theta] = \mathbb{E}[Z|\mathcal{B}_\theta]$ a.s. More generally, show for sub – σ -algebra, $\mathcal{G} \subset \mathcal{B}$, show $\mathbb{E}\left[Z\circ\theta|\theta^{-1}\mathcal{G}\right] = \mathbb{E}[Z|\mathcal{G}]\circ\theta$ a.s.

Solution to Exercise (16.1). First observe that $\mathbb{E}[Z|\mathcal{G}]\circ\theta$ is $\mathcal{G}$ – measurable being the composition of $(\Omega,\theta^{-1}\mathcal{G}) \xrightarrow{\theta} (\Omega,\mathcal{G}) \xrightarrow{\mathbb{E}[Z|\mathcal{G}]} (\mathbb{R},\mathcal{B}_\mathbb{R})$. Now let $A \in \mathcal{G}$, then

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[Z\circ\theta|\theta^{-1}\mathcal{G}\right]:\theta^{-1}A\right] &= \mathbb{E}\left[Z\circ\theta:\theta^{-1}A\right] = \mathbb{E}\left[(Z\cdot 1_A)\circ\theta\right] \\ &= \mathbb{E}[Z\cdot 1_A] = \mathbb{E}\left[\mathbb{E}[Z|\mathcal{G}]\,1_A\right] \\ &= \mathbb{E}\left[(\mathbb{E}[Z|\mathcal{G}]\,1_A)\circ\theta\right] = \mathbb{E}\left[\mathbb{E}[Z|\mathcal{G}]\circ\theta:\theta^{-1}A\right].\end{aligned}$$

As $A \in \mathcal{G}$ is arbitrary, it follows that $\mathbb{E}\left[Z\circ\theta|\theta^{-1}\mathcal{G}\right] = \mathbb{E}[Z|\mathcal{G}]\circ\theta$ a.s. Taking $\mathcal{G} = \mathcal{B}_\theta$ then shows,

$$\mathbb{E}[Z\circ\theta|\mathcal{B}_\theta] = \mathbb{E}\left[Z\circ\theta|\theta^{-1}\mathcal{B}_\theta\right] = \mathbb{E}[Z|\mathcal{B}_\theta]\circ\theta = \mathbb{E}[Z|\mathcal{B}_\theta] \text{ a.s.}$$

Exercise 16.2. Let $1 \le p < \infty$. Following the ideas introduced in Exercises 14.8 and 14.9, show

$$L^p(P) - \lim_{n\to\infty} \frac{f + f\circ\theta + \dots + f\circ\theta^{n-1}}{n} = \mathbb{E}_{\mathcal{B}_\theta}[f] \text{ for all } f \in L^p(\Omega,\mathcal{B},P).$$

(Some of these ideas will again be used in the proof of Theorem 16.9 below.)

Definition 16.7. *A sequence of random variables $\xi = \{\xi_k\}_{k=1}^\infty$ is a* ***stationary*** *if $(\xi_2,\xi_3,,\dots) \overset{d}{=} (\xi_1,\xi_2,\dots)$.*

If we temporarily let

$$\theta(x_1,x_2,x_3,\dots) = \theta(x_2,x_3,\dots) \text{ for } (x_1,x_2,x_3,\dots) \in \mathbb{R}^\mathbb{N}, \tag{16.2}$$

the stationarity condition states that $\theta\xi \overset{d}{=} \xi$. Equivalently if $\mu = \mathrm{Law}_P(\xi_1,\xi_2,\dots)$ on $\left(\mathbb{R}^\mathbb{N},\mathcal{B}_\mathbb{R}^{\otimes\mathbb{N}}\right)$, then $\xi = \{\xi_k\}_{k=1}^\infty$ is stationary iff $\mu\circ\theta^{-1} = \mu$. Let us also observe that ξ is stationary implies $\theta^2\xi \overset{d}{=} \theta\xi \overset{d}{=} \xi$ and $\theta^3\xi \overset{d}{=} \theta\xi \overset{d}{=} \xi$, etc. so that $\theta^n\xi \overset{d}{=} \xi$ for all $n \in \mathbb{N}$.[1] In what follows for $x \in (x_1,x_2,x_3,\dots) \in \mathbb{R}^\mathbb{N}$ we will let $S_0(x) = 0$,

$$\begin{aligned} S_n(x) &= x_1 + x_2 + \dots + x_n, \text{ and} \\ S_n^* &:= \max(S_1,S_2,\dots,S_n)\end{aligned}$$

for all $n \in \mathbb{N}$.

Lemma 16.8 (Maximal Ergodic Lemma). *Suppose $\xi := \{\xi_k\}_{k=1}^\infty$ is a stationary sequence and $S_n(\xi) = \xi_1 + \dots + \xi_n$ as above, then*

$$\mathbb{E}\left[\xi_1 : \sup_n S_n(\xi) > 0\right] \ge 0. \tag{16.3}$$

Proof. In this proof, θ will be as in Eq. (16.2). If $1 \le k \le n$, then

$$S_k(\xi) = \xi_1 + S_{k-1}(\theta\xi) \le \xi_1 + S_{k-1}^*(\theta\xi) \le \xi_1 + S_n^*(\theta\xi) = \xi_1 + [S_n^*(\theta\xi)]_+$$

and therefore, $S_n^*(\xi) \le \xi_1 + [S_n^*(\theta\xi)]_+$. So we may conclude that

$$\begin{aligned}\mathbb{E}[\xi_1 : S_n^*(\xi) > 0] &\ge \mathbb{E}\left[S_n^*(\xi) - [S_n^*(\theta\xi)]_+ : S_n^* > 0\right] \\ &= \mathbb{E}\left[[S_n^*(\xi)]_+ - [S_n^*(\theta\xi)]_+ 1_{S_n^*>0}\right] \\ &\ge \mathbb{E}\left[[S_n^*(\xi)]_+ - [S_n^*(\theta\xi)]_+\right] = \mathbb{E}[S_n^*(\xi)]_+ - \mathbb{E}[S_n^*(\theta\xi)]_+ = 0,\end{aligned}$$

wherein we used $\xi \overset{d}{=} \theta\xi$ for the last equality. Letting $n \to \infty$ making use of the MCT and the observation that $\{S_n^*(\xi) > 0\} \uparrow \{\sup_n S_n(\xi) > 0\}$ gives Eq. (16.3). ■

[1] In other words if $\{\xi_k\}_{k=1}^\infty$ is stationary, then by lopping off the first random variable on each side of the identity, $(\xi_2,\xi_3,,\dots) \overset{d}{=} (\xi_1,\xi_2,\dots)$, implies that

$$(\xi_3,\xi_4,\dots) \overset{d}{=} (\xi_2,\xi_3,,\dots) \overset{d}{=} (\xi_1,\xi_2,\dots).$$

Continuing this way inductively shows that stationarity is equivalent to $(\xi_n,\xi_{n+1},\xi_{n+2},\dots) \overset{d}{=} (\xi_1,\xi_2,\dots)$ for all $n \in \mathbb{N}$.

Theorem 16.9 (Birkoff's Ergodic Theorem). *Suppose that* $f \in L^1(\Omega, \mathcal{B}, P)$ *or* $f \geq 0$ *and is* $\mathcal{B}$ *– measurable, then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1} = \mathbb{E}\left[f|\mathcal{B}_\theta\right] \ \ a.s.. \tag{16.4}$$

Moreover if $f \in L^p(\Omega, \mathcal{B}, P)$ *for some* $1 \leq p < \infty$ *then the convergence in Eq. (16.4) holds in* L^p *as well.*

Proof. Let us begin with the general observation that if $\xi = (\xi_1, \xi_2, \dots)$ is a sequence of random variables such that $\xi_i \circ \theta = \xi_{i+1}$ for $i = 1, 2, \dots,$ then ξ is stationary. This is because,

$$(\xi_1, \xi_2, \dots) \overset{d}{=} (\xi_1, \xi_2, \dots)\circ\theta = (\xi_1\circ\theta, \xi_2\circ\theta, \dots) = (\xi_2, \xi_3, \dots).$$

We will first prove Eq. (16.4) under the assumption that $f \in L^1(P)$. We now let $g := \mathbb{E}[f|\mathcal{B}_\theta]$ and $\xi_k := f\circ\theta^{k-1} - g$ for all $k \in \mathbb{N}$. Since g is $\mathcal{B}_\theta$ – measurable we know that $g\circ\theta = g$ and therefore,

$$\xi_k\circ\theta = \left(f\circ\theta^{k-1} - g\right)\circ\theta = f\circ\theta^k - g = \xi_{k+1}$$

and therefore $\xi = (\xi_1, \xi_2, \dots)$ is stationary. To simplify notation let us write S_n for $S_n(\xi) = \xi_1 + \dots + \xi_n$. To finish the proof we need to show that $\lim_{n\to\infty}\frac{S_n}{n} = 0$ a.s. for then,

$$\frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1} = \frac{1}{n}S_n + g \to g = \mathbb{E}[f|\mathcal{B}_\theta] \text{ a.s.}$$

In order to show $\lim_{n\to\infty}\frac{S_n}{n} = 0$ a.s. it suffices to show $M(\xi) := \limsup_{n\to\infty}\frac{S_n(\xi)}{n} \leq 0$ a.s. If we can do this we can also show that $M(-\xi) = \limsup_{n\to\infty}\frac{-S_n(\xi)}{n} \leq 0$, i.e.

$$\liminf_{n\to\infty}\frac{S_n(\xi)}{n} \geq 0 \geq \limsup_{n\to\infty}\frac{S_n(\xi)}{n} \text{ a.s.}$$

which shows that $\lim_{n\to\infty}\frac{S_n}{n} = 0$ a.s. Finally in order to prove $M(\xi) \leq 0$ a.s. it suffices to show $P(M(\xi) > \varepsilon) = 0$ for all $\varepsilon > 0$. This is what we will do now.

Since $S_n\circ\theta = S_{n+1} - \xi_1$ we have so that

$$M(\xi)\circ\theta = \limsup_{n\to\infty}\frac{1}{n}(S_{n+1} - \xi_1) = \limsup_{n\to\infty}\left[\frac{n+1}{n}\cdot\frac{1}{n+1}S_{n+1}\right] = M(\xi).$$

Thus $M(\xi)$ is an invariant function and therefore $A_\varepsilon := \{M(\xi) > \varepsilon\} \in \mathcal{B}_\theta$. Using $\mathbb{E}[\xi_1|\mathcal{B}_\theta] = \mathbb{E}[f - g|\mathcal{B}_\theta] = g - g = 0$ a.s. it follows that

$$\begin{aligned} 0 &= \mathbb{E}\left[\mathbb{E}[\xi_1|\mathcal{B}_\theta] : M(\xi) > \varepsilon\right] = \mathbb{E}[\xi_1 : M(\xi - \varepsilon) > 0] \\ &= \mathbb{E}[\xi_1 - \varepsilon : M(\xi - \varepsilon) > 0] + \varepsilon P(A_\varepsilon). \end{aligned}$$

If we now define $\xi_n^\varepsilon := (\xi_n - \varepsilon)1_{A_\varepsilon}$, which is still stationary since

$$\xi_n^\varepsilon\circ\theta = (\xi_n\circ\theta - \varepsilon)1_{A_\varepsilon}\circ\theta = (\xi_{n+1} - \varepsilon)1_{A_\varepsilon} = \xi_{n+1}^\varepsilon,$$

then it is easily verified[2] that

$$A_\varepsilon = \{M(\xi - \varepsilon) > 0\} = \left\{\sup_n S_n(\xi^\varepsilon) > 0\right\}.$$

Therefore by an application of the maximal ergodic Lemma 16.8 we have,

$$-\varepsilon P(M(\xi) > \varepsilon) = \mathbb{E}[\xi_1 - \varepsilon : A_\varepsilon] = \mathbb{E}\left[\xi_1^\varepsilon : \sup_n S_n(\xi^\varepsilon) > 0\right] \geq 0$$

which shows $P(M(\xi) > \varepsilon) = 0$.

Now suppose that $f \in L^p(P)$. To prove the L^p – convergence of the limit in Eq. (16.4) it suffices by Corollary 12.47 to show $\left\{\left|\frac{1}{n}S_n(\eta)\right|^p\right\}_{n=1}^\infty$ is uniformly integrable. This can be done as in the second solution to Exercise 12.6 (Resnick § **6.7**, #5). Here are the details.

First observe that $\{|\eta_k|^p\}_{k=1}^\infty$ are uniformly integrable. Indeed, by stationarity,

$$\mathbb{E}[|\eta_k|^p : |\eta_k|^p \geq a] = \mathbb{E}[|\eta_1|^p : |\eta_1|^p \geq a]$$

and therefore

$$\sup_k \mathbb{E}[|\eta_k|^p : |\eta_k|^p \geq a] = \mathbb{E}[|\eta_1|^p : |\eta_1|^p \geq a] \overset{\text{DCT}}{\to} 0 \text{ as } a \to \infty.$$

Thus if $\varepsilon > 0$ is given we may find (see Proposition 12.42) $\delta > 0$ such that $\mathbb{E}[|\eta_k|^p : A] \leq \varepsilon$ whenever $A \in \mathcal{B}$ with $P(A) \leq \delta$. Then for such an A, we have (using Jensen's inequality relative to normalized counting measure on $\{1, 2, \dots, n\}$),

$$\mathbb{E}\left[\left|\frac{1}{n}S_n(\eta)\right|^p : A\right] \leq \mathbb{E}\left[\frac{1}{n}S_n(|\eta|^p) : A\right] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[|\eta_k|^p : A] \leq \frac{1}{n}n\varepsilon = \varepsilon.$$

[2] Since $A_\varepsilon \subset \{\sup S_n/n > \varepsilon\}$, it follows that

$$\begin{aligned} A_\varepsilon &= \left\{\sup\frac{S_n}{n} > \varepsilon\right\}\cap A_\varepsilon = \{\sup S_n - n\varepsilon > 0\}\cap A_\varepsilon \\ &= \{\sup S_n(\xi - \varepsilon) > 0\}\cap A_\varepsilon = \{\sup S_n(\xi - \varepsilon)1_{A_\varepsilon} > 0\} \\ &= \left\{\sup_n S_n(\xi^\varepsilon) > 0\right\}. \end{aligned}$$

Another application of Proposition 12.42 shows $\left\{\left|\frac{1}{n}S_n(\eta)\right|^p\right\}_{n=1}^\infty$ is uniformly integrable as

$$\sup_n \mathbb{E}\left|\frac{1}{n}S_n(\eta)\right|^p \leq \sup_n \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left[|\eta_k|^p\right] = \mathbb{E}\left|\eta_1\right|^p < \infty.$$

Finally we need to consider the case where $f \geq 0$ but $f \notin L^1(P)$. As before, let $g = \mathbb{E}[f|\mathcal{B}_\theta] \geq 0$. For $r \in (0,\infty)$ and let $f_r := f \cdot 1_{g\leq r}$. We then have

$$\mathbb{E}[f_r|\mathcal{B}_\theta] = \mathbb{E}[f \cdot 1_{g\leq r}|\mathcal{B}_\theta] = 1_{g\leq r}\mathbb{E}[f \cdot |\mathcal{B}_\theta] = 1_{g\leq r}\cdot g$$

and in particular, $\mathbb{E}f_r = \mathbb{E}(1_{g\leq r}g) \leq r < \infty$. Thus by the L^1 – case already proved,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f_r \circ \theta^{k-1} = 1_{g\leq r}\cdot g \text{ a.s.}$$

On the other hand, since g is θ –invariant, we see that $f_r \circ \theta^k = f \circ \theta^k \cdot 1_{g\leq r}$ and therefore

$$\frac{1}{n}\sum_{k=1}^n f_r \circ \theta^{k-1} = \left(\frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1}\right) 1_{g\leq r}.$$

Using these identities and the fact that $r < \infty$ was arbitrary we may conclude that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} = g \text{ a.s. on } \{g < \infty\}. \tag{16.5}$$

To take care of the set where $\{g = \infty\}$, again let $r \in (0,\infty)$ but now take $f_r = f \wedge r \leq f$. It then follows that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} \geq \liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^n \left[f_r \circ \theta^{k-1}\right] = \mathbb{E}[f \wedge r|\mathcal{B}_\theta].$$

Letting $r \uparrow \infty$ and using the cMCT implies,

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} \geq \mathbb{E}[f|\mathcal{B}_\theta] = g$$

and therefore $\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} = \infty$ a.s. on $\{g = \infty\}$. This then shows that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} = \infty = g \text{ a.s. on } \{g = \infty\}.$$

which combined with Eq. (16.5) completes the proof. ∎

As a corollary we have the following version of the strong law of large numbers, also see Theorems 20.30 and Example 18.78 below for other proofs.

Theorem 16.10 (Kolmogorov's Strong Law of Large Numbers). *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables and let $S_n := X_1 + \cdots + X_n$. If X_n are integrable or $X_n \geq 0$, then*

$$\lim_{n\to\infty}\frac{1}{n}S_n = \mathbb{E}X_1 \text{ a.s.}$$

and $\frac{1}{n}S_n \to \mathbb{E}X_1$ in $L^1(P)$ when $\mathbb{E}|X_n| < \infty$.

Proof. We may assume that $\Omega = \mathbb{R}^\mathbb{N}$, $\mathcal{B}$ is the product σ – algebra, and $P = \mu^{\otimes\mathbb{N}}$ where $\mu = \mathrm{Law}_P(X_1)$. In this model, $X_n(\omega) = \omega_n$ for all $\omega \in \Omega$ and we take $\theta : \Omega \to \Omega$ as in Eq. (16.2). Wit this notation we have $X_n = X_1 \circ \theta^{n-1}$ and therefore, $S_n = \sum_{k=1}^n X_1 \circ \theta^{k-1}$. So by Birkoff's ergodic theorem $\lim_{n\to\infty}\frac{1}{n}S_n = \mathbb{E}[X_1|\mathcal{B}_\theta] =: g$ a.s.

If $A \in \mathcal{B}_\theta$, then $A = \theta^{-n}(A) \in \sigma(X_{n+1}, X_{n+2}, \dots)$ and therefore $A \in \mathcal{T} = \cap_n \sigma(X_{n+1}, X_{n+2}, \dots)$ – the tail σ – algebra. However by Kolmogorov's 0 - 1 law (Proposition 10.50), we know that $\mathcal{T}$ is almost trivial and therefore so is $\mathcal{B}_\theta$. Hence we may conclude that $g = c$ a.s. where $c \in [0,\infty]$ is a constant, see Lemma 10.49.

If $X_1 \geq 0$ a.s. and $\mathbb{E}X_1 = \infty$ then we must $c = \mathbb{E}[X_1|\mathcal{B}_\theta] = \infty$ a.s. for if $c < \infty$, then $\mathbb{E}X_1 = \mathbb{E}[\mathbb{E}[X_1|\mathcal{B}_\theta]] = \mathbb{E}[c] < \infty$. When $X_1 \in L^1(P)$, the convergence in Birkoff's ergodic theorem is also in L^1 and therefore we may conclude that

$$c = \mathbb{E}c = \lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n}S_n\right] = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[S_n] = \mathbb{E}X_1.$$

Thus we have shown in all cases that $\lim_{n\to\infty}\frac{1}{n}S_n = \mathbb{E}[X_1|\mathcal{B}_\theta] = \mathbb{E}X_1$ a.s. ∎

Part III

Stochastic Processes I

In the sequel $(\Omega, \mathcal{B}, P)$ will be a probability space and $(S, \mathcal{S})$ will denote a measurable space which we refer to as **state space**. If we say that $f : \Omega \to S$ is a function we will always assume that it is $\mathcal{B}/\mathcal{S}$ – measurable. We also let $\mathcal{S}_b$ denote the bounded $\mathcal{S}/\mathcal{B}_{\mathbb{R}}$ – measurable functions from S to $\mathbb{R}$. On occasion we will assume that $(S, \mathcal{S})$ is a standard Borel space in order to have available to us the existence of regular conditional distributions (see Remark 14.16 and Theorem 14.32) and the use of Kolmogorov's extension Theorem 17.54 for proving the existence of Markov processes.

In the rest of this book we will devote most of our time to studying **stochastic processes**, i.e. a collection of random variables or more generally random functions, $X := \{X_t : \Omega \to S\}_{t\in T}$, indexed by some parameter space, T. The weakest description of such a stochastic process will be through its "finite dimensional distributions."

Definition 16.11. *Given a stochastic process, $X := \{X_t : \Omega \to S\}_{t\in T}$, and a finite subset, $\Lambda \subset T$, we say that $\nu_\Lambda := \mathrm{Law}_P\left(\{X_t\}_{t\in\Lambda}\right)$ on $\left(S^\Lambda, \mathcal{S}^{\otimes\Lambda}\right)$ is a finite dimensional distribution of X.*

Unless T is a countable or finite set or X_t has some continuity properties in t, knowledge of the finite dimensional distributions alone is not going to be adequate for our purposes, however it is a starting point. For now we are going to restrict our attention to the case where $T = \mathbb{N}_0$ or $T = \mathbb{R}_+ := [0, \infty)$ ($t \in T$ is typically interpreted as a time). Later in this part we will further restrict attention to stochastic processes indexed by $\mathbb{N}_0$ leaving the technically more complicated case where $T = \mathbb{R}_+$ to later parts of the book.

Definition 16.12. *An increasing (i.e. non-decreasing) sequence $\{\mathcal{B}_t\}_{t\in T}$ of sub-σ-algebras of $\mathcal{B}$ is called a **filtration**. We will let $\mathcal{B}_\infty := \vee_{t\in T}\mathcal{B}_t := \sigma\left(\cup_{t\in T}\mathcal{B}_t\right)$. A four-tuple, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in T}, P\right)$, where $(\Omega, \mathcal{B}, P)$ is a probability space and $\{\mathcal{B}_t\}_{t\in T}$ is a filtration is called a **filtered probability space.** We say that a stochastic process, $\{X_t\}_{t\in T}$, of random functions from $\Omega \to S$ is **adapted** to the filtration if X_t is $\mathcal{B}_t/\mathcal{S}$ – measurable for every $t \in T$.*

A typical way to make a filtration is to start with a stochastic process $\{X_t\}_{t\in T}$ and then define $\mathcal{B}_t^X := \sigma\left(X_s : s \le t\right)$. Clearly $\{X_t\}_{t\in T}$ will always be adapted to this filtration.

In this part of the book we are going to study stochastic processes with certain dependency structures. This will take us to the notion of Markov processes and martingales. Before starting our study of Markov processes it will be helpful to record a few more facts about probability kernels.

Given a probability kernel, Q, on $S \times S$ (so $Q : S \times \mathcal{S} \to [0,1]$), we may associate a linear transformation, $T = T_Q : \mathcal{S}_b \to \mathcal{S}_b$ defined by

$$(Tf)(x) = Q(x, f) = \int_S Q(x, dy) f(y) \text{ for all } f \in \mathcal{S}_b. \tag{16.6}$$

It is easy to check that T satisfies;

1. $T1 = 1$,
2. $Tf \ge 0$ if $0 \le f \in \mathcal{S}_b$,
3. if $f_n \in \mathcal{S}_b$ and $f_n \to f$ boundedly then $Tf_n \to Tf$ boundedly as well.

Notice that an operator $T : \mathcal{S}_b \to \mathcal{S}_b$ satisfying conditions 1. and 2. above also satisfies $Tf \le Tg$ and Tf is real is f. Indeed if $f = f_+ - f_-$ is real then

$$Tf = T(f_+ - f_-) = Tf_+ - Tf_-$$

with $0 \le Tf_\pm \in \mathbb{R}$ and if $f \le g$ then $0 \le f - g$ which implies

$$Tf - Tg = T(f - g) \ge 0.$$

As $\pm f \le |f|$ when f is real, we have $\pm Tf \le T|f|$ and therefore $|Tf| \le T|f|$. More generally if f is complex and $x \in S$, we may choose $\theta \in \mathbb{R}$ such that $e^{i\theta}(Tf)(x) \ge 0$ and therefore,

$$\begin{aligned} |(Tf)(x)| &= e^{i\theta}(Tf)(x) = \left(T\left[e^{i\theta}f\right]\right)(x) \\ &= \left(T\,\mathrm{Re}\left[e^{i\theta}f\right]\right)(x) + i\left(T\,\mathrm{Im}\left[e^{i\theta}f\right]\right)(x). \end{aligned}$$

Furthermore we must have $\left(T\,\mathrm{Im}\left[e^{i\theta}f\right]\right)(x) = 0$ and using $\mathrm{Re}\left[e^{i\theta}f\right] \le |f|$ we find,

$$|(Tf)(x)| = \left(T\,\mathrm{Re}\left[e^{i\theta}f\right]\right)(x) \le (T|f|)(x).$$

As $x \in S$ was arbitrary we have shown that $|Tf| \le T|f|$. Thus if $|f| \le M$ for some $0 \le M < \infty$ we may conclude,

$$|Tf| \le T|f| \le T(M \cdot 1) = MT1 = M.$$

Proposition 16.13. *If $T : \mathcal{S}_b \to \mathcal{S}_b$ is a linear transformation satisfying the three properties listed after Eq. (16.6), then $Q(x, A) := (T1_A)(x)$ for all $A \in \mathcal{S}$ and $x \in S$ is a probability kernel such that Eq. (16.6) holds.*

The proof of this proposition is straightforward and will be left to the reader. Let me just remark that if $Q(x, A) := (T1_A)(x)$ for all $x \in S$ and $A \in \mathcal{S}$ then $Tf = Q(\cdot, f)$ for all simple functions in $\mathcal{S}_b$ and then by approximation for all $f \in \mathcal{S}_b$.

Corollary 16.14. *If Q_1 and Q_2 are two probability kernels on $(S, \mathcal{S}) \times (S, \mathcal{S})$, then $T_{Q_1}T_{Q_2} = T_Q$ where Q is the probability kernel given by*

$$\begin{aligned} Q(x, A) &= (T_{Q_1}T_{Q_2}1_A)(x) = Q_1(x, Q_2(\cdot, A)) \\ &= \int_S Q_1(x, dy)\, Q_2(y, A) \end{aligned}$$

for all $A \in \mathcal{S}$ and $x \in S$. We will denote Q by Q_1Q_2.

From now on we will identify the probability kernel $Q : S \times \mathcal{S} \to [0,1]$ with the linear transformation $T = T_Q$ and simply write Qf for $Q(\cdot, f)$. The last construction that we need involving probability kernels is the following extension of the notion of product measure.

Proposition 16.15. *Suppose that ν is a probability measure on $(S, \mathcal{S})$ and $Q_k : S \times \mathcal{S} \to [0,1]$ are probability kernels on $(S, \mathcal{S}) \times (S, \mathcal{S})$ for $1 \leq k \leq n$. Then there exists a probability measure μ on $\left(S^{n+1}, \mathcal{S}^{\otimes(n+1)}\right)$ such that for all $f \in \mathcal{S}_b^{\otimes(n+1)}$ we have*

$$\mu(f) = \int_S d\nu(x_0) \int_S Q_1(x_0, dx_1) \int_S Q_2(x_1, dx_2) \cdot \\ \dots \cdot \int_S Q_n(x_{n-1}, dx_n) f(x_0, \dots, x_n). \quad (16.7)$$

Part of the assertion here is that all functions appearing are bounded and measurable so that all of the above integrals make sense. We will denote μ in the future by,

$$d\mu(x_0, \dots, x_n) = d\nu(x_0) Q_1(x_0, dx_1) Q_2(x_1, dx_2) \dots Q_n(x_{n-1}, dx_n).$$

Proof. The fact that all of the iterated integrals make sense in Eq. (16.7) follows from Exercise 14.3, the measurability statements in Fubini's theorem, and induction. The measure μ is defined by setting $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{S}^{\otimes(n+1)}$. It is a simple matter to check that μ is a measure on $\left(S^{n+1}, \mathcal{S}^{\otimes(n+1)}\right)$ and that $\int_S f d\mu$ agrees with the right side of Eq. (16.7) for all $f \in \mathcal{S}_b^{\otimes(n+1)}$. ■

Remark 16.16. As usual the measure μ is determined by its value on product functions of the form $f(x_0, \dots, x_n) = \prod_{i=0}^n f_i(x_i)$ with $f_i \in \mathcal{S}_b$. For such a function we have

$$\mu(f) = \mathbb{E}_\nu\left[f_0 Q_1 M_{f_1} Q_2 M_{f_2} \dots Q_{n-1} M_{f_{n-1}} Q_n f_n\right]$$

where $M_f : \mathcal{S}_b \to \mathcal{S}_b$ is defined by $M_f g = fg$, i.e. M_f is multiplication by f.

17

The Markov Property

For purposes of this section, $T = \mathbb{N}_0$ or $\mathbb{R}_+$, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in T}, P\right)$ is a filtered probability space, and $(S, \mathcal{S})$ be a measurable space. We will often write $t \geq 0$ to mean that $t \in T$. Thus we will often denote a stochastic process by $\{X_t\}_{t\geq 0}$ instead of $\{X_t\}_{t\in T}$.

Definition 17.1 (The Markov Property). *A stochastic process $\{X_t : \Omega \to S\}_{t\in T}$ is said to satisfy the* ***Markov property*** *if X_t is adapted and*

$$\mathbb{E}_{\mathcal{B}_s} f(X_t) := \mathbb{E}\left[f(X_t) | \mathcal{B}_s\right] = \mathbb{E}\left[f(X_t) | X_s\right] \text{ a.s. for all } 0 \leq s < t \tag{17.1}$$

and for every $f \in \mathcal{S}_b$.

If Eq. (17.1) holds then by the factorization Lemma 6.40 there exists $F \in \mathcal{S}_b$ such that $F(X_s) = \mathbb{E}\left[f(X_t) | X_s\right]$. Conversely if we want to verify Eq. (17.1) it suffices to find an $F \in \mathcal{S}_b$ such that $\mathbb{E}_{\mathcal{B}_s} f(X_t) = F(X_s)$ a.s. This is because, by the tower property of conditional expectation,

$$\mathbb{E}_{\mathcal{B}_s} f(X_t) = F(X_s) = \mathbb{E}\left[F(X_s) | X_s\right] = \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_s} f(X_t) | X_s\right] = \mathbb{E}\left[f(X_t) | X_s\right] \text{ a.s.} \tag{17.2}$$

Poetically speaking as stochastic process with the Markov property is forgetful in the sense that knowing the positions of the process up to some time $s \leq t$ does not give any more information about the position of the process, X_t, at time t than knowing where the process was at time s. We will in fact show (Theorem 17.4 below) that given X_s what the process did before time s is independent of the what it will do after time s.

Lemma 17.2. *If $\{X_t\}_{t\geq 0}$ satisfies the Markov property relative to the filtration $\{\mathcal{B}_t\}_{t\geq 0}$ it also satisfies the Markov property relative to $\left\{\mathcal{B}_t^X = \sigma(X_s : s \leq t)\right\}_{t\geq 0}$.*

Proof. It is clear that $\{X_t\}_{t\in T}$ is $\mathcal{B}_t^X$ – adapted and that $\sigma(X_s) \subset \mathcal{B}_s^X \subset \mathcal{B}_s$ for all $s \in T$. Therefore using the tower property of conditional expectation we have,

$$\mathbb{E}_{\mathcal{B}_s^X} f(X_t) = \mathbb{E}_{\mathcal{B}_s^X} \mathbb{E}_{\mathcal{B}_s} f(X_t) = \mathbb{E}_{\mathcal{B}_s^X} \mathbb{E}_{\sigma(X_s)} f(X_t) = \mathbb{E}_{\sigma(X_s)} f(X_t).$$

■

Remark 17.3. If $T = \mathbb{N}_0$, a stochastic process $\{X_n\}_{n\geq 0}$ is Markov iff for all $f \in \mathcal{S}_b$,

$$\mathbb{E}\left[f(X_{m+1}) | \mathcal{B}_m\right] = \mathbb{E}\left[f(X_{m+1}) | X_m\right] \text{ a.s. for all } m \geq 0 \tag{17.3}$$

Indeed if Eq. (17.3) holds for all m, we may use induction on n to show

$$\mathbb{E}\left[f(X_n) | \mathcal{B}_m\right] = \mathbb{E}\left[f(X_n) | X_m\right] \text{ a.s. for all } n \geq m. \tag{17.4}$$

It is clear that Eq. (17.4) holds for $n = m$ and $n = m + 1$. So now suppose Eq. (17.4) holds for a given $n \geq m$. Using Eq. (17.3) with $m = n$ implies

$$\mathbb{E}_{\mathcal{B}_n} f(X_{n+1}) = \mathbb{E}\left[f(X_{n+1}) | X_n\right] = F(X_n)$$

for some $F \in \mathcal{S}_b$. Thus by the tower property of conditional expectations and the induction hypothesis,

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_m} f(X_{n+1}) &= \mathbb{E}_{\mathcal{B}_m} \mathbb{E}_{\mathcal{B}_n} f(X_{n+1}) = \mathbb{E}_{\mathcal{B}_m} F(X_n) = \mathbb{E}\left[F(X_n) | X_m\right] \\ &= \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_n} f(X_{n+1}) | X_m\right] = \mathbb{E}\left[f(X_{n+1}) | X_m\right].\end{aligned}$$

The next theorem and Exercise 17.1 shows that a stochastic process has the Markov property iff it has the property that given its present state, its past and future are independent.

Theorem 17.4 (Markov Independence). *Suppose that $\{X_t\}_{t\in T}$ is an adapted stochastic process with the Markov property and let $\mathcal{F}_s := \sigma(X_t : t \geq s)$ be the future σ – algebra. Then $\mathcal{B}_s$ is independent of $\mathcal{F}_s$ given X_s which we abbreviate as $\mathcal{B}_s \perp\!\!\!\perp_{X_s} \mathcal{F}_s$. In more detail we are asserting the*

$$P(A \cap B | X_s) = P(A | X_s) \cdot P(B | X_s) \text{ a.s.}$$

for all $A \in \mathcal{B}_s$ and $B \in \mathcal{F}_s$ or equivalently that

$$\mathbb{E}\left[FG | X_s\right] = \mathbb{E}\left[F | X_s\right] \cdot \mathbb{E}\left[G | X_s\right] \text{ a.s.} \tag{17.5}$$

for all $F \in (\mathcal{B}_s)_b$ and $G \in (\mathcal{F}_s)_b$ and $s \in T$.

Proof. Suppose first that $G = \prod_{i=1}^{n} g_i(X_{t_i})$ with $s < t_1 < t_2 < \dots < t_n$ and $g_i \in \mathcal{S}_b$. Then by the Markov property and the tower property of conditional expectations,

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_s}[G] &= \mathbb{E}_{\mathcal{B}_s}\left[\prod_{i=1}^{n} g_i(X_{t_i})\right] = \mathbb{E}_{\mathcal{B}_s}\mathbb{E}_{\mathcal{B}_{t_{n-1}}}\left[\prod_{i=1}^{n} g_i(X_{t_i})\right] \\ &= \mathbb{E}_{\mathcal{B}_s}\prod_{i=1}^{n-1} g_i(X_{t_i}) \cdot \mathbb{E}_{\mathcal{B}_{t_{n-1}}} g_n(X_{t_n}) \\ &= \mathbb{E}_{\mathcal{B}_s}\prod_{i=1}^{n-1} g_i(X_{t_i}) \cdot \mathbb{E}\left[g_n(X_{t_n}) | X_{t_{n-1}}\right] = \mathbb{E}_{\mathcal{B}_s}\prod_{i=1}^{n-1} \tilde{g}_i(X_{t_i})\end{aligned}$$

where $\tilde{g}_i = g_i$ for $i < n-1$ and $\tilde{g}_{n-1} = g_{n-1} \cdot g$ where g is chosen so that $\mathbb{E}\left[g_n(X_{t_n}) | X_{t_{n-1}}\right] = g\left(X_{t_{n-1}}\right)$ a.s.. Continuing this way inductively we learn that $\mathbb{E}_{\mathcal{B}_s}[G] = F(X_s)$ a.s. for some $F \in \mathcal{S}_b$ and therefore,

$$\mathbb{E}[G|X_s] = \mathbb{E}[\mathbb{E}_{\mathcal{B}_s} G|X_s] = \mathbb{E}[F(X_s)|X_s] = F(X_s) = \mathbb{E}_{\mathcal{B}_s}[G] \text{ a.s.}$$

Now suppose that $G = g_0(X_s)\prod_{i=1}^{n} g_i(X_{t_i})$ where $g_i \in \mathcal{S}_b$ and $s < t_1 < t_2 < \dots < t_n$ and $F \in (\mathcal{B}_s)_b$. Then

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_s}[FG] &= F \cdot \mathbb{E}_{\mathcal{B}_s} G = F \cdot g_0(X_s)\,\mathbb{E}_{\mathcal{B}_s}\left[\prod_{i=1}^{n} g_i(X_{t_i})\right] \\ &= F \cdot g_0(X_s)\,\mathbb{E}\left[\prod_{i=1}^{n} g_i(X_{t_i}) | X_s\right] \\ &= F \cdot \mathbb{E}\left[g_0(X_s)\prod_{i=1}^{n} g_i(X_{t_i}) | X_s\right] = F \cdot \mathbb{E}[G|X_s].\end{aligned}$$

We may now condition this equation on X_s to arrive at Eq. (17.5) for product functions, G, as above. An application of the multiplicative system Theorem 8.2 may now be used to show that Eq. (17.5) holds for general $G \in (\mathcal{F}_s)_b$. ■

Exercise 17.1. Suppose that $\{X_t\}_{t\geq 0}$ is an adapted stochastic process such that Eq. (17.5) holds for all $F \in (\mathcal{B}_s)_b$ and $G \in (\mathcal{F}_s)_b$ and $s \in T$. Show that $\{X_t\}_{t\in T}$ has the Markov property.

17.1 Markov Processes

If S is a standard Borel space (i.e. S is isomorphic to a Borel subset of $[0,1]$), we may a find regular conditional probability kernels, $Q_{s,t}$ on $(S,\mathcal{S})\times(S,\mathcal{S}) \to [0,1]$ for all $0 \leq s < t$ such that

$$\mathbb{E}[f(X_t)|X_s] = Q_{s,t}(X_s; f) = (Q_{s,t}f)(X_s) \text{ a.s.} \tag{17.6}$$

Moreover by the Markov property if $0 \leq \sigma < s < t$, then

$$\begin{aligned}(Q_{\sigma,t}f)(X_\sigma) &= \mathbb{E}[f(X_t)|X_\sigma] = \mathbb{E}[\mathbb{E}_{\mathcal{B}_s} f(X_t)|X_\sigma] \\ &= \mathbb{E}[(Q_{st}f)(X_s)|X_\sigma] = (Q_{\sigma,s}Q_{st}f)(X_\sigma) \;\; P-\text{a.s.} \\ &= Q_{\sigma,s}(X_\sigma; Q_{s,t}(\cdot; f)),\;\; P-\text{a.s.}\end{aligned}$$

If we let $\mu_t := \mathrm{Law}_P(X_t) : \mathcal{S} \to [0,1]$ for all $t \in T$, we have just shown that for every $f \in \mathcal{S}_b$, that

$$Q_{\sigma,t}(\cdot; f) = Q_{\sigma,s}(\cdot; Q_{s,t}(\cdot; f)) \qquad \mu_\sigma - \text{a.s.} \tag{17.7}$$

In the sequel we want to assume that such kernels exists and that Eq. (17.7) holds for everywhere not just μ_σ – a.s. Thus we make the following definitions.

Definition 17.5 (Markov transition kernels). *We say a collection of probability kernels, $\{Q_{s,t}\}_{0\leq s\leq t<\infty}$, on $S \times S$ are* ***Markov transition kernels*** *if $Q_{s,s}(x,dy) = \delta_x(dy)$ (as an operator $Q_{s,s} = I_{\mathcal{S}_b}$) for all $s \in T$ and the* ***Chapmann-Kolmogorov equations*** *hold;*

$$Q_{\sigma,t} = Q_{\sigma,s}Q_{s,t} \text{ for all } 0 \leq \sigma \leq s \leq t. \tag{17.8}$$

Recall that Eq. (17.8) is equivalent to

$$Q_{\sigma,t}(x,A) = \int_S Q_{\sigma,s}(x,dy)\,Q_{s,t}(y,A) \text{ for all } x \in S \text{ and } A \in \mathcal{S} \tag{17.9}$$

or

$$Q_{\sigma,t}(x; f) = Q_{\sigma,s}(x; Q_{s,t}(\cdot; f)) \text{ for all } x \in S \text{ and } f \in \mathcal{S}_b. \tag{17.10}$$

Thus Markov transition kernels should satisfy Eq. (17.7) everywhere not just almost everywhere.

The reader should keep in mind that $Q_{\sigma,t}(x,A)$ represents the jump probability of starting at x at time σ and ending up in $A \in \mathcal{S}$ at time t. With this in mind, $Q_{\sigma,s}(x,dy)\,Q_{s,t}(y,A)$ intuitively is the probability of jumping from x at time s to y at time t followed by a jump into A at time u. Thus Eq. (17.9) states that averaging these probabilities over the intermediate location (y) of the particle at time t gives the jump probability of starting at x at time s and ending up in $A \in \mathcal{S}$ at time t. This interpretation is rigorously true when S is a finite or countable set.

Definition 17.6 (Markov process). *A* ***Markov process*** *is an adapted stochastic process, $\{X_t : \Omega \to S\}_{t\geq 0}$, with the Markov property such that there are Markov transition kernels $\{Q_{s,t}\}_{0\leq s\leq t<\infty}$, on $S \times S$ such that Eq. (17.6) holds.*

Definition 17.7. *A stochastic process, $\left\{X_t : \Omega \to S := \mathbb{R}^d\right\}_{t\in T}$, has* ***independent increments*** *if for all finite subsets, $\Lambda = \{0 \le t_0 < t_1 < \cdots < t_n\} \subset T$ the random variables $\{X_0\} \cup \left\{X_{t_k} - X_{t_{k-1}}\right\}_{k=1}^n$ are independent. We refer to $X_t - X_s$ for $s < t$ as an* ***increment*** *of X.*

Exercise 17.2. Suppose that $\left\{X_t : \Omega \to S := \mathbb{R}^d\right\}_{t\in T}$ is a stochastic process with independent increments and let $\mathcal{B}_t := \mathcal{B}_t^X$ for all $t \in T$. Show, for all $0 \le s < t$, that $(X_t - X_s)$ is independent of $\mathcal{B}_s^X$ and then use this to show $\{X_t\}_{t\in T}$ is a Markov process with transition kernels defined by $0 \le s \le t$,

$$Q_{s,t}(x, A) := \mathbb{E}\left[1_A(x + X_t - X_s)\right] \text{ for all } A \in \mathcal{S} \text{ and } x \in \mathbb{R}^d. \tag{17.11}$$

You should verify that $\{Q_{s,t}\}_{0\le s\le t}$ are indeed Markov transition kernels, i.e. satisfy the Chapmann-Kolmogorov equations.

Example 17.8 (Random Walks). Suppose that $\left\{\xi_n : \Omega \to S := \mathbb{R}^d\right\}_{n=0}^\infty$ are independent random vectors and $X_m := \sum_{k=0}^m \xi_k$ and $\mathcal{B}_m := \sigma(\xi_0, \dots, \xi_m)$ for each $m \in T = \mathbb{N}_0$. Then $\{X_m\}_{m\ge0}$ has independent increments and therefore has the Markov property with Markov transition kernels being given by

$$\begin{aligned} Q_{s,t}(x, f) &= \mathbb{E}\left[f(x + X_t - X_s)\right] \\ &= \mathbb{E}\left[f\left(x + \sum_{s<k\le t} \xi_k\right)\right] \end{aligned}$$

or in other words,

$$Q_{s,t}(x, \cdot) = \mathrm{Law}_P\left(x + \sum_{s<k\le t} \xi_k\right).$$

The one step transition kernels are determined by

$$(Q_{n,n+1} f)(x) = \mathbb{E}\left[f(x + \xi_{n+1})\right] \text{ for } n \in \mathbb{N}_0.$$

Exercise 17.3. Let us now suppose that $\{\xi_n : \Omega \to S\}_{n=0}^\infty$ are independent random functions where $(S, \mathcal{S})$ is a general a measurable space, $\mathcal{B}_n := \sigma(\xi_0, \xi_1, \dots, \xi_n)$ for $n \ge 0$, $u_n : S \times S \to S$ are measurable functions for $n \ge 1$, and $X_n : \Omega \to S$ for $n \in \mathbb{N}_0$ are defined by $X_0 = \xi_0$ and then inductively for $n \ge 1$ by

$$X_{n+1} = u_{n+1}(X_n, \xi_{n+1}) \text{ for } n \ge 0.$$

Convince yourself that for $0 \le m < n$ there is a measurable function, $\varphi_{n,m} : S^{n-m+1} \to S$ determined by the $\{u_k\}$ such that $X_n = \varphi_{n,m}(X_m, \xi_{m+1}, \dots, \xi_n)$. (You need not write the proof of this assertion in your solution.) In particular, $X_n = \varphi_{n,0}(\xi_0, \dots, \xi_n)$ is $\mathcal{B}_n/\mathcal{S}$ – measurable so that $X = \{X_n\}_{n\ge0}$ is adapted. Show $\{X_n\}_{n\ge0}$ is a Markov process with transition kernels,

$$Q_{m,n}(x, \cdot) = \mathrm{Law}_P(\varphi_{n,m}(x, \xi_{m+1}, \dots, \xi_n)) \text{ for all } 0 \le m \le n$$

where (by definition) $Q_{m,m}(x, \cdot) = \delta_x(\cdot)$. Please explicitly verify that $\{Q_{m,n}\}_{0\le m\le n}$ are Markov transition kernels, i.e. satisfy the Chapmann-Kolmogorov equations.

Remark 17.9. Suppose that $T = \mathbb{N}_0$ and $\{Q_{m,n} : 0 \le m \le n\}$ are Markov transition kernels on $S \times S$. Since

$$Q_{m,n} = Q_{m,m+1} Q_{m+1,m+2} \dots Q_{n-1,n}, \tag{17.12}$$

it follows that the $Q_{m,n}$ are uniquely determined by knowing the one step transition kernels, $\{Q_{n,n+1}\}_{n=0}^\infty$. Conversely if $\{Q_{n,n+1}\}_{n=0}^\infty$ are arbitrarily given probability kernels on $S \times S$ and $Q_{m,n}$ are defined as in Eq. (17.12), then the resulting $\{Q_{m,n} : 0 \le m \le n\}$ are Markov transition kernels on $S \times S$. Moreover if S is a countable set, then we may let

$$q_{m,n}(x, y) := Q_{m,n}(x, \{y\}) = P(X_n = y | X_m = x) \text{ for all } x, y \in S \tag{17.13}$$

so that

$$Q_{m,n}(x, A) = \sum_{y\in A} q_{m,n}(x, y).$$

In this case it is easily checked that

$$\begin{aligned} &q_{m,n}(x, y) \\ &\quad= \sum_{x_i\in S : m<i<n} q_{m,m+1}(x, x_{m+1})\, q_{m+1,m+2}(x_{m+1}, x_{m+2}) \dots q_{n-1,n}(x_{n-1}, y). \end{aligned} \tag{17.14}$$

The reader should observe that this is simply matrix multiplication!

Exercise 17.4 (Polya's Urn). Suppose that an urn contains r red balls and g green balls. At each time $(t \in T = \mathbb{N}_0)$ we draw a ball out, then replace it and add c more balls of the color drawn. It is reasonable to model this as a Markov process with $S := \mathbb{N}_0 \times \mathbb{N}_0$ and $X_n := (r_n, g_n) \in S$ being the number of red and green balls respectively in the earn at time n. Find

$$q_{n,n+1}((r, g), (r', g')) = P(X_{n+1} = (r', g') | X_n = (r, g))$$

for this model.

Theorem 17.10 (Finite Dimensional Distributions). *Suppose that* $X = \{X_t\}_{t\geq 0}$ *is a Markov process with Markov transition kernels* $\{Q_{s,t}\}_{0\leq s\leq t}$. *Further let* $\nu := \mathrm{Law}_P(X_0)$, *then for all* $0 = t_0 < t_1 < t_2 < \cdots < t_n$ *we have*

$$\mathrm{Law}_P\left(X_{t_0}, X_{t_1}, \dots, X_{t_n}\right)(dx_0, dx_1, \dots, dx_n) = d\nu(x_0)\prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \tag{17.15}$$

or equivalently,

$$\mathbb{E}\left[f\left(X_{t_0}, X_{t_1}, \dots, X_{t_n}\right)\right] = \int_{S^{n+1}} f(x_0, x_1, \dots, x_n)\, d\nu(x_0)\prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \tag{17.16}$$

for all $f \in \mathcal{S}_b^{\otimes(n+1)}$.

Proof. Because of the multiplicative system Theorem 8.2, it suffices to prove Eq. (17.16) for functions of the form $f(x_1, \dots, x_n) = \prod_{i=0}^n f_i(x_i)$ where $f_i \in \mathcal{S}_b$. The proof is now easily completed by induction on n. It is true for $n = 0$ by definition of ν. Now assume it is true for some $n-1 \geq 0$. We then have, making use of the inductive hypothesis, that

$$\begin{aligned}
&\mathbb{E}\left[f\left(X_{t_0}, X_{t_1}, \dots, X_{t_n}\right)\right] \\
&= \mathbb{E}\mathbb{E}_{\mathcal{B}_{t_{n-1}}}\left[\prod_{i=0}^n f_i\left(X_{t_i}\right)\right] \\
&= \mathbb{E}\left[Q_{t_{n-1},t_n}\left(X_{t_{n-1}}, f_n\right)\cdot\prod_{i=0}^{n-1} f_i\left(X_{t_i}\right)\right] \\
&= \int_{S^n} Q_{t_{n-1},t_n}(x_{n-1}, f_n)\cdot\prod_{i=0}^{n-1} f_i(x_i)\, d\nu(x_0)\prod_{i=1}^{n-1} Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \int_{S^n}\left[\int_S Q_{t_{n-1},t_n}(x_{n-1}, dx_n)\, f_n(x_n)\right]\cdot\prod_{i=0}^{n-1} f_i(x_i)\, d\nu(x_0)\prod_{i=1}^{n-1} Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \int_{S^{n+1}} f(x_0, x_1, \dots, x_n)\, d\nu(x_0)\prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i)
\end{aligned}$$

as desired. ∎

Theorem 17.11 (Existence of Markov processes). *Suppose that* $\{Q_{s,t}\}_{0\leq s\leq t}$ *are Markov transition kernels on a standard Borel space,* $(S, \mathcal{S})$. *Let* $\Omega := S^T$, $X_t : \Omega \to S$ *be the projection map,* $X_t(\omega) = \omega(t)$ *and* $\mathcal{B}_t = \mathcal{B}_t^X = \sigma(X_s : s \leq t)$ *for all* $t \in T$ *and* $\mathcal{B} := \mathcal{S}^{\otimes T} = \sigma(X_t : t \in T)$. *Then to each probability measure,* ν, *on* $(S, \mathcal{S})$ *there exists a unique probability measure* P_ν *on* $(\Omega, \mathcal{B})$ *such that 1)* $\mathrm{Law}_{P_\nu}(X_0) = \nu$ *and 2)* $\{X_t\}_{t\geq 0}$ *is a Markov process having* $\{Q_{s,t}\}_{0\leq s\leq t}$ *as its Markov transition kernels.*

Proof. This is mainly an exercise in applying Kolmogorov's extension Theorem 17.54 as described in the appendix to this chapter. I will only briefly sketch the proof here.

For each $\Lambda = \{0 = t_0 < t_1 < t_2 < \cdots < t_n\} \subset T$, let P_Λ be the measure on $\left(S^{n+1}, \mathcal{S}^{\otimes(n+1)}\right)$ defined by

$$dP_\Lambda(x_0, x_1, \dots, x_n) = d\nu(x_0)\prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i).$$

Using the Chapman-Kolmogorov equations one shows that the $\{P_\Lambda\}_{\Lambda\subset_f T}$ ($\Lambda \subset_f T$ denotes a finite subset of T) are consistently defined measures as described in the statement of Theorem 17.54. Therefore it follows by an application of that theorem that there exists a unique measure P_ν on $(\Omega, \mathcal{B})$ such that

$$\mathrm{Law}_{P_\nu}(X|_\Lambda) = P_\Lambda \text{ for all } \Lambda \subset_f T. \tag{17.17}$$

In light of Theorem 17.10, in order to finish the proof we need only show that $\{X_t\}_{t\geq 0}$ is a Markov process having $\{Q_{s,t}\}_{0\leq s\leq t}$ as its Markov transition kernels. Since if this is this case it finite dimensional distributions must be given as in Eq. (17.17) and therefore P_ν is uniquely determined. So let us now verify the desired Markov property.

Again let $\Lambda = \{0 = t_0 < t_1 < t_2 < \cdots < t_n\} \subset T$ with $t_{n-1} = s < t = t_n$ and suppose that $f(x_0, \dots, x_n) = h(x_0, \dots, x_{n-1})\, g(x_n)$ with $h \in \mathcal{S}_b^{\otimes n}$ and $f \in \mathcal{S}_b$. By the definition of P_ν we then have (writing $\mathbb{E}_\nu$ for $\mathbb{E}_{P_\nu}$),

$$\begin{aligned}
&\mathbb{E}_\nu\left[h\left(X_{t_0}, X_{t_1}, \dots, X_{t_{n-1}}\right) g(X_t)\right] \\
&= \int_{S^{n+1}} h(x_0, x_1, \dots, x_{n-1})\, g(x_n)\, d\nu(x_0)\prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \int_{S^n} h(x_0, x_1, \dots, x_{n-1})\, Q_{t_{n-1},t_n}(x_{n-1}, g)\, d\nu(x_0)\prod_{i=1}^{n-1} Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \mathbb{E}_\nu\left[h\left(X_{t_0}, X_{t_1}, \dots, X_{t_{n-1}}\right) Q_{t_{n-1},t_n}\left(X_{t_{n-1}}, g\right)\right] \\
&= \mathbb{E}_\nu\left[h\left(X_{t_0}, X_{t_1}, \dots, X_{t_{n-1}}\right) Q_{s,t}(X_s, g)\right]
\end{aligned}$$

It then follows by an application of the multiplicative system theorem that

$$\mathbb{E}_\nu\left[Hg(X_t)\right] = \mathbb{E}_\nu\left[HQ_{s,t}(X_s, g)\right] \text{ for all } H \in (\mathcal{B}_s)_b$$

and therefore that

$$\mathbb{E}_\nu\left[g(X_t)\,|\mathcal{B}_s\right] = Q_{s,t}(X_s, g) \text{ a.s.}$$

■

We are now going to specialize to the more manageable class of "time homogeneous" Markov processes.

Definition 17.12. *We say that a collection of Markov transition kernels, $\{Q_{s,t}\}_{0\le s\le t}$ are **time homogeneous** if $Q_{s,t} = Q_{0,t-s}$ for all $0 \le s \le t$. In this case we usually let $Q_t := Q_{0,t-s}$. The condition that $Q_{s,s}(x,\cdot) = \delta_x$ now reduces to $Q_0(x,\cdot) = \delta_x$ and the Chapmann-Kolmogorov equations reduce to*

$$Q_sQ_t = Q_{s+t} \text{ for all } s,t \ge 0, \tag{17.18}$$

i.e.

$$\int_S Q_s(x,dy)\,Q_t(y,A) = Q_{s+t}(x,A) \text{ for all } s,t\ge 0,\ x\in S, \text{ and } A\in\mathcal{S}. \tag{17.19}$$

*A collection of operators $\{Q_t\}_{t\ge 0}$ with $Q_0 = Id$ satisfying Eq. (17.18) is called a **one parameter semi-group.***

Definition 17.13. *A Markov process is **time homogeneous** if it has time homogeneous Markov transition kernels. In this case we will have,*

$$\mathbb{E}[f(X_t)|\mathcal{B}_s] = Q_{t-s}(X_s,f) = (Q_{t-s}f)(X_s) \text{ a.s.} \tag{17.20}$$

for all $0\le s\le t$ and $f\in\mathcal{S}_b$.

Theorem 17.14 (The time homogeneous Markov property). *Suppose that $(S,\mathcal{S})$ is a measurable space, $Q_t : S\times\mathcal{S}\to[0,1]$ are time homogeneous Markov transition kernels, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\ge 0}\right)$ is a filtered measure space, $\{X_t:\Omega\to S\}_{t\ge 0}$ are adapted functions, and for each $x\in S$ there exists a probability measure, P_x on $(\Omega,\mathcal{B})$ such that;*

1. *$X_0(\omega) = x$ for P_x – a.e. ω and*
2. *$\{X_t\}_{t\ge 0}$ is a time homogeneous Markov process with transition kernels $\{Q_t\}_{t\ge 0}$ relative to P_x.*

Let us further suppose that P is any probability measure on $(\Omega,\mathcal{B})$ such that $\{X_t\}_{t\ge 0}$ is a time homogeneous Markov process with transition kernels being $\{Q_t\}$. Then for all $F\in\mathcal{S}_b^{\otimes T}$ and $t\ge 0$ we have that $S\ni x\to\mathbb{E}_xF(X)$ is $\mathcal{S}/\mathcal{B}_{\mathbb{R}}$ – measurable and

$$\mathbb{E}_P[F(X_{t+\cdot})|\mathcal{B}_t] = \mathbb{E}_P[F(X_{t+\cdot})|X_t] = \mathbb{E}_{X_t}[F(X)] \quad P \text{ – a.s.} \tag{17.21}$$

***Warning:** In this equation $\mathbb{E}_{X_t}$ does not denote $\mathbb{E}_{\sigma(X_t)} = \mathbb{E}[\cdot|X_t]$ but instead[1] it means the composition of X_t with the function $S\ni x\to\mathbb{E}_x[F(X)]\in\mathbb{R}$. In more detail we are saying,*

$$\begin{aligned}\mathbb{E}_P[F(X_{t+\cdot})|\mathcal{B}_t](\omega) &= \mathbb{E}_{X_t(\omega)}[F(X)]\\ &= \int_\Omega F(X(\omega'))\ P_{X_t(\omega)}(d\omega').\end{aligned}$$

Proof. Let $\nu_t = \mathrm{Law}_P(X_t)$. If $g\in\mathcal{S}_b$, $f\in\mathcal{S}_b^{\otimes(n+1)}$, and $F(X) := f(X_{t_0},\dots,X_{t_n})$, then

$$\begin{aligned}\mathbb{E}_P[g(X_t)F(X_{t+\cdot})] &= \mathbb{E}_P[g(X_t)f(X_{t_0+t},\dots,X_{t_n+t})]\\ &= \int g(x_0)f(x_0,\dots,x_n)\,d\nu_t(x_0)\prod_{j=1}^n Q_{t_j-t_{j-1}}(x_{j-1},dx_j)\\ &= \int d\nu_t(x_0)\,g(x_0)\,\mathbb{E}_{x_0}f(X_{t_0},\dots,X_{t_n})\\ &= \int d\nu_t(x_0)\,g(x_0)\,\mathbb{E}_{x_0}F(X) = \mathbb{E}_x[g(X_t)\,\mathbb{E}_{X_t}F(X)].\end{aligned}$$

An application of the multiplicative systems Theorem 8.2 shows this equation is valid for all $F\in\mathcal{S}_b^{\otimes T}$ and this tells us that

$$\mathbb{E}_P[F(X_{t+\cdot})|X_t] = \mathbb{E}_{X_t}F(X)\ \ P_x - \text{a.s.}$$

for all $F\in\mathcal{S}_b^{\otimes T}$.

Now suppose that $G\in(\mathcal{B}_t)_b$ and $F\in\mathcal{S}_b^{\otimes T}$. As $F(X_{t+\cdot})\in(\mathcal{F}_t)_b$ it follows by Theorem 17.4 that

$$\mathbb{E}_P[G\cdot F(X_{t+\cdot})|X_t] = \mathbb{E}_P[G|X_t]\cdot\mathbb{E}_P[F(X_{t+\cdot})|X_t] \text{ a.s.}$$

Thus we may conclude that

$$\begin{aligned}\mathbb{E}_P[G\cdot F(X_{t+\cdot})] &= \mathbb{E}_P[\mathbb{E}_P[G|X_t]\cdot\mathbb{E}_P[F(X_{t+\cdot})|X_t]]\\ &= \mathbb{E}_P[\mathbb{E}_P[G|X_t]\cdot\mathbb{E}_{X_t}F(X)]\\ &= \mathbb{E}_P[\mathbb{E}_P[\mathbb{E}_{X_t}F(X)\cdot G|X_t]] = \mathbb{E}_P[\mathbb{E}_{X_t}[F(X)]\cdot G].\end{aligned}$$

This being valid for all $G\in(\mathcal{B}_t)_b$ is equivalent to Eq. (17.21). ■

[1] Unfortunately we now have a lot of different meanings for $\mathbb{E}_\xi$ depending on what ξ happens to be. So if $\xi = P$ is a measure then $\mathbb{E}_P$ stands for expectation relative to P. If $\xi = \mathcal{G}$ is a σ – algebra it stands for conditional expectation relative to $\mathcal{G}$ and a given probability measure which not indicated in the notation. Finally if $x\in S$ we are writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$.

Remark 17.15. Admittedly Theorem 17.14 is a bit hard to parse on first reading. Therefore it is useful to rephrase what it says in the case that the state space, S, is finite or countable and $x \in S$ and $t > 0$ are such that $P(X_t = x) > 0$. Under these additional hypothesis we may combine Theorems 17.4 and 17.14 to find $\{X_s\}_{s \le t}$ and $\{X_s\}_{s \ge t}$ are $P(\cdot|X_t = x)$ – independent and moreover,

$$\mathrm{Law}_{P(\cdot|X_t=x)}(X_{t+\cdot}) = \mathrm{Law}_{P_x}(X_\cdot). \tag{17.22}$$

Last assertion simply states that given $X_t = x$ the process, X, after time t behaves just like the process starting afresh from x.

17.2 Discrete Time Homogeneous Markov Processes

The proof of the following easy lemma is left to the reader.

Lemma 17.16. *If $Q_n : S \times \mathcal{S} \to [0,1]$ for $n \in \mathbb{N}_0$ are time homogeneous Markov kernels then $Q_n = Q^n$ where $Q := Q_1$ and $Q^0 := I$. Conversely if Q is a probability kernel on $S \times S$ then $Q_n := Q^n$ for $n \in \mathbb{N}_0$ are time homogeneous Markov kernels.*

Example 17.17 (Random Walks Revisited). Suppose that $\xi_0 : \Omega \to S := \mathbb{R}^d$ is independent of $\left\{\xi_n : \Omega \to S := \mathbb{R}^d\right\}_{n=1}^\infty$ which are now assumed to be i.i.d. If $X_m = \sum_{k=0}^m \xi_k$ is as in Example 17.8, then $\{X_m\}_{m \ge 0}$ is a time homogeneous Markov process with

$$Q_m(x, \cdot) = \mathrm{Law}_P(X_m - X_0)$$

and the one step transition kernel, $Q = Q_1$, is given by

$$Qf(x) = Q(x, f) = \mathbb{E}[f(x + \xi_1)] = \int_S f(x + y)\, d\rho(y)$$

where $\rho := \mathrm{Law}_P(\xi_1)$. For example if $d = 1$ and $P(\xi_i = 1) = p$ and $P(\xi_i = -1) = q := 1 - p$ for some $0 \le p \le 1$, then we may take $S = \mathbb{Z}$ and we then have

$$Qf(x) = Q(x, f) = pf(x + 1) + qf(x - 1).$$

Example 17.18 (Ehrenfest Urn Model). Let a beaker filled with a particle fluid mixture be divided into two parts A and B by a semipermeable membrane. Let $X_n = (\#$ of particles in $A)$ which we assume evolves by choosing a particle at random from $A \cup B$ and then replacing this particle in the opposite bin from which it was found. Modeling $\{X_n\}$ as a Markov process we find,

$$P(X_{n+1} = j \mid X_n = i) = \begin{cases} 0 & \text{if } j \notin \{i-1, i+1\} \\ \frac{i}{N} & \text{if } \quad j = i - 1 \\ \frac{N-i}{N} & \text{if } \quad j = i + 1 \end{cases} =: q(i, j)$$

As these probabilities do not depend on n, $\{X_n\}$ is a time homogeneous Markov chain.

Exercise 17.5. Consider a rat in a maze consisting of 7 rooms which is laid out as in the following figure.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & & \end{bmatrix}$$

In this figure rooms are connected by either vertical or horizontal adjacent passages only, so that 1 is connected to 2 and 4 but not to 5 and 7 is only connected to 4. At each time $t \in \mathbb{N}_0$ the rat moves from her current room to one of the adjacent rooms with equal probability (the rat always changes rooms at each time step). Find the one step 7×7 transition matrix, q, with entries given by $q(i, j) := P(X_{n+1} = j | X_n = i)$, where X_n denotes the room the rat is in at time n.

Solution to Exercise (17.5). The rat moves to an adjacent room from nearest neighbor locations probability being $1/D$ where D is the number of doors in the room where the rat is currently located. The transition matrix is therefore,

$$q = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}. \tag{17.23}$$

and the corresponding jump diagram is given in Figure 17.1.

Exercise 17.6 (2 - step MC). Consider the following simple (i.e. no-brainer) two state "game" consisting of moving between two sites labeled 1 and 2. At each site you find a coin with sides labeled 1 and 2. The probability of flipping a 2 at site 1 is $a \in (0, 1)$ and a 1 at site 2 is $b \in (0, 1)$. If you are at site i at time n, then you flip the coin at this site and move or stay at the current site as indicated by coin toss. We summarize this scheme by the "jump diagram" of Figure 17.2. It is reasonable to suppose that your location, X_n, at time n is modeled by a

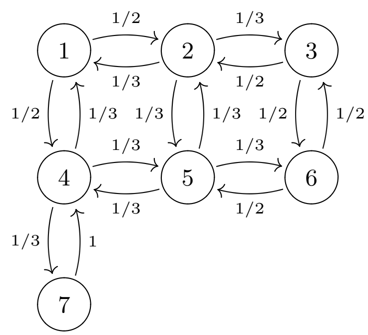


Fig. 17.1. The jump diagram for our rat in the maze.

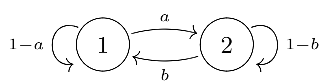


Fig. 17.2. The generic jump diagram for a two state Markov chain.

Markov process with state space, $S = \{1, 2\}$. Explain (briefly) why this is a time homogeneous chain and find the one step transition probabilities,

$$q(i,j) = P(X_{n+1} = j | X_n = i) \text{ for } i, j \in S.$$

Use your result and basic linear (matrix) algebra to compute, $\lim_{n\to\infty} P(X_n = 1)$. Your answer should be independent of the possible starting distributions, $\nu = (\nu_1, \nu_2)$ for X_0 where $\nu_i := P(X_0 = i)$.

17.3 Continuous time homogeneous Markov processes

An analogous (to Lemma 17.16) "infinitesimal description" of time homogeneous Markov kernels in the continuous time case can involve a considerable number of technicalities. Nevertheless, in this section we are going to ignore these difficulties in order to give a general impression of how the story goes. We will cover more precisely the missing details later.

So let $\{Q_t\}_{t\in\mathbb{R}_+}$ be time homogeneous collection of Markov transition kernels. We define the **infinitesimal generator** of $\{Q_t\}_{t\geq 0}$ by,

$$Af := \frac{d}{dt}|_{0^+} Q_t f = \lim_{t\downarrow 0} \frac{Q_t f - f}{t}. \tag{17.24}$$

For now we make the (often unreasonable assumption) that the limit in Eq. (17.24) holds for all $f \in \mathcal{S}_b$. This assumption is OK when S is a finite or

sometimes even a countable state space. For more complicated states spaces we will have to restrict the set of $f \in \mathcal{S}_b$ that we consider when computing Af by Eq. (17.24). You should get a feeling for this issue by working through Exercise 17.8 which involves "Brownian motion."

Since we are assuming $\frac{d}{dt}|_{0^+} Q_t f$ exists we must also have $\lim_{t\downarrow 0} Q_t f = f$. More generally for $t, h > 0$, using the semi-group property, we have

$$Q_{t+h} - Q_t = (Q_h - I) Q_t = Q_t (Q_h - I) \tag{17.25}$$

and therefore,

$$Q_{t+h} f - Q_t f = (Q_h - I) Q_t f \to 0 \text{ as } h \downarrow 0$$

so that $Q_t f$ is right continuous. Similarly,

$$Q_{t-h} - Q_t = -(Q_h - I) Q_{t-h} = -Q_{t-h} (Q_h - I) \tag{17.26}$$

and

$$|Q_{t-h} f - Q_t f| = |Q_{t-h} (Q_h - I) f| \leq |Q_{t-h} |(Q_h - I) f|| \leq \sup_S |(Q_h - I) f|$$

which will tend to zero as $h \downarrow 0$ provided $Q_h f \to f$ uniformly (another fantasy in general). With this as justification we will assume that $t \to Q_t f$ is continuous in t.

Taking Eq. (17.25) divided by h and Eq. (17.26) divided by $-h$ and then letting $h \downarrow 0$ implies,

$$\left(\frac{d}{dt}\right)_+ Q_t f = A Q_t f = Q_t A f$$

and

$$\left(\frac{d}{dt}\right)_- Q_t = A Q_t = Q_t A.$$

where $\left(\frac{d}{dt}\right)_+$ and $\left(\frac{d}{dt}\right)_-$ denote the right and left derivatives at t. So in principle we can expect that $\{Q_t\}_{t\geq 0}$ is uniquely determined by its infinitesimal generator A by solving the differential equation,

$$\frac{d}{dt} Q_t = A Q_t = Q_t A \text{ with } Q_0 = Id. \tag{17.27}$$

Assuming all of this works out as sketched, it is now reasonable to denote Q_t by e^{tA}. Let us now give a few examples to illustrate the discussion above.

Example 17.19. Suppose that $S = \{1, 2, \ldots, n\}$ and Q_t is a Markov-semi-group with infinitesimal generator, A, so that $\frac{d}{dt} Q_t = A Q_t = Q_t A$. By assumption

$Q_t(i,j) \geq 0$ for all $i,j \in S$ and $\sum_{j=1}^n Q_t(i,j) = 1$ for all $i \in S$. We may write this last condition as $Q_t\mathbf{1} = \mathbf{1}$ for all $t \geq 0$ where $\mathbf{1}$ denotes the vector in $\mathbb{R}^n$ with all entries being 1. Differentiating $Q_t\mathbf{1} = \mathbf{1}$ at $t = 0$ shows that $A\mathbf{1} = 0$, i.e. $\sum_{j=1}^n A_{ij} = 0$ for all $i \in S$. Since

$$A_{ij} = \lim_{t \downarrow 0} \frac{Q_t(i,j) - \delta_{ij}}{t}$$

if $i \neq j$ we will have,

$$A_{ij} = \lim_{t \downarrow 0} \frac{Q_t(i,j)}{t} \geq 0.$$

Thus we have shown the infinitesimal generator, A, of Q_t must satisfy $A_{ij} \geq 0$ for all $i \neq j$ and $\sum_{j=1}^n A_{ij} = 0$ for all $i \in S$. You are asked to prove the converse in Exercise 17.7. So an explicit example of an infinitesimal generator when $S = \{1,2,3\}$ is

$$A = \begin{pmatrix} -3 & 1 & 2 \\ 4 & -6 & 2 \\ 7 & 1 & -8 \end{pmatrix}.$$

Exercise 17.7. Suppose that $S = \{1,2,\dots,n\}$ and A is a matrix such that $A_{ij} \geq 0$ for $i \neq j$ and $\sum_{j=1}^n A_{ij} = 0$ for all i. Show

$$Q_t = e^{tA} := \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n \tag{17.28}$$

is a time homogeneous Markov kernel.

Hints: 1. To show $Q_t(i,j) \geq 0$ for all $t \geq 0$ and $i,j \in S$, write $Q_t = e^{-t\lambda} e^{t(\lambda I + A)}$ where $\lambda > 0$ is chosen so that $\lambda I + A$ has only non-negative entries. 2. To show $\sum_{j \in S} Q_t(i,j) = 1$, compute $\frac{d}{dt} Q_t\mathbf{1}$.

Example 17.20 (Poisson Process). By Exercise 17.2, it follows that Poisson process, $\{N_t \in S := \mathbb{N}_0\}_{t \geq 0}$ with intensity λ has the Markov property. For all $0 \leq s \leq t$ we have,

$$\begin{aligned} P(N_t = y | N_s = x) &= P(N_s + N_t - N_s = y | N_s = x) \\ &= P(N_s + N_t - N_s = y | N_s = x) \\ &= P(N_t - N_s = y - x | N_s = x) \\ &= 1_{y \geq x} \frac{(\lambda(t-s))^{y-x}}{(y-x)!} e^{-\lambda(t-s)} =: q_{t-s}(x,y). \end{aligned}$$

With this notation it follows that

$$P(f(N_t) | N_s) = (Q_{t-s} f)(N_s)$$

where

$$\begin{aligned} Q_t f(x) &= \sum_{y \in S} q_t(x,y) f(y) \\ &= \sum_{y \in S} 1_{y \geq x} \frac{(\lambda t)^{y-x}}{(y-x)!} e^{-\lambda t} f(y) \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} f(x+n). \end{aligned} \tag{17.29}$$

In particular $\{N_t\}_{t \geq 0}$ is a time homogeneous Markov process. It is easy (but technically unnecessary) to directly verify the semi-group property;

$$(q_t q_s)(x,z) := \sum_{y \in S} q_t(x,y) q_s(y,z) = q_{s+t}(x,z).$$

This can be done using the binomial theorem as follows;

$$\begin{aligned} \sum_{y \in S} q_t(x,y) q_s(y,z) &= \sum_{z \in S} 1_{y \geq x} \frac{(\lambda t)^{y-x}}{(y-x)!} e^{-\lambda t} \cdot 1_{z \geq y} \frac{(\lambda s)^{z-y}}{(z-y)!} e^{-\lambda s} \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \cdot 1_{z \geq x+n} \frac{(\lambda s)^{z-x-n}}{(z-x-n)!} e^{-\lambda s} \\ &= 1_{z \geq x} e^{-\lambda(t+s)} \sum_{n=0}^{z-x} \frac{(\lambda t)^n}{n!} \frac{(\lambda s)^{z-x-n}}{(z-x-n)!} \\ &= 1_{z \geq x} e^{-\lambda(t+s)} \frac{(\lambda(t+s))^{z-x}}{(z-x)!} = q_{s+t}(x,z). \end{aligned}$$

To identify infinitesimal generator, $A = \frac{d}{dt}|_{0^+} Q_t$, in this example observe that

$$\begin{aligned} \frac{d}{dt} Q_t f(x) &= \frac{d}{dt} \left[e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f(x+n) \right] \\ &= -\lambda Q_t f(x) + \lambda e^{-\lambda t} \sum_{n=0}^{\infty} \frac{n(\lambda t)^{n-1}}{n!} f(x+n) \\ &= -\lambda Q_t f(x) + \lambda e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f(x+n+1) \\ &= -\lambda Q_t f(x) + \lambda (Q_t f)(x+1) \\ &= Q_t [\lambda f(\cdot + 1) - \lambda f(\cdot)](x) \end{aligned}$$

and hence

$$Af(x) = \lambda (f(x+1) - f(x)).$$

Finally let us try to solve Eq. (17.27) in order to recover Q_t from A. Formally we can hope that $Q_t = e^{tA}$ where e^{tA} is given as its power series expansion. To simplify the computation it convenient to write $A = \lambda(T - I)$ where $If = f$ and $Tf = f(\cdot + 1)$. Since I and T commute we further expect

$$e^{tA} = e^{\lambda t(T-I)} = e^{-\lambda t I} e^{\lambda t T} = e^{-\lambda t} e^{\lambda t T}$$

where

$$\begin{aligned}\left(e^{\lambda t T} f\right)(x) &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} (T^n f)(x) \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f(x+n).\end{aligned}$$

Putting this all together we find

$$\left(e^{tA} f\right)(x) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f(x+n)$$

which is indeed in agreement with $Q_t f(x)$ as we saw in Eq. (17.29).

Definition 17.21 (Brownian Motion). *Let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}, P\right)$ be a filtered probability space. A real valued adapted process, $\{X_t : \Omega \to S = \mathbb{R}\}_{t\in\mathbb{R}_+}$, is called a* ***Brownian motion*** *if;*

1. *$\{X_t\}_{t\in\mathbb{R}_+}$ has independent increments,*
2. *for $0 \le s < t$, $X_t - X_s \overset{d}{=} N(0, t-s)$, i.e. $X_t - X_s$ is a normal mean zero random variable with variance $(t-s)$,*
3. *$t \to X_t(\omega)$ is continuous for all $\omega \in \Omega$.*

Exercise 17.8 (Brownian Motion). Assuming a Brownian motion $\{B_t\}_{t\ge0}$ exists as described in Definition 17.21 show;

1. The process is a time homogeneous Markov process with transition kernels given by;
$$Q_t(x, dy) = q_t(x, y)\, dy \tag{17.30}$$
where
$$q_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y-x|^2}. \tag{17.31}$$
2. Show by direct computation that $Q_t Q_s = Q_{t+s}$ for all $s, t > 0$. **Hint:** one of the many ways to do this it to use basic facts you have already proved about sums of independent Gaussian random variables along with the identity,
$$(Q_t f)(x) = \mathbb{E}\left[f\left(x + \sqrt{t} Z\right)\right],$$
where $Z \overset{d}{=} N(0,1)$.
3. Show by direct computation that $q_t(x, y)$ satisfies the **heat equation**,
$$\frac{d}{dt} q_t(x, y) = \frac{1}{2} \frac{d^2}{dx^2} q_t(x, y) = \frac{1}{2} \frac{d^2}{dy^2} q_t(x, y) \text{ for } t > 0.$$
4. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function with compact support. Show
$$\frac{d}{dt} Q_t f = A Q_t f = Q_t A f \text{ for all } t > 0,$$
where
$$Af(x) = \frac{1}{2} f''(x).$$

By combining Exercise 17.8 with Theorem 17.11 proves the following corollary.

Corollary 17.22. *There exits a Markov process $\{B_t\}_{t\ge0}$ satisfying properties 1. and 2. of Definition 17.21.*

To get the path continuity property of Brownian motion requires additional arguments which we will do in a number of ways later, see Theorems 26.3, 26.7, and 34.5. Modulo technical details, Exercise 17.8 shows that $A = \frac{1}{2}\frac{d^2}{dx^2}$ is the infinitesimal generator of Brownian motion, i.e. of Q_t in Eqs. (17.30) and (17.31). The technical details we have ignored involve the proper function spaces in which to carry out these computations along with a proper description of the domain of the operator A. We will have to postpone these somewhat delicate issues until later. By the way it is no longer necessarily a good idea to try to recover Q_t as $\sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$ in this example since in order for $\sum_{n=0}^{\infty} \frac{t^n}{n!} A^n f$ to make sense one needs to assume that f is a least C^∞ and even this will not guarantee convergence of the sum.

A Lévy process is a generalization of this type a process. A **Lévy process** is a process with independent stationary increments which has right continuous paths.

Example 17.23. If $\{N_t\}_{t\ge0}$ is a Poisson process and $\{B_t\}_{t\ge0}$ is a Brownian motion which is independent of $\{N_t\}_{t\ge0}$, then $X_t = B_t + N_t$ is a Lévy process, i.e. has independent stationary increments and is right continuous. The process X is a time homogeneous Markov process with Markov transition kernels given by;

$$\begin{aligned}(Q_t f)(x) = \mathbb{E} f(x + N_t + B_t) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y|^2} \mathbb{E}\left[f(x + y + N_t)\right] dy \\ &= \frac{e^{-\lambda t}}{\sqrt{2\pi t}} \sum_{n=0}^{\infty} \int_{\mathbb{R}} e^{-\frac{1}{2t}|y|^2} \frac{(\lambda t)^n}{n!} f(x + y + n)\, dy \\ &= \frac{e^{-\lambda t}}{\sqrt{2\pi t}} \sum_{n=0}^{\infty} \int_{\mathbb{R}} e^{-\frac{1}{2t}|y-n|^2} \frac{(\lambda t)^n}{n!} f(x + y)\, dy \\ &= \int_{\mathbb{R}} q_t(x, y) f(y)\, dy\end{aligned}$$

where

$$q_t(x, y) = \frac{e^{-\lambda t}}{\sqrt{2\pi t}} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\frac{1}{2t}|y-n|^2}.$$

The infinitesimal generator, $A = \frac{d}{dt}|_{0^+} Q_t$ of this process satisfies,

$$(Af)(x) = \frac{1}{2} f''(x) + \lambda \left(f(x + 1) - f(x)\right)$$

at least for all $f \in C_c^2(\mathbb{R})$. This example will be significantly generalized in Theorem 17.28 below.

In order to continue with giving examples in the continuous time case we will need a simple measure theoretic result.

Lemma 17.24. *If Ω is a set, $\Omega_0 \subset \Omega$, and $\mathcal{B}_0$ is a σ – algebra on Ω_0, then $\tilde{\mathcal{B}}_0 := \{A \subset \Omega : A \cap \Omega_0 \in \mathcal{B}_0\}$ is a σ – algebra on Ω. Moreover, $f : \Omega \to \mathbb{R}$ is $\tilde{\mathcal{B}}_0$ – measurable iff $f|_{\Omega_0}$ is $\mathcal{B}_0$ measurable.*

Proof. It is clear that $\emptyset, \Omega \in \tilde{\mathcal{B}}_0$ and that $\tilde{\mathcal{B}}_0$ is closed under countable unions since $A_n \in \tilde{\mathcal{B}}_0$ iff $A_n \cap \Omega_0 \in \mathcal{B}_0$ which implies $[\cup A_n] \cap \Omega_0 = \cup [A_n \cap \Omega_0] \in \mathcal{B}_0$ and this implies that $[\cup A_n] \in \tilde{\mathcal{B}}_0$. Lastly if $A \in \tilde{\mathcal{B}}_0$ then $A \cap \Omega_0 \in \tilde{\mathcal{B}}_0$ implies that

$$A^c \cap \Omega_0 = \Omega_0 \setminus A = \Omega_0 \setminus [A \cap \Omega_0] \in \mathcal{B}_0$$

and therefore $A \in \tilde{\mathcal{B}}_0$.

For the second assertion, let us observe that for $W \in \mathcal{B}_{\mathbb{R}}$ we have

$$f^{-1}(W) \cap \Omega_0 = f|_{\Omega_0}^{-1}(W)$$

so that $f^{-1}(W) \in \tilde{\mathcal{B}}_0$ iff $f^{-1}(W) \cap \Omega_0 \in \mathcal{B}_0$ iff $f|_{\Omega_0}^{-1}(W) \in \mathcal{B}_0$. It now clearly follows that $f : \Omega \to \mathbb{R}$ is $\tilde{\mathcal{B}}_0$ – measurable iff $f|_{\Omega_0}$ is $\mathcal{B}_0$ measurable. ■

Definition 17.25. *Suppose that $\Omega = \sum_{n=0}^{\infty} \Omega_n$ and $\mathcal{B}_n$ is a σ – algebra on Ω_n for all n. Then we let $\oplus_{n=0}^{\infty} \mathcal{B}_n =: \mathcal{B}$ be the σ – algebra on Ω such that $A \subset \Omega$ is measurable iff $A \cap \Omega_n \in \mathcal{B}_n$ for all n. That is to say $\mathcal{B} = \cap_{n=0}^{\infty} \mathcal{B}'_n$ with $\mathcal{B}'_n = \{A \subset \Omega : A \cap \Omega_n \in \mathcal{B}_n\}$.*

From Lemma 17.24 it follows that $f : \Omega \to \mathbb{R}$ is $\oplus_{n=0}^{\infty} \mathcal{B}_n$ – measurable iff $f^{-1}(W) \in \tilde{\mathcal{B}}_n$ for all n iff $f|_{\Omega_n}^{-1}(W) \in \mathcal{B}_n$ for all n iff $f|_{\Omega_n}$ is $\mathcal{B}_n$ – measurable for all n. We in fact do not really use any properties of Ω_n for these statements it is not even necessary for n to run over a countable index set!

Theorem 17.26. *Suppose that $\left(\Omega, \left\{\tilde{\mathcal{B}}_n\right\}_{n \in \mathbb{N}_0}, \mathcal{B}, P, \{Y_n : \Omega \to S\}_{n \in \mathbb{N}_0}\right)$ is a time homogeneous Markov chain with one step transition kernel, $\tilde{Q}$. Further suppose that $\{N_t\}_{t \geq 0}$ is a Poisson process with parameter λ which is independent of $\tilde{\mathcal{B}}_\infty := \vee_n \tilde{\mathcal{B}}_n$. Let $\mathcal{B}_t$ be the σ – algebra on Ω such that*

$$[\mathcal{B}_t]_{N_t = n} := \left[\sigma\left(N_s : s \leq t \ \& \ \tilde{\mathcal{B}}_n\right)\right]_{N_t = n}.$$

To be more explicit, $A \subset \Omega$ is in $\mathcal{B}_t$ iff

$$A \cap \{N_t = n\} \in \sigma\left(N_s : s \leq t \ \& \ \tilde{\mathcal{B}}_n\right) \text{ for all } n \in \mathbb{N}_0.$$

Finally let $X_t := Y_{N_t}$ for $t \in \mathbb{R}_+$. Then $\{\mathcal{B}_t\}_{t \in \mathbb{R}_+}$ is a filtration, $\{X_t\}_{t \geq 0}$ is adapted to this filtration and is time homogeneous Markov process with transition semi-group given by

$$Q_t = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \tilde{Q}^n = e^{-\lambda t} e^{\lambda t \tilde{Q}} = e^{t\lambda(\tilde{Q} - I)}.$$

Proof. Let us begin by showing that $\mathcal{B}_t$ is increasing. If $0 \leq s < t$ and $A \in \mathcal{B}_s$ then $A \cap \{N_s = m\} \in \sigma\left(N_r : r \leq s \ \& \ \tilde{\mathcal{B}}_m\right)$ therefore, for $n \geq m$, we have

$$A \cap \{N_s = m\} \cap \{N_t = n\} \in \sigma\left(N_r : r \leq t \ \& \ \tilde{\mathcal{B}}_m\right) \subset \sigma\left(N_r : r \leq t \ \& \ \tilde{\mathcal{B}}_n\right)$$

and therefore,

$$A \cap \{N_t = n\} = \cup_{m \leq n} \left[A \cap \{N_s = m\} \cap \{N_t = n\}\right] \in \sigma\left(N_r : r \leq t \ \& \ \tilde{\mathcal{B}}_n\right)$$

for all $n \in \mathbb{N}_0$. Thus we have shown $A \in \mathcal{B}_t$ and therefore $\mathcal{B}_s \subset \mathcal{B}_t$ for all $s \leq t$. Since $X_t|_{N_t = n} = Y_n|_{N_t = n}$ is $\sigma\left(N_r : r \leq t \ \& \ \tilde{\mathcal{B}}_n\right)$ – measurable for all $n \in \mathbb{N}_0$, it follows by Lemma 17.24 that X_t is $\mathcal{B}_t$ – measurable.

We now need to show that Markov property. To this end let $t \geq s$, $f \in \mathcal{S}_b$ and $g \in (\mathcal{B}_s)_b$. Then for each $n \in \mathbb{N}_0$, we have $g|_{N_s=n} \in \sigma\left(N_r : r \leq s \;\&\; \tilde{\mathcal{B}}_n\right)$ and therefore,

$$\begin{aligned}
\mathbb{E}\left[f(X_t)\, g\right] &= \sum_{m,n=0}^{\infty} \mathbb{E}\left[f(X_t) \cdot g : N_s = n,\ N_t - N_s = m\right] \\
&= \sum_{m,n=0}^{\infty} \mathbb{E}\left[f(Y_{n+m})\, 1_{N_t - N_s = m} \cdot g \cdot 1_{N_s = n}\right] \\
&= \sum_{m,n=0}^{\infty} P(N_t - N_s = m)\, \mathbb{E}\left[f(Y_{n+m}) \cdot g \cdot 1_{N_s = n}\right] \\
&= \sum_{m,n=0}^{\infty} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^m}{m!} \mathbb{E}\left[\left(\tilde{Q}^m f\right)(Y_n) \cdot g \cdot 1_{N_s=n}\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[(Q_{t-s} f)(Y_n) \cdot g \cdot 1_{N_s=n}\right] = \mathbb{E}\left[(Q_{t-s} f)(Y_{N_s}) \cdot g\right] \\
&= \mathbb{E}\left[(Q_{t-s} f)(X_s) \cdot g\right].
\end{aligned}$$

This shows that

$$\mathbb{E}\left[f(X_t) \,|\, \mathcal{B}_s\right] = (Q_{t-s} f)(X_s) \;\; P-\text{a.s.}$$

which completes the proof. ∎

Corollary 17.27. *Suppose that S is countable or finite set and $a : S \times S \to \mathbb{R}$ is a function such that $a(x,y) \geq 0$ for all $x \neq y$, and there exists $\lambda < \infty$ such that*

$$a_x := \sum_{y \neq x} a(x,y) \leq \lambda \text{ for all } x \in S.$$

Let us now define $a(x,x) = -a_x$ and $A : \mathcal{S}_b \to \mathcal{S}_b$ by

$$Af(x) := \sum_{y \in S} a(x,y) f(y) = \sum_{y \neq x} a(x,y)\left[f(y) - f(x)\right] \text{ for all } x \in S.$$

Then;

1. *The functions, $\hat{q} : S \times S \to [0,1]$ defined by*

$$\hat{q}(x,y) := \begin{cases} \lambda^{-1} a(x,y) & \text{if } x \neq y \\ 1 - \lambda^{-1} a_x & \text{if } x = y \end{cases}$$

   are the matrix elements of a one step Markov transition kernel, $\tilde{Q}$, i.e. $\tilde{Q} : \mathcal{S}_b \to \mathcal{S}_b$ is defined by

$$\tilde{Q} f(x) = \sum_{y \in S} q(x,y) f(y).$$

   In this operator theoretic notation we have $A = \lambda\left(\tilde{Q} - I\right)$.

2. *If $\{Y_n\}_{n=0}^{\infty}$ is a time homogeneous Markov chain with one step transition kernel, $\tilde{Q}$, and $\{N_t\}_{t\geq 0}$ is an independent Poisson process with intensity λ, then $X_t := Y_{N_t}$ is a time homogeneous Markov process with transition kernels,*

$$Q_t = e^{tA} = e^{-t\lambda} e^{t\lambda\tilde{Q}}.$$

   In particular, A is the infinitesimal generator of a Markov transition semi-group.

Notice that

$$0 \leq e^{t\lambda\tilde{Q}}(x,y) = \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \tilde{Q}^n(x,y)$$

and

$$\begin{aligned}
\sum_{y \in S} e^{t\lambda\tilde{Q}}(x,y) &= \sum_{y \in S} \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \tilde{Q}^n(x,y) \\
&= \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \sum_{y \in S} \tilde{Q}^n(x,y) = \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} 1 = e^{t\lambda}
\end{aligned}$$

so that $Q_t(x,y) \geq 0$ for all x, y and $\sum_{y \in S} Q_t(x,y) = 1$ for all $x \in S$ as must be the case.

The compound Poisson process in the next theorem gives another example of a Lévy process an example of the construction in Theorem 17.26. (The reader should compare the following result with Theorem 23.36 below.)

Theorem 17.28 (Compound Poisson Process). *Suppose that $\{Z_i\}_{i=1}^{\infty}$ are i.i.d. random vectors in $\mathbb{R}^d$ and $\{N_t\}_{t\geq 0}$ be an independent Poisson process with intensity λ. Further let $Z_0 : \Omega \to \mathbb{R}^d$ be independent of $\{N_t\}_{t\geq 0}$ and the $\{Z_i\}_{i=1}^{\infty}$ and then define, for $t \in \mathbb{R}_+$,*

$$\mathcal{B}_t = \oplus_{n=0}^{\infty} \left[\sigma(N_s : s \leq t, Z_0, \dots, Z_n)\right]_{\{N_t = n\}}$$

and $X_t := S_{N_t}$ where $S_n := Z_0 + Z_1 + \cdots + Z_n$. Then $\{\mathcal{B}_t\}_{t\geq 0}$ is a filtration (i.e. it is increasing), $\{X_t\}_{t\geq 0}$ is a $\mathcal{B}_t$ – adapted process such that for all $0 \leq s < t$, $X_t - X_s$ is independent of $\mathcal{B}_s$. The increments are stationary and therefore $\{X_t\}_{t\geq 0}$ is a Lévy process. The time homogeneous transition kernel is given by

$$\begin{aligned}(Q_t f)(x) &= \mathbb{E}\left[f\left(x + Z_1 + \dots + Z_{N_t}\right)\right] \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \mathbb{E}\left[f\left(x + Z_1 + \dots + Z_n\right)\right].\end{aligned}$$

If we define $\left(\tilde{Q} f\right)(x) := \mathbb{E}\left[f\left(x + Z_1\right)\right]$, *the above equation may be written as,*

$$Q_t = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \tilde{Q}^n = e^{\lambda t\left(\tilde{Q} - I\right)}.$$

Proof. Let us begin by showing that $\mathcal{B}_t$ is increasing. First observe that

$$[\sigma(N_s : s \le t, Z_0, \dots, Z_n)]_{\{N_t = n\}} = \{A \subset \{N_t = n\} : A \in \sigma(N_s : s \le t, Z_0, \dots, Z_n)\}.$$

If $0 \le s < t$ and $A \in \mathcal{B}_s$ then $A \cap \{N_s = m\} \in \sigma(N_r : r \le s, Z_0, \dots, Z_m)$ therefore, for $n \ge m$, we have

$$A \cap \{N_s = m\} \cap \{N_t = n\} \in \sigma(N_r : r \le t, Z_0, \dots, Z_m) \subset \sigma(N_r : r \le t, Z_0, \dots, Z_n)$$

and we may conclude that

$$A \cap \{N_t = n\} = \cup_{m \le n} [A \cap \{N_s = m\} \cap \{N_t = n\}] \in \sigma(N_r : r \le t, Z_0, \dots, Z_n)$$

for all $n \in \mathbb{N}_0$. Thus we have shown $A \in \mathcal{B}_t$ and therefore $\mathcal{B}_s \subset \mathcal{B}_t$ for all $s \le t$. Since

$$X_t|_{N_t = n} = [Z_0 + Z_1 + \dots + Z_n]\,|_{N_t = n}$$

is $[\sigma(N_s : s \le t, Z_0, \dots, Z_n)]_{\{N_t = n\}}$ – measurable for all $n \in \mathbb{N}_0$, it follows by Lemma 17.24 that X_t is $\mathcal{B}_t$ – measurable.

We now show that $X_t - X_s$ is independent of $\mathcal{B}_s$ for all $t \ge s$. To this end, let $f \in (\mathcal{B}_{\mathbb{R}^d})_b$ and $g \in (\mathcal{B}_s)_b$. Then for each $n \in \mathbb{N}_0$, we have

$$g|_{N_s = n} = G_n\left(\{N_r\}_{r \le s}, Z_0, \dots, Z_n\right)$$

while

$$(X_t - X_s)\,|_{N_s = n} = Z_{n+1} + \dots + Z_{N_t}.$$

Therefore we have,

$$\mathbb{E}[f(X_t - X_s) \cdot g] = \sum_{m,n=0}^{\infty} \mathbb{E}[f(X_t - X_s) \cdot g : N_s = n,\ N_t - N_s = m]$$

and if we let $a_{m,n}$ be the summand on the right side of this equation we have,

$$\begin{aligned}a_{m,n} &= \mathbb{E}\left[f\left(Z_{n+1} + \dots + Z_{n+m}\right) \cdot G_n\left(\{N_r\}_{r \le s}, Z_0, \dots, Z_n\right) : N_s = n,\ N_t - N_s = m\right] \\ &= \mathbb{E}\left[f\left(Z_{n+1} + \dots + Z_{n+m}\right) 1_{N_t - N_s = m} \cdot G_n\left(\{N_r\}_{r \le s}, Z_0, \dots, Z_n\right) 1_{N_s = n}\right] \\ &= \mathbb{E}\left[f\left(Z_{n+1} + \dots + Z_{n+m}\right) 1_{N_t - N_s = m}\right] \cdot \mathbb{E}\left[G_n\left(\{N_r\}_{r \le s}, Z_0, \dots, Z_n\right) 1_{N_s = n}\right] \\ &= e^{-\lambda(t-s)} \frac{(\lambda(t-s))^m}{m!} \mathbb{E}\left[f\left(Z_1 + \dots + Z_m\right)\right] \cdot \mathbb{E}\left[G_n\left(\{N_r\}_{r \le s}, Z_0, \dots, Z_n\right) 1_{N_s = n}\right]\end{aligned}$$

Therefore it follows that

$$\begin{aligned}\mathbb{E}[f(X_t - X_s) \cdot g] &= \sum_{m,n=0}^{\infty} a_{m,n} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{m,n} \\ &= (Q_{t-s} f)(0) \cdot \sum_{n=0}^{\infty} \mathbb{E}\left[G_n\left(\{N_r\}_{r \le s}, Z_0, \dots, Z_n\right) 1_{N_s = n}\right] \\ &= (Q_{t-s} f)(0) \cdot \mathbb{E}[g]\end{aligned}$$

from which it follows that $X_t - X_s$ is independent of $\mathcal{B}_s$ and

$$\mathbb{E}[f(X_t - X_s)] = (Q_{t-s} f)(0).$$

(This equation shows that the distribution of the increments is stationary.) We now know by Exercise 17.2 that $\{X_t\}$ is a Markov process and the transition kernel is given by

$$(Q_{t-s} f)(x) = \mathbb{E}[f(x + X_t - X_s)] = (Q_{t-s} f(x + \cdot))(0)$$

as described above. ∎

17.4 First Step Analysis and Hitting Probabilities

In this section we suppose that $T = \mathbb{N}_0$, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in T}\right)$ is a filtered measures space, $X_t : \Omega \to S$ is a $\mathcal{B}_t/\mathcal{S}$ – measurable function for all $t \in T$, $Q : S \times \mathcal{S} \to [0,1]$ is a Markov-transition kernel, and for each $x \in S$ there exists a probability, P_x, on $(\Omega, \mathcal{B})$ such that $P_x(X_0 = x) = 1$ and $\{X_t\}_{t \ge 0}$ is a time homogenous Markov process with Q as its one step Markov transition kernel. To shorten notation we will write $\mathbb{E}_x$ for the expectation relative to the measure P_x.

Definition 17.29 (Hitting times). *For* $B \in \mathcal{S}$, *let*

$$T_B(X) := \min\{n \ge 0 : X_n \in B\}$$

with the convention that $\min \emptyset = \infty$. *We call* $T_B(X) = T_B(X_0, X_1, \dots)$ *the* ***first hitting time*** *of* B *by* $X = \{X_n\}_n$.

Notation 17.30 *For $A \in \mathcal{S}$, let $Q_A : A \times \mathcal{S}_A \to [0,1]$ be the restriction of Q to A, so that $Q_A(x,C) := Q(x,C)$ for all $x \in A$ and $C \in \mathcal{S}_A$. As with probability kernels we may identify Q_A with an operator from $(\mathcal{S}_A)_b$ to itself via,*

$$(Q_A f)(x) = \int_A Q(x,dy) f(y) \text{ for all } x \in A \text{ and } f \in (\mathcal{S}_A)_b .$$

Theorem 17.31. *Let n denote a non-negative integer. If $h : B \to \mathbb{R}$ is measurable and either bounded or non-negative, then*

$$\mathbb{E}_x [h(X_n) : T_B = n] = \left(Q_A^{n-1} Q[1_B h]\right)(x)$$

and

$$\mathbb{E}_x [h(X_{T_B}) : T_B < \infty] = \left(\sum_{n=0}^{\infty} Q_A^n Q[1_B h]\right)(x). \tag{17.32}$$

If $g : A \to \mathbb{R}_+$ is a measurable function, then for all $x \in A$ and $n \in \mathbb{N}_0$,

$$\mathbb{E}_x [g(X_n) 1_{n<T_B}] = (Q_A^n g)(x).$$

In particular we have

$$\mathbb{E}_x \left[\sum_{n<T_B} g(X_n)\right] = \sum_{n=0}^{\infty} (Q_A^n g)(x) =: u(x), \tag{17.33}$$

where by convention, $\sum_{n<T_B} g(X_n) = 0$ when $T_B = 0$.

Proof. Let $x \in A$. In computing each of these quantities we will use;

$$\begin{aligned} \{T_B > n\} &= \{X_i \in A \text{ for } 0 \le i \le n\} \text{ and} \\ \{T_B = n\} &= \{X_i \in A \text{ for } 0 \le i \le n-1\} \cap \{X_n \in B\}. \end{aligned}$$

From the second identity above it follows that for

$$\begin{aligned} \mathbb{E}_x [h(X_n) : T_B = n] &= \mathbb{E}_x \left[h(X_n) : (X_1, \dots, X_{n-1}) \in A^{n-1},\, X_n \in B\right] \\ &= \sum_{n=1}^{\infty} \int_{A^{n-1} \times B} \prod_{j=1}^{n} Q(x_{j-1}, dx_j) h(x_n) \\ &= \left(Q_A^{n-1} Q[1_B h]\right)(x) \end{aligned}$$

and therefore

$$\begin{aligned} \mathbb{E}_x [h(X_{T_B}) : T_B < \infty] &= \sum_{n=1}^{\infty} \mathbb{E}_x [h(X_n) : T_B = n] \\ &= \sum_{n=1}^{\infty} Q_A^{n-1} Q[1_B h] = \sum_{n=0}^{\infty} Q_A^n Q[1_B h]. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}_x [g(X_n) 1_{n<T_B}] &= \int_{A^n} Q(x,dx_1) Q(x_1,dx_2) \dots Q(x_{n-1},dx_n) g(x_n) \\ &= (Q_A^n g)(x) \end{aligned}$$

and therefore,

$$\begin{aligned} \mathbb{E}_x \left[\sum_{n=0}^{\infty} g(X_n) 1_{n<T_B}\right] &= \sum_{n=0}^{\infty} \mathbb{E}_x [g(X_n) 1_{n<T_B}] \\ &= \sum_{n=0}^{\infty} (Q_A^n g)(x). \end{aligned}$$

■

In practice it is not so easy to sum the series in Eqs. (17.32) and (17.33). Thus we would like to have another way to compute these quantities. Since $\sum_{n=0}^{\infty} Q_A^n$ is a geometric series, we expect that

$$\sum_{n=0}^{\infty} Q_A^n = (I - Q_A)^{-1}$$

which is basically correct at least when $(I - Q_A)$ is invertible. This suggests that if $u(x) = \mathbb{E}_x [h(X_{T_B}) : T_B < \infty]$, then (see Eq. (17.32))

$$u = Q_A u + Q[1_B h] \text{ on } A, \tag{17.34}$$

and if $u(x) = \mathbb{E}_x \left[\sum_{n<T_B} g(X_n)\right]$, then (see Eq. (17.33))

$$u = Q_A u + g \text{ on } A. \tag{17.35}$$

That these equations are valid is the content of Corollaries 17.33 and 17.34 below which we will prove using the "first step" analysis in the next theorem. We will give another direct proof in Theorem 17.39 below as well.

Theorem 17.32 (First step analysis). *Let us keep the assumptions in Theorem 17.14 and add the further assumption that $T = \mathbb{N}_0$. Then for all $F \in \mathcal{S}_b^{\otimes \mathbb{N}}$ or $F : S^{\mathbb{N}_0} \to [0,\infty]$ measurable;*

$$\mathbb{E}_x [F(X_0, X_1, \dots)] = \int_S Q(x,dy) \mathbb{E}_y F(x, X_0, X_1, \dots). \tag{17.36}$$

This equation can be iterated to show more generally that

$$\mathbb{E}_x [F(X_0, X_1, \dots)] = \int_{S^n} \prod_{j=1}^{n} Q(x_{j-1}, dx_j) \mathbb{E}_{x_n} [F(x_0, x_1, \dots, x_{n-1}, X_0, X_1, \dots)] \tag{17.37}$$

where $x_0 := x$.

Proof. Since $X_0(\omega) = x$ for P_x – a.e. ω, we have $F(X_0, X_1, \dots) = F(x, X_1, X_2, \dots)$ a.s. Therefore by Theorem 17.14 we know that

$$\mathbb{E}_x[F(X_0, X_1, \dots)|\mathcal{B}_1] = \mathbb{E}_x[F(x, X_1, X_2, \dots)|\mathcal{B}_1] = \mathbb{E}_{X_1}F(x, X_0, X_1, \dots).$$

Taking expectations of this equation shows,

$$\begin{aligned}\mathbb{E}_x[F(X_0, X_1, \dots)] &= \mathbb{E}_x[\mathbb{E}_{X_1}F(x, X_0, X_1, \dots)] \\ &= \int_S Q(x, dy)\,\mathbb{E}_y F(x, X_0, X_1, \dots).\end{aligned}$$

■

Corollary 17.33. *Suppose that $B \in \mathcal{S}$, $A := B^c \in \mathcal{S}$, $h : B \to \mathbb{R}$ is a measurable function which is either bounded or non-negative, and*

$$u(x) := \mathbb{E}_x[h(X_{T_B}) : T_B < \infty] \text{ for } x \in A.$$

Then $u : A \to \mathbb{R}$ satisfies Eq. (17.34), i.e. $u = Q_A u + Q[1_B h]$ on A or in more detail

$$u(x) = \int_A Q(x, dy)\,u(y) + \int_B Q(x, dy)\,h(y) \text{ for all } x \in A.$$

In particular, when $h \equiv 1$, $u(x) = P_x(T_B < \infty)$ is a solution to the equation,

$$u = Q_A u + Q1_B \text{ on } A. \tag{17.38}$$

Proof. To shorten the notation we will use the convention that $h(X_{T_B}) = 0$ if $T_B = \infty$ so that we may simply write $u(x) := \mathbb{E}_x[h(X_{T_B})]$. Let

$$F(X_0, X_1, \dots) = h\left(X_{T_B(X)}\right) = h\left(X_{T_B(X)}\right)1_{T_B(X)<\infty},$$

then for $x \in A$ we have $F(x, X_0, X_1, \dots) = F(X_0, X_1, \dots)$. Therefore by the first step analysis (Theorem 17.32) we learn

$$\begin{aligned}u(x) = \mathbb{E}_x h\left(X_{T_B(X)}\right) &= \mathbb{E}_x F(x, X_1, \dots) = \int_S Q(x, dy)\,\mathbb{E}_y F(x, X_0, X_1, \dots) \\ &= \int_S Q(x, dy)\,\mathbb{E}_y F(X_0, X_1, \dots) = \int_S Q(x, dy)\,\mathbb{E}_y\left[h\left(X_{T_B(X)}\right)\right] \\ &= \int_A Q(x, dy)\,\mathbb{E}_y\left[h\left(X_{T_B(X)}\right)\right] + \int_B Q(x, dy)\,h(y) \\ &= \int_A Q(x, dy)\,u(y) + \int_B Q(x, dy)\,h(y) \\ &= (Q_A u)(x) + (Q1_B)(x).\end{aligned}$$

■

Corollary 17.34. *Suppose that $B \in \mathcal{S}$, $A := B^c \in \mathcal{S}$, $g : A \to [0, \infty]$ is a measurable function. Further let $u(x) := \mathbb{E}_x\left[\sum_{n<T_B} g(X_n)\right]$. Then $u(x)$ satisfies Eq. (17.35), i.e. $u = Q_A u + g$ on A or in more detail,*

$$u(x) = \int_A Q(x, dy)\,u(y) + g(x) \text{ for all } x \in A.$$

In particular if we take $g \equiv 1$ in this equation we learn that

$$\mathbb{E}_x T_B = \int_A Q(x, dy)\,\mathbb{E}_y T_B + 1 \text{ for all } x \in A.$$

Proof. Let $F(X_0, X_1, \dots) = \sum_{n<T_B(X_0, X_1, \dots)} g(X_n)$ be the sum of the values of g along the chain before its first exit from A, i.e. entrance into B. With this interpretation in mind, if $x \in A$, it is easy to see that

$$\begin{aligned}F(x, X_0, X_1, \dots) &= \begin{cases} g(x) & \text{if } X_0 \in B \\ g(x) + F(X_0, X_1, \dots) & \text{if } X_0 \in A \end{cases} \\ &= g(x) + 1_{X_0 \in A} \cdot F(X_0, X_1, \dots).\end{aligned}$$

Therefore by the first step analysis (Theorem 17.32) it follows that

$$\begin{aligned}u(x) = \mathbb{E}_x F(X_0, X_1, \dots) &= \int_S Q(x, dy)\,\mathbb{E}_y F(x, X_0, X_1, \dots) \\ &= \int_S Q(x, dy)\,\mathbb{E}_y[g(x) + 1_{X_0 \in A} \cdot F(X_0, X_1, \dots)] \\ &= g(x) + \int_A Q(x, dy)\,\mathbb{E}_y[F(X_0, X_1, \dots)] \\ &= g(x) + \int_A Q(x, dy)\,u(y).\end{aligned}$$

■

The problem with Corollaries 17.33 and 17.34 is that the solutions to Eqs. (17.34) and (17.35) may not be unique as we will see in the next examples. Theorem 17.39 below will explain when these ambiguities may occur and how to deal with them when they do.

Example 17.35 (Biased random walks I). Let $p \in (1/2, 1)$ and consider the biased random walk $\{S_n\}_{n\geq 0}$ on the $S = \mathbb{Z}$ where $S_n = X_0 + X_1 + \dots + X_n$, $\{X_i\}_{i=1}^\infty$ are i.i.d. with $P(X_i = 1) = p \in (0, 1)$ and $P(X_i = -1) = q := 1 - p$, and $X_0 = x$ for some $x \in \mathbb{Z}$. Let $B := \{0\}$ and $u(x) := P_x(T_B < \infty)$. Clearly $u(0) = 0$ and by the first step analysis,

$$u(x) = pu(x+1) + qu(x-1) \text{ for } x \neq 0. \tag{17.39}$$

From Exercise 17.10 below, we know that the general solution to Eq. (17.39) is of the form

$$u(x) = a\lambda_+^x + b\lambda_-^x$$

where $\lambda_\pm$ are the roots for the characteristic polynomial, $p\lambda^2 - \lambda + q = 0$. since constants solve Eq. (17.39) we know that one root is 1 as is easily verified. The other root[2] is q/p. Thus the general solution is of the form, $w(x) = a + b(q/p)^x$. In all case we are going to choose a and b so that $0 = u(0) = w(0)$ (i.e. $a+b=0$) so that $w(x) = a + (1-a)(q/p)^x$. For $x > 0$ we choose $a = a_+$ so that $w_+(x) := a_+ + (1-a_+)(q/p)^x$ satisfies $w_+(1) = u(1)$ and for $x < 0$ we choose $a = a_-$ so that $w_-(x) := a_- + (1-a_-)(q/p)^x$ satisfies $w_-(-1) = u(-1)$. With these choice we will have $u(x) = w_+(x)$ for $x \geq 0$ and $u(x) = w_-(x)$ for $x \leq 0$ – see Exercise 17.10 and Remark 17.37. Observe that

$$u(1) = a_+ + (1-a_+)(q/p) \implies a_+ = \frac{u(1) - (q/p)}{1 - (q/p)}$$

and

$$u(-1) = a_- + (1-a_-)(p/q) \implies a_- = \frac{(p/q) - u(-1)}{(p/q) - 1}.$$

Case 1. $x < 0$: As $x \to -\infty$, we will have $|u(x)| \to \infty$ unless $a_- = 1$. Thus we must take $a_- = 1$ and we have shown,

$$P_x(T_0 < \infty) = w_-(x) = 1 \text{ for all } x < 0.$$

Case 2. $x > 0$: For $n \in \mathbb{N}_0$, let $T_n = \min\{m : X_m = n\}$ be the first time X hits n. By the MCT we have,

$$P_x(T_0 < \infty) = \lim_{n\to\infty} P_x(T_0 < T_n).$$

So we will now try to compute $u(x) = P_x(T_0 < T_n)$. By the first step analysis (take $B = \{0, n\}$ and $h(0) = 1$ and $h(n) = 0$ in Corollary 17.33) we will still have that $u(x)$ satisfies Eq. (17.39) for $0 < x < n$ but now the boundary conditions are $u(0) = 1$ and $u(n) = 0$. Accordingly $u(x)$ for $0 \leq x \leq n$ is still of the form given in Eq. (17.39) but we may now determine $a = a_n$ using the boundary condition

$$0 = u(n) = a + (1-a)(q/p)^n = (q/p)^n + a(1 - (q/p)^n)$$

from which it follows that

[2] Indeed,

$$p\left(\frac{q}{p}\right)^2 - \frac{q}{p} + q = \frac{q}{p}[q - 1 + p] = 0.$$

$$a_n = \frac{(q/p)^n}{(q/p)^n - 1} \to 0 \text{ as } n \to \infty.$$

Thus we have shown

$$\begin{aligned} P_x(T_0 < T_n) &= \frac{(q/p)^n}{(q/p)^n - 1} + \left(1 - \frac{(q/p)^n}{(q/p)^n - 1}\right)(q/p)^x \\ &= \frac{(q/p)^n - (q/p)^x}{(q/p)^n - 1} \\ &= \frac{(q/p)^x - (q/p)^n}{1 - (q/p)^n} \to (q/p)^x \text{ as } n \to \infty \end{aligned}$$

and therefore, since $T_n \uparrow \infty$ P_x – a.s. as $n \uparrow \infty$,

$$P_x(T_0 < \infty) = (q/p)^x \text{ for all } x > 0.$$

Example 17.36 (Biased random walks II). Continue the notation in Example 17.35. Let us now try to compute $\mathbb{E}_x T_0$. Since $P_x(T_0 = \infty) > 0$ for $x > 0$ we already know that $\mathbb{E}_x T_0 = \infty$ for all $x > 0$. Nevertheless we will deduce this fact again here.

Letting $u(x) = \mathbb{E}_x T_0$ it follows by the first step analysis that, for $x \neq 0$,

$$\begin{aligned} u(x) &= p[1 + u(x+1)] + q[1 + u(x-1)] \\ &= pu(x+1) + qu(x-1) + 1 \end{aligned} \tag{17.40}$$

with $u(0) = 0$. Notice $u(x) = \infty$ is a solution to this equation while if $u(a) < \infty$ for some $a \neq 0$ then Eq. (17.40) implies that $u(x) < \infty$ for all $x \neq 0$ with the same sign as a.

A particular solution to this equation may be found by trying $u(x) = \alpha x$ to learn,

$$\alpha x = p\alpha(x+1) + q\alpha(x-1) + 1 = \alpha x + \alpha(p-q) + 1$$

which is valid for all x provided $\alpha = (q-p)^{-1}$. The general **finite** solution to Eq. (17.40) is therefore,

$$u(x) = (q-p)^{-1}x + a + b(q/p)^x. \tag{17.41}$$

Using the boundary condition, $u(0) = 0$ allows us to conclude that $a + b = 0$ and therefore,

$$u(x) = u_a(x) = (q-p)^{-1}x + a[1 - (q/p)^x]. \tag{17.42}$$

Notice that $u_a(x) \to -\infty$ as $x \to +\infty$ no matter how a is chosen and therefore we must conclude that the desired solution to Eq. (17.40) is $u(x) = \infty$ for $x > 0$ as we already mentioned.

The question now is for $x<0$. Is it again the case that $u(x)=\infty$ or is $u(x)=u_a(x)$ for some $a\in\mathbb{R}$. Since $\lim_{x\to\infty}u_a(x)=-\infty$ unless $a\le 0$, we may restrict our attention to $a\le 0$. To work out which $a\le 0$ is correct observe by MCT that

$$\mathbb{E}_x T_0=\lim_{n\to-\infty}\mathbb{E}_x\left[T_n\wedge T_0\right]=\lim_{n\to-\infty}\mathbb{E}_x\left[T_{\{n,0\}}\right].$$

So let $n\in\mathbb{Z}$ with $n<0$ be fixed for the moment. By item 8. of Theorem 17.39 we may conclude that $u(x):=\mathbb{E}_x\left[T_{\{n,0\}}\right]<\infty$ for all $n\le x\le 0$. Then by the first step analysis, $u(x)$ satisfies Eq. (17.40) for $n<x<0$ and has boundary conditions $u(n)=0=u(0)$. Using the boundary condition $u(n)=0$ to determine $a=a_n$ in Eq. (17.42) implies,

$$0=u_a(n)=(q-p)^{-1}n+a\left[1-(q/p)^n\right]$$

so that

$$a=a_n=\frac{n}{\left(1-(q/p)^n\right)(p-q)}\to 0 \text{ as } n\to-\infty.$$

Thus we conclude that

$$\begin{aligned}\mathbb{E}_x T_0&=\lim_{n\to-\infty}\mathbb{E}_x\left[T_n\wedge T_0\right]=\lim_{n\to-\infty}u_{a_n}(x)\\&=\frac{x}{q-p}=\frac{|x|}{p-q}\text{ for }x<0.\end{aligned}$$

Remark 17.37 (More on the boundary conditions). If we were to use Corollary 17.34 directly to derive Eq. (17.40) in the case that $u(x):=\mathbb{E}_x\left[T_{\{n,0\}}\right]<\infty$ we for all $0\le x\le n$. we would find, for $x\neq 0$, that

$$u(x)=\sum_{y\notin\{n,0\}}q(x,y)\,u(y)+1$$

which implies that $u(x)$ satisfies Eq. (17.40) for $n<x<0$ provided $u(n)$ and $u(0)$ are taken to be equal to zero. Let us again choose a and b

$$w(x):=(q-p)^{-1}x+a+b\,(q/p)^x$$

satisfies $w(0)=0$ and $w(-1)=u(-1)$. Then both w and u satisfy Eq. (17.40) for $n<x\le 0$ and agree at 0 and -1 and therefore are equal[3] for $n\le x\le 0$ and in particular $0=u(n)=w(n)$. Thus correct boundary conditions on w in order for $w=u$ are $w(0)=w(n)=0$ as we have used above.

[3] Observe from Eq. (17.40) we have for $x\neq 0$ that,

$$u(x-1)=q^{-1}\left[u(x)-pu(x+1)-1\right].$$

From this equation it follows easily that $u(x)$ for $x\le 0$ is determined by its values at $x=0$ and $x=-1$.

Definition 17.38. *Suppose $(A,\mathcal{A})$ is a measurable space. A* ***sub-probability kernel*** *on $(A,\mathcal{A})$ is a function $\rho:A\times\mathcal{A}\to[0,1]$ such that $\rho(\cdot,C)$ is $\mathcal{A}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $C\in\mathcal{A}$ and $\rho(x,\cdot):\mathcal{A}\to[0,1]$ is a measure for all $x\in A$.*

As with probability kernels we will identify ρ with the linear map, $\rho:\mathcal{A}_b\to\mathcal{A}_b$ given by

$$(\rho f)(x)=\rho(x,f)=\int_A f(y)\,\rho(x,dy).$$

Of course we have in mind that $\mathcal{A}=\mathcal{S}_A$ and $\rho=Q_A$. In the following lemma let $\|g\|_\infty:=\sup_{x\in A}|g(x)|$ for all $g\in\mathcal{A}_b$.

Theorem 17.39. *Let ρ be a sub-probability kernel on a measurable space $(A,\mathcal{A})$ and define $u_n(x):=(\rho^n 1)(x)$ for all $x\in A$ and $n\in\mathbb{N}_0$. Then;*

1. *u_n is a decreasing sequence so that $u:=\lim_{n\to\infty}u_n$ exists and is in $\mathcal{A}_b$. (When $\rho=Q_A$, $u_n(x)=P_x(T_B>n)\downarrow u(x)=P(T_B=\infty)$ as $n\to\infty$.)*
2. *The function u satisfies $\rho u=u$.*
3. *If $w\in\mathcal{A}_b$ and $\rho w=w$ then $|w|\le\|w\|_\infty u$. In particular the equation, $\rho w=w$, has a non-zero solution $w\in\mathcal{A}_b$ iff $u\neq 0$.*
4. *If $u=0$ and $g\in\mathcal{A}_b$, then there is at most one $w\in\mathcal{A}_b$ such that $w=\rho w+g$.*
5. *Let*
$$U:=\sum_{n=0}^\infty u_n=\sum_{n=0}^\infty\rho^n 1:A\to[0,\infty] \tag{17.43}$$
and suppose that $U(x)<\infty$ for all $x\in A$. Then for each $g\in\mathcal{S}_b$,
$$w=\sum_{n=0}^\infty\rho^n g \tag{17.44}$$
is absolutely convergent,
$$|w|\le\|g\|_\infty U, \tag{17.45}$$
$\rho(x,|w|)<\infty$ for all $x\in A$, and w solves $w=\rho w+g$. Moreover if v also solves $v=\rho v+g$ and $|v|\le CU$ for some $C<\infty$ then $v=w$.
Observe that when $\rho=Q_A$,
$$U(x)=\sum_{n=0}^\infty P_x(T_B>n)=\sum_{n=0}^\infty\mathbb{E}_x\left(1_{T_B>n}\right)=\mathbb{E}_x\left(\sum_{n=0}^\infty 1_{T_B>n}\right)=\mathbb{E}_x\left[T_B\right].$$
6. *If $g:A\to[0,\infty]$ is any measurable function then*
$$w:=\sum_{n=0}^\infty\rho^n g:A\to[0,\infty]$$
is a solution to $w=\rho w+g$. (It may be that $w\equiv\infty$ though!) Moreover if $v:A\to[0,\infty]$ satisfies $v=\rho v+g$ then $w\le v$. Thus w is the minimal non-negative solution to $v=\rho v+g$.

7. *If there exists $\alpha < 1$ such that $u \le \alpha$ on A then $u = 0$. (When $\rho = Q_A$, this state that $P_x(T_B = \infty) \le \alpha$ for all $x \in A$ implies $P_x(T_A = \infty) = 0$ for all $x \in A$.)*
8. *If there exists an $\alpha < 1$ and an $n \in \mathbb{N}$ such that $u_n = \rho^n 1 \le \alpha$ on A, then there exists $C < \infty$ such that*
$$u_k(x) = \left(\rho^k 1\right)(x) \le C\beta^k \text{ for all } x \in A \text{ and } k \in \mathbb{N}_0$$
where $\beta := \alpha^{1/n} < 1$. In particular, $U \le C(1-\beta)^{-1}$ and $u = 0$ under this assumption.
(When $\rho = Q_A$ this assertion states; if $P_x(T_B > n) \le \alpha$ for all $\alpha \in A$, then $P_x(T_B > k) \le C\beta^k$ and $\mathbb{E}_x T_B \le C(1-\beta)^{-1}$ for all $k \in \mathbb{N}_0$.)

Proof. We will prove each item in turn.

1. First observe that $u_1(x) = \rho(x, A) \le 1 = u_0(x)$ and therefore,
$$u_{n+1} = \rho^{n+1} 1 = \rho^n u_1 \le \rho^n 1 = u_n.$$
We now let $u := \lim_{n\to\infty} u_n$ so that $u : A \to [0,1]$.
2. Using DCT we may let $n \to \infty$ in the identity, $\rho u_n = u_{n+1}$ in order to show $\rho u = u$.
3. If $w \in \mathcal{A}_b$ with $\rho w = w$, then
$$|w| = |\rho^n w| \le \rho^n |w| \le \|w\|_\infty \rho^n 1 = \|w\|_\infty \cdot u_n.$$
Letting $n \to \infty$ shows that $|w| \le \|w\|_\infty u$.
4. If $w_i \in \mathcal{A}_b$ solves $w_i = \rho w_i + g$ for $i = 1, 2$ then $w := w_2 - w_1$ satisfies $w = \rho w$ and therefore $|w| \le Cu = 0$.
5. Let $U := \sum_{n=0}^\infty u_n = \sum_{n=0}^\infty \rho^n 1 : A \to [0, \infty]$ and suppose $U(x) < \infty$ for all $x \in A$. Then $u_n(x) \to 0$ as $n \to \infty$ and so bounded solutions to $\rho u = u$ are necessarily zero. Moreover we have, for all $k \in \mathbb{N}_0$, that
$$\rho^k U = \sum_{n=0}^\infty \rho^k u_n = \sum_{n=0}^\infty u_{n+k} = \sum_{n=k}^\infty u_n \le U. \tag{17.46}$$
Since the tails of convergent series tend to zero it follows that $\lim_{k\to\infty} \rho^k U = 0$.
Now if $g \in \mathcal{S}_b$, we have
$$\sum_{n=0}^\infty |\rho^n g| \le \sum_{n=0}^\infty \rho^n |g| \le \sum_{n=0}^\infty \rho^n \|g\|_\infty = \|g\|_\infty \cdot U < \infty \tag{17.47}$$
and therefore $\sum_{n=0}^\infty \rho^n g$ is absolutely convergent. Making use of Eqs. (17.46) and (17.47) we see that
$$\sum_{n=1}^\infty \rho |\rho^n g| \le \|g\|_\infty \cdot \rho U \le \|g\|_\infty U < \infty$$
and therefore (using DCT),
$$w = \sum_{n=0}^\infty \rho^n g = g + \sum_{n=1}^\infty \rho^n g$$
$$= g + \rho \sum_{n=1}^\infty \rho^{n-1} g = g + \rho w,$$
i.e. w solves $w = g + \rho w$.
If $v : A \to \mathbb{R}$ is measurable such that $|v| \le CU$ and $v = g + \rho v$, then $y := w - v$ solves $y = \rho y$ with $|y| \le (C + \|g\|_\infty) U$. It follows that
$$|y| = |\rho^n y| \le (C + \|g\|_\infty) \rho^n U \to 0 \text{ as } n \to \infty,$$
i.e. $0 = y = w - v$.
6. If $g \ge 0$ we may always define w by Eq. (17.44) allowing for $w(x) = \infty$ for some or even all $x \in A$. As in the proof of the previous item (with DCT being replaced by MCT), it follows that $w = \rho w + g$. If $v \ge 0$ also solves $v = g + \rho v$, then
$$v = g + \rho(g + \rho v) = g + \rho g + \rho^2 v$$
and more generally by induction we have
$$v = \sum_{k=0}^n \rho^k g + \rho^{n+1} v \ge \sum_{k=0}^n \rho^k g.$$
Letting $n \to \infty$ in this last equation shows that $v \ge w$.
7. If $u \le \alpha < 1$ on A, then by item 3. with $w = u$ we find that
$$u \le \|u\|_\infty \cdot u \le \alpha u$$
which clearly implies $u = 0$.
8. If $u_n \le \alpha < 1$, then for any $m \in \mathbb{N}$ we have,
$$u_{n+m} = \rho^m u_n \le \alpha \rho^m 1 = \alpha u_m.$$
Taking $m = kn$ in this inequality shows, $u_{(k+1)n} \le \alpha u_{kn}$. Thus a simple induction argument shows $u_{kn} \le \alpha^k$ for all $k \in \mathbb{N}_0$. For general $l \in \mathbb{N}_0$ we write $l = kn + r$ with $0 \le r < n$. We then have,
$$u_l = u_{kn+r} \le u_{kn} \le \alpha^k = \alpha^{\frac{l-r}{n}} = C\alpha^{l/n}$$
where $C = \alpha^{-\frac{n-1}{n}}$.

■

Corollary 17.40. *If $h : B \to [0,\infty]$ is measurable, then $u(x) := \mathbb{E}_x[h(X_{T_B}) : T_B < \infty]$ is the unique minimal non-negative solution to Eq. (17.34) while if $g : A \to [0,\infty]$ is measurable, then $u(x) = \mathbb{E}_x\left[\sum_{n<T_B} g(X_n)\right]$ is the unique minimal non-negative solution to Eq. (17.35).*

Exercise 17.9. Keeping the notation of Example 17.35 and 17.36. Use Corollary 17.40 to show again that $P_x(T_B < \infty) = (q/p)^x$ for all $x > 0$ and $\mathbb{E}_x T_0 = x/(q-p)$ for $x < 0$. You should do so without making use of the extraneous hitting times, T_n for $n \neq 0$.

Corollary 17.41. *If $P_x(T_B = \infty) = 0$ for all $x \in A$ and $h : B \to \mathbb{R}$ is a bounded measurable function, then $u(x) := \mathbb{E}_x[h(X_{T_B})]$ is the **unique** solution to Eq. (17.34).*

Corollary 17.42. *Suppose now that $A = B^c$ is a finite subset of S and there exists an $\alpha \in (0,1)$ such that $P_x(T_B = \infty) \le \alpha$ for all $x \in A$. Then there exists $C < \infty$ and $\beta \in (0,1)$ such that $P_x(T_B > n) \le C\beta^n$.*

Proof. We know that

$$\lim_{n\to\infty} P_x(T_B > n) = P_x(T_B = \infty) \le \alpha \text{ for all } x \in A.$$

Therefore if $\tilde{\alpha} \in (\alpha, 1)$, using the fact that A is a finite set, there exists an n sufficiently large such that $P_x(T_B > n) \le \tilde{\alpha}$ for all $x \in A$. The result now follows from item 8. of Theorem 17.39. ■

17.5 Finite state space chains

In this subsection I would like to write out the above theorems in the special case where S is a finite set. In this case we will let $q(x,y) := Q(x,\{y\})$ so that

$$(Qf)(x) = \sum_{y\in S} q(x,y) f(y).$$

Thus if we view $f : S \to \mathbb{R}$ as a column vector and Q to be the matrix with $q(x,y)$ in the x^{th} – row and y^{th} – column, then Qf is simply matrix multiplication. As above we now suppose that S is partitioned into two nonempty subsets B and $A = B^c$. We further assume that $P_x(T_B < \infty) > 0$ for all $x \in A$, i.e. it is possible with positive probability for the chain $\{X_n\}_{n=0}^\infty$ to visit B when started from any point in A. Because of Corollary 17.42 we know that in fact there exists $C < \infty$ and $\beta \in (0,1)$ such that $P_x(T_B > n) \le C\beta^n$ for all $n \in \mathbb{N}_0$. In particular it follows that $\mathbb{E}_x T_B < \infty$ and $P_x(T_B < \infty) = 1$ for all $x \in A$.

If we let $Q_A = Q_{A,A}$ be the matrix with entries, $Q_A = (q(x,y))_{x,y\in A}$ and I be the corresponding identity matrix, then $(Q_A - I)^{-1}$ exits according to Theorem 17.39. Let us further let $R = Q_{A,B}$ be the matrix with entries, $(q(x,y))_{x\in A \text{ and } y\in B}$. Thus Q decomposes as

$$Q = \begin{array}{cc} \begin{array}{cc} A & B \end{array} & \\ \begin{bmatrix} Q_A & R \\ * & * \end{bmatrix} & \begin{array}{c} A \\ B \end{array} \end{array}.$$

To summarize, Q_A is Q with the rows and columns indexed by B deleted and R is the Q – matrix with the columns indexed by A deleted and rows indexed by B being deleted. Given a function $h : B \to \mathbb{R}$ let $(Rh)(x) = \sum_{y\in B} q(x,y) h(y)$ for all $x \in A$ which again may be thought of as matrix multiplication.

Theorem 17.43. *Let us continue to use the notation and assumptions as described above. If $h : B \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are given functions, then for all $x \in A$ we have;*

$$\mathbb{E}_x[h(X_{T_B})] = \left[(I - Q_A)^{-1} Rh\right](x) \text{ and}$$

$$\mathbb{E}_x\left[\sum_{n<T_B} g(X_n)\right] = \left[(I - Q_A)^{-1} g\right](x).$$

Remark 17.44. Here is a story to go along with the above scenario. Suppose that $g(x)$ is the toll you have to pay for visiting a site $x \in A$ while $h(y)$ is the amount of prize money you get when landing on a point in B. Then $\mathbb{E}_x\left[\sum_{0\le n<T} g(X_n)\right]$ is the expected toll you have to pay before your first exit from A while $\mathbb{E}_x[h(X_T)]$ is your expected winnings upon exiting B.

Here are some typical choices for h and g.

1. If $y \in B$ and $h = \delta_y$, then
$$P_x(X_{T_B} = y) = \left[(I - Q_A)^{-1} R\delta_y\right](x) = \left[(I - Q_A)^{-1} R\right]_{x,y}.$$
2. If $y \in A$ and $g = \delta_y$, then
$$\sum_{n<T_B} g(X_n) = \sum_{n<T_B} \delta_y(X_n) = \#\text{ visits to before hitting } B$$
and hence
$$\begin{aligned} \mathbb{E}_x(\#\text{ visits to before hitting } B) &= \left[(I - Q_A)^{-1} \delta_y\right](x) \\ &= (I - Q_A)^{-1}_{xy}. \end{aligned}$$

3. If $g = \mathbf{1}$, i.e. $g(y) = 1$ for all $y \in A$, then $\sum_{n<T_B} g(X_n) = T_B$ and we find,
$$\mathbb{E}_x T_B = \left[(I - Q_A)^{-1}\mathbf{1}\right]_x = \sum_{y\in A} (I - Q_A)^{-1}_{xy},$$
where $\mathbb{E}_x T_B$ is the expected hitting time of B when starting from x.

Example 17.45. Let us continue the rat in the maze Exercise 17.5 and now suppose that room 3 contains food while room 7 contains a mouse trap.
$$\begin{bmatrix} 1 & 2 & 3\text{ (food)} \\ 4 & 5 & 6 \\ 7\text{ (trap)} & & \end{bmatrix}.$$
We would like to compute the probability that the rat reaches the food before he is trapped. To answer this question we let $A = \{1,2,4,5,6\}$, $B = \{3,7\}$, and $T := T_B$ be the first hitting time of B. Then deleting the 3 and 7 rows of q in Eq. (17.23) leaves the matrix,
$$\begin{array}{c c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix} \end{array}.$$
Deleting the 3 and 7 columns from this matrix gives
$$Q_A = \begin{array}{c c} \begin{matrix} 1 & 2 & 4 & 5 & 6 \end{matrix} & \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix} \end{array}$$
and deleting the $1,2,4,5,$ and 6 columns gives
$$R = Q_{A,B} = \begin{array}{c c} \begin{matrix} 3 & 7 \end{matrix} & \\ \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix} \end{array}.$$
Therefore,
$$I - Q_A = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{3} & 1 & 0 & -\frac{1}{3} & 0 \\ -\frac{1}{3} & 0 & 1 & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix},$$
and using a computer algebra package we find
$$(I - Q_A)^{-1} = \begin{array}{c c} \begin{matrix} 1 & 2 & 4 & 5 & 6 \end{matrix} & \\ \begin{bmatrix} \frac{11}{6} & \frac{5}{4} & \frac{5}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{7}{4} & \frac{3}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{4} & 1 & \frac{1}{3} \\ \frac{2}{3} & 1 & 1 & 2 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{4}{3} \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix} \end{array}.$$
In particular we may conclude,
$$\begin{bmatrix} \mathbb{E}_1 T \\ \mathbb{E}_2 T \\ \mathbb{E}_4 T \\ \mathbb{E}_5 T \\ \mathbb{E}_6 T \end{bmatrix} = (I - Q_A)^{-1}\mathbf{1} = \begin{bmatrix} \frac{17}{3} \\ \frac{14}{3} \\ \frac{14}{3} \\ \frac{16}{3} \\ \frac{11}{3} \end{bmatrix},$$
and
$$\begin{bmatrix} P_1(X_T = 3) & P_1(X_T = 7) \\ P_2(X_T = 3) & P_2(X_T = 3) \\ P_4(X_T = 3) & P_4(X_T = 3) \\ P_5(X_T = 3) & P_5(X_T = 3) \\ P_6(X_T = 3) & P_6(X_T = 7) \end{bmatrix} = (I - Q_A)^{-1} R = \begin{array}{c c} \begin{matrix} 3 & 7 \end{matrix} & \\ \begin{bmatrix} \frac{7}{12} & \frac{5}{12} \\ \frac{3}{4} & \frac{1}{4} \\ \frac{5}{12} & \frac{7}{12} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{5}{6} & \frac{1}{6} \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix} \end{array}.$$
.

Since the event of hitting 3 before 7 is the same as the event $\{X_T = 3\}$, the desired hitting probabilities are
$$\begin{bmatrix} P_1(X_T = 3) \\ P_2(X_T = 3) \\ P_4(X_T = 3) \\ P_5(X_T = 3) \\ P_6(X_T = 3) \end{bmatrix} = \begin{bmatrix} \frac{7}{12} \\ \frac{3}{4} \\ \frac{5}{12} \\ \frac{2}{3} \\ \frac{5}{6} \end{bmatrix}.$$
We can also derive these hitting probabilities from scratch using the first step analysis. In order to do this let
$$h_i = P_i(X_T = 3) = P_i(X_n \text{ hits 3 (food) before 7(trapped)}).$$

By the first step analysis we will have,

$$\begin{aligned} h_i &= \sum_j P_i\left(X_T = 3 | X_1 = j\right) P_i\left(X_1 = j\right) \\ &= \sum_j q\left(i,j\right) P_i\left(X_T = 3 | X_1 = j\right) \\ &= \sum_j q\left(i,j\right) P_j\left(X_T = 3\right) \\ &= \sum_j q\left(i,j\right) h_j \end{aligned}$$

where $h_3 = 1$ and $h_7 = 0$. Looking at the jump diagram (Figure 17.3) we easily

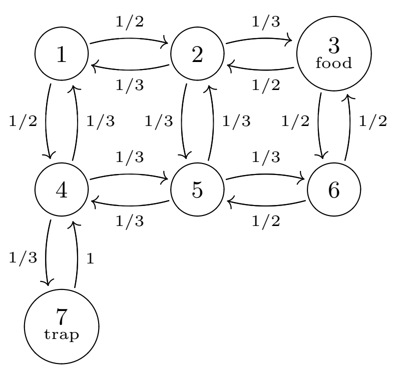


Fig. 17.3. The jump diagram for our proverbial rat in the maze.

find

$$\begin{aligned} h_1 &= \frac{1}{2}\left(h_2 + h_4\right) \\ h_2 &= \frac{1}{3}\left(h_1 + h_3 + h_5\right) = \frac{1}{3}\left(h_1 + 1 + h_5\right) \\ h_4 &= \frac{1}{3}\left(h_1 + h_5 + h_7\right) = \frac{1}{3}\left(h_1 + h_5\right) \\ h_5 &= \frac{1}{3}\left(h_2 + h_4 + h_6\right) \\ h_6 &= \frac{1}{2}\left(h_3 + h_5\right) = \frac{1}{2}\left(1 + h_5\right) \end{aligned}$$

and the solutions to these equations are (as seen before) given by

$$\left[h_1 = \frac{7}{12}, h_2 = \frac{3}{4}, h_4 = \frac{5}{12}, h_5 = \frac{2}{3}, h_6 = \frac{5}{6}\right]. \tag{17.48}$$

Similarly, if

$$k_i := P_i\left(X_T = 7\right) = P_i\left(X_n \text{ is trapped before dinner}\right),$$

we need only use the above equations with h replaced by k and now taking $k_3 = 0$ and $k_7 = 1$ to find,

$$\begin{aligned} k_1 &= \frac{1}{2}\left(k_2 + k_4\right) \\ k_2 &= \frac{1}{3}\left(k_1 + k_5\right) \\ k_4 &= \frac{1}{3}\left(k_1 + k_5 + 1\right) \\ k_5 &= \frac{1}{3}\left(k_2 + k_4 + k_6\right) \\ k_6 &= \frac{1}{2}k_5 \end{aligned}$$

and then solve to find,

$$\left[k_1 = \frac{5}{12}, k_2 = \frac{1}{4}, k_4 = \frac{7}{12}, k_5 = \frac{1}{3}, k_6 = \frac{1}{6}\right]. \tag{17.49}$$

Notice that the sum of the hitting probabilities in Eqs. (17.48) and (17.49) add up to 1 as they should.

17.5.1 Invariant distributions and return times

For this subsection suppose that $S = \{1, 2, \ldots, n\}$ and Q_{ij} is a Markov matrix. To each state $i \in S$, let

$$\tau_i := \min\{n \geq 1 : X_n = i\} \tag{17.50}$$

be the **first passage time of the chain to site** i.

Proposition 17.46. *The Markov matrix Q has an invariant distribution.*

Proof. If $\mathbf{1} := \left[1\ 1\ \ldots\ 1\right]^{\mathrm{tr}}$, then $Q\mathbf{1} = \mathbf{1}$ from which it follows that

$$0 = \det\left(Q - I\right) = \det\left(Q^{\mathrm{tr}} - I\right).$$

Therefore there exists a non-zero row vector ν such that $Q^{\mathrm{tr}}\nu^{\mathrm{tr}} = \nu^{\mathrm{tr}}$ or equivalently that $\nu Q = \nu$. At this point we would be done if we knew that $\nu_i \geq 0$ for all i – but we don't. So let $\pi_i := |\nu_i|$ and observe that

$$\pi_i = |\nu_i| = \left|\sum_{k=1}^n \nu_k Q_{ki}\right| \le \sum_{k=1}^n |\nu_k| Q_{ki} \le \sum_{k=1}^n \pi_k Q_{ki}.$$

We now claim that in fact $\pi = \pi Q$. If this were not the case we would have $\pi_i < \sum_{k=1}^n \pi_k Q_{ki}$ for some i and therefore

$$0 < \sum_{i=1}^n \pi_i < \sum_{i=1}^n \sum_{k=1}^n \pi_k Q_{ki} = \sum_{k=1}^n \sum_{i=1}^n \pi_k Q_{ki} = \sum_{k=1}^n \pi_k$$

which is a contradiction. So all that is left to do is normalize π_i so $\sum_{i=1}^n \pi_i = 1$ and we are done. ■

We are now going to assume that Q is **irreducible** which means that for all $i \neq j$ there exists $n \in \mathbb{N}$ such that $Q_{ij}^n > 0$. Alternatively put this implies that $P_i(T_j < \infty) = P_i(\tau_j < \infty) > 0$ for all $i \neq j$. By Corollary 17.42 we know that $\mathbb{E}_i[\tau_j] = \mathbb{E}_i T_j < \infty$ for all $i \neq j$ and it is not too hard to see that $\mathbb{E}_i \tau_i < \infty$ also holds. The fact that $\mathbb{E}_i \tau_i < \infty$ for all $i \in S$ will come out of the proof of the next proposition as well.

Proposition 17.47. *If Q is irreducible, then there is precisely one invariant distribution, π, which is given by $\pi_i = 1/(\mathbb{E}_i \tau_i) > 0$ for all $i \in S$.*

Proof. We begin by using the first step analysis to write equations for $\mathbb{E}_i[\tau_j]$ as follows:

$$\begin{aligned}\mathbb{E}_i[\tau_j] &= \sum_{k=1}^n \mathbb{E}_i[\tau_j | X_1 = k] Q_{ik} = \sum_{k \neq j} \mathbb{E}_i[\tau_j | X_1 = k] Q_{ik} + Q_{ij} 1 \\ &= \sum_{k \neq j} (\mathbb{E}_k[\tau_j] + 1) Q_{ik} + Q_{ij} 1 = \sum_{k \neq j} \mathbb{E}_k[\tau_j] Q_{ik} + 1.\end{aligned}$$

and therefore,

$$\mathbb{E}_i[\tau_j] = \sum_{k \neq j} Q_{ik} \mathbb{E}_k[\tau_j] + 1. \tag{17.51}$$

Now suppose that π is any invariant distribution for Q, then multiplying Eq. (17.51) by π_i and summing on i shows

$$\begin{aligned}\sum_{i=1}^n \pi_i \mathbb{E}_i[\tau_j] &= \sum_{i=1}^n \pi_i \sum_{k \neq j} Q_{ik} \mathbb{E}_k[\tau_j] + \sum_{i=1}^n \pi_i 1 \\ &= \sum_{k \neq j} \pi_k \mathbb{E}_k[\tau_j] + 1.\end{aligned}$$

Since $\sum_{k \neq j} \pi_k \mathbb{E}_k[\tau_j] < \infty$ we may cancel it from both sides of this equation in order to learn $\pi_j \mathbb{E}_j[\tau_j] = 1$. ■

We may use Eq. (17.51) to compute $\mathbb{E}_i[\tau_j]$ in examples. To do this, fix j and set $v_i := \mathbb{E}_i \tau_j$. Then Eq. (17.51) states that $v = Q^{(j)} v + \mathbf{1}$ where $Q^{(j)}$ denotes Q with the j^{th} – column replaced by all zeros. Thus we have

$$(\mathbb{E}_i \tau_j)_{i=1}^n = \left(I - Q^{(j)}\right)^{-1} \mathbf{1}, \tag{17.52}$$

i.e.

$$\begin{bmatrix} \mathbb{E}_1 \tau_j \\ \vdots \\ \mathbb{E}_n \tau_j \end{bmatrix} = \left(I - Q^{(j)}\right)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{17.53}$$

17.5.2 Some worked examples

Example 17.48. Let $S = \{1, 2\}$ and $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with jump diagram in Figure 17.4. In this case $Q^{2n} = I$ while $Q^{2n+1} = Q$ and therefore $\lim_{n \to \infty} Q^n$ does not

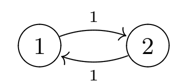


Fig. 17.4. A non-random chain.

exist. On the other hand it is easy to see that the invariant distribution, π, for Q is $\pi = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ and, moreover,

$$\frac{Q + Q^2 + \cdots + Q^N}{N} \to \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \end{bmatrix}.$$

Let us compute

$$\begin{bmatrix} \mathbb{E}_1 \tau_1 \\ \mathbb{E}_2 \tau_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbb{E}_1 \tau_2 \\ \mathbb{E}_2 \tau_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

so that indeed, $\pi_1 = 1/\mathbb{E}_1 \tau_1$ and $\pi_2 = 1/\mathbb{E}_2 \tau_2$. Of course $\tau_1 = 2$ (P_1 -a.s.) and $\tau_2 = 2$ (P_2 -a.s.) so that it is obvious that $\mathbb{E}_1 \tau_1 = \mathbb{E}_2 \tau_2 = 2$.

Example 17.49. Again let $S = \{1, 2\}$ and $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ with jump diagram in Figure 17.5. In this case the chain is not irreducible and every $\pi = [a\ b]$ with $a + b = 1$ and $a, b \ge 0$ is an invariant distribution.

Fig. 17.5. A simple non-irreducible chain.

Example 17.50. Suppose that $S = \{1,2,3\}$, and

$$Q = \begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix}$$

has the jump graph given by 17.6. Notice that $Q_{11}^2 > 0$ and $Q_{11}^3 > 0$ that Q is

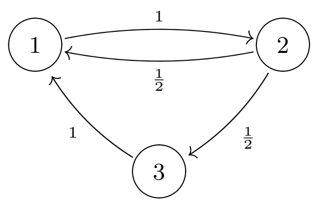


Fig. 17.6. A simple 3 state jump diagram.

"aperiodic." We now find the invariant distribution,

$$\operatorname{Nul}(Q-I)^{\mathrm{tr}} = \operatorname{Nul}\begin{bmatrix} -1 & \frac{1}{2} & 1 \\ 1 & -1 & 0 \\ 0 & \frac{1}{2} & -1 \end{bmatrix} = \mathbb{R}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.$$

Therefore the invariant distribution is given by

$$\pi = \frac{1}{5}\begin{bmatrix} 2 & 2 & 1 \end{bmatrix}.$$

Let us now observe that

$$Q^2 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$Q^3 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$Q^{20} = \begin{bmatrix} \frac{409}{1024} & \frac{205}{512} & \frac{205}{1024} \\ \frac{205}{512} & \frac{409}{1024} & \frac{205}{1024} \\ \frac{205}{512} & \frac{205}{512} & \frac{51}{256} \end{bmatrix} = \begin{bmatrix} 0.39941 & 0.40039 & 0.20020 \\ 0.40039 & 0.39941 & 0.20020 \\ 0.40039 & 0.40039 & 0.19922 \end{bmatrix}.$$

Let us also compute $\mathbb{E}_2\tau_3$ via,

$$\begin{bmatrix} \mathbb{E}_1\tau_3 \\ \mathbb{E}_2\tau_3 \\ \mathbb{E}_3\tau_3 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$$

so that

$$\frac{1}{\mathbb{E}_3\tau_3} = \frac{1}{5} = \pi_3.$$

Example 17.51. The transition matrix,

$$Q = \begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array} \begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix}$$

is represented by the jump diagram in Figure 17.7. This chain is aperiodic. We

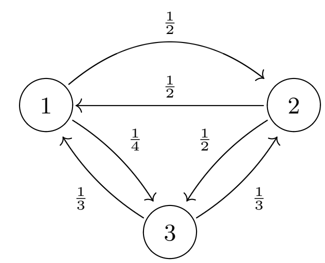


Fig. 17.7. In the above diagram there are jumps from 1 to 1 with probability 1/4 and jumps from 3 to 3 with probability 1/3 which are not explicitly shown but must be inferred by conservation of probability.

find the invariant distribution as,

$$\begin{aligned} \operatorname{Nul}(Q-I)^{\mathrm{tr}} &= \operatorname{Nul}\left(\begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)^{\mathrm{tr}} \\ &= \operatorname{Nul}\left(\begin{bmatrix} -\frac{3}{4} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & -1 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & -\frac{2}{3} \end{bmatrix}\right) = \mathbb{R}\begin{bmatrix} 1 \\ \frac{5}{6} \\ 1 \end{bmatrix} = \mathbb{R}\begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix} \end{aligned}$$

$$\pi = \frac{1}{17}\begin{bmatrix} 6 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 0.35294 & 0.29412 & 0.35294 \end{bmatrix}.$$

In this case

$$Q^{10} = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}^{10} = \begin{bmatrix} 0.352\,98 & 0.294\,04 & 0.352\,98 \\ 0.352\,89 & 0.294\,23 & 0.352\,89 \\ 0.352\,95 & 0.294\,1 & 0.352\,95 \end{bmatrix}.$$

Let us also compute

$$\begin{bmatrix} \mathbb{E}_1\tau_2 \\ \mathbb{E}_2\tau_2 \\ \mathbb{E}_3\tau_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/4 & 0 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{11}{5} \\ \frac{17}{5} \\ \frac{13}{5} \end{bmatrix}$$

so that

$$1/\mathbb{E}_2\tau_2 = 5/17 = \pi_2.$$

Example 17.52. Consider the following Markov matrix,

$$Q = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 0 & 3/4 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1/4 & 3/4 & 0 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \end{array}$$

with jump diagram in Figure 17.8. Since this matrix is doubly stochastic (i.e

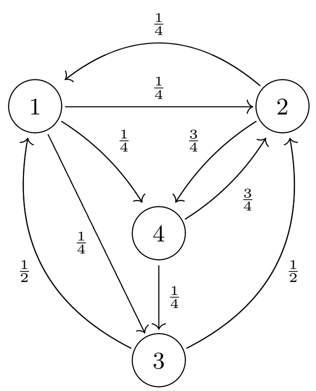


Fig. 17.8. The jump diagram for Q.

$\sum_{i=1}^4 Q_{ij} = 1$ for all j as well as $\sum_{j=1}^4 Q_{ij} = 1$ for all i), it is easy to check that $\pi = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$. Let us compute $\mathbb{E}_3\tau_3$ as follows

$$\begin{bmatrix} \mathbb{E}_1\tau_3 \\ \mathbb{E}_2\tau_3 \\ \mathbb{E}_3\tau_3 \\ \mathbb{E}_4\tau_3 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/4 & 1/4 & 0 & 1/4 \\ 1/4 & 0 & 0 & 3/4 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{50}{17} \\ \frac{52}{17} \\ 4 \\ \frac{30}{17} \end{bmatrix}$$

so that $\mathbb{E}_3\tau_3 = 4 = 1/\pi_4$ as it should be. Similarly,

$$\begin{bmatrix} \mathbb{E}_1\tau_2 \\ \mathbb{E}_2\tau_2 \\ \mathbb{E}_3\tau_2 \\ \mathbb{E}_4\tau_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/4 & 0 & 1/4 & 1/4 \\ 1/4 & 0 & 0 & 3/4 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 3/4 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{54}{17} \\ 4 \\ \frac{44}{17} \\ \frac{50}{17} \end{bmatrix}$$

and again $\mathbb{E}_2\tau_2 = 4 = 1/\pi_2$.

17.5.3 Exercises

Exercise 17.10 (2nd order recurrence relations). Let a, b, c be real numbers with $a \neq 0 \neq c$, $\alpha, \beta \in \mathbb{Z} \cup \{\pm\infty\}$ with $\alpha < \beta$, and suppose $\{u(x) : x \in [\alpha, \beta] \cap \mathbb{Z}\}$ solves the second order homogeneous recurrence relation:

$$au(x+1) + bu(x) + cu(x-1) = 0 \tag{17.54}$$

for $\alpha < x < \beta$. Show:

1. for any $\lambda \in \mathbb{C}$,
$$a\lambda^{x+1} + b\lambda^x + c\lambda^{x-1} = \lambda^{x-1} p(\lambda) \tag{17.55}$$
where $p(\lambda) = a\lambda^2 + b\lambda + c$ is the **characteristic polynomial** associated to Eq. (17.54).
Let $\lambda_\pm = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ be the roots of $p(\lambda)$ and suppose for the moment that $b^2 - 4ac \neq 0$. From Eq. (17.54) it follows that for any choice of $A_\pm \in \mathbb{R}$, the function,
$$w(x) := A_+\lambda_+^x + A_-\lambda_-^x,$$
solves Eq. (17.54) for all $x \in \mathbb{Z}$.

2. Show there is a unique choice of constants, $A_\pm \in \mathbb{R}$, such that the function $u(x)$ is given by
$$u(x) := A_+\lambda_+^x + A_-\lambda_-^x \text{ for all } \alpha \le x \le \beta.$$
3. Now suppose that $b^2 = 4ac$ and $\lambda_0 := -b/(2a)$ is the double root of $p(\lambda)$. Show for any choice of A_0 and A_1 in $\mathbb{R}$ that
$$w(x) := (A_0 + A_1 x)\lambda_0^x$$
solves Eq. (17.54) for all $x \in \mathbb{Z}$. **Hint:** Differentiate Eq. (17.55) with respect to λ and then set $\lambda = \lambda_0$.
4. Again show that any function u solving Eq. (17.54) is of the form $u(x) = (A_0 + A_1 x)\lambda_0^x$ for $\alpha \le x \le \beta$ for some unique choice of constants $A_0, A_1 \in \mathbb{R}$.

In the next couple of exercises you are going to use first step analysis to show that a simple unbiased random walk on $\mathbb{Z}$ is null recurrent. We let $\{X_n\}_{n=0}^\infty$ be the Markov chain with values in $\mathbb{Z}$ with transition probabilities given by
$$P(X_{n+1} = x \pm 1 | X_n = x) = 1/2 \text{ for all } n \in \mathbb{N}_0 \text{ and } x \in \mathbb{Z}.$$
Further let $a, b \in \mathbb{Z}$ with $a < 0 < b$ and
$$T_{a,b} := \min\{n : X_n \in \{a,b\}\} \text{ and } T_b := \inf\{n : X_n = b\}.$$
We know by Corollary[4] 17.42 that $\mathbb{E}_0[T_{a,b}] < \infty$ from which it follows that $P(T_{a,b} < \infty) = 1$ for all $a < 0 < b$.

Exercise 17.11. Let $w_x := P_x\left(X_{T_{a,b}} = b\right) := P\left(X_{T_{a,b}} = b | X_0 = x\right)$.

1. Use first step analysis to show for $a < x < b$ that
$$w_x = \frac{1}{2}(w_{x+1} + w_{x-1}) \tag{17.56}$$
provided we define $w_a = 0$ and $w_b = 1$.
2. Use the results of Exercise 17.10 to show
$$P_x\left(X_{T_{a,b}} = b\right) = w_x = \frac{1}{b-a}(x-a). \tag{17.57}$$
3. Let
$$T_b := \begin{cases} \min\{n : X_n = b\} & \text{if} \quad \{X_n\} \text{ hits } b \\ \infty & \text{otherwise} \end{cases}$$
be the first time $\{X_n\}$ hits b. Explain why, $\left\{X_{T_{a,b}} = b\right\} \subset \{T_b < \infty\}$ and use this along with Eq. (17.57) to conclude[5] that $P_x(T_b < \infty) = 1$ for all $x < b$. (By symmetry this result holds true for all $x \in \mathbb{Z}$.)

[4] Apply this corollary to finite walk in $[a,b] \cap \mathbb{Z}$.
[5] The fact that $P_j(T_b < \infty) = 1$ is also follows from Example 10.55 above.

Exercise 17.12. The goal of this exercise is to give a second proof of the fact that $P_x(T_b < \infty) = 1$. Here is the outline:

1. Let $w_x := P_x(T_b < \infty)$. Again use first step analysis to show that w_x satisfies Eq. (17.56) for all x with $w_b = 1$.
2. Use Exercise 17.10 to show that there is a constant, c, such that
$$w_x = c(x-b) + 1 \text{ for all } x \in \mathbb{Z}.$$
3. Explain why c must be zero to again show that $P_x(T_b < \infty) = 1$ for all $x \in \mathbb{Z}$.

Exercise 17.13. Let $T = T_{a,b}$ and $u_x := \mathbb{E}_x T := \mathbb{E}[T | X_0 = x]$.

1. Use first step analysis to show for $a < x < b$ that
$$u_x = \frac{1}{2}(u_{x+1} + u_{x-1}) + 1 \tag{17.58}$$
with the convention that $u_a = 0 = u_b$.
2. Show that
$$u_x = A_0 + A_1 x - x^2 \tag{17.59}$$
solves Eq. (17.58) for any choice of constants A_0 and A_1.
3. Choose A_0 and A_1 so that u_x satisfies the boundary conditions, $u_a = 0 = u_b$. Use this to conclude that
$$\mathbb{E}_x T_{a,b} = -ab + (b+a)x - x^2 = -a(b-x) + bx - x^2. \tag{17.60}$$

Remark 17.53. Notice that $T_{a,b} \uparrow T_b = \inf\{n : X_n = b\}$ as $a \downarrow -\infty$, and so passing to the limit as $a \downarrow -\infty$ in Eq. (17.60) shows
$$\mathbb{E}_x T_b = \infty \text{ for all } x < b.$$
Combining the last couple of exercises together shows that $\{X_n\}$ is "null - recurrent."

Exercise 17.14. Let $T = T_b$. The goal of this exercise is to give a second proof of the fact and $u_x := \mathbb{E}_x T = \infty$ for all $x \neq b$. Here is the outline. Let $u_x := \mathbb{E}_x T \in [0, \infty] = [0, \infty) \cup \{\infty\}$.

1. Note that $u_b = 0$ and, by a first step analysis, that u_x satisfies Eq. (17.58) for all $x \neq b$ – allowing for the possibility that some of the u_x may be infinite.
2. Argue, using Eq. (17.58), that if $u_x < \infty$ for some $x < b$ then $u_y < \infty$ for all $y < b$. Similarly, if $u_x < \infty$ for some $x > b$ then $u_y < \infty$ for all $y > b$.
3. If $u_x < \infty$ for all $x > b$ then u_x must be of the form in Eq. (17.59) for some A_0 and A_1 in $\mathbb{R}$ such that $u_b = 0$. However, this would imply, $u_x = \mathbb{E}_x T \to -\infty$ as $x \to \infty$ which is impossible since $\mathbb{E}_x T \ge 0$ for all x. Thus we must conclude that $\mathbb{E}_x T = u_x = \infty$ for all $x > b$. (A similar argument works if we assume that $u_x < \infty$ for all $x < b$.)

17.6 Appendix: Kolmogorov's extension theorem II

The Kolmogorov extension Theorem 9.51 generalizes to the case where $\mathbb{N}$ is replaced by an arbitrary index set, T. Let us set up the notation for this theorem. Let T be an arbitrary index set, $\{(S_t, \mathcal{S}_t)\}_{t\in T}$ be a collection of standard Borel spaces, $S = \prod_{t\in T} S_t$, $\mathcal{S} := \otimes_{t\in T}\mathcal{S}_t$, and for $\Lambda \subset T$ let

$$\left(S_\Lambda := \prod_{t\in\Lambda} S_t, \mathcal{S}_\Lambda := \otimes_{t\in\Lambda}\mathcal{S}_t\right)$$

and $X_\Lambda : S \to S_\Lambda$ be the projection map, $X_\Lambda(x) := x|_\Lambda$. If $\Lambda \subset \Lambda' \subset T$, also let $X_{\Lambda,\Lambda'} : S_{\Lambda'} \to S_\Lambda$ be the projection map, $X_{\Lambda,\Lambda'}(x) := x|_\Lambda$ for all $x \in S_{\Lambda'}$.

Theorem 17.54 (Kolmogorov). *For each $\Lambda \subset_f T$ (i.e. $\Lambda \subset T$ and $\#(\Lambda) < \infty$), let μ_Λ be a probability measure on $(S_\Lambda, \mathcal{S}_\Lambda)$. We further suppose $\{\mu_\Lambda\}_{\Lambda\subset_f T}$ satisfy the following compatibility relations;*

$$\mu_{\Lambda'} \circ X^{-1}_{\Lambda,\Lambda'} = \mu_\Lambda \text{ for all } \Lambda \subset \Lambda' \subset_f T. \tag{17.61}$$

Then there exists a unique probability measure, P, on $(S, \mathcal{S})$ such that $P\circ X_\Lambda^{-1} = \mu_\Lambda$ for all $\Lambda \subset_f T$.

Proof. (For slight variation on the proof of this theorem given here, see Exercise 17.16.) Let

$$\mathcal{A} := \cup_{\Lambda\subset_f T} X_\Lambda^{-1}(\mathcal{S}_\Lambda)$$

and for $A = X_\Lambda^{-1}(A') \in \mathcal{A}$, let $P(A) := \mu_\Lambda(A')$. The compatibility conditions in Eq. (17.61) imply P is a well defined finitely additive measure on the algebra, $\mathcal{A}$. We now complete the proof by showing P is continuous on $\mathcal{A}$.

To this end, suppose $A_n := X^{-1}_{\Lambda_n}(A'_n) \in \mathcal{A}$ with $A_n \downarrow \emptyset$ as $n \to \infty$. Let $\Lambda := \cup_{n=1}^\infty \Lambda_n$ – a countable subset of T. Owing to Theorem 9.51, there is a unique probability measure, P_Λ, on $(S_\Lambda, \mathcal{S}_\Lambda)$ such that $P_\Lambda\left(X_\Gamma^{-1}(A)\right) = \mu_\Gamma(A)$ for all $\Gamma \subset_f \Lambda$ and $A \in \mathcal{S}_\Gamma$. Hence if we let $\tilde{A}_n := X^{-1}_{\Lambda,\Lambda_n}(A_n)$, we then have

$$P(A_n) = \mu_{\Lambda_n}(A'_n) = P_\Lambda\left(\tilde{A}_n\right)$$

with $\tilde{A}_n \downarrow \emptyset$ as $n \to \infty$. Since P_Λ is a measure, we may conclude

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P_\Lambda\left(\tilde{A}_n\right) = 0.$$

■

Exercise 17.15. Let us write $\Lambda \subset_c T$ to mean $\Lambda \subset T$ and Λ is at most countable. Show

$$\mathcal{S} = \cup_{\Lambda\subset_c T} X_\Lambda^{-1}(\mathcal{S}_\Lambda). \tag{17.62}$$

Hint: Verify Eq. (17.62) by showing $\mathcal{S}_0 := \cup_{\Lambda\subset_c T} X_\Lambda^{-1}(\mathcal{S}_\Lambda)$ is a σ – algebra.

Exercise 17.16. For each $\Lambda \subset T$, let $\mathcal{S}'_\Lambda := X_\Lambda^{-1}(\mathcal{S}_\Lambda) = \sigma(X_i : i \in \Lambda) \subset \mathcal{S}$. Show;

1. if $U, V \subset T$ then $\mathcal{S}'_U \cap \mathcal{S}'_V = \mathcal{S}'_{U\cap V}$.
2. By Theorem 9.51, if $U, V \subset_c T$, there exists unique probability measures, P_U and P_V on $\mathcal{S}'_U$ and $\mathcal{S}'_V$ respectively such that $P_U \circ X_\Lambda^{-1} = \mu_\Lambda$ for all $\Lambda \subset_f U$ and $P_V \circ X_\Lambda^{-1} = \mu_\Lambda$ for all $\Lambda \subset_f V$. Show $P_U = P_V$ on $\mathcal{S}'_U \cap \mathcal{S}'_V$. Hence for any $A \in \mathcal{S}$ we may define $P(A) := P_U(A)$ provided $A \in \mathcal{S}'_U$.
3. Show P defined in the previous item is a countably additive measure on $\mathcal{S}$.

17.7 Removing the standard Borel restriction

Theorem 17.55. *Let $\{(S_n, \mathcal{S}_n)\}_{n\in\mathbb{N}_0}$ be a collection of measurable spaces, $S = \prod_{n=0}^\infty S_n$ and $\mathcal{S} := \otimes_{n=0}^\infty \mathcal{S}_n$. Moreover for each $n \in \mathbb{N}_0$ let $S^n := S_0 \times \cdots \times S_n$ and $\mathcal{S}^n := \mathcal{S}_0 \otimes \cdots \otimes \mathcal{S}_n$. We further suppose that μ_0 is a given probability measure on $(S_0, \mathcal{S}_0)$ and $T_n : S^{n-1} \times \mathcal{S}_n \to [0,1]$ for $n = 1, 2, \ldots$ are give probability kernels on $S^{n-1} \times S_n$. Finally let μ_n be the probability measure on $(S^n, \mathcal{S}^n)$ defined inductively by,*

$$\mu_n(dx_0, \ldots, dx_n) = \mu_{n-1}(dx_0, \ldots, dx_{n-1}) T_n(x_0, \ldots, x_{n-1}, dx_n) \ \forall\, n \in \mathbb{N}. \tag{17.63}$$

Then there exists a unique probability measure, P on $(S, \mathcal{S})$ such that

$$P(f) = \int_{S^n} F d\mu_n$$

whenever $f(x) = F(x_0, \ldots, x_n)$ for some $F \in (\mathcal{S}_n)_b$.

Remark 17.56 (Heuristic proof). Before giving the formal proof of this theorem let me indicate the main ideas. Let $X_i : S \to S_i$ be the projection maps and $\mathcal{B}_n := \sigma(X_0, \ldots, X_n)$. If P exists, then

$$\begin{aligned} P[F(X_0, \ldots, X_{n+1})|\mathcal{B}_n] &= T_n(X_0, \ldots, X_n; F(X_0, \ldots, X_n, \cdot)) \\ &= (TF(X_0, \ldots, X_n, \cdot))(X_0, \ldots, X_n). \end{aligned}$$

Indeed,

$$\begin{aligned} &\mathbb{E}[T_n(X_0, \ldots, X_n; F(X_0, \ldots, X_n, \cdot)) G(X_0, \ldots, X_n)] \\ &\quad = \int T_n(x_0, \ldots, x_n, dx_{n+1}) F(x_0, \ldots, x_n, x_{n+1}) G(x_0, \ldots, x_n) d\mu_n(x_0, \ldots, x_n) \\ &\quad = \int F(x_0, \ldots, x_n, x_{n+1}) G(x_0, \ldots, x_n) d\mu_{n+1}(x_0, \ldots, x_n, x_{n+1}) \\ &\quad = \mathbb{E}[F(X_0, \ldots, X_{n+1}) G(X_0, \ldots, X_n)]. \end{aligned}$$

Now suppose that $f_n = F_n(X_0, \dots, X_n)$ is a decreasing sequence of functions such that $\lim_{n\to\infty} P(f_n) =: \varepsilon > 0$. Letting $f_\infty := \lim_{n\to\infty} f_n$ we would have $f_n \geq f_\infty$ for all n and therefore $f_n \geq \mathbb{E}[f_\infty|\mathcal{B}_n] := \bar{f}_n$. We also use

$$\begin{aligned}
&\bar{f}_n(X_0, X_1, \dots, X_n) \\
&\quad = \mathbb{E}[f_\infty|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[f_\infty|\mathcal{B}_{n+1}]|\mathcal{B}_n] \\
&\quad = \mathbb{E}[\bar{f}_{n+1}|\mathcal{B}_n] = \int \bar{f}_{n+1}(X_0, X_1, \dots, x_{n+1}) T_{n+1}(X_0, \dots, X_n, dx_{n+1})
\end{aligned}$$

and $P(\bar{f}_n) = P(f_\infty) = \lim_{m\to\infty} P(f_m) = \varepsilon > 0$ (we only use the case where $n = 0$ here). Since $P(\bar{f}_0(X_0)) = \varepsilon > 0$, there exists $x_0 \in S_0$ such that

$$\varepsilon \leq \bar{f}_0(x_0) = \mathbb{E}[\bar{f}_1|\mathcal{B}_0] = \int \bar{f}_1(x_0, x_1) T_1(x_0, dx_1)$$

and so similarly there exists $x_1 \in S_1$ such that

$$\varepsilon \leq \bar{f}_1(x_0, x_1) = \int \bar{f}_2(x_0, x_1, x_2) T_2(x_0, x_1, dx_2).$$

Again it follows that there must exists an $x_2 \in S_2$ such that $\varepsilon \leq \bar{f}_2(x_0, x_1, x_2)$. We continue on this way to find and $x \in S$ such that

$$f_n(x) \geq \bar{f}_n(x_0, \dots, x_n) \geq \varepsilon \text{ for all } n.$$

Thus if $P(f_n) \downarrow \varepsilon > 0$ then $\lim_{n\to\infty} f_n(x) \geq \varepsilon \neq 0$ as desired.

Proof. Now onto the formal proof. Let $\mathbb{S}$ denote the space of finitely based bounded cylinder functions on S, i.e. functions of the form $f(x) = F(x_0, \dots, x_n)$ with $F \in \mathcal{S}_b^n$. For such an f we define

$$I(f) := P_n(F).$$

It is easy to check that I is a well defined positive linear functional on $\mathbb{S}$.

Now suppose that $0 \leq f_n \in \mathbb{S}$ are forms a decreasing sequence of functions such that $\lim_{n\to\infty} I(f_n) = \varepsilon > 0$. We wish to show that $\lim_{n\to\infty} f_n(x) \neq 0$ for every $x \in S$. By assumption, $f_n(x) = F_n(x_0, \dots, x_{N_n})$ for some $N_n \in \mathbb{N}$ of which we may assume $N_0 < N_1 < N_2 < \dots$. Moreover if $N_0 = 2 < N_1 = 5 < N_2 = 7 < \dots$, we may replace $(f_0, f_1, \dots)$ by

$$(g_0, g_1, g_2, \dots) = (1, 1, f_0, f_0, f_0, f_1, f_1, f_2, \dots).$$

Noting that $\lim_{n\to\infty} g_n = \lim_{n\to\infty} f_n$, $\lim_{n\to\infty} I(g_n) = I(f_n)$, and $g_n(x) = G_n(x_0, \dots, x_n)$ for some $G_n \in \mathcal{S}_b^n$, we may now assume that $f_n(x) = F_n(x_0, \dots, x_n)$ with $F_n \in \mathcal{S}_b^n$.

For any $k \leq n$ let

$$F_n^k(x_0, \dots, x_k) := \int \cdots \int F_n(x_0, \dots, x_n) \prod_{l=k}^{n-1} T_l(x_0, \dots, x_l, dx_{l+1})$$

which is an explicit version of $P_n[F_n(x_0, \dots, x_n)|x_0, \dots, x_k] “=” \mathbb{E}[f_n|\mathcal{B}_k](x)$. By construction of the measures P_n it follows that

$$P_k F_n^k = P_n F_n = I(f_n) \text{ for all } k \leq n. \tag{17.64}$$

Since

$$F_n(x_0, \dots, x_n) = f_n(x) \leq f_{n+1}(x) = F_{n+1}(x_0, \dots, x_n, x_{n+1}),$$

it follows that

$$\begin{aligned}
F_n^k(x_0, \dots, x_k) &= \int F_n(x_0, \dots, x_n) \prod_{l=k}^{n} T_l(x_0, \dots, x_l, dx_{l+1}) \\
&\leq \int F_{n+1}(x_0, \dots, x_n, x_{n+1}) \prod_{l=k}^{n} T_l(x_0, \dots, x_l, dx_{l+1}) \\
&= F_{n+1}^k(x_0, \dots, x_k).
\end{aligned}$$

Thus we may define $F^k(x_0, \dots, x_k) :=\downarrow \lim_{n\to\infty} F_n^k(x_0, \dots, x_k)$ which is formally equal to $\mathbb{E}[f|\mathcal{B}_k](x)$. Hence we expect that

$$F^k(x_0, \dots, x_k) = \int F^{k+1}(x_0, \dots, x_k, x_{k+1}) T_k(x_0, \dots, x_k, dx_{k+1}) \tag{17.65}$$

by the tower property for conditional expectations. This is indeed that case since,

$$\begin{aligned}
&\int F^{k+1}(x_0, \dots, x_k, x_{k+1}) T_k(x_0, \dots, x_k, dx_{k+1}) \\
&\quad = \lim_{n\to\infty} \int F_n^{k+1}(x_0, \dots, x_k, x_{k+1}) T_k(x_0, \dots, x_k, dx_{k+1})
\end{aligned}$$

while

$$\begin{aligned}
&\int F_n^{k+1}(x_0, \dots, x_k, x_{k+1}) T_k(x_0, \dots, x_k, dx_{k+1}) \\
&\quad = \int \left[\int \cdots \int F_n(x_0, \dots, x_n) \prod_{l=k+1}^{n} T_l(x_0, \dots, x_l, dx_{l+1})\right] T_k(x_0, \dots, x_k, dx_{k+1}) \\
&\quad = \int \cdots \int F_n(x_0, \dots, x_n) \prod_{l=k}^{n} T_l(x_0, \dots, x_l, dx_{l+1}) \\
&\quad = F_n^k(x_0, \dots, x_k).
\end{aligned}$$

We may now pass to the limit as $n \to \infty$ in Eq. (17.64) to find

$$P_k\left(F^k\right) = \varepsilon > 0 \text{ for all } k.$$

For $k = 0$ it follows that $F^0(x_0) \ge \varepsilon > 0$ for some $x_0 \in S_0$ for otherwise $P_0(F_0) < \varepsilon$. But

$$\varepsilon \le F^0(x_0) = \int F^1(x_0, x_1)\, T_1(x_0, dx_1)$$

and so there exists x_1 such that

$$\varepsilon \le F^1(x_0, x_1) = \int F^2(x_0, x_1, x_2)\, T_2(x_0, x_1, dx_2)$$

and hence there exists x_2 such that $\varepsilon \le F^2(x_0, x_1, x_2)$, etc. etc. Thus in the end we find an $x = (x_0, x_1, \dots) \in S$ such that $F^k(x_0, \dots, x_n) \ge \varepsilon$ for all k. Finally recall that

$$F_n^k(x_0, \dots, x_k) \ge F^k(x_0, \dots, x_k) \ge \varepsilon \text{ for all } k \le n.$$

Taking $k = n$ then implies,

$$f_n(x) = F_n^n(x_0, \dots, x_n) \ge F^n(x_0, \dots, x_n) \ge \varepsilon \text{ for all } n.$$

Therefore we have constructed a $x \in S$ such that $f(x) = \lim_{n\to\infty} f_n(x) \ge \varepsilon > 0$.

We may now use the Caratheodory extension theorem to show that P extends to a countably additive measure on $(S, \mathcal{S})$. Indeed suppose $A_n \in \mathcal{A}(X_i : i \in \mathbb{N}_0)$. If $A_n \downarrow \emptyset$ then $1_{A_n} \downarrow 0$ and by what we have just proved,

$$P(A_n) = P(1_{A_n}) \downarrow 0 \text{ as } n \to \infty.$$

■

Corollary 17.57 (Infinite Product Measures). *Let $\{(S_n, \mathcal{S}_n, \mu_n)\}_{n\in\mathbb{N}_0}$ be a collection of measurable spaces, then there exists P on $(S, \mathcal{S})$ such that*

$$P(f) = P(f) = \int_{S^n} F(x_0, \dots, x_n)\, d\nu_0(x_0) \dots d\nu_n(x_n)$$

whenever $f(x) = F(x_0, \dots, x_n)$ for some $F \in (\mathcal{S}_n)_b$.

Proof. Let $\mu_0 = \nu_0$ and

$$T_n(x_0, \dots, x_{n-1}, dx_{n+1}) = v_n(dx_n).$$

Then in this case we will have

$$\mu_n(dx_0, \dots, dx_n) = d\nu_0(x_0)\, d\nu_1(dx_1) \dots \nu_n(dx_n)$$

as desired. ■

17.8 *Appendix: More Probability Kernel Constructions

Lemma 17.58. *Suppose that $(X, \mathcal{M})$, $(Y, \mathcal{F})$, and $(Z, \mathcal{B})$ are measurable spaces and $Q : X \times \mathcal{F} \to [0,1]$ and $R : Y \times \mathcal{B} \to [0,1]$ are probability kernels. Then for every bounded measurable function, $F : (Y \times Z, \mathcal{F} \otimes \mathcal{B}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$, the map*

$$y \to \int_Z R(y, dz)\, F(y, z)$$

is measurable. Moreover, if we define $P(x; A)$ for $A \in \mathcal{F} \otimes \mathcal{B}$ and $x \in X$ by

$$P(x, A) = \int_Y Q(x, dy) \int_Z R(y, dz)\, 1_A(y, z),$$

then $P : X \times \mathcal{F} \otimes \mathcal{B} \to [0,1]$ is a probability kernel such that

$$P(x, F) = \int_Y Q(x, dy) \int_Z R(y, dz)\, F(y, z)$$

for all bounded measurable functions, $F : (Y \times Z, \mathcal{F} \otimes \mathcal{B}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$. We will denote the kernel P by $Q \otimes R$ and write

$$(Q \otimes R)(x, dy, dz) = Q(x, dy)\, R(y, dz).$$

Moreover if $S(z, dw)$ is another probability kernel, then $((Q \otimes R) \otimes S) = (Q \otimes (R \otimes S))$.

Proof. A routine exercise in using the multiplicative systems theorem. To verify the last assertion it suffices to consider the kernels on sets of the form $A \times B \times C$ in which case,

$$\begin{aligned}
&(Q \otimes (R \otimes S))(x, A \times B \times C) \\
&\quad = \int_Y Q(x, dy) \int_{Z\times W} RS(y, dz, dw)\, 1_{A\times B\times C}(y, z, w) \\
&\quad = \int_Y Q(x, dy)\, 1_A(y) \int_{Z\times W} RS(y; B \times C) \\
&\quad = \int_Y Q(x, dy)\, 1_A(y) \int_{Z\times W} R(y, dz)\, S(z, dw)\, 1_{B\times C}(z, w) \\
&\quad = \int_Y Q(x, dy)\, 1_A(y) \int_Z R(y, dz)\, S(z, C)\, 1_B(z)
\end{aligned}$$

while

$$
\begin{aligned}
&((Q\otimes R)\otimes S)(x,A\times B\times C)\\
&\quad=\int_{Y\times Z} QR(x,dy,dz)\int_{Z\times W} S(z,dw)\,1_{A\times B\times C}(y,z,w)\\
&\quad=\int_{Y\times Z} QR(x,dy,dz)\,1_{A\times B}(y,z)\,S(z,C)\\
&\quad=\int_{Y} Q(x,dy)\int_{Z} R(y,dz)\,1_{A\times B}(y,z)\,S(z,C)\\
&\quad=\int_{Y} Q(x,dy)\,1_{A}(y)\int_{Z} R(y,dz)\,S(z,C)\,1_{B}(z).
\end{aligned}
$$

■

Corollary 17.59. *Keeping the notation in Lemma 17.58, let QR be the probability kernel given by $QR(x,dz)=\int_Y Q(x,dy)\,R(y,dz)$ so that*

$$QR(x;B)=Q\otimes R(x;Y\times B).$$

Then we have $Q(RS)=(QR)S$.

Proof. Let $C\in\mathcal{B}_W$, then

$$
\begin{aligned}
Q(RS)(x;C)&=Q\otimes(RS)(x;Y\times C)=\int_Y Q(x,dy)\,(RS)(y;C)\\
&=\int_Y Q(x,dy)\,(R\otimes S)(y;Z\times C)=[Q\otimes(R\otimes S)](Y\times Z\times C).
\end{aligned}
$$

Similarly one shows that

$$(QR)S(x;C)=[(Q\otimes R)\otimes S](Y\times Z\times C)$$

and then the result follows from Lemma 17.58. ■

(Sub and Super) Martingales

Let us start with a reminder of a few key notions that were already introduced in Chapter 17. As usual we will let $(S, \mathcal{S})$ denote a measurable space called **state space.** (Often in this chapter we will take $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.) As in Chapter 17, we will fix a **filtered probability space,** $\left(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n\in\mathbb{N}_0}, P\right)$, i.e. $\mathcal{B}_n \subset \mathcal{B}_{n+1} \subset \mathcal{B}$ for all $n = 0, 1, 2 \dots$. We further define

$$\mathcal{B}_\infty := \vee_{n=0}^\infty \mathcal{B}_n := \sigma\left(\cup_{n=0}^\infty \mathcal{B}_n\right) \subset \mathcal{B}. \tag{18.1}$$

Also recall that a sequence of random functions, $Y_n : \Omega \to S$ for $n \in \mathbb{N}_0$, are said to be **adapted** to the filtration if Y_n is $\mathcal{B}_n/\mathcal{S}$ – measurable for all n.

Definition 18.1. *Let $X := \{X_n\}_{n=0}^\infty$ is a be an adapted sequence of integrable random variables. Then;*

1. *X is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – **martingale** if $\mathbb{E}[X_{n+1}|\mathcal{B}_n] = X_n$ a.s. for all $n \in \mathbb{N}_0$.*
2. *X is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – **submartingale** if $\mathbb{E}[X_{n+1}|\mathcal{B}_n] \ge X_n$ a.s. for all $n \in \mathbb{N}_0$.*
3. *X is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – **supermartingale** if $\mathbb{E}[X_{n+1}|\mathcal{B}_n] \le X_n$ a.s. for all $n \in \mathbb{N}_0$.*

It is often fruitful to view X_n as your earnings at time n while playing some game of chance. In this interpretation, your expected earnings at time $n+1$ given the history of the game up to time n is the same, greater than, less than your earnings at time n if $X = \{X_n\}_{n=0}^\infty$ is a martingale, submartingale or supermartingale respectively. In this interpretation, martingales are fair games, submartingales are games which are favorable to the gambler (unfavorable to the casino), and supermartingales are games which are unfavorable to the gambler (favorable to the casino), see Example 18.4.

By induction one shows that X is a supermartingale, martingale, or submartingale iff

$$\mathbb{E}[X_m|\mathcal{B}_n] \overset{\le}{\underset{\ge}{=}} X_n \text{ a.s for all } m \ge n, \tag{18.2}$$

to be read from top to bottom respectively. This last equation may also be expressed as

$$\mathbb{E}[X_m|\mathcal{B}_n] \overset{\le}{\underset{\ge}{=}} X_{m\wedge n} \text{ a.s for all } m, n \in \mathbb{N}_0. \tag{18.3}$$

The reader should also note that $\mathbb{E}[X_n]$ is decreasing, constant, or increasing respectively. The next lemma shows that we may shrink the filtration, $\{\mathcal{B}_n\}_{n=0}^\infty$, within limits and still have X retain the property of being a supermartingale, martingale, or submartingale.

Lemma 18.2 (Shrinking the filtration). *Suppose that X is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively and $\{\mathcal{B}'_n\}_{n=0}^\infty$ is another filtration such that $\sigma(X_0, \dots, X_n) \subset \mathcal{B}'_n \subset \mathcal{B}_n$ for all n. Then X is a $\{\mathcal{B}'_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively.*

Proof. Since $\{X_n\}_{n=0}^\infty$ is adapted to $\{\mathcal{B}_n\}_{n=0}^\infty$ and $\sigma(X_0, \dots, X_n) \subset \mathcal{B}'_n \subset \mathcal{B}_n$, for all n,

$$\mathbb{E}_{\mathcal{B}'_n} X_{n+1} = \mathbb{E}_{\mathcal{B}'_n}\mathbb{E}_{\mathcal{B}_n} X_{n+1} \overset{\le}{\underset{\ge}{=}} \mathbb{E}_{\mathcal{B}'_n} X_n = X_n,$$

when X is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively – read from top to bottom. ■

Enlarging the filtration is another matter all together. In what follows we will simply say X is a supermartingale, martingale, submartingale if it is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale.

18.1 (Sub and Super) Martingale Examples

Example 18.3. Suppose that $\{Z_n\}_{n=0}^\infty$ are independent integrable random variables such that $\mathbb{E}Z_n = 0$ for all $n \ge 1$. Then $S_n := \sum_{k=0}^n Z_k$ is a martingale relative to the filtration, $\mathcal{B}_n^Z := \sigma(Z_0, \dots, Z_n)$. Indeed,

$$\mathbb{E}[S_{n+1} - S_n|\mathcal{B}_n] = \mathbb{E}[Z_{n+1}|\mathcal{B}_n] = \mathbb{E}Z_{n+1} = 0.$$

This same computation also shows that $\{S_n\}_{n\ge0}$ is a submartingale if $\mathbb{E}Z_n \ge 0$ and supermartingale if $\mathbb{E}Z_n \le 0$ for all n.

Exercise 18.1. Construct an example of a martingale, $\{M_n\}_{n=0}^\infty$ such that $\mathbb{E}|M_n| \to \infty$ as $n \to \infty$.

Example 18.4 (Setting the odds). Let S be a finite set (think of the outcomes of a spinner, or dice, or a roulette wheel) and $p : S \to (0,1)$ be a probability function[1]. Let $\{Z_n\}_{n=1}^\infty$ be random functions with values in S such that $p(s) := P(Z_n = s)$ for all $s \in S$. (Z_n represents the outcome of the n^{th} – game.) Also let $\alpha : S \to [0,\infty)$ be the house's payoff function, i.e. for each dollar you (the gambler) bets on $s \in S$, the house will pay $\alpha(s)$ dollars back if s is rolled. Further let $W : \Omega \to \mathbb{W}$ be measurable function into some other measure space, $(\mathbb{W}, \mathcal{F})$ which is to represent your random (or not so random) "whims.". We now assume that Z_n is independent of $(W, Z_1, \dots, Z_{n-1})$ for each n, i.e. the dice are not influenced by the previous plays or your whims. If we let $\mathcal{B}_n := \sigma(W, Z_1, \dots, Z_n)$ with $\mathcal{B}_0 = \sigma(W_0)$, then we are assuming the Z_n is independent of $\mathcal{B}_{n-1}$ for each $n \in \mathbb{N}$.

As a gambler, you are allowed to choose **before** the n^{th} – game is played, the amounts $\left(\{C_n(s)\}_{s\in S}\right)$ that you want to bet on each of the possible outcomes of the n^{th} – game. Assuming the you are not clairvoyant (i.e. can not see the future), these amounts may be random but must be $\mathcal{B}_{n-1}$ – measurable, that is $C_n(s) = C_n(W, Z_1, \dots, Z_{n-1}, s)$, i.e. $\{C_n(s)\}_{n=1}^\infty$ is "previsible" process (see Definition 18.5 below). Thus if X_0 denotes your initial wealth (assumed to be a non-random quantity) and X_n denotes your wealth just after the n^{th} – game is played, then

$$X_n - X_{n-1} = -\sum_{s\in S} C_n(s) + C_n(Z_n)\,\alpha(Z_n)$$

where $-\sum_{s\in S} C_n(s)$ is your total bet on the n^{th} – game and $C_n(Z_n)\,\alpha(Z_n)$ represents the house's payoff to you for the n^{th} – game. Therefore it follows that

$$X_n = X_0 + \sum_{k=1}^n \left[-\sum_{s\in S} C_k(s) + C_n(Z_k)\,\alpha(Z_k)\right],$$

X_n is $\mathcal{B}_n$ – measurable for each n, and

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_{n-1}}[X_n - X_{n-1}] &= -\sum_{s\in S} C_n(s) + \mathbb{E}_{\mathcal{B}_{n-1}}[C_n(Z_n)\,\alpha(Z_n)] \\ &= -\sum_{s\in S} C_n(s) + \sum_{s\in S} C_n(s)\,\alpha(s)\,p(s) \\ &= \sum_{s\in S} C_n(s)\,(\alpha(s)\,p(s) - 1).\end{aligned}$$

Thus it follows, that no matter the choice of the betting "strategy," $\{C_n(s) : s \in S\}_{n=1}^\infty$, we will have

$$\mathbb{E}_{\mathcal{B}_{n-1}}[X_n - X_{n-1}] = \begin{cases} \geq 0 \text{ if } \alpha(\cdot)\,p(\cdot) \geq 1 \\ = 0 \text{ if } \alpha(\cdot)\,p(\cdot) = 1 \\ \leq 0 \text{ if } \alpha(\cdot)\,p(\cdot) \leq 1 \end{cases},$$

that is $\{X_n\}_{n\geq 0}$ is a sub-martingale, martingale, or supermartingale depending on whether $\alpha \cdot p \geq 1$, $\alpha \cdot p = 1$, or $\alpha \cdot p \leq 1$.

Moral: If the Casino wants to be guaranteed to make money on average, it had better choose $\alpha : S \to [0,\infty)$ such that $\alpha(s) < 1/p(s)$ for all $s \in S$. In this case the expected earnings of the gambler will be decreasing which means the expected earnings of the Casino will be increasing.

Definition 18.5. *We say* $\{C_n : \Omega \to S\}_{n=1}^\infty$ *is* ***predictable or previsible*** *if each* C_n *is* $\mathcal{B}_{n-1}/\mathcal{S}$ *– measurable for all* $n \in \mathbb{N}$.

A typical example is when $\{X_n : \Omega \to S\}_{n=0}^\infty$ is a sequence of measurable functions on a probability space $(\Omega, \mathcal{B}, P)$ and $\mathcal{B}_n := \sigma(X_0, \dots, X_n)$. An application of Lemma 14.1 shows that a sequence of random variables, $\{Y_n\}_{n=0}^\infty$, is adapted to the filtration iff there are $\mathcal{S}^{\otimes(n+1)}/\mathcal{B}_\mathbb{R}$ – measurable functions, $f_n : S^{n+1} \to \mathbb{R}$, such that $Y_n = f_n(X_0, \dots, X_n)$ for all $n \in \mathbb{N}_0$ and a sequence of random variables, $\{Z_n\}_{n=1}^\infty$, is predictable iff there exists, there are measurable functions, $f_n : \mathbb{R}^n \to \mathbb{R}$ such that $Z_n = f_n(X_0, \dots, X_{n-1})$ for all $n \in \mathbb{N}$.

Example 18.6. Suppose that $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and $X \in L^1(\Omega, \mathcal{B}, P)$. Then $X_n := \mathbb{E}[X|\mathcal{B}_n]$ is a martingale. Indeed, by the tower property of conditional expectations,

$$\mathbb{E}[X_{n+1}|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_{n+1}]\,|\mathcal{B}_n] = \mathbb{E}[X|\mathcal{B}_n] = X_n \text{ a.s.}$$

Example 18.7. Suppose that $\Omega = [0,1]$, $\mathcal{B} = \mathcal{B}_{[0,1]}$, and $P = m$ – Lebesgue measure. Let $\mathcal{P}_n = \left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right\}_{k=1}^{2^n-1} \cup \left\{\left[0, \frac{1}{2^n}\right]\right\}$ and $\mathcal{B}_n := \sigma(\mathcal{P}_n)$ for each $n \in \mathbb{N}$. Then $M_n := 2^n 1_{(0,2^{-n}]}$ for $n \in \mathbb{N}$ is a martingale (Exercise 18.2) such that $\mathbb{E}|M_n| = 1$ for all n. However, there is no $X \in L^1(\Omega, \mathcal{B}, P)$ such that $M_n = \mathbb{E}[X|\mathcal{B}_n]$. To verify this last assertion, suppose such an X existed. Let . We would then have for $2^n > k > 0$ and any $m > n$, that

$$\begin{aligned}\mathbb{E}\left[X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_m} X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] \\ &= \mathbb{E}\left[M_m : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] = 0.\end{aligned}$$

Using $\mathbb{E}[X : A] = 0$ for all A in the π – system, $\mathcal{Q} := \cup_{n=1}^\infty \left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] : 0 < k < 2^n\right\}$, an application of the π – λ theorem shows $\mathbb{E}[X : A] = 0$ for all $A \in \sigma(\mathcal{Q}) = \mathcal{B}$. Therefore $X = 0$ a.s. by Proposition 7.22. But this is impossible since $1 = \mathbb{E}M_n = \mathbb{E}X$.

[1] To be concrete, take $S = \{2, \dots, 12\}$ representing the possible values for the sums of the upward pointing faces of two dice. Assuming the dice are independent and fair then determines $p : S \to (0,1)$. For example $p(2) = p(12) = 1/36$, $p(3) = p(11) = 1/18$, $p(7) = 1/6$, etc.

Moral: not all L^1 – bounded martingales are of the form in example 18.6. Proposition 18.8 shows what is missing from this martingale in order for it to be of the form in Example 18.6. See the comments after Example 18.11 for another example of an L^1 – bounded martingale which is not of the form in example 18.6.

Exercise 18.2. Show that $M_n := 2^n 1_{(0,2^{-n}]}$ for $n \in \mathbb{N}$ as defined in Example 18.7 is a martingale.

Proposition 18.8. *Suppose $1 \leq p < \infty$ and $X \in L^p(\Omega, \mathcal{B}, P)$. Then the collection of random variables, $\Gamma := \{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \subset \mathcal{B}\}$ is a bounded subset of $L^p(\Omega, \mathcal{B}, P)$ which is also uniformly integrable.*

Proof. Since $\mathbb{E}_{\mathcal{G}}$ is a contraction on all L^p – spaces it follows that Γ is bounded in L^p with

$$\sup_{\mathcal{G}\subset\mathcal{B}} \|\mathbb{E}[X|\mathcal{G}]\|_p \leq \|X\|_p.$$

For the $p > 1$ the uniform integrability of Γ follows directly from Lemma 12.48.

We now concentrate on the $p = 1$ case. Recall that $|\mathbb{E}_{\mathcal{G}} X| \leq \mathbb{E}_{\mathcal{G}} |X|$ a.s. and therefore,

$$\mathbb{E}[|\mathbb{E}_{\mathcal{G}} X| : |\mathbb{E}_{\mathcal{G}} X| \geq a] \leq \mathbb{E}[|X| : |\mathbb{E}_{\mathcal{G}} X| \geq a] \text{ for all } a > 0.$$

But by Chebyshev's inequality,

$$P(|\mathbb{E}_{\mathcal{G}} X| \geq a) \leq \frac{1}{a}\mathbb{E}|\mathbb{E}_{\mathcal{G}} X| \leq \frac{1}{a}\mathbb{E}|X|.$$

Since $\{|X|\}$ is uniformly integrable, it follows from Proposition 12.42 that, by choosing a sufficiently large, $\mathbb{E}[|X| : |\mathbb{E}_{\mathcal{G}} X| \geq a]$ is as small as we please uniformly in $\mathcal{G} \subset \mathcal{B}$ and therefore,

$$\lim_{a\to\infty} \sup_{\mathcal{G}\subset\mathcal{B}} \mathbb{E}[|\mathbb{E}_{\mathcal{G}} X| : |\mathbb{E}_{\mathcal{G}} X| \geq a] = 0.$$

■

Example 18.9. This example generalizes Example 18.7. Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and Q is another probability measure on $(\Omega, \mathcal{B})$. Let us assume that $Q|_{\mathcal{B}_n} \ll P|_{\mathcal{B}_n}$ for all n, which by the Raydon-Nikodym Theorem 15.8, implies there exists $0 \leq X_n \in L^1(\Omega, \mathcal{B}_n, P)$ with $\mathbb{E}X_n = 1$ such that $dQ|_{\mathcal{B}_n} = X_n dP|_{\mathcal{B}_n}$, or equivalently put, for any $B \in \mathcal{B}_n$ we have

$$Q(B) = \int_B X_n dP = \mathbb{E}[X_n : B].$$

Since $B \in \mathcal{B}_n \subset \mathcal{B}_{n+1}$, we also have $\mathbb{E}[X_{n+1} : B] = Q(B) = \mathbb{E}[X_n : B]$ for all $B \in \mathcal{B}_n$ and hence $\mathbb{E}[X_{n+1}|\mathcal{B}_n] = X_n$ a.s., i.e. $X = \{X_n\}_{n=0}^\infty$ is a positive martingale.

Example 18.7 is of this form with $Q = \delta_0$. Notice that $\delta_0|_{\mathcal{B}_n} \ll m|_{\mathcal{B}_n}$ for all $n < \infty$ while $\delta_0 \perp m$ on $\mathcal{B}_{[0,1]} = \mathcal{B}_\infty$. See Section 19.3 for more in the direction of this example.

Lemma 18.10. *Let $X := \{X_n\}_{n=0}^\infty$ be an adapted process of integrable random variables on a filtered probability space, $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ and let $d_n := X_n - X_{n-1}$ with $X_{-1} := \mathbb{E}X_0$. Then X is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{B}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{B}_n] \geq 0$ or $\mathbb{E}[d_{n+1}|\mathcal{B}_n] \leq 0$ respectively) for all $n \in \mathbb{N}_0$.*

Conversely if $\{d_n\}_{n=1}^\infty$ is an adapted sequence of integrable random variables and X_0 is a $\mathcal{B}_0$ -measurable integrable random variable. Then $X_n = X_0 + \sum_{j=1}^n d_j$ is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{B}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{B}_n] \geq 0$ or $\mathbb{E}[d_{n+1}|\mathcal{B}_n] \leq 0$ respectively) for all $n \in \mathbb{N}$.

Proof. We prove the assertions for martingales only, the other all being similar. Clearly X is a martingale iff

$$0 = \mathbb{E}[X_{n+1}|\mathcal{B}_n] - X_n = \mathbb{E}[X_{n+1} - X_n|\mathcal{B}_n] = \mathbb{E}[d_{n+1}|\mathcal{B}_n].$$

The second assertion is an easy consequence of the first assertion. ■

Example 18.11. Suppose that $\{Z_n\}_{n=0}^\infty$ is a sequence of independent integrable random variables, $X_n = Z_0 \dots Z_n$, and $\mathcal{B}_n := \sigma(Z_0, \cdots, Z_n)$. (Observe that $\mathbb{E}|X_n| = \prod_{k=0}^n \mathbb{E}|Z_k| < \infty$.) Since

$$\mathbb{E}[X_{n+1}|\mathcal{B}_n] = \mathbb{E}[X_n Z_{n+1}|\mathcal{B}_n] = X_n \mathbb{E}[Z_{n+1}|\mathcal{B}_n] = X_n \cdot \mathbb{E}[Z_{n+1}] \text{ a.s.},$$

it follows that $\{X_n\}_{n=0}^\infty$ is a martingale if $\mathbb{E}Z_n = 1$. If we further assume, for all n, that $Z_n \geq 0$ so that $X_n \geq 0$, then $\{X_n\}_{n=0}^\infty$ is a supermartingale (submartingale) provided $\mathbb{E}Z_n \leq 1$ ($\mathbb{E}Z_n \geq 1$) for all n.

Let us specialize the above example even more by taking $Z_n \overset{d}{=} p + U$ where $p \geq 0$ and U is the uniform distribution on $[0, 1]$. In this case we have by the strong law of large numbers that

$$\frac{1}{n}\ln X_n = \frac{1}{n}\sum_{k=0}^n \ln Z_k \to \mathbb{E}[\ln(p + U)] \text{ a.s.}$$

An elementary computation shows

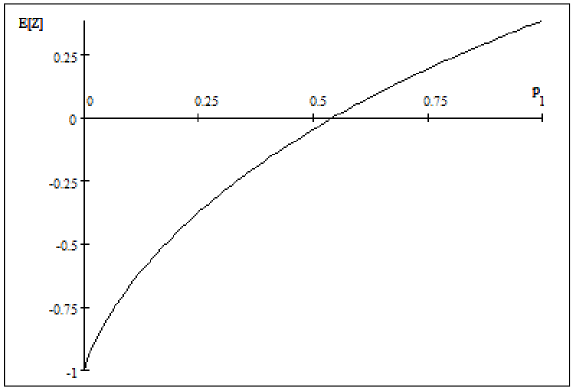


Fig. 18.1. The graph of $\mathbb{E}[\ln(p+U)]$ as a function of p. This function has a zero at $p = p_c \cong 0.542\,21$.

$$\mathbb{E}[\ln(p+U)] = \int_0^1 \ln(p+x)\,dx = \int_p^{p+1} \ln(p+x)\,dx$$
$$= (x\ln x - x)_{x=p}^{x=p+1} = (p+1)\ln(p+1) - p\ln p - 1$$

Hence we may conclude that

$$X_n \to \lim_{n\to\infty} \exp(n\mathbb{E}[\ln(p+U)]) = \begin{cases} 0 & \text{if } p < p_c \\ ? & \text{if } p = p_c \\ \infty & \text{if } p > p_c \end{cases} \text{ a.s.}$$

Notice that $\mathbb{E}Z_n = p + 1/2$ and therefore X_n is a martingale precisely when $p = 1/2$ and is a sub-martingale for $p > 1/2$. So for $1/2 < p < p_c$, $\{X_n\}_{n=1}^\infty$ is a positive sub-martingale, $\mathbb{E}X_n = (p+1/2)^{n+1} \to \infty$ yet $\lim_{n\to\infty} X_n = 0$ a.s. Have a look at the excel file (Product_positive-(sub)martingales.xls) in order to construct sample paths for the $\{X_n\}_{n=0}^\infty$.

Proposition 18.12. *Suppose that $X = \{X_n\}_{n=0}^\infty$ is a martingale and φ is a convex function such that $\varphi(X_n) \in L^1$ for all n. Then $\varphi(X) = \{\varphi(X_n)\}_{n=0}^\infty$ is a submartingale. If φ is also assumed to be increasing, it suffices to assume that X is a submartingale in order to conclude that $\varphi(X)$ is a submartingale. (For example if X is a positive submartingale, $p \in (1,\infty)$, and $\mathbb{E}X_n^p < \infty$ for all n, then $X^p := \{X_n^p\}_{n=0}^\infty$ is another positive submartingale.*

Proof. When X is a martingale, by the conditional Jensen's inequality 14.25,

$$\varphi(X_n) = \varphi(\mathbb{E}_{\mathcal{B}_n} X_{n+1}) \le \mathbb{E}_{\mathcal{B}_n}[\varphi(X_{n+1})]$$

which shows $\varphi(X)$ is a submartingale. Similarly, if X is a submartingale and φ is convex and increasing, then φ preserves the inequality, $X_n \le \mathbb{E}_{\mathcal{B}_n} X_{n+1}$, and hence

$$\varphi(X_n) \le \varphi(\mathbb{E}_{\mathcal{B}_n} X_{n+1}) \le \mathbb{E}_{\mathcal{B}_n}[\varphi(X_{n+1})]$$

so again $\varphi(X)$ is a submartingale. ■

Proposition 18.13 (Markov Chains and Martingales). *Suppose that $\left(\Omega, \mathcal{B}, \{\mathcal{B}\}_{n\in\mathbb{N}_0}, \{X_n : \Omega \to S\}_{n\ge 0}, Q, P\right)$ is a time homogeneous Markov chain and $f : \mathbb{N}_0 \times S \to \mathbb{R}$ be measurable function which is either non-negative or satisfies $\mathbb{E}[|f(n, X_n)|] < \infty$ for all n and let $Z_n := f(n, X_n)$. Then $\{Z_n\}_{n=0}^\infty$ is a (sub-martingale) martingale if $(Qf(n+1,\cdot) \le f(n\cdot))$ $Qf(n+1,\cdot) = f(n,\cdot)$ for all $n \ge 0$. In particular if $f : S \to \mathbb{R}$ is a function such that $(Qf \le f)$ $Qf = f$ then $Z_n = f(X_n)$ is a (sub-martingale) martingale. (Also see Exercise 18.5 below.)*

Proof. Using the Markov property and the definition of Q, we have

$$\mathbb{E}[Z_{n+1}|\mathcal{B}_n] = \mathbb{E}[f(n+1, X_{n+1})|\mathcal{B}_n] = [Qf(n+1,\cdot)](X_n).$$

The latter expression is (less than or equal) equal to Z_n if $(Qf(n+1,\cdot) \le f(n\cdot))$ $Qf(n+1,\cdot) = f(n,\cdot)$ for all $n \ge 0$. ■

One way to find solutions to the equation $Qf(n+1,\cdot) = f(n,\cdot)$ at least for a finite number of n is to let $g : S \to \mathbb{R}$ be an arbitrary function and $T \in \mathbb{N}$ be given and then define

$$f(n,y) := \left(Q^{T-n} g\right)(y) \text{ for } 0 \le n \le T.$$

Then $Qf(n+1,\cdot) = Q\left(Q^{T-n-1}g\right) = Q^{T-n}g = f(n,\cdot)$ and we will have that

$$Z_n = f(n, X_n) = \left(Q^{T-n}g\right)(X_n)$$

is a Martingale for $0 \le n \le T$. If $f(n,\cdot)$ satisfies $Qf(n+1,\cdot) = f(n,\cdot)$ for all n then we must have, with $f_0 := f(0,\cdot)$,

$$f(n,\cdot) = Q^{-n} f_0$$

where $Q^{-1}g$ denotes a function h solving $Qh = g$. In general Q is not invertible and hence there may be no solution to $Qh = g$ or there might be many solutions.

In special cases one can often make sense of these expressions as you will see Exercise 18.5. In this exercise we will continue the notation in Exercise 17.35 where $S = \mathbb{Z}$, $S_n = X_0 + X_1 + \cdots + X_n$, where $\{X_i\}_{i=1}^\infty$ are i.i.d. with $P(X_i = 1) = p \in (0,1)$ and $P(X_i = -1) = q := 1 - p$, and X_0 is S –valued random variable independent of $\{X_i\}_{i=1}^\infty$. Recall that $\{S_n\}_{n=0}^\infty$ is a time homogeneous Markov chain with transition kernel determined by

$Qf(x) = pf(x+1) + qf(x-1)$. As we have seen if $f(x) = a + b(q/p)^x$, then $Qf = f$ and therefore

$$M_n = a + b(q/p)^{S_n}$$

is a Martingale for all $a, b \in \mathbb{R}$.

Now suppose that $\lambda \neq 0$ and observe that $Q\lambda^x = \left(p\lambda + q\lambda^{-1}\right)\lambda^x$. Thus it follows that we may set $Q^{-1}\lambda^x = \left(p\lambda + q\lambda^{-1}\right)^{-1}\lambda^x$ and therefore conclude that

$$f(n,x) := Q^{-n}\lambda^x = \left(p\lambda + q\lambda^{-1}\right)^{-n}\lambda^x$$

satisfies $Qf(n+1,\cdot) = f(n,\cdot)$. So if we suppose that X_0 is a bounded so that S_n is bounded for all n, we will have $\left\{M_n = \left(p\lambda + q\lambda^{-1}\right)^{-n}\lambda^{S_n}\right\}_{n\geq 0}$ is a martingale for all $\lambda \neq 0$.

Exercise 18.3. For $\theta \in \mathbb{R}$ let

$$f_\theta(n,x) := Q^{-n}e^{\theta x} = \left(pe^\theta + qe^{-\theta}\right)^{-n}e^{\theta x}$$

so that $Qf_\theta(n+1,\cdot) = f_\theta(n,\cdot)$ for all $\theta \in \mathbb{R}$. Compute;

1. $f_\theta^{(k)}(n,x) := \left(\frac{d}{d\theta}\right)^k f_\theta(n,x)$ for $k = 1, 2$.
2. Use your results to show,

$$M_n^{(1)} := S_n - n(p-q)$$

and

$$M_n^{(2)} := (S_n - n(p-q))^2 - 4npq$$

are martingales.

(If you are ambitious you might also find $M_n^{(3)}$.)

Remark 18.14. If $\{M_n(\theta)\}_{n=0}^\infty$ is a martingale depending differentiability on a parameter $\theta \in \mathbb{R}$. Then for all $A \in \mathcal{B}_n$,

$$\mathbb{E}\left[\frac{d}{d\theta}M_{n+1}(\theta) : A\right] = \frac{d}{d\theta}\mathbb{E}\left[M_{n+1}(\theta) : A\right] = \frac{d}{d\theta}\mathbb{E}\left[M_n(\theta) : A\right] = \mathbb{E}\left[\frac{d}{d\theta}M_n(\theta) : A\right]$$

provided it is permissible to interchange $\frac{d}{d\theta}$ with the expectations in this equation. Thus under "suitable" hypothesis, we will have $\left\{\frac{d}{d\theta}M_n(\theta)\right\}_{n\geq 0}$ is another martingale.

18.2 Decompositions

Notation 18.15 *Given a sequence $\{Z_k\}_{k=0}^\infty$, let $\Delta_k Z := Z_k - Z_{k-1}$ for $k = 1, 2, \ldots$.*

Lemma 18.16 (Doob Decomposition). *Each adapted sequence, $\{Z_n\}_{n=0}^\infty$, of integrable random variables has a unique decomposition,*

$$Z_n = M_n + A_n \tag{18.4}$$

where $\{M_n\}_{n=0}^\infty$ is a martingale and A_n is a predictable process such that $A_0 = 0$. Moreover this decomposition is given by $A_0 = 0$,

$$A_n := \sum_{k=1}^n \mathbb{E}_{\mathcal{B}_{k-1}}\left[\Delta_k Z\right] \text{ for } n \geq 1 \tag{18.5}$$

and

$$M_n = Z_n - A_n = Z_n - \sum_{k=1}^n \mathbb{E}_{\mathcal{B}_{k-1}}\left[\Delta_k Z\right] \tag{18.6}$$

$$= Z_0 + \sum_{k=1}^n \left(Z_k - \mathbb{E}_{\mathcal{B}_{k-1}}Z_k\right). \tag{18.7}$$

In particular, $\{Z_n\}_{n=0}^\infty$ is a submartingale (supermartingale) iff A_n is increasing (decreasing) almost surely.

Proof. Assuming Z_n has a decomposition as in Eq. (18.4), then

$$\mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}Z\right] = \mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}M + \Delta_{n+1}A\right] = \Delta_{n+1}A \tag{18.8}$$

wherein we have used M is a martingale and A is predictable so that $\mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}M\right] = 0$ and $\mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}A\right] = \Delta_{n+1}A$. Hence we must define, for $m \geq 1$,

$$A_n := \sum_{k=1}^n \Delta_k A = \sum_{k=1}^n \mathbb{E}_{\mathcal{B}_{k-1}}\left[\Delta_k Z\right]$$

which is a predictable process. This proves the uniqueness of the decomposition and the validity of Eq. (18.5).

For existence, from Eq. (18.5) it follows that

$$\mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}Z\right] = \Delta_{n+1}A = \mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}A\right].$$

Hence, if we define $M_n := Z_n - A_n$, then

$$\mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}M\right] = \mathbb{E}_{\mathcal{B}_n}\left[\Delta_{n+1}Z - \Delta_{n+1}A\right] = 0$$

and hence $\{M_n\}_{n=0}^\infty$ is a martingale. Moreover, Eq. (18.7) follows from Eq. (18.6) since,

$$M_n = Z_0 + \sum_{k=1}^n \left(\Delta_k Z - \mathbb{E}_{\mathcal{B}_{k-1}}[\Delta_k Z]\right)$$

and

$$\begin{aligned}\Delta_k Z - \mathbb{E}_{\mathcal{B}_{k-1}}[\Delta_k Z] &= Z_k - Z_{k-1} - \mathbb{E}_{\mathcal{B}_{k-1}}[Z_k - Z_{k-1}] \\ &= Z_k - Z_{k-1} - \left(\mathbb{E}_{\mathcal{B}_{k-1}} Z_k - Z_{k-1}\right) = Z_k - \mathbb{E}_{\mathcal{B}_{k-1}} Z_k.\end{aligned}$$

■

Remark 18.17. Suppose that $X = \{X_n\}_{n=0}^\infty$ is a submartingale and $X_n = M_n + A_n$ is it Doob decomposition. Then $A_\infty =\uparrow \lim_{n\to\infty} A_n$ exists a.s.,

$$\mathbb{E}A_n = \mathbb{E}[X_n - M_n] = \mathbb{E}X_n - \mathbb{E}M_0 = \mathbb{E}[X_n - X_0] \tag{18.9}$$

and hence by MCT,

$$\mathbb{E}A_\infty =\uparrow \lim_{n\to\infty} \mathbb{E}[X_n - X_0]. \tag{18.10}$$

Hence if $\lim_{n\to\infty} \mathbb{E}[X_n - X_0] = \sup_n \mathbb{E}[X_n - X_0] < \infty$, then $\mathbb{E}A_\infty < \infty$ and so by DCT, $A_n \to A_\infty$ in $L^1(\Omega, \mathcal{B}, P)$. In particular if $\sup_n \mathbb{E}|X_n| < \infty$, we may conclude that $\{X_n\}_{n=0}^\infty$ is $L^1(\Omega, \mathcal{B}, P)$ convergent iff $\{M_n\}_{n=0}^\infty$ is $L^1(\Omega, \mathcal{B}, P)$ convergent. (We will see below in Corollary 18.54 that $X_\infty := \lim_{n\to\infty} X_n$ and $M_\infty := \lim_{n\to\infty} M_n$ exist almost surely under the assumption that $\sup_n \mathbb{E}|X_n| < \infty$.)

Example 18.18. Suppose that $N = \{N_n\}_{n=0}^\infty$ is a square integrable martingale, i.e. $\mathbb{E}N_n^2 < \infty$ for all n. Then from Proposition 18.12, $X := \left\{X_n = N_n^2\right\}_{n=0}^\infty$ is a positive submartingale. In this case

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_{k-1}} \Delta_k X &= \mathbb{E}_{\mathcal{B}_{k-1}}\left(N_k^2 - N_{k-1}^2\right) = \mathbb{E}_{\mathcal{B}_{k-1}}[(N_k - N_{k-1})(N_k + N_{k-1})] \\ &= \mathbb{E}_{\mathcal{B}_{k-1}}[(N_k - N_{k-1})(N_k - N_{k-1})] \\ &= \mathbb{E}_{\mathcal{B}_{k-1}}(N_k - N_{k-1})^2\end{aligned}$$

wherein the second to last equality we have used

$$\mathbb{E}_{\mathcal{B}_{k-1}}[(N_k - N_{k-1}) N_{k-1}] = N_{k-1}\mathbb{E}_{\mathcal{B}_{k-1}}(N_k - N_{k-1}) = 0 \text{ a.s.}$$

in order to change $(N_k + N_{k-1})$ to $(N_k - N_{k-1})$. Hence the increasing predictable process, A_n, in the Doob decomposition may be written as

$$A_n = \sum_{k\le n} \mathbb{E}_{\mathcal{B}_{k-1}} \Delta_k X = \sum_{k\le n} \mathbb{E}_{\mathcal{B}_{k-1}} (\Delta_k N)^2. \tag{18.11}$$

Exercise 18.4 (Very similar to above example?). Suppose $\{M_n\}_{n=0}^\infty$ is a square integrable martingale. Show;

1. $\mathbb{E}\left[M_{n+1}^2 - M_n^2 | \mathcal{B}_n\right] = \mathbb{E}\left[(M_{n+1} - M_n)^2 | \mathcal{B}_n\right]$. Conclude from this that the Doob decomposition of M_n^2 is of the form,
$$M_n^2 = N_n + A_n$$
where
$$A_n := \sum_{1\le k\le n} \mathbb{E}\left[(M_k - M_{k-1})^2 | \mathcal{B}_{k-1}\right].$$
2. If we further assume that $M_k - M_{k-1}$ is independent of $\mathcal{B}_{k-1}$ for all $k = 1, 2, \ldots$, explain why,
$$A_n = \sum_{1\le k\le n} \mathbb{E}(M_k - M_{k-1})^2.$$

The next exercise shows how to characterize Markov processes via martingales.

Exercise 18.5 (Martingale problem I). Suppose that $\{X_n\}_{n=0}^\infty$ is an $(S, \mathcal{S})$ – valued adapted process on some filtered probability space $\left(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n\in\mathbb{N}_0}, P\right)$ and Q is a probability kernel on S. To each $f : S \to \mathbb{R}$ which is bounded and measurable, let

$$M_n^f := f(X_n) - \sum_{k<n} (Qf(X_k) - f(X_k)) = f(X_n) - \sum_{k<n} ((Q - I) f)(X_k).$$

Show;

1. If $\{X_n\}_{n\ge0}$ is a time homogeneous Markov chain with transition kernel, Q, then $\left\{M_n^f\right\}_{n\ge0}$ is a martingale for each $f \in \mathcal{S}_b$.
2. Conversely if $\left\{M_n^f\right\}_{n\ge0}$ is a martingale for each $f \in \mathcal{S}_b$, then $\{X_n\}_{n\ge0}$ is a time homogeneous Markov chain with transition kernel, Q.

Remark 18.19. If X is a real valued random variable, then $X = X^+ - X^-$, $|X| = X^+ + X^-$, $X^+ \le |X| = 2X^+ - X$, so that

$$\mathbb{E}X^+ \le \mathbb{E}|X| = 2\mathbb{E}X^+ - \mathbb{E}X.$$

Hence if $\{X_n\}_{n=0}^\infty$ is a submartingale then

$$\mathbb{E}X_n^+ \le \mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \le 2\mathbb{E}X_n^+ - \mathbb{E}X_0$$

from which it follows that

$$\sup_n \mathbb{E}X_n^+ \le \sup_n \mathbb{E}|X_n| \le 2\sup_n \mathbb{E}X_n^+ - \mathbb{E}X_0. \tag{18.12}$$

In particular, an integrable submartingale $\{X_n\}_{n=0}^\infty$ is $L^1(P)$ bounded iff $\{X_n^+\}_{n=0}^\infty$ is $L^1(P)$ bounded.

Theorem 18.20 (Krickeberg Decomposition). *Suppose that X is an integrable submartingale such that $C := \sup_n \mathbb{E}[X_n^+] < \infty$ or equivalently $\sup_n \mathbb{E}|X_n| < \infty$, see Eq. (18.12). Then*

$$M_n :=\uparrow \lim_{p\to\infty} \mathbb{E}\left[X_p^+|\mathcal{B}_n\right] \text{ exists a.s.,}$$

$M = \{M_n\}_{n=0}^\infty$ is a positive martingale, $Y = \{Y_n\}_{n=0}^\infty$ with $Y_n := X_n - M_n$ is a positive supermartingale, and hence $X_n = M_n - Y_n$. So X can be decomposed into the difference of a positive martingale and a positive supermartingale.

Proof. From Proposition 18.12 we know that $X^+ = \{X_n^+\}$ is a still a positive submartingale. Therefore for each $n \in \mathbb{N}$, and $p \ge n$,

$$\mathbb{E}_{\mathcal{B}_n}\left[X_{p+1}^+\right] = \mathbb{E}_{\mathcal{B}_n}\mathbb{E}_{\mathcal{B}_p}\left[X_{p+1}^+\right] \ge \mathbb{E}_{\mathcal{B}_n}X_p^+ \text{ a.s.}$$

Therefore $\mathbb{E}_{\mathcal{B}_n}X_p^+$ is increasing in p for $p \ge n$ and therefore, $M_n := \lim_{p\to\infty}\mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right]$ exists in $[0,\infty]$. By Fatou's lemma, we know that

$$\mathbb{E}M_n \le \liminf_{p\to\infty}\mathbb{E}\left[\mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right]\right] \le \liminf_{p\to\infty}\mathbb{E}\left[X_p^+\right] = C < \infty$$

which shows M is integrable. By cMCT and the tower property of conditional expectation,

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n}M_{n+1} &= \mathbb{E}_{\mathcal{B}_n}\lim_{p\to\infty}\mathbb{E}_{\mathcal{B}_{n+1}}\left[X_p^+\right] = \lim_{p\to\infty}\mathbb{E}_{\mathcal{B}_n}\mathbb{E}_{\mathcal{B}_{n+1}}\left[X_p^+\right] \\ &= \lim_{p\to\infty}\mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] = M_n \text{ a.s.,}\end{aligned}$$

which shows $M = \{M_n\}$ is a martingale.

We now define $Y_n := M_n - X_n$. Using the submartingale property of X^+ implies,

$$\begin{aligned}Y_n = M_n - X_n &= \lim_{p\to\infty}\mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] - X_n = \lim_{p\to\infty}\mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] - X_n^+ + X_n^- \\ &= \lim_{p\to\infty}\mathbb{E}_{\mathcal{B}_n}\left[X_p^+ - X_n^+\right] + X_n^- \ge 0 \text{ a.s..}\end{aligned}$$

Moreover,

$$\mathbb{E}\left[Y_{n+1}|\mathcal{B}_n\right] = \mathbb{E}\left[M_{n+1} - X_{n+1}|\mathcal{B}_n\right] = M_n - \mathbb{E}\left[X_{n+1}|\mathcal{B}_n\right] \ge M_n - X_n = Y_n$$

wherein we have use M is a martingale in the second equality and X is submartingale the last inequality. ∎

18.3 Stopping Times

Definition 18.21. *Again let $\{\mathcal{B}_n\}_{n=0}^\infty$ be a filtration on $(\Omega,\mathcal{B})$ and assume that $\mathcal{B} = \mathcal{B}_\infty := \vee_{n=0}^\infty \mathcal{B}_n := \sigma\left(\cup_{n=0}^\infty \mathcal{B}_n\right)$. A function, $\tau : \Omega \to \bar{\mathbb{N}} := \mathbb{N}\cup\{0,\infty\}$ is said to be a stopping time if $\{\tau \le n\} \in \mathcal{B}_n$ for all $n \in \bar{\mathbb{N}}$. Equivalently put, $\tau : \Omega \to \bar{\mathbb{N}}$ is a stopping time iff the process, $n \to 1_{\tau\le n}$ is adapted.*

Lemma 18.22. *Let $\{\mathcal{B}_n\}_{n=0}^\infty$ be a filtration on $(\Omega,\mathcal{B})$ and $\tau : \Omega \to \bar{\mathbb{N}}$ be a function. Then the following are equivalent;*

1. *τ is a stopping time.*
2. *$\{\tau \le n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.*
3. *$\{\tau > n\} = \{\tau \ge n+1\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.*
4. *$\{\tau = n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.*

Moreover if any of these conditions hold for $n \in \mathbb{N}_0$ then they also hold for $n = \infty$.

Proof. (1. $\iff$ 2.) Observe that if $\{\tau \le n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$, then $\{\tau < \infty\} = \cup_{n=1}^\infty \{\tau \le n\} \in \mathcal{B}_\infty$ and therefore $\{\tau = \infty\} = \{\tau < \infty\}^c \in \mathcal{B}_\infty$ and hence $\{\tau \le \infty\} = \{\tau < \infty\}\cup\{\tau = \infty\} \in \mathcal{B}_\infty$. Hence in order to check that τ is a stopping time, it suffices to show $\{\tau \le n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.

The equivalence of 2., 3., and 4. follows from the identities

$$\begin{aligned}\{\tau > n\}^c &= \{\tau \le n\}, \\ \{\tau = n\} &= \{\tau \le n\} \setminus \{\tau \le n-1\}, \text{ and} \\ \{\tau \le n\} &= \cup_{k=0}^n \{\tau = k\}\end{aligned}$$

from which we conclude that 2. $\implies$ 3. $\implies$ 4. $\implies$ 1. ∎

Clearly any constant function, $\tau : \Omega \to \bar{\mathbb{N}}$, is a stopping time. The reader should also observe that if $\mathcal{B}_n = \sigma(X_0,\dots,X_n)$, then $\tau : \Omega \to \bar{\mathbb{N}}$ is a stopping time iff, for each $n \in \mathbb{N}_0$ there exists a measurable function, $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ such that $1_{\{\tau=n\}} = f_n(X_0,\dots,X_n)$. In other words, if $\tau(\omega) = n$ and ω' is any other point in Ω such that $X_k(\omega) = X_k(\omega')$ for $k \le n$ then $\tau(\omega') = n$. Here is another common example of a stopping time.

Example 18.23 (Hitting times). Let $(S,\mathcal{S})$ be a state space, $X := \{X_n : \Omega \to S\}_{n=0}^\infty$ be an adapted process on the filtered space, $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty)$ and $A \in \mathcal{S}$. Then the **first hitting time of** A,

$$\tau := \inf\{n \in \mathbb{N}_0 : X_n \in A\},$$

(with convention that $\inf\emptyset = \infty$) is a stopping time. To see this, observe that

$$\{\tau = n\} = \{X_0 \in A^c,\dots,X_{n-1} \in A^c, X_n \in A\} \in \sigma(X_0,\dots,X_n) \subset \mathcal{B}_n.$$

More generally if σ is a stopping time, then the **first hitting time after** σ,

$$\tau := \inf\{k \geq \sigma : X_k \in A\},$$

is also a stopping time. Indeed,

$$\begin{aligned}\{\tau = n\} &= \{\sigma \leq n\} \cap \{X_\sigma \notin A, \dots, X_{n-1} \notin A, X_n \in A\} \\ &= \cup_{0 \leq k \leq n} \{\sigma = k\} \cap \{X_k \notin A, \dots, X_{n-1} \notin A, X_n \in A\}\end{aligned}$$

which is in $\mathcal{B}_n$ for all n. Here we use the convention that

$$\{X_k \notin A, \dots, X_{n-1} \notin A, X_n \in A\} = \{X_n \in A\} \text{ if } k = n.$$

On the other hand the last hitting time, $\tau = \sup\{n \in \mathbb{N}_0 : X_n \in A\}$, of a set A is typically not a stopping time. Indeed, in this case

$$\{\tau = n\} = \{X_n \in A, X_{n+1} \notin A, X_{n+2} \notin A, \dots\} \in \sigma(X_n, X_{n+1}, \dots)$$

which typically will not be in $\mathcal{B}_n$.

Proposition 18.24 (New Stopping Times from Old). *Let $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty)$ be a filtered measure space and suppose σ, τ, and $\{\tau_n\}_{n=1}^\infty$ are all stopping times. Then*

1. *$\tau \wedge \sigma$, $\tau \vee \sigma$, $\tau + \sigma$ are all stopping times.*
2. *If $\tau_k \uparrow \tau_\infty$ or $\tau_k \downarrow \tau_\infty$, then τ_∞ is a stopping time.*
3. *In general, $\sup_k \tau_k = \lim_{k\to\infty} \max\{\tau_1, \dots, \tau_k\}$ and $\inf_k \tau_k = \lim_{k\to\infty} \min\{\tau_1, \dots, \tau_k\}$ are also stopping times.*

Proof.

1. Since $\{\tau \wedge \sigma > n\} = \{\tau > n\} \cap \{\sigma > n\} \in \mathcal{B}_n$, $\{\tau \vee \sigma \leq n\} = \{\tau \leq n\} \cap \{\sigma \leq n\} \in \mathcal{B}_n$ for all n, and
$$\{\tau + \sigma = n\} = \cup_{k=0}^n \{\tau = k, \sigma = n - k\} \in \mathcal{B}_n$$
for all n, $\tau \wedge \sigma$, $\tau \vee \sigma$, $\tau + \sigma$ are all stopping times.
2. If $\tau_k \uparrow \tau_\infty$, then $\{\tau_\infty \leq n\} = \cap_k \{\tau_k \leq n\} \in \mathcal{B}_n$ and so τ_∞ is a stopping time. Similarly, if $\tau_k \downarrow \tau_\infty$, then $\{\tau_\infty > n\} = \cap_k \{\tau_k > n\} \in \mathcal{B}_n$ and so τ_∞ is a stopping time. (Recall that $\{\tau_\infty > n\} = \{\tau_\infty \geq n + 1\}$.)
3. This follows from items 1. and 2.

■

Lemma 18.25. *If τ is a stopping time, then the processes, $f_n := 1_{\{\tau \leq n\}}$, and $f_n := 1_{\{\tau = n\}}$ are adapted and $f_n := 1_{\{\tau < n\}}$ is predictable. Moreover, if σ and τ are two stopping times, then $f_n := 1_{\sigma < n \leq \tau}$ is predictable.*

Proof. These are all trivial to prove. For example, if $f_n := 1_{\sigma < n \leq \tau}$, then f_n is $\mathcal{B}_{n-1}$ measurable since,

$$\{\sigma < n \leq \tau\} = \{\sigma < n\} \cap \{n \leq \tau\} = \{\sigma < n\} \cap \{\tau < n\}^c \in \mathcal{B}_{n-1}.$$

■

Notation 18.26 (Stochastic intervals) *If $\sigma, \tau : \Omega \to \bar{\mathbb{N}}$, let*

$$(\sigma, \tau] := \{(\omega, n) \in \Omega \times \bar{\mathbb{N}} : \sigma(\omega) < n \leq \tau(\omega)\}$$

and we will write $1_{(\sigma,\tau]}$ for the process, $1_{\sigma < n \leq \tau}$.

Our next goal is to define the "stopped" σ – algebra, $\mathcal{B}_\tau$. To motivate the upcoming definition, suppose $X_n : \Omega \to \mathbb{R}$ are given functions for all $n \in \mathbb{N}_0$, $\mathcal{B}_n := \sigma(X_0, \dots, X_n)$, and $\tau : \Omega \to \mathbb{N}_0$ is a $\mathcal{B}_\cdot$ – stopping time. Recalling that a function $Y : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ measurable iff $Y(\omega) = f_n(X_0(\omega), \dots X_n(\omega))$ for some measurable function, $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$, it is reasonable to suggest that Y is $\mathcal{B}_\tau$ measurable iff $Y(\omega) = f_{\tau(\omega)}\left(X_0(\omega), \dots X_{\tau(\omega)}(\omega)\right)$, where $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ are measurable random variables. If this is the case, then we would have $1_{\tau=n} Y = f_n(X_0, \dots, X_n)$ is $\mathcal{B}_n$ – measurable for all n. Hence we should define $A \subset \Omega$ to be in $\mathcal{B}_\tau$ iff 1_A is $\mathcal{B}_\tau$ measurable iff $1_{\tau=n} 1_A$ is $\mathcal{B}_n$ measurable for all n which happens iff $\{\tau = n\} \cap A \in \mathcal{B}_n$ for all n.

Definition 18.27 (Stopped σ – algebra). *Given a stopping time τ on a filtered measure space $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty)$ with $\mathcal{B}_\infty := \vee_{n=0}^\infty \mathcal{B}_n := \sigma(\cup_{n=0}^\infty \mathcal{B}_n)$, let*

$$\mathcal{B}_\tau := \{A \subset \Omega : \{\tau = n\} \cap A \in \mathcal{B}_n \text{ for all } n \leq \infty\}. \tag{18.13}$$

Lemma 18.28. *Suppose σ and τ are stopping times.*

1. *A set, $A \subset \Omega$ is in $\mathcal{B}_\tau$ iff $A \cap \{\tau \leq n\} \in \mathcal{B}_n$ for all $n \leq \infty$.*
2. *$\mathcal{B}_\tau$ is a sub-σ-algebra of $\mathcal{B}_\infty$.*
3. *If $\sigma \leq \tau$, then $\mathcal{B}_\sigma \subset \mathcal{B}_\tau$.*

Proof. 1. Since

$$\begin{aligned}A \cap \{\tau \leq n\} &= \cup_{k \leq n} [A \cap \{\tau \leq k\}] \text{ and} \\ A \cap \{\tau = n\} &= [A \cap \{\tau \leq n\}] \setminus [A \cap \{\tau \leq n - 1\}],\end{aligned}$$

it easily follows that $A \subset \Omega$ is in $\mathcal{B}_\tau$ iff $A \cap \{\tau \leq n\} \in \mathcal{B}_n$ for all $n \leq \infty$.

2. Since $\Omega \cap \{\tau \leq n\} = \{\tau \leq n\} \in \mathcal{B}_n$ for all n, it follows that $\Omega \in \mathcal{B}_\tau$. If $A \in \mathcal{B}_\tau$, then, for all $n \in \mathbb{N}_0$,

$$A^c \cap \{\tau \leq n\} = \{\tau \leq n\} \setminus A = \{\tau \leq n\} \setminus [A \cap \{\tau \leq n\}] \in \mathcal{B}_n.$$

This shows $A^c \in \mathcal{B}_\tau$. Similarly if $\{A_k\}_{k=1}^\infty \subset \mathcal{B}_\tau$, then

$$\{\tau \le n\} \cap (\cap_{k=1}^\infty A_k) = \cap_{k=1}^\infty (\{\tau \le n\} \cap A_k) \in \mathcal{B}_n$$

and hence $\cap_{k=1}^\infty A_k \in \mathcal{B}_\tau$. This completes the proof the $\mathcal{B}_\tau$ is a σ – algebra. Since $A = A \cap \{\tau \le \infty\}$, it also follows that $\mathcal{B}_\tau \subset \mathcal{B}_\infty$.

3. Now suppose that $\sigma \le \tau$ and $A \in \mathcal{B}_\sigma$. Since $A \cap \{\sigma \le n\}$ and $\{\tau \le n\}$ are in $\mathcal{B}_n$ for all $n \le \infty$, we find

$$A \cap \{\tau \le n\} = [A \cap \{\sigma \le n\}] \cap \{\tau \le n\} \in \mathcal{B}_n \ \forall\ n \le \infty$$

which shows $A \in \mathcal{B}_\tau$. ■

Proposition 18.29 ($\mathcal{B}_\tau$ – **measurable random variables).** *Let $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty)$ be a filtered measure space. Let τ be a stopping time and $Z : \Omega \to \mathbb{R}$ be a function. Then the following are equivalent;*

1. *Z is $\mathcal{B}_\tau$ – measurable,*
2. *$1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable for all $n \le \infty$,*
3. *$1_{\{\tau = n\}} Z$ is $\mathcal{B}_n$ – measurable for all $n \le \infty$.*
4. *There exists, $Y_n : \Omega \to \mathbb{R}$ which are $\mathcal{B}_n$ – measurable for all $n \le \infty$ such that*
$$Z = Y_\tau = \sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau = n\}} Y_n.$$

Proof. 1. $\implies$ 2. By definition, if $A \in \mathcal{B}_\tau$, then $1_{\{\tau \le n\}} 1_A = 1_{\{\tau \le n\} \cap A}$ is $\mathcal{B}_n$ – measurable for all $n \le \infty$. Consequently any simple $\mathcal{B}_\tau$ – measurable function, Z, satisfies $1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable for all n. So by the usual limiting argument (Theorem 6.39), it follows that $1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable for all n for any $\mathcal{B}_\tau$ – measurable function, Z.

2. $\implies$ 3. This property follows from the identity,

$$1_{\{\tau = n\}} Z = 1_{\{\tau \le n\}} Z - 1_{\{\tau < n\}} Z.$$

3. $\implies$ 4. Simply take $Y_n = 1_{\{\tau = n\}} Z$.

4. $\implies$ 1. Since $Z = \sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau = n\}} Y_n$, it suffices to show $1_{\{\tau = n\}} Y_n$ is $\mathcal{B}_\tau$ – measurable if Y_n is $\mathcal{B}_n$ – measurable. Further, by the usual limiting arguments using Theorem 6.39, it suffices to assume that $Y_n = 1_A$ for some $A \in \mathcal{B}_n$. In this case $1_{\{\tau = n\}} Y_n = 1_{A \cap \{\tau = n\}}$. Hence we must show $A \cap \{\tau = n\} \in \mathcal{B}_\tau$ which indeed is true because

$$A \cap \{\tau = n\} \cap \{\tau = k\} = \begin{cases} \emptyset \in \mathcal{B}_k & \text{if } k \ne n \\ A \cap \{\tau = n\} \in \mathcal{B}_k & \text{if } k = n \end{cases}.$$

Alternatively proof for 1. $\implies$ 2. If Z is $\mathcal{B}_\tau$ measurable, then $\{Z \in B\} \cap \{\tau \le n\} \in \mathcal{B}_n$ for all $n \le \infty$ and $B \in \mathcal{B}_\mathbb{R}$. Hence if $B \in \mathcal{B}_\mathbb{R}$ with $0 \notin B$, then

$$\left\{1_{\{\tau \le n\}} Z \in B\right\} = \{Z \in B\} \cap \{\tau \le n\} \in \mathcal{B}_n \text{ for all } n$$

and similarly,

$$\left\{1_{\{\tau \le n\}} Z = 0\right\}^c = \left\{1_{\{\tau \le n\}} Z \ne 0\right\} = \{Z \ne 0\} \cap \{\tau \le n\} \in \mathcal{B}_n \text{ for all } n.$$

From these two observations, it follows that $\left\{1_{\{\tau \le n\}} Z \in B\right\} \in \mathcal{B}_n$ for all $B \in \mathcal{B}_\mathbb{R}$ and therefore, $1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable. ■

Exercise 18.6. Suppose τ is a stopping time, $(S, \mathcal{S})$ is a measurable space, and $Z : \Omega \to S$ is a function. Show that Z is $\mathcal{B}_\tau / \mathcal{S}$ measurable iff $Z|_{\{\tau = n\}}$ is $(\mathcal{B}_n)_{\{\tau = n\}} / \mathcal{S}$ – measurable for all $n \in \mathbb{N}_0$.

Lemma 18.30 ($\mathcal{B}_\sigma$ – **conditioning).** *Suppose σ is a stopping time and $Z \in L^1(\Omega, \mathcal{B}, P)$ or $Z \ge 0$, then*

$$\mathbb{E}[Z|\mathcal{B}_\sigma] = \sum_{n \le \infty} 1_{\sigma = n} \mathbb{E}[Z|\mathcal{B}_n] = Y_\sigma \tag{18.14}$$

where

$$Y_n := \mathbb{E}[Z|\mathcal{B}_n] \text{ for all } n \in \bar{\mathbb{N}}. \tag{18.15}$$

Proof. By Proposition 18.29, Y_σ is $\mathcal{B}_\sigma$ – measurable. Moreover if Z is integrable, then

$$\begin{aligned}
\sum_{n \le \infty} \mathbb{E}\left[1_{\{\sigma = n\}} |Y_n|\right] &= \sum_{n \le \infty} \mathbb{E} 1_{\{\sigma = n\}} |\mathbb{E}[Z|\mathcal{B}_n]| \\
&\le \sum_{n \le \infty} \mathbb{E}\left[1_{\{\sigma = n\}} \mathbb{E}[|Z| |\mathcal{B}_n]\right] \\
&= \sum_{n \le \infty} \mathbb{E}\left[\mathbb{E}\left[1_{\{\sigma = n\}} |Z| |\mathcal{B}_n\right]\right] \\
&= \sum_{n \le \infty} \mathbb{E}\left[1_{\{\sigma = n\}} |Z|\right] = \mathbb{E}|Z| < \infty
\end{aligned} \tag{18.16}$$

and therefore

$$\begin{aligned}
\mathbb{E}|Y_\sigma| &= \mathbb{E}\left|\sum_{n \le \infty} \left[1_{\{\sigma = n\}} Y_n\right]\right| \\
&\le \sum_{n \le \infty} \mathbb{E}\left[1_{\{\sigma = n\}} |Y_n|\right] \le \mathbb{E}|Z| < \infty.
\end{aligned}$$

Furthermore if $A \in \mathcal{B}_\sigma$, then

$$\begin{aligned}\mathbb{E}[Z:A] &= \sum_{n\le\infty}\mathbb{E}[Z:A\cap\{\sigma=n\}] = \sum_{n\le\infty}\mathbb{E}[Y_n:A\cap\{\sigma=n\}]\\ &= \sum_{n\le\infty}\mathbb{E}\left[1_{\{\sigma=n\}}Y_n:A\right] = \mathbb{E}\left[\sum_{n\le\infty}1_{\{\sigma=n\}}Y_n:A\right]\\ &= \mathbb{E}[Y_\sigma:A],\end{aligned}$$

wherein the interchange of the sum and the expectation in the second to last equality is justified by the estimate in 18.16 or by the fact that everything in sight is positive when $Z\ge 0$. ■

Exercise 18.7. Suppose σ and τ are two stopping times. Show;

1. $\{\sigma<\tau\}$, $\{\sigma=\tau\}$, and $\{\sigma\le\tau\}$ are all in $\mathcal{B}_\sigma\cap\mathcal{B}_\tau$,
2. $\mathcal{B}_{\sigma\wedge\tau}=\mathcal{B}_\sigma\cap\mathcal{B}_\tau$,
3. $\mathcal{B}_{\sigma\vee\tau}=\mathcal{B}_\sigma\vee\mathcal{B}_\tau:=\sigma(\mathcal{B}_\sigma\cup\mathcal{B}_\tau)$, and
4. $\mathcal{B}_\sigma=\mathcal{B}_{\sigma\wedge\tau}$ on C where C is any one of the following three sets; $\{\sigma\le\tau\}$, $\{\sigma<\tau\}$, or $\{\sigma=\tau\}$.

For sake of completeness (as it will be needed in the next exercise), let me check for you directly that $\{\sigma\le\tau\}\in\mathcal{B}_{\sigma\wedge\tau}$. This will be the case iff $\{\sigma\le\tau\}\cap\{\sigma\wedge\tau=n\}\in\mathcal{B}_n$ for all $n\in\mathbb{N}_0$. This is however true since,

$$\{\sigma\le\tau\}\cap\{\sigma\wedge\tau=n\}=\{\sigma\le\tau\}\cap\{\sigma=n\}=\{n\le\tau\}\cap\{\sigma=n\}\in\mathcal{B}_n.$$

Similarly one shows that $\{\sigma<\tau\}$, $\{\tau\le\sigma\}$, $\{\tau<\sigma\}$, and $\{\sigma=\tau\}$ are in $\mathcal{B}_{\sigma\wedge\tau}$.

Theorem 18.31 (Tower Property II). *Let $X\in L^1(\Omega,\mathcal{B},P)$ or $X:\Omega\to[0,\infty]$ be a $\mathcal{B}$ – measurable function. Then given* ***any*** *two stopping times, σ and τ, show*

$$\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}X=\mathbb{E}_{\mathcal{B}_\tau}\mathbb{E}_{\mathcal{B}_\sigma}X=\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X. \tag{18.17}$$

Proof. As usual it suffices to consider the case where $X\ge 0$ and this case there will be now convergence issues to worry about.

First Proof. Notice that

$$1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\tau}=\sum_{n\le\infty}1_{\tau\le\sigma}1_{\tau=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau\le\sigma}\sum_{n\le\infty}1_{\tau\wedge\sigma=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}$$

and similarly,

$$1_{\tau<\sigma}\mathbb{E}_{\mathcal{B}_\tau}=\sum_{n\le\infty}1_{\tau<\sigma}1_{\tau=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau<\sigma}\sum_{n\le\infty}1_{\tau\wedge\sigma=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau<\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}.$$

Using these remarks and the fact that $\{\tau\le\sigma\}$ and $\{\tau>\sigma\}$ are both in $\mathcal{B}_{\sigma\wedge\tau}=\mathcal{B}_\sigma\cap\mathcal{B}_\tau$ we find;

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}&=\mathbb{E}_{\mathcal{B}_\sigma}(1_{\tau\le\sigma}+1_{\tau>\sigma})\mathbb{E}_{\mathcal{B}_\tau}=\mathbb{E}_{\mathcal{B}_\sigma}1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}\\ &=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}\mathbb{E}_{\mathcal{B}_\tau}\\ &=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}=\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}.\end{aligned}$$

Second Proof. In this proof we are going to make use of the localization Lemma 14.23. Since $\mathcal{B}_{\sigma\wedge\tau}\subset\mathcal{B}_\sigma$, it follows by item 4. of Exercise 18.7 that $\mathcal{B}_\sigma=\mathcal{B}_{\sigma\wedge\tau}$ on $\{\sigma\le\tau\}$ and on $\{\sigma<\tau\}$. We will actually use the first statement in the form, $\mathcal{B}_\tau=\mathcal{B}_{\sigma\wedge\tau}$ on $\{\tau\le\sigma\}$. From Lemma 18.30, we have

$$\begin{aligned}1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\tau}&=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}\text{ and}\\ 1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}&=1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}.\end{aligned}$$

Using these relations and the basic properties of conditional expectation we arrive at,

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}X&=\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}[1_{\tau\le\sigma}X+1_{\tau>\sigma}X]\\ &=\mathbb{E}_{\mathcal{B}_\sigma}[1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\tau}X]+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}X\\ &=\mathbb{E}_{\mathcal{B}_\sigma}[1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}X]+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}\mathbb{E}_{\mathcal{B}_\tau}X\\ &=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\sigma}[\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}X]+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X\\ &=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}X+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X=\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X\text{ a.s.}\end{aligned}$$

■

Exercise 18.8. Show, by example, that it is not necessarily true that

$$\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_2}=\mathbb{E}_{\mathcal{G}_1\wedge\mathcal{G}_2}$$

for arbitrary $\mathcal{G}_1$ and $\mathcal{G}_2$ – sub-sigma algebras of $\mathcal{B}$.

Hint: it suffices to take $(\Omega,\mathcal{B},P)$ with $\Omega=\{1,2,3\}$, $\mathcal{B}=2^\Omega$, and $P(\{j\})=\frac{1}{3}$ for $j=1,2,3$.

Exercise 18.9 (Geometry of commuting projections). Suppose that H is a Hilbert space and $H_i\subset H$ for $i=1,2$ are two closed subspaces. Let $P_i=P_{H_i}$ denote orthogonal projection onto H_i and $P=P_M$ be orthogonal projection onto $M:=H_1\cap H_2$. Show;

1. Suppose there exists $M_0\subset H_1\cap H_2$ such that $M_1\perp M_2$ where $M_i=\{h\in H_i:h\perp M_0\}$ so that $H_1=M_0\overset{\perp}{\oplus}M_1$ and $H_2=M_0\overset{\perp}{\oplus}M_2$. Then $M_0=H_1\cap H_2$ and $P_1P_2=P=P_2P_1$.
2. If $P_1P_2=P_2P_1$, then $P_1P_2=P=P_2P_1$. Moreover if we let $M_0=H_1\cap H_2$ and M_i be as above, then $M_1\perp M_2$.

Solution to Exercise (18.9). 1. The assumptions imply that $P_i = P_{M_0} + P_{M_i}$ for $i = 1, 2$. Moreover since $M_1 \perp M_2$ and $M_i \perp M_0$ for $i = 1, 2$, it follows that $P_{M_0} P_{M_i} = P_{M_i} P_{M_0} = 0$ and $P_{M_1} P_{M_2} = 0 = P_{M_2} P_{M_1}$. Therefore,

$$P_1 P_2 = (P_{M_0} + P_{M_1})(P_{M_0} + P_{M_2}) = P_{M_0}^2 = P_{M_0}$$

and similarly $P_2 P_1 = P_{M_0}$. Finally since $P_1 P_2 (H) \supset P_1 P_2 (M) = M$ it follows that $M \subset P_{M_0}(H) = M_0 \subset M$ and so $M = M_0$.

2. Let $Q := P_1 P_2 = P_2 P_1$ and notice that $Qh \in H_1$ and $Qh \in H_2$ for all $h \in H$, i.e. $Q(H) \subset M$, $Q|_M = I_M$ and so $Q(H) = M$, and that $Q^* = (P_1 P_2)^* = P_2^* P_1^* = P_2 P_1 = Q$. Therefore it follows that $Q = P_M$ as claimed.

We now let $M_i = \{h \in H_i : h \perp M_0\}$ where $M_0 = M = H_1 \cap H_2$. Then $M_i = (I - P_M) H_i$ and for $h_i \in H_i$ we have, using $P_M = P_1 P_2$,

$$\begin{aligned}((I - P_M) h_1, (I - P_M) h_2) &= (h_1, (I - P_M) h_2) \\ &= (h_1, h_2) - (h_1, P_1 P_2 h_2) \\ &= (h_1, h_2) - (P_1 h_1, P_2 h_2) = (h_1, h_2) - (h_1, h_2) = 0.\end{aligned}$$

Hence it follows that $M_1 \perp M_2$.

Alternative proof. Let P_i be orthogonal projection onto H_i, P_0 be orthogonal projection onto $H_0 := H_1 \cap H_2$, and Q_i be orthogonal projection onto M_i. Then $P_i = P_0 + Q_i$ for $i = 1, 2$. Indeed, we $P_0 Q_i = Q_i P_0 = 0$ so that

$$(P_0 + Q_i)^2 = P_0^2 + Q_i^2 = P_0 + Q_i$$

and $(P_0 + Q_i)^2 = P_0 + Q_i$ which shows that $P_0 + Q_i$ is an orthogonal projection. Moreover it is easy to check that $\mathrm{Ran}(P_0 + Q_i) = H_i$ so that $P_i = P_0 + Q_i$ as claimed. Having said this we have in general that

$$P_1 P_2 = (P_0 + Q_1)(P_0 + Q_2) = P_0 + Q_1 Q_2$$

and similarly that $P_2 P_1 = P_0 + Q_2 Q_1$ from which it follows that $P_1 P_2 = P_2 P_1$ iff $Q_1 Q_2 = Q_2 Q_1$. Lastly if $Q_1 Q_2 = Q_2 Q_1$ then for all $x \in H$ we will have

$$Q_1 Q_2 x = Q_2 Q_1 x \in M_1 \cap M_2 = \{0\}$$

which shows that $Q_1 Q_2 = Q_2 Q_1 = 0$. Thus we have shown that $Q_1 Q_2 = Q_2 Q_1$ iff $Q_1 Q_2 = 0 = Q_2 Q_1$ which happens iff $M_1 \perp M_2$.

Exercise 18.10. Let σ and τ be stopping times and apply the results of Exercise 18.9 with $M_0 := L^2(\Omega, \mathcal{B}_{\sigma\wedge\tau}, P)$, $H_1 := L^2(\Omega, \mathcal{B}_\sigma, P)$, and $H_2 = L^2(\Omega, \mathcal{B}_\tau, P)$ to give another proof of Theorem 18.31.

Solution to Exercise (18.9). In order to apply Exercise 18.9, we need to show if $X \in H_1$ and $Y \in H_2$ are both orthogonal to M_0 then $X \perp Y$, i.e.

$\mathbb{E}[XY] = 0$. In order to compute this last expectation, let $\bar{X}_n := \mathbb{E}[X|\mathcal{B}_n]$ and $\bar{Y}_n := \mathbb{E}[Y|\mathcal{B}_n]$ for all $n \le \infty$ so that (by Lemma 18.30)

$$X = \mathbb{E}[X|\mathcal{B}_\sigma] = \sum_{n\le\infty} 1_{\sigma=n} \mathbb{E}[X|\mathcal{B}_n] = \sum_{n\le\infty} 1_{\sigma=n} \bar{X}_n = \bar{X}_\sigma$$

and similarly $Y = \bar{Y}_\tau$. We then have,

$$\begin{aligned} XY = \bar{X}_\sigma \bar{Y}_\tau &= \bar{X}_\sigma \bar{Y}_\tau (1_{\sigma\le\tau} + 1_{\sigma>\tau}) = 1_{\sigma\le\tau} \bar{X}_\sigma \bar{Y}_\tau + \bar{X}_\sigma \bar{Y}_\tau 1_{\sigma>\tau} \\ &= \left(1_{\sigma\le\tau} \bar{X}_{\sigma\wedge\tau}\right) \bar{Y}_\tau + \bar{X}_\sigma \left(\bar{Y}_{\sigma\wedge\tau} 1_{\sigma>\tau}\right) \\ &= \left(1_{\sigma\le\tau} \bar{X}_{\sigma\wedge\tau}\right) Y + X \left(\bar{Y}_{\sigma\wedge\tau} 1_{\sigma>\tau}\right). \end{aligned}$$

Since $\left(1_{\sigma\le\tau} \bar{X}_{\sigma\wedge\tau}\right)$ and $\left(\bar{Y}_{\sigma\wedge\tau} 1_{\sigma>\tau}\right)$ are both in $M_0 = L^2(\Omega, \mathcal{B}_{\sigma\wedge\tau}, P)$ (notice that $L^2(P) \ni \bar{X}_{\sigma\wedge\tau} = \mathbb{E}[X|\mathcal{B}_{\sigma\wedge\tau}]$ by Lemma 18.30 again) and both X and Y are orthogonal to M_0 by assumption, we may conclude

$$\mathbb{E}[XY] = \mathbb{E}\left[\left(1_{\sigma\le\tau} \bar{X}_{\sigma\wedge\tau}\right) Y\right] + \mathbb{E}\left[X \left(\bar{Y}_{\sigma\wedge\tau} 1_{\sigma>\tau}\right)\right] = 0 + 0 = 0.$$

Having checked the hypothesis of item 1. of Exercise 18.9 we may conclude that

$$\mathbb{E}_{\mathcal{B}_\sigma} \mathbb{E}_{\mathcal{B}_\tau} = \mathbb{E}_{\mathcal{B}_\tau} \mathbb{E}_{\mathcal{B}_\sigma} = \mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}$$

on $L^2(\Omega, \mathcal{B}, P)$. This then extends to $L^1(\Omega, \mathcal{B}, P)$ by the standard limiting arguments we used in constructing conditional expectations.

18.4 Stochastic Integrals and Optional Stopping

Notation 18.32 *Suppose that* $\{c_n\}_{n=1}^\infty$ *and* $\{x_n\}_{n=0}^\infty$ *are two sequences of numbers, let* $c \cdot \Delta x = \{(c \cdot \Delta x)_n\}_{n\in\mathbb{N}_0}$ *denote the sequence of numbers defined by* $(c \cdot \Delta x)_0 = 0$ *and*

$$(c \cdot \Delta x)_n = \sum_{j=1}^n c_j (x_j - x_{j-1}) = \sum_{j=1}^n c_j \Delta_j x \text{ for } n \ge 1.$$

(For convenience of notation later we will interpret $\sum_{j=1}^0 c_j \Delta_j x = 0$*.)*

For a gambling interpretation of $(c \cdot \Delta x)_n$, let x_j represent the price of a stock at time j. Suppose that you, the investor, then buys c_{j-1} shares at time $j-1$ and then sells these shares back at time j. With this interpretation, $c_{j-1} \Delta_j x$ represents your profit (or loss if negative) in the time interval from $j-1$ to j and $(c \cdot \Delta x)_n$ represents your profit (or loss) from time 0 to time n. By the way, if you want to buy 5 shares of the stock at time $n = 3$ and then sell them all at time 9, you would take $c_k = 5 \cdot 1_{3<k\le 9}$ so that

$$(c \cdot \Delta x)_9 = 5 \cdot \sum_{3<k\le 9} \Delta_k x = 5 \cdot (x_9 - x_3)$$

would represent your profit (loss) for this transaction. The next example formalizes this notion.

Example 18.33. Suppose that $0 \le \sigma \le \tau$ where $\sigma, \tau \in \bar{\mathbb{N}}_0$ and let $c_n := 1_{\sigma<n\le\tau}$. Then

$$\begin{aligned}(c \cdot \Delta x)_n &= \sum_{j=1}^{n} 1_{\sigma<j\le\tau} (x_j - x_{j-1}) = \sum_{j=1}^{\infty} 1_{\sigma<j\le\tau\wedge n} (x_j - x_{j-1}) \\ &= \sum_{j=1}^{\infty} 1_{\sigma\wedge n<j\le\tau\wedge n} (x_j - x_{j-1}) = x_{\tau\wedge n} - x_{\sigma\wedge n}.\end{aligned}$$

More generally if $\sigma, \tau \in \bar{\mathbb{N}}_0$ are arbitrary and $c_n := 1_{\sigma<n\le\tau}$ we will have $c_n := 1_{\sigma\wedge\tau<n\le\tau}$ and therefore

$$(c \cdot \Delta x)_n = x_{\tau\wedge n} - x_{\sigma\wedge\tau\wedge n}.$$

Proposition 18.34 (The Discrete Stochastic Integral). *Let $X = \{X_n\}_{n=0}^{\infty}$ be an adapted integrable process, i.e. $\mathbb{E}|X_n| < \infty$ for all n. If X is a martingale and $\{C_n\}_{n=1}^{\infty}$ is a predictable sequence of bounded random variables, then $\{(C \cdot \Delta X)_n\}_{n=1}^{\infty}$ is still a martingale. If $X := \{X_n\}_{n=0}^{\infty}$ is a submartingale (supermartingale) (necessarily real valued) and $C_n \ge 0$, then $\{(C \cdot \Delta X)_n\}_{n=1}^{\infty}$ is a submartingale (supermartingale).*

Conversely if X is an adapted process of integrable functions such that $\mathbb{E}[(C \cdot \Delta X)_n] = 0$ for all bounded predictable processes, $\{C_n\}_{n=1}^{\infty}$, then X is a martingale. Similarly if X is real valued adapted process such that

$$\mathbb{E}[(C \cdot \Delta X)_n] \begin{matrix}\le\\=\\\ge\end{matrix} 0 \tag{18.18}$$

for all n and for all bounded, non-negative predictable processes, C, then X is a supermartingale, martingale, or submartingale respectively. (In other words, X is a sub-martingale if no matter what your (non-negative) betting strategy is you will make money on average.)

Proof. For any adapted process X, we have

$$\begin{aligned}\mathbb{E}\left[(C \cdot \Delta X)_{n+1} |\mathcal{B}_n\right] &= \mathbb{E}\left[(C \cdot \Delta X)_n + C_{n+1} (X_{n+1} - X_n) |\mathcal{B}_n\right] \\ &= (C \cdot \Delta X)_n + C_{n+1}\mathbb{E}\left[(X_{n+1} - X_n) |\mathcal{B}_n\right].\end{aligned} \tag{18.19}$$

The first assertions easily follow from this identity.

Now suppose that X is an adapted process of integrable functions such that $\mathbb{E}[(C \cdot \Delta X)_n] = 0$ for all bounded predictable processes, $\{C_n\}_{n=1}^{\infty}$. Taking expectations of Eq. (18.19) then allows us to conclude that

$$\mathbb{E}\left[C_{n+1}\mathbb{E}\left[(X_{n+1} - X_n) |\mathcal{B}_n\right]\right] = 0$$

for all bounded $\mathcal{B}_n$ – measurable random variables, C_{n+1}. Taking $C_{n+1} := \operatorname{sgn}(\mathbb{E}[(X_{n+1} - X_n) |\mathcal{B}_n])$ shows $|\mathbb{E}[(X_{n+1} - X_n) |\mathcal{B}_n]| = 0$ a.s. and hence X is a martingale. Similarly, if for all non-negative, predictable C, Eq. (18.18) holds for all $n \ge 1$, and $C_n \ge 0$, then taking $A \in \mathcal{B}_n$ and $C_k = \delta_{k,n+1} 1_A$ in Eq. (18.12) allows us to conclude that

$$\mathbb{E}[X_{n+1} - X_n : A] = \mathbb{E}\left[(C \cdot \Delta X)_{n+1}\right] \begin{matrix}\le\\=\\\ge\end{matrix} 0,$$

i.e. X is a supermartingale, martingale, or submartingale respectively. ■

Example 18.35. Suppose that $\{X_n\}_{n=0}^{\infty}$ are mean zero independent integrable random variables and $f_k : \mathbb{R}^k \to \mathbb{R}$ are bounded measurable functions for $k \in \mathbb{N}$. Then $\{Y_n\}_{n=0}^{\infty}$, defined by $Y_0 = 0$ and

$$Y_n := \sum_{k=1}^{n} f_k (X_0, \dots, X_{k-1}) (X_k - X_{k-1}) \text{ for } n \in \mathbb{N}, \tag{18.20}$$

is a martingale sequence relative to $\left\{\mathcal{B}_n^X\right\}_{n\ge0}$.

Notation 18.36 *Given an adapted process, X, and a stopping time τ, let $X_n^\tau := X_{\tau\wedge n}$. We call $X^\tau := \{X_n^\tau\}_{n=0}^{\infty}$ the **process X stopped by** τ.*

Observe that

$$|X_n^\tau| = |X_{\tau\wedge n}| = \left|\sum_{0\le k\le n} 1_{\tau=k} X_k\right| \le \sum_{0\le k\le n} 1_{\tau=k} |X_k| \le \sum_{0\le k\le n} |X_k|,$$

so that $X_n^\tau \in L^1(P)$ for all n provided $X_n \in L^1(P)$ for all n.

Example 18.37. Suppose that $X = \{X_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale, with $\mathbb{E}|X_n| < \infty$ and let σ and τ be stopping times. Then for any $A \in \mathcal{B}_\sigma$, the process $C_n := 1_A \cdot 1_{\sigma<n\le\tau}$ is predictable since for all $n \in \mathbb{N}$ we have

$$\begin{aligned}A \cap \{\sigma < n \le \tau\} &= (A \cap \{\sigma < n\}) \cap \{n \le \tau\} \\ &= (A \cap \{\sigma \le n-1\}) \cap \{\tau \le n-1\}^c \in \mathcal{B}_{n-1}.\end{aligned}$$

Therefore by Proposition 18.34, $\{(C \cdot \Delta X)_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale respectively where

$$\begin{aligned}(C \cdot \Delta X)_n &= \sum_{k=1}^{n} 1_A \cdot 1_{\sigma < k \le \tau} \Delta_k X = 1_A \cdot \sum_{k=1}^{n} 1_{\sigma \wedge \tau < k \le \tau} \Delta_k X \\ &= \sum_{k=1}^{\infty} 1_A \cdot 1_{\sigma \wedge \tau \wedge n < k \le \tau \wedge n} \Delta_k X = 1_A \left(X_n^{\tau} - X_n^{\sigma \wedge \tau}\right).\end{aligned}$$

Theorem 18.38 (Optional stopping theorem). *Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale with either $\mathbb{E}|X_n| < \infty$ for all n or $X_n \ge 0$ for all n. Then for every stopping time, τ, X^{τ} is a $\{\mathcal{B}_n\}_{n=0}^{\infty}$ – supermartingale, martingale, or submartingale respectively.*

Proof. When $\mathbb{E}|X_n| < \infty$ for all $n \ge 0$ we may take $\sigma = 0$ and $A = \Omega$ in Example 18.37 in order to learn that $\{X_n^{\tau} - X_0\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale respectively and therefore so is $\{X_n^{\tau} = X_0 + X_n^{\tau} - X_0\}_{n=0}^{\infty}$. When X_n is only non-negative we have to give a different proof which does not involve any subtractions (which might be undefined).

For the second proof we simply observe that $1_{\tau \le n} X_{\tau} = \sum_{k=0}^{n} 1_{\tau = k} X_k$ is $\mathcal{B}_n$ measurable, $\{\tau > n\} \in \mathcal{B}_n$, and

$$X_{\tau \wedge (n+1)} = 1_{\tau \le n} X_{\tau} + 1_{\tau > n} X_{n+1}.$$

Therefore

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n}\left[X_{(n+1)}^{\tau}\right] = \mathbb{E}_{\mathcal{B}_n}\left[X_{\tau \wedge (n+1)}\right] &= 1_{\tau \le n} X_{\tau} + 1_{\tau > n} \mathbb{E}_{\mathcal{B}_n} X_{n+1} \\ &\begin{Bmatrix} \le \\ = \\ \ge \end{Bmatrix} 1_{\tau \le n} X_{\tau} + 1_{\tau > n} X_n = X_{\tau \wedge n},\end{aligned}$$

where the top, middle, bottom (in)equality holds depending on whether X is a supermartingale, martingale, or submartingale respectively. (This second proof works for both cases at once. For another proof see Remark 18.40.) ■

Theorem 18.39 (Optional sampling theorem I). *Suppose that σ and τ are two stopping times and τ is bounded, i.e. there exists $N \in \mathbb{N}$ such that $\tau \le N < \infty$ a.s. If $X = \{X_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale, with either $\mathbb{E}|X_n| < \infty$ of $X_n \ge 0$ for all $0 \le n \le N$, then*

$$\mathbb{E}[X_{\tau}|\mathcal{B}_{\sigma}] \underset{\ge}{\overset{\le}{=}} X_{\sigma \wedge \tau} \text{ a.s.} \tag{18.21}$$

respectively[2] *from top to bottom.*

[2] This is the natural generalization of Eq. (18.3) to the stopping time setting.

Proof. First suppose that $\mathbb{E}|X_n| < \infty$ for $0 \le n \le N$ and let $A \in \mathcal{B}_{\sigma}$. From Example 18.37 we know that $1_A\left(X_n^{\tau} - X_n^{\sigma \wedge \tau}\right)$ is a supermartingale, martingale, or submartingale respectively and in particular for all $n \in \mathbb{N}_0$ we have

$$\mathbb{E}\left[1_A\left(X_n^{\tau} - X_n^{\sigma \wedge \tau}\right)\right] \underset{\ge}{\overset{\le}{=}} 0 \text{ respectively.}$$

Taking $n = N$ in this equation using $\sigma \wedge \tau \le \tau \le N$ then implies, for all $A \in \mathcal{B}_{\sigma}$, that

$$\mathbb{E}\left[(X_{\tau} - X_{\sigma \wedge \tau}) : A\right] \underset{\ge}{\overset{\le}{=}} 0 \text{ respectively}$$

and this is equivalent to Eq. (18.21).

When we only assume that $X_n \ge 0$ for all n we again have to give a different proof which avoids subtractions which may be undefined. One way to do this is to use Theorem 18.38 in order to conclude that X^{τ} is a supermartingale, martingale, or submartingale respectively and in particular that

$$\mathbb{E}[X_{\tau}|\mathcal{B}_n] = \mathbb{E}[X_N^{\tau}|\mathcal{B}_n] \underset{\ge}{\overset{\le}{=}} X_{n \wedge N}^{\tau} \text{ for all } n \le \infty.$$

Combining this result with Lemma 18.30 then implies

$$\mathbb{E}[X_{\tau}|\mathcal{B}_{\sigma}] = \sum_{n \le \infty} 1_{\sigma = n} \mathbb{E}[X_{\tau}|\mathcal{B}_n] \underset{\ge}{\overset{\le}{=}} \sum_{n \le \infty} 1_{\sigma = n} X_{n \wedge N}^{\tau} = X_{\sigma \wedge N}^{\tau} = X_{\sigma \wedge \tau}. \tag{18.22}$$

(This second proof again covers both cases at once!) ■

Exercise 18.11. Give another proof of Theorem 18.39 when $\mathbb{E}|X_n| < \infty$ by using the tower property in Theorem 18.31 along with the Doob decomposition of Lemma 18.16.

Solution to Exercise (18.11). First suppose X is a martingale in which case $X_n = \mathbb{E}_{\mathcal{B}_n} X_N$ for all $n \le N$ and hence by Lemma 18.16

$$X_{\tau} = \sum_{n \le N} 1_{\tau = n} X_n = \sum_{n \le N} 1_{\tau = n} \mathbb{E}_{\mathcal{B}_n} X_N = \sum_{n \le \infty} 1_{\tau = n} \mathbb{E}_{\mathcal{B}_n} X_N = \mathbb{E}_{\mathcal{B}_{\tau}} X_N.$$

Therefore, by Theorem 18.31

$$\mathbb{E}_{\mathcal{B}_{\sigma}} X_{\tau} = \mathbb{E}_{\mathcal{B}_{\sigma}} \mathbb{E}_{\mathcal{B}_{\tau}} X_N = \mathbb{E}_{\mathcal{B}_{\sigma \wedge \tau}} X_N = X_{\sigma \wedge \tau}.$$

Now suppose that X is a submartingale. By the Doob decomposition of Lemma 18.16, $X_n = M_n + A_n$ where M is a martingale and A is an increasing predictable process. In this case we have

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_\sigma} X_\tau &= \mathbb{E}_{\mathcal{B}_\sigma} M_\tau + \mathbb{E}_{\mathcal{B}_\sigma} A_\tau = M_{\sigma\wedge\tau} + \mathbb{E}_{\mathcal{B}_\sigma} A_\tau \\ &\geq M_{\sigma\wedge\tau} + \mathbb{E}_{\mathcal{B}_\sigma} A_{\sigma\wedge\tau} = M_{\sigma\wedge\tau} + A_{\sigma\wedge\tau} = X_{\sigma\wedge\tau}.\end{aligned}$$

The supermartingale case follows from the submartingale result just proved applied to $-X$.

Exercise 18.12. Give yet another (full) proof of Theorem 18.39 using the following outline;

1. Show by induction on n starting with $n = N$ that
$$\mathbb{E}[X_\tau|\mathcal{B}_n] \underset{\geq}{\overset{\leq}{=}} X_{\tau\wedge n} \text{ a.s. for all } 0 \leq n \leq N. \tag{18.23}$$
2. Observe the above inequality holds as an equality for $n > N$ as well.
3. Combine this result with Lemma 18.16 to complete the proof.

This argument makes it clear why we must at least initially assume that $\tau \leq N$ for some $N \in \mathbb{N}$. To relax this restriction will require a limiting argument which will be the topic of Section 18.8 below.

Solution to Exercise (18.12). To keep the notation manageable I will give the proof in the case that $\{X_n\}$ is a submartingale. Since $\tau \leq N$ everywhere, X_τ is $\mathcal{B}_\tau \subset \mathcal{B}_N$ measurable, it follows that Eq. (18.23) holds for all $n \geq N$. Now consider $n = N-1$. Using

$$X_\tau = 1_{\tau=N} X_N + 1_{\tau\leq N-1} X_\tau$$

where $\{\tau = N\} = \{\tau \leq N-1\}^c \in \mathcal{B}_{N-1}$ and $1_{\tau\leq N-1} X_\tau$ is $\mathcal{B}_{N-1}$ – measurable, we learn,

$$\begin{aligned}\mathbb{E}[X_\tau|\mathcal{B}_{N-1}] &= 1_{\tau=N}\mathbb{E}[X_N|\mathcal{B}_{N-1}] + 1_{\tau\leq N-1} X_\tau \\ &\geq 1_{\tau=N} X_{N-1} + 1_{\tau\leq N-1} X_\tau = X_{\tau\wedge(N-1)}.\end{aligned}$$

Applying this same argument with τ replaced by $\tau\wedge(N-k) \leq N-k$ shows

$$\mathbb{E}\left[X_{\tau\wedge(N-k)}|\mathcal{B}_{N-k-1}\right] \geq X_{\tau\wedge(N-k)\wedge(N-k-1)} = X_{\tau\wedge(N-k-1)}$$

for any $0 \leq k \leq N-1$. Combining these observations with the tower property of conditional expectations allows us to conclude that

$$\mathbb{E}[X_\tau|\mathcal{B}_{N-k}] \geq X_{\tau\wedge(N-k)}.$$

For example,

$$\begin{aligned}\mathbb{E}[X_\tau|\mathcal{B}_{N-3}] &= \mathbb{E}[\mathbb{E}[X_\tau|\mathcal{B}_{N-1}]|\mathcal{B}_{N-3}] \\ &\geq \mathbb{E}\left[X_{\tau\wedge(N-1)}|\mathcal{B}_{N-3}\right] = \mathbb{E}\left[\mathbb{E}\left[X_{\tau\wedge(N-1)}|\mathcal{B}_{N-2}\right]|\mathcal{B}_{N-3}\right] \\ &\geq \mathbb{E}\left[X_{\tau\wedge(N-2)}|\mathcal{B}_{N-3}\right] \geq X_{\tau\wedge(N-3)}.\end{aligned}$$

Thus we have now verified Eq. (18.23) for all $n \in \bar{\mathbb{N}}_0$.

We now combine this result with Lemma 18.16 to learn,

$$\mathbb{E}[X_\tau|\mathcal{B}_\sigma] = \sum_{n\leq\infty} 1_{\sigma=n}\mathbb{E}[X_\tau|\mathcal{B}_n] \geq \sum_{n\leq\infty} 1_{\sigma=n} X_{\tau\wedge n} = X_{\tau\wedge\sigma}.$$

Remark 18.40. Theorem 18.39 can be used to give a simple proof of the Optional stopping Theorem 18.38. For example, if $X = \{X_n\}_{n=0}^\infty$ is a submartingale and τ is a stopping time, then

$$\mathbb{E}_{\mathcal{B}_n} X_{\tau\wedge(n+1)} \geq X_{[\tau\wedge(n+1)]\wedge n} = X_{\tau\wedge n},$$

i.e. X^τ is a submartingale.

18.5 Submartingale Maximal Inequalities

Notation 18.41 (Running Maximum) *If $X = \{X_n\}_{n=0}^\infty$ is a sequence of (extended) real numbers, we let*

$$X_N^* := \max\{X_0, \dots, X_N\}. \tag{18.24}$$

Proposition 18.42 (Maximal Inequalities of Bernstein and Lévy). *Let $\{X_n\}$ be a submartingale on a filtered probability space, $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$. Then*[3] *for any $a \geq 0$ and $N \in \mathbb{N}$,*

$$aP(X_N^* \geq a) \leq \mathbb{E}[X_N : X_N^* \geq a] \leq \mathbb{E}\left[X_N^+\right], \tag{18.25}$$

$$aP\left(\min_{n\leq N} X_n \leq -a\right) \leq \mathbb{E}\left[X_N : \min_{k\leq N} X_k > -a\right] - \mathbb{E}[X_0] \tag{18.26}$$

$$\leq \mathbb{E}\left[X_N^+\right] - \mathbb{E}[X_0], \tag{18.27}$$

and

$$aP(X_N^* \geq a) \leq 2\mathbb{E}\left[X_N^+\right] - \mathbb{E}[X_0]. \tag{18.28}$$

[3] The first inequality is the most important.

Proof. Let $\tau := \inf\{n : X_n \geq a\}$ and observe that

$$X_N^* \geq X_\tau \geq a \text{ on } \{\tau \leq N\} = \{X_N^* \geq a\}. \tag{18.29}$$

Since $\{\tau \leq N\} \in B_{\tau\wedge N}$, it follows by the optional sampling Theorem 18.39 that

$$\mathbb{E}[X_\tau : \tau \leq N] = \mathbb{E}[X_{\tau\wedge N} : \tau \leq N] \leq \mathbb{E}[X_N : \tau \leq N]$$

which combined with Eq. (18.29) implies,

$$a \cdot P(X_N^* \geq a) = a \cdot P(\tau \leq N) \leq \mathbb{E}[X_\tau : \tau \leq N] \leq \mathbb{E}[X_N : X_N^* \geq a],$$

i.e. Eq. (18.25) holds.

More generally if X is **any** integrable process and τ is the random time defined by, $\tau := \inf\{n : X_n \geq a\}$ we still have Eq. (18.29) and

$$\begin{aligned} aP(X_N^* \geq a) &= \mathbb{E}[a : \tau \leq N] \leq \mathbb{E}[X_\tau : \tau \leq N] && (18.30)\\ &= \mathbb{E}[X_N : \tau \leq N] - \mathbb{E}[X_N - X_\tau : \tau \leq N] \\ &= \mathbb{E}[X_N : \tau \leq N] - \mathbb{E}[X_N - X_{\tau\wedge N}]. && (18.31)\end{aligned}$$

Let me emphasize again that in deriving Eq. (18.31), we have **not** used any special properties (not even adaptedness) of X. If X is now assumed to be a submartingale, by the optional sampling Theorem 18.39, $\mathbb{E}_{\mathcal{B}_{\tau\wedge N}} X_N \geq X_{\tau\wedge N}$ and in particular $\mathbb{E}[X_N - X_{\tau\wedge N}] \geq 0$. Combining this observation with Eq. (18.31) and Eq. (18.29) again gives Eq. (18.25).

Secondly we may apply Eq. (18.31) with X_n replaced by $-X_n$ to find

$$\begin{aligned} aP\left(\min_{n\leq N} X_n \leq -a\right) &= aP\left(-\min_{n\leq N} X_n \geq a\right) = aP\left(\max_{n\leq N}(-X_n) \geq a\right) \\ &\leq -\mathbb{E}[X_N : \tau \leq N] + \mathbb{E}[X_N - X_{\tau\wedge N}] && (18.32)\end{aligned}$$

where now,

$$\tau := \inf\{n : -X_n \geq a\} = \inf\{n : X_n \leq -a\}.$$

By the optional sampling Theorem 18.39, $\mathbb{E}[X_{\tau\wedge N} - X_0] \geq 0$ and adding this to right side of Eq. (18.32) gives the estimate

$$\begin{aligned} aP\left(\min_{n\leq N} X_n \leq -a\right) &\leq -\mathbb{E}[X_N : \tau \leq N] + \mathbb{E}[X_N - X_{\tau\wedge N}] + \mathbb{E}[X_{\tau\wedge N} - X_0] \\ &\leq \mathbb{E}[X_N - X_0] - \mathbb{E}[X_N : \tau \leq N] \\ &= \mathbb{E}[X_N : \tau > N] - \mathbb{E}[X_0] \\ &= \mathbb{E}\left[X_N : \min_{k\leq N} X_k > -a\right] - \mathbb{E}[X_0]\end{aligned}$$

which proves Eq. (18.26) and hence Eq. (18.27). Adding Eqs. (18.25) and (18.27) gives the estimate in Eq. (18.28). ∎

Remark 18.43. It is of course possible to give a direct proof of Proposition 18.42. For example,

$$\begin{aligned} \mathbb{E}\left[X_N : \max_{n\leq N} X_n \geq a\right] &= \sum_{k=1}^N \mathbb{E}[X_N : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &\geq \sum_{k=1}^N \mathbb{E}[X_k : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &\geq \sum_{k=1}^N \mathbb{E}[a : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &= aP\left(\max_{n\leq N} X_n \geq a\right)\end{aligned}$$

which proves Eq. (18.25).

Corollary 18.44. *Suppose that $\{Y_n\}_{n=1}^\infty$ is a non-negative supermartingale, $a > 0$ and $N \in \mathbb{N}$, then*

$$aP\left(\max_{n\leq N} Y_n \geq a\right) \leq \mathbb{E}[Y_0 \wedge a] - \mathbb{E}\left[Y_N : \max_{n\leq N} Y_n < a\right] \leq \mathbb{E}[Y_0 \wedge a]. \tag{18.33}$$

Proof. Let $X_n := -Y_n$ in Eq. (18.26) to learn

$$aP\left(\min_{n\leq N}(-Y_n) \leq -a\right) \leq \mathbb{E}\left[-Y_N : \min_{n\leq N}(-Y_n) > -a\right] + \mathbb{E}[Y_0]$$

or equivalently that

$$aP\left(\max_{n\leq N} Y_n \geq a\right) \leq \mathbb{E}[Y_0] - \mathbb{E}\left[Y_N : \max_{n\leq N} Y_n < a\right] \leq \mathbb{E}[Y_0]. \tag{18.34}$$

Since $\varphi_a(x) := a \wedge x$ is concave and nondecreasing, it follows by Jensen's inequality that

$$\mathbb{E}[\varphi_a(Y_n)|\mathcal{B}_m] \leq \varphi_a(\mathbb{E}[Y_n|\mathcal{B}_m]) \leq \varphi_a(Y_n) \text{ for all } m \leq n.$$

In this way we see that $\varphi_a(Y_n) = Y_n \wedge a$ is a supermartingale as well. Applying Eq. (18.34) with Y_n replaced by $Y_n \wedge a$ proves Eq. (18.33). ∎

Lemma 18.45. *Suppose that X and Y are two non-negative random variables such that $P(Y \geq y) \leq \frac{1}{y}\mathbb{E}[X : Y \geq y]$ for all $y > 0$. Then for all $p \in (1,\infty)$,*

$$\mathbb{E}Y^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}X^p. \tag{18.35}$$

Proof. We will begin by proving Eq. (18.35) under the additional assumption that $Y \in L^p(\Omega, \mathcal{B}, P)$. Since

$$\begin{aligned}
\mathbb{E}Y^p &= p\mathbb{E}\int_0^\infty 1_{y\le Y} \cdot y^{p-1}dy = p\int_0^\infty \mathbb{E}\left[1_{y\le Y}\right] \cdot y^{p-1}dy \\
&= p\int_0^\infty P(Y \ge y) \cdot y^{p-1}dy \le p\int_0^\infty \frac{1}{y}\mathbb{E}\left[X : Y \ge y\right] \cdot y^{p-1}dy \\
&= p\mathbb{E}\int_0^\infty X1_{y\le Y} \cdot y^{p-2}dy = \frac{p}{p-1}\mathbb{E}\left[XY^{p-1}\right].
\end{aligned}$$

Now apply Hölder's inequality, with $q = p(p-1)^{-1}$, to find

$$\mathbb{E}\left[XY^{p-1}\right] \le \|X\|_p \cdot \left\|Y^{p-1}\right\|_q = \|X\|_p \cdot \left[\mathbb{E}|Y|^p\right]^{1/q}.$$

Combining thew two inequalities shows and solving for $\|Y\|_p$ shows $\|Y\|_p \le \frac{p}{p-1}\|X\|_p$ which proves Eq. (18.35) under the additional restriction of Y being in $L^p(\Omega, \mathcal{B}, P)$.

To remove the integrability restriction on Y, for $M > 0$ let $Z := Y \wedge M$ and observe that

$$P(Z \ge y) = P(Y \ge y) \le \frac{1}{y}\mathbb{E}\left[X : Y \ge y\right] = \frac{1}{y}\mathbb{E}\left[X : Z \ge y\right] \text{ if } y \le M$$

while

$$P(Z \ge y) = 0 = \frac{1}{y}\mathbb{E}\left[X : Z \ge y\right] \text{ if } y > M.$$

Since Z is bounded, the special case just proved shows

$$\mathbb{E}\left[(Y \wedge M)^p\right] = \mathbb{E}Z^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X^p.$$

We may now use the MCT to pass to the limit, $M \uparrow \infty$, and hence conclude that Eq. (18.35) holds in general. ∎

Corollary 18.46 (Doob's Inequality). *If $X = \{X_n\}_{n=0}^\infty$ be a non-negative submartingale and $1 < p < \infty$, then*

$$\mathbb{E}X_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X_N^p. \tag{18.36}$$

Proof. Equation 18.36 follows by applying Lemma 18.45 with the aid of Proposition 18.42. ∎

Corollary 18.47 (Doob's Inequality). *If $\{M_n\}_{n=0}^\infty$ is a martingale and $1 < p < \infty$, then for all $a > 0$,*

$$P\left(|M|_N^* \ge a\right) \le \frac{1}{a}\mathbb{E}\left[|M|_N : M_N^* \ge a\right] \le \frac{1}{a}\mathbb{E}\left[|M_N|\right] \tag{18.37}$$

and

$$\mathbb{E}|M|_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_N|^p. \tag{18.38}$$

Proof. By the conditional Jensen's inequality, it follows that $X_n := |M_n|$ is a submartingale. Hence Eq. (18.37) follows from Eq. (18.25) and Eq. (18.38) follows from Eq. (18.36). ∎

Example 18.48. Let $\{X_n\}$ be a sequence of independent integrable random variables with mean zero, $S_0 = 0$, $S_n := X_1 + \dots + X_n$ for $n \in \mathbb{N}$, and $|S|_n^* = \max_{j\le n}|S_j|$. Since $\{S_n\}_{n=0}^\infty$ is a martingale, by cJensen's inequality, $\{|S_n|^p\}_{n=1}^\infty$ is a (possibly extended) submartingale for any $p \in [1, \infty)$. Therefore an application of Eq. (18.25) of Proposition 18.42 show

$$P\left(|S|_N^* \ge \alpha\right) = P\left(|S|_N^{*p} \ge \alpha^p\right) \le \frac{1}{\alpha^p}\mathbb{E}\left[|S_N|^p : S_N^* \ge \alpha\right].$$

(When $p = 2$, this is Kolmogorov's inequality in Theorem 20.42 below.) From Corollary 18.47 we also know that

$$\mathbb{E}|S|_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|S_N|^p.$$

In particular when $p = 2$, this inequality becomes,

$$\mathbb{E}|S|_N^{*2} \le 4 \cdot \mathbb{E}|S_N|^2 = 4 \cdot \sum_{n=1}^N \mathbb{E}|X_n|^2.$$

18.6 Submartingale Upcrossing Inequality and Convergence Theorems

The main results of this section are consequences of the following example and lemma which say that the optimal strategy for betting on a sub-martingale is to go "all in." Any other strategy, including buy low and sell high, will not fare better (on average) than going *all in.*

Example 18.49. Suppose that $\{X_n\}_{n=0}^\infty$ represents the value of a stock which is known to be a sub-martingale. At time $n-1$ you are allowed buy $C_n \in [0,1]$

shares of the stock which you will then sell at time n. Your net gain (loss) in this transaction is $C_n X_n - C_n X_{n-1} = C_n \Delta_n X$ and your wealth at time n will be

$$W_n = W_0 + \sum_{k=1}^{n} C_k \Delta_k X.$$

The next lemma asserts that the way to maximize your expected gain is to choose $C_k = 1$ for all k, i.e. buy the maximum amount of stock you can at each stage. We will refer to this as the **all in** strategy..

Lemma 18.50 ("All In"). *If $\{X_n\}_{n=0}^{\infty}$ is a sub-martingale and $\{C_k\}_{k=1}^{\infty}$ is a previsible process with values in $[0,1]$, then*

$$\mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right) \leq \mathbb{E}\left[X_n - X_0\right]$$

with equality when $C_k = 1$ for all k, i.e. the optimal strategy is to go all in.

Proof. Notice that $\{1 - C_k\}_{k=1}^{\infty}$ is a previsible non-negative process and therefore by Proposition 18.34,

$$\mathbb{E}\left(\sum_{k=1}^{n} \left(1 - C_k\right) \Delta_k X\right) \geq 0.$$

Since

$$X_n - X_0 = \sum_{k=1}^{n} \Delta_k X = \sum_{k=1}^{n} C_k \Delta_k X + \sum_{k=1}^{n} \left(1 - C_k\right) \Delta_k X,$$

it follows that

$$\mathbb{E}\left[X_n - X_0\right] = \mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right) + \mathbb{E}\left(\sum_{k=1}^{n} \left(1 - C_k\right) \Delta_k X\right) \geq \mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right).$$

■

We are now going to apply Lemma 18.50 to the time honored gambling strategy of buying low and selling high in order to prove the important "upcrossing" inequality of Doob, see Theorem 18.51. To be more precise, suppose that $\{X_n\}_{n=0}^{\infty}$ is a sub-martingale representing a stock price and $-\infty < a < b < \infty$ are given numbers. The (sub-optimal) strategy we wish to employ is to buy the stock when it first drops below a and then sell the first time it rises above b and then repeat this strategy over and over again.

Given a function, $\mathbb{N}_0 \ni n \to X_n \in \mathbb{R}$ and $-\infty < a < b < \infty$, let

$$\begin{aligned}
\tau_0 = 0,\ \tau_1 &= \inf\{n \geq \tau_0 : X_n \leq a\} \\
\tau_2 = \inf\{n \geq \tau_1 : X_n \geq b\},\ \tau_3 &:= \inf\{n \geq \tau_2 : X_n \leq a\} \\
&\vdots \\
\tau_{2k} = \inf\{n \geq \tau_{2k-1} : X_n \geq b\},\ \tau_{2k+1} &:= \inf\{n \geq \tau_{2k} : X_n \leq a\} \qquad (18.39) \\
&\vdots
\end{aligned}$$

with the usual convention that $\inf \emptyset = \infty$ in the definitions above, see Figures 18.2 and 18.3.

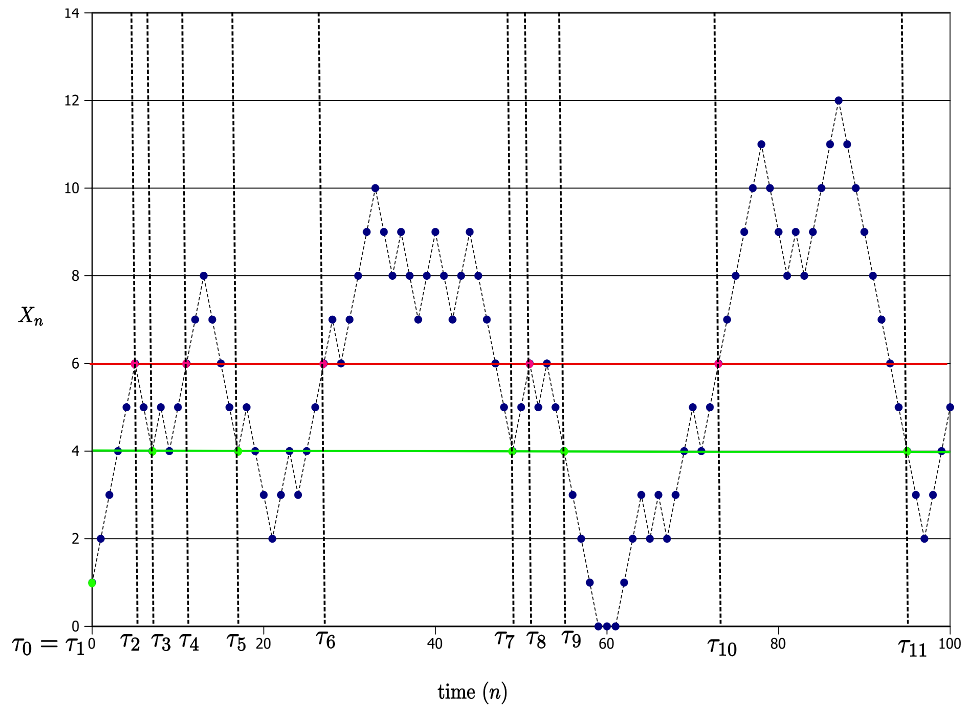


Fig. 18.2. A sample path or the positive part of a random walk with level crossing of $a = 1$ and $b = 2$ being marked off.

In terms of these stopping time our betting strategy may be describe as,

$$C_n = \sum_{k=1}^{\infty} 1_{\tau_{2k-1} < n \leq \tau_{2k}} \text{ for } n \in \mathbb{N}, \qquad (18.40)$$

see Figure 18.3 for a more intuitive description of $\{C_n\}_{n=1}^\infty$.

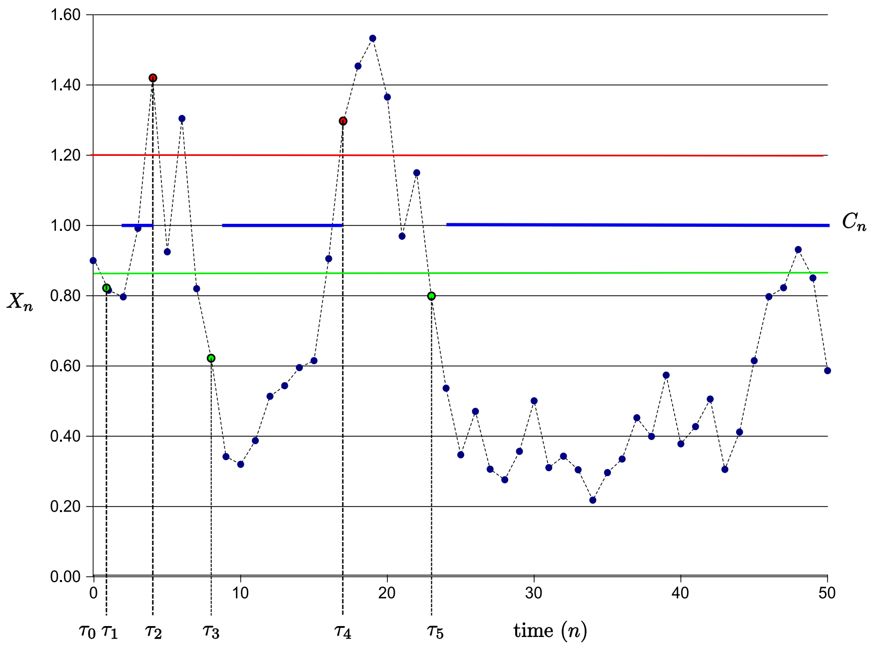


Fig. 18.3. In this figure we are taking $a = 0.85$ and $b = 1.20$. There are two upcrossings and we imaging buying below0.85 and selling above 1.20. The graph of C_n is given in blue in the above figure.

Observe that $\tau_{n+1} \geq \tau_n + 1$ for all $n \geq 1$ and hence $\tau_n \geq n - 1$ for all $n \geq 1$. Further, for each $N \in \bar{\mathbb{N}}$ let

$$U_N^X(a,b) = \max\{k : \tau_{2k} \leq N\} \tag{18.41}$$

be the **number of upcrossings of** X **across** $[a,b]$ in the time interval, $[0,N]$.

In Figure 18.3 you will notice that there are two upcrossings and at the end we are holding a stock for a loss of no more than $(a - X_N)_+$. In this example $X_0 = 0.90$ and we do not purchase a stock until time 1, i.e. $C_n = 1$ for the first time at $n = 2$. On the other hand if $X_0 < a$, then on the **first** upcrossing we would be guaranteed to make at least

$$b - X_0 = b - a + a - X_0 = b - a + (a - X_0)_+.$$

With these observations in mind, if there is at least one upcrossing, then

$$W_N := \sum_{k=1}^N C_k \Delta_k X \geq (b-a)\, U_N^X(a,b) + (a - X_0)_+ - (a - X_N)_+ \tag{18.42}$$

$$= (b-a)\, U_N^X(a,b) + (X_0 - a)_- - (X_N - a)_-. \tag{18.43}$$

In words the inequality in Eq. (18.43) states that our net gain in buying at or below a and selling at or above b is at least equal to $(b-a)$ times the number of times we buy low and sell high plus a possible bonus for buying below a at time 0 and a penalty for holding the stock below a at the end of the day. The key inequality in Eq. (18.43) may also be verified when no upcrossings occur. Here are the three case to consider.

1. If $X_n > a$ for all $0 \leq n \leq N$, then $C_n = 0$ for all n so $W_N = 0$ while $(X_0 - a)_- - (X_N - a)_- = 0 - 0 = 0$ as well.
2. If $X_0 \leq a$ and $X_n < b$ for all $0 \leq n \leq N$, then $C_n = 1$ for all n so that
$$\begin{aligned} W_N = X_N - X_0 &= (X_N - a) - (X_0 - a) \\ &= (X_N - a) + (X_0 - a)_- \geq -(X_N - a)_- + (X_0 - a)_-. \end{aligned}$$
3. If $X_0 > a$, but $\tau_1 \leq N$ and $X_n < b$ for all $0 \leq n \leq N$, then
$$W_N = X_N - X_{\tau_1} \geq X_N - a \geq -(X_N - a)_- = -(X_N - a)_- + (X_0 - a)_-.$$

Theorem 18.51 (Doob's Upcrossing Inequality). *If $\{X_n\}_{n=0}^\infty$ is a submartingale and $-\infty < a < b < \infty$, then for all $N \in \mathbb{N}$,*

$$\mathbb{E}\left[U_N^X(a,b)\right] \leq \frac{1}{b-a}\left[\mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+\right].$$

Proof. First Proof. Let $\{C_k\}_{k=1}^\infty$ be the buy low sell high strategy defined in Eq. (18.40). Taking expectations of the inequality in Eq. (18.43) making use of Lemma 18.50 implies,

$$\begin{aligned} \mathbb{E}[X_N - a - (X_0 - a)] &= \mathbb{E}[X_N - X_0] \geq \mathbb{E}[(C \cdot \Delta X)_N] \\ &\geq (b-a)\,\mathbb{E}U_N^X(a,b) + \mathbb{E}(X_0 - a)_- - \mathbb{E}(X_N - a)_-. \end{aligned}$$

The result follows from this inequality and the fact that $(X_n - a) = (X_n - a)_+ - (X_n - a)_-$. ■

*Remark 18.52 (*Second Proof).* Here is a variant on the above proof which may safely be skipped. We first suppose that $X_n \geq 0$, $a = 0$ and $b > 0$. Let

$$\begin{aligned} \tau_0 = 0,\ \tau_1 &= \inf\{n \geq \tau_0 : X_n = 0\} \\ \tau_2 = \inf\{n \geq \tau_1 : X_n \geq b\},\ \tau_3 &:= \inf\{n \geq \tau_2 : X_n = 0\} \\ &\vdots \\ \tau_{2k} = \inf\{n \geq \tau_{2k-1} : X_n \geq b\},\ \tau_{2k+1} &:= \inf\{n \geq \tau_{2k} : X_n = 0\} \\ &\vdots \end{aligned}$$

a sequence of stopping times. Suppose that N is given and we choose k such that $2k > N$. Then we know that $\tau_{2k} \geq N$. Thus if we let $\tau_n' := \tau_n \wedge N$, we know that $\tau_n' = N$ for all $n \geq 2k$. Therefore,

$$\begin{aligned}X_N - X_0 &= \sum_{n=1}^{2k} \left(X_{\tau_n'} - X_{\tau_{n-1}'}\right) \\ &= \sum_{n=1}^{k} \left(X_{\tau_{2n}'} - X_{\tau_{2n-1}'}\right) + \sum_{n=1}^{k} \left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'}\right) \\ &\geq bU_N^X(0,b) + \sum_{n=1}^{k} \left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'}\right), \end{aligned} \tag{18.44}$$

wherein we have used $X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \geq b$ if there were an upcrossing in the interval $\left[\tau_{2n-1}', \tau_{2n}'\right]$ and $X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \geq 0$ otherwise,[4] see Figure 18.4. Taking expectations of Eq. (18.44) implies

$$\mathbb{E}X_N - \mathbb{E}X_0 \geq b\mathbb{E}U_N^X(0,b) + \sum_{n=1}^{k} \mathbb{E}\left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'}\right) \geq b\mathbb{E}U_N^X(0,b)$$

wherein we have used the optional sampling theorem to guarantee,

$$\mathbb{E}\left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'}\right) \geq 0.$$

If X is a general submartingale and $-\infty < a < b < \infty$, we know by cJensen's inequality (with $\varphi(x) = (x-a)_+$ which is convex and increasing) that $(X_n - a)_+$ is still a sub-martingale and moreover

$$U_N^X(a,b) = U^{(X-a)_+}(0, b-a)$$

and therefore

$$\begin{aligned}(b-a)\,\mathbb{E}\left[U_N^X(a,b)\right] &= (b-a)\,\mathbb{E}\left[U^{(X-a)_+}(0,b-a)\right] \\ &\leq \mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+ .\end{aligned}$$

The second proof is now complete, nevertheless it is worth contemplating a bit how is that $\mathbb{E}\left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'}\right) \geq 0$ given that are strategy being employed is now to buy high and sell low. On $\{\tau_{2n-1} \leq N\}$, $X_{\tau_{2n-1}} - X_{\tau_{2n-2}} \leq 0 - b = -b$ and therefore,

[4] If $\tau_{2n-1} \geq N$, then $X_{\tau_{2n}'} - X_{\tau_{2n-1}'} = X_N - X_N = 0$, while if $\tau_{2n-1} < N$, $X_{\tau_{2n}'} - X_{\tau_{2n-1}'} = X_{\tau_{2n}'} - 0 \geq 0$.

$$\begin{aligned}0 \leq& \mathbb{E}\left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'}\right) \\ &= \mathbb{E}\left(X_{\tau_{2n-1}} - X_{\tau_{2n-2}} : \tau_{2n-1} \leq N\right) + \mathbb{E}\left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'} : \tau_{2n-1} > N\right) \\ &\leq -bP(\tau_{2n-1} \leq N) + \mathbb{E}\left(X_N - X_{\tau_{2n-2}'} : \tau_{2n-1} > N\right).\end{aligned}$$

Therefore we must have

$$\mathbb{E}\left(X_N - X_{\tau_{2n-2} \wedge N} : \tau_{2n-1} > N\right) \geq bP(\tau_{2n-1} \leq N)$$

so that X_N must be sufficiently large sufficiently often on the set where $\tau_{2n-1} > N$.

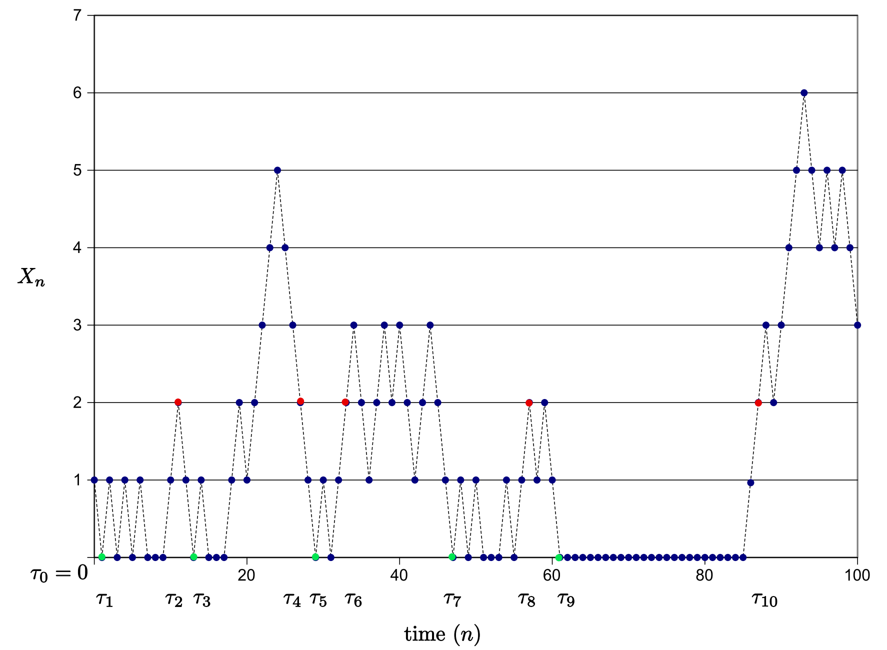


Fig. 18.4. A sample path of a positive submartingale along with stopping times τ_{2j-1} and τ_{2j} which are the successive hitting times of 0 and 2 respectively. If we take $N = 70$ in this case, then observe that Notice that $X_{\tau_8 \wedge 70} - X_{\tau_7 \wedge 70} \geq 2$ while $X_{\tau_{10} \wedge 70} - X_{\tau_9 \wedge 70} = 0$

Lemma 18.53. *Suppose $X = \{X_n\}_{n=0}^\infty$ is a sequence of extended real numbers such that $U_\infty^X(a,b) < \infty$ for all $a, b \in \mathbb{Q}$ with $a < b$. Then $X_\infty := \lim_{n\to\infty} X_n$ exists in $\bar{\mathbb{R}}$.*

Proof. If $\lim_{n\to\infty} X_n$ does not exists in $\bar{\mathbb{R}}$, then there would exists $a, b \in \mathbb{Q}$ such that

$$\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n$$

and for this choice of a and b, we must have $X_n < a$ and $X_n > b$ infinitely often. Therefore, $U_\infty^X(a,b) = \infty$. ∎

Corollary 18.54. *Suppose $\{X_n\}_{n=0}^\infty$ is an integrable submartingale such that $\sup_n \mathbb{E}X_n^+ < \infty$ (or equivalently $C := \sup_n \mathbb{E}|X_n| < \infty$, see Remark 18.19), then $X_\infty := \lim_{n\to\infty} X_n$ exists in $\mathbb{R}$ a.s. and $X_\infty \in L^1(\Omega, \mathcal{B}, P)$. Moreover $\{X_n\}_{n\in\bar{\mathbb{N}}_0}$ is a submartingale (that is we also have $X_n \le \mathbb{E}[X_\infty|\mathcal{B}_n]$ a.s. for all n), iff $\{X_n^+\}_{n=1}^\infty$ is uniformly integrable.*

Proof. For any $-\infty < a < b < \infty$, by Doob's upcrossing inequality (Theorem 18.51) and the MCT,

$$\mathbb{E}\left[U_\infty^X(a,b)\right] \le \frac{1}{b-a}\left[\sup_N \mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0-a)_+\right] < \infty$$

where

$$U_\infty^X(a,b) := \lim_{N\to\infty} U_N^X(a,b)$$

is the total number of upcrossings of X across $[a,b]$.[5] In particular it follows that

$$\Omega_0 := \cap\left\{U_\infty^X(a,b) < \infty : a,b\in\mathbb{Q} \text{ with } a<b\right\}$$

has probability one. Hence by Lemma 18.53, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists in $\bar{\mathbb{R}}$. By Fatou's lemma we know that

$$\mathbb{E}[|X_\infty|] = \mathbb{E}\left[\liminf_{n\to\infty}|X_n|\right] \le \liminf_{n\to\infty}\mathbb{E}[|X_n|] \le C < \infty \tag{18.45}$$

and therefore that $X_\infty \in \mathbb{R}$ a.s.

Since (as we have already shown) $X_n^+ \to X_\infty^+$ a.s., if $\{X_n^+\}_{n=1}^\infty$ is uniformly integrable, then $X_n^+ \to X_\infty^+$ in $L^1(P)$ by Vitalli's convergence Theorem 12.44. Therefore for $A \in \mathcal{B}_n$ we have by Fatou's lemma that

$$\begin{aligned}\mathbb{E}[X_n 1_A] &\le \limsup_{m\to\infty}\mathbb{E}[X_m 1_A] = \limsup_{m\to\infty}\left(\mathbb{E}\left[X_m^+1_A\right] - \mathbb{E}\left[X_m^- 1_A\right]\right)\\ &= \mathbb{E}\left[X_\infty^+ 1_A\right] - \liminf_{m\to\infty}\mathbb{E}\left[X_m^- 1_A\right] \le \mathbb{E}\left[X_\infty^+1_A\right] - \mathbb{E}\left[\liminf_{m\to\infty} X_m^- 1_A\right]\\ &= \mathbb{E}\left[X_\infty^+1_A\right] - \mathbb{E}\left[X_\infty^- 1_A\right] = \mathbb{E}[X_\infty 1_A].\end{aligned}$$

Since $A \in \mathcal{B}_n$ was arbitrary we may conclude that $X_n \le \mathbb{E}[X_\infty|\mathcal{B}_n]$ a.s. for n.

Conversely if we suppose that $X_n \le \mathbb{E}[X_\infty|\mathcal{B}_n]$ a.s. for n, then by cJensen's inequality (with $\varphi(x) = x \vee 0$ being an increaing convex function),

$$X_n^+ \le (\mathbb{E}[X_\infty|\mathcal{B}_n])^+ \le \mathbb{E}\left[X_\infty^+|\mathcal{B}_n\right] \text{ a.s. for all } n$$

and therefore $\{X_n^+\}_{n=1}^\infty$ is uniformly integrable by Proposition 18.8 and Exercise 12.5.

Second Proof. We may also give another proof of the first assertion based on the Krickeberg decomposition Theorem 18.20 and the supermartingale convergence Corollary 18.63 below. Indeed, by the Krickeberg decomposition Theorem 18.20, $X_n = M_n - Y_n$ where M is a positive martingale and Y is a positive supermartingale. Hence by two applications of Corollary 18.63 we may conclude that

$$X_\infty = \lim_{n\to\infty} X_n = \lim_{n\to\infty} M_n - \lim_{n\to\infty} Y_n$$

exists in $\mathbb{R}$ almost surely. ∎

Remark 18.55. If $\{X_n\}_{n=0}^\infty$ is a submartingale such that $\{X_n^+\}_{n=0}^\infty$ is uniformly integrable, it does not necessarily follows that $\{X_n\}_{n=0}^\infty$ is uniformly integrable. Indeed, let $X_n = -M_n$ where M_n is the non-uniformly integrable martingale in Example 18.7. Then X_n is a negative (sub)martingale and hence $X_n^+ \equiv 0$ is uniformly integrable but $\{X_n\}_{n=0}^\infty$ is **not** uniformly integrable. This also shows that assuming the positive part of a martingale is uniformly integrable is not sufficient to show the martingale itself is uniformly integrable. Keep in mind in this example that $\lim_{n\to\infty} X_n = 0$ a.s. while $\mathbb{E}X_n = 1$ for all n and so clearly $\lim_{n\to\infty}\mathbb{E}X_n = 1 \ne 0 = \mathbb{E}[\lim_{n\to\infty} X_n]$ in this case.

Notation 18.56 *Given a probability space, $(\Omega, \mathcal{B}, P)$ and $A, B \in \mathcal{B}$, we say $A = B$ a.s. iff $P(A \triangle B) = 0$ or equivalently iff $1_A = 1_B$ a.s.*

Corollary 18.57 (Localizing Corollary 18.54). *Suppose $M = \{M_n\}_{n=0}^\infty$ is a martingale and $c < \infty$ such that $\Delta_n M \le c$ a.s. for all n. Then*

$$\left\{\lim_{n\to\infty} M_n \text{ exists in } \mathbb{R}\right\} = \left\{\sup_n M_n < \infty\right\} \text{ a.s.}$$

Proof. Let $\tau_a := \inf\{n : M_n \ge a\}$ for all $a \in \mathbb{N}$. Then by the optional stopping theorem, $n \to M_n^{\tau_a}$ is still a martingale. Since $M_n^{\tau_a} \le a + c$,[6] it follows that $\mathbb{E}(M_n^{\tau_a})_+ \le a + c < \infty$ for all n. Hence we may apply Corollary 18.54 to conclude, $\lim_{n\to\infty} M_n^{\tau_a} = M_\infty^{\tau_a}$ exists in $\mathbb{R}$ almost surely. Therefore $n \to M_n$ is convergent in $\mathbb{R}$ almost surely on the set

$$\cup_a\{M^{\tau_a} = M\} = \left\{\sup_n M_n < \infty\right\}.$$

Conversely if $n \to M_n$ is convergent in $\mathbb{R}$, then $\sup_n M_n < \infty$. ∎

[5] Notice that $(X_N - a)_+ \le |X_N - a| \le |X_N| + a$ so that $\sup_N \mathbb{E}(X_N - a)_+ \le C + a < \infty$.

[6] If $n < \tau_a$ then $M_n < a$ and if $n \ge \tau_a$ then $M_n^{\tau_a} = M_{\tau_a} \le M_{\tau_a - 1} + c < a + c$.

Corollary 18.58. *Suppose $M = \{M_n\}_{n=0}^\infty$ is a martingale, and $c < \infty$ such that $|\Delta_n M| \le c$ a.s. for all n. Let*

$$C := \left\{ \lim_{n\to\infty} M_n \text{ exists in } \mathbb{R} \right\} \text{ and}$$
$$D := \left\{ \limsup_{n\to\infty} M_n = \infty \text{ and } \liminf_{n\to\infty} M_n = -\infty \right\}.$$

Then, $P(C \cup D) = 1$. (In words, either $\lim_{n\to\infty} M_n$ exists in $\mathbb{R}$ or $\{M_n\}_{n=1}^\infty$ is "wildly" oscillating as $n \to \infty$.)

Proof. Since both M and $-M$ satisfy the hypothesis of Corollary 18.57, we may conclude that (almost surely),

$$C = \left\{ \sup_n M_n < \infty \right\} = \left\{ \inf_n M_n > -\infty \right\} \text{ a.s.}$$

and hence almost surely,

$$\begin{aligned} C^c &= \left\{ \sup_n M_n = \infty \right\} = \left\{ \inf_n M_n = -\infty \right\} \\ &= \left\{ \sup_n M_n = \infty \right\} \cap \left\{ \inf_n M_n = -\infty \right\} = D. \end{aligned}$$

■

Corollary 18.59. *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and $A_n \in \mathcal{B}_n$ for all n. Then*

$$\left\{ \sum_n 1_{A_n} = \infty \right\} = \{A_n \text{ i.o.}\} = \left\{ \sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty \right\} \text{ a.s.} \tag{18.46}$$

Proof. Let $\Delta_n M := 1_{A_n} - \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}]$ so that $\mathbb{E}[\Delta_n M|\mathcal{B}_{n-1}] = 0$ for all n. Thus if

$$M_n := \sum_{k\le n} \Delta_n M = \sum_{k\le n} \left(1_{A_n} - \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}]\right),$$

then M is a martingale with $|\Delta_n M| \le 1$ for all n. Let C and C be as in Corollary 18.58. Since $\{A_n \text{ i.o.}\} = \{\sum_n 1_{A_n} = \infty\}$, it follows that

$$\{A_n \text{ i.o.}\} = \left\{ \sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty \right\} \text{ a.s. on } C.$$

Moreover, on $\{\sup_n M_n = \infty\}$ we must have $\sum_n 1_{A_n} = \infty$ and on $\{\inf_n M_n = -\infty\}$ that $\sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty$ and so

$$\sum_n 1_{A_n} = \infty \text{ and } \sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty \text{ a.s. on } D.$$

Thus it follows that Eq. (18.46) holds on $C \cup D$ a.s. which completes the proof since $\Omega = C \cup D$ a.s.. ■

See Durrett [14, Chapter 4.3] for more in this direction.

18.7 *Supermartingale inequalities

As the optional sampling theorem was our basic tool for deriving submartingale inequalities, the following optional switching lemma will be our basic tool for deriving positive supermartingale inequalities.

Lemma 18.60 (Optional switching lemma). *Suppose that X and Y are two supermartingales and τ is a stopping time such that $X_\tau \ge Y_\tau$ on $\{\tau < \infty\}$. Then*

$$Z_n = 1_{n<\tau} X_n + 1_{n\ge\tau} Y_n = \begin{cases} X_n & \text{if } n < \tau \\ Y_n & \text{if } n \ge \tau \end{cases}$$

is again a supermartingale. (In short we can switch from X to Y at time, τ, provided $Y \le X$ at the switching time, τ.) This lemma is valid if $X_n, Y_n \in L^1(\Omega, \mathcal{B}_n, P)$ for all n or if both $X_n, Y_n \ge 0$ for all n. In the latter case, we should be using the extended notion of conditional expectations.

Proof. We begin by observing,

$$\begin{aligned} Z_{n+1} &= 1_{n+1<\tau} X_{n+1} + 1_{n+1\ge\tau} Y_{n+1} \\ &= 1_{n+1<\tau} X_{n+1} + 1_{n\ge\tau} Y_{n+1} + 1_{\tau=n+1} Y_{n+1} \\ &\le 1_{n+1<\tau} X_{n+1} + 1_{n\ge\tau} Y_{n+1} + 1_{\tau=n+1} X_{n+1} \\ &= 1_{n<\tau} X_{n+1} + 1_{n\ge\tau} Y_{n+1}. \end{aligned}$$

Since $\{n < \tau\}$ and $\{n \ge \tau\}$ are $\mathcal{B}_n$ – measurable, it now follows from the supermartingale property of X and Y that

$$\begin{aligned} \mathbb{E}_{\mathcal{B}_n} Z_{n+1} &\le \mathbb{E}_{\mathcal{B}_n}\left[1_{n<\tau} X_{n+1} + 1_{n\ge\tau} Y_{n+1}\right] \\ &= 1_{n<\tau} \mathbb{E}_{\mathcal{B}_n}[X_{n+1}] + 1_{n\ge\tau} \mathbb{E}_{\mathcal{B}_n}[Y_{n+1}] \\ &\le 1_{n<\tau} X_n + 1_{n\ge\tau} Y_n = Z_n. \end{aligned}$$

■

18.7.1 Maximal Inequalities

Theorem 18.61 (Supermartingale maximal inequality). *Let X be a positive supermartingale (in the extended sense) and $a \in \mathcal{B}_0$ with $a \geq 0$, then*

$$aP\left[\sup_n X_n \geq a | \mathcal{B}_0\right] \leq a \wedge X_0 \tag{18.47}$$

and moreover

$$P\left[\sup_n X_n = \infty | \mathcal{B}_0\right] = 0 \text{ on } \{X_0 < \infty\}. \tag{18.48}$$

In particular if $X_0 < \infty$ a.s. then $\sup_n X_n < \infty$ a.s.

Proof. Proof. Simply apply Corollary 18.44 with $Y_n = ((2a) \wedge X_n) \cdot 1_A$ where $A \in \mathcal{B}_0$ to find

$$a\mathbb{E}\left(P\left[\sup_n X_n \geq a | \mathcal{B}_0\right] : A\right) = aP\left[\sup_n X_n \geq a : A\right] \leq \mathbb{E}\left[a \wedge X_0 : A\right].$$

Since this holds for all $A \in \mathcal{B}_0$, Eq. (18.47) follows.

Second Proof. Let $\tau := \inf\{n : X_n \geq a\}$ which is a stopping time since,

$$\{\tau \leq n\} = \{X_n \geq a\} \in \mathcal{B}_n \text{ for all } n.$$

Since $X_\tau \geq a$ on $\{\tau < \infty\}$ and $Y_n := a$ is a supermartingale, it follows by the switching Lemma 18.60 that

$$Z_n := 1_{n<\tau} X_n + a 1_{n \geq \tau}$$

is a supermartingale (in the extended sense). In particular it follows

$$aP(\tau \leq n | \mathcal{B}_0) = \mathbb{E}_{\mathcal{B}_0}\left[a 1_{n \geq \tau}\right] \leq \mathbb{E}_{\mathcal{B}_0} Z_n \leq Z_0,$$

and

$$Z_0 = 1_{0<\tau}\ X_0 + a 1_{\tau=0} = 1_{X_0<a} X_0 + 1_{X_0 \geq a} a = a \wedge X_0.$$

Therefore, using the cMCT,

$$\begin{aligned} aP\left[\sup_n X_n \geq a | \mathcal{B}_0\right] &= aP\left[\tau < \infty | \mathcal{B}_0\right] = \lim_{n\to\infty} aP(\tau \leq n | \mathcal{B}_0) \\ &\leq Z_0 = a \wedge X_0 \end{aligned}$$

which proves Eq. (18.47).

For the last assertion, take $a > 0$ to be constant in Eq. (18.47) and then use the cDCT to let $a \uparrow \infty$ to conclude

$$P\left[\sup_n X_n = \infty | \mathcal{B}_0\right] = \lim_{a\uparrow\infty} P\left[\sup_n X_n \geq a | \mathcal{B}_0\right] \leq \lim_{a\uparrow\infty} 1 \wedge \frac{X_0}{a} = 1_{X_0=\infty}.$$

Multiplying this equation by $1_{X_0<\infty}$ and then taking expectations implies

$$\mathbb{E}\left[1_{\sup_n X_n=\infty} 1_{X_0<\infty}\right] = \mathbb{E}\left[1_{X_0=\infty} 1_{X_0<\infty}\right] = 0$$

which implies $1_{\sup_n X_n=\infty} 1_{X_0<\infty} = 0$ a.s., i.e. $\sup_n X_n < \infty$ a.s. on $\{X_0 < \infty\}$. ∎

18.7.2 The upcrossing inequality and convergence result

Theorem 18.62 (Dubin's Upcrossing Inequality). *Suppose $X = \{X_n\}_{n=0}^\infty$ is a positive supermartingale and $0 < a < b < \infty$. Then*

$$P\left(U_\infty^X(a,b) \geq k | \mathcal{B}_0\right) \leq \left(\frac{a}{b}\right)^k \left(1 \wedge \frac{X_0}{a}\right), \text{ for } k \geq 1 \tag{18.49}$$

and $U_\infty(a,b) < \infty$ a.s. and in fact

$$\mathbb{E}\left[U_\infty^X(a,b)\right] \leq \frac{1}{b/a - 1} = \frac{a}{b-a} < \infty.$$

Proof. Since

$$U_N^X(a,b) = U_N^{X/a}(1, b/a),$$

it suffices to consider the case where $a = 1$ and $b > 1$. Let τ_n be the stopping times defined in Eq. (18.39) with $a = 1$ and $b > 1$, i.e.

$$\begin{aligned} \tau_0 = 0,\ &\tau_1 = \inf\{n \geq \tau_0 : X_n \leq 1\} \\ \tau_2 = \inf\{n \geq \tau_1 : X_n \geq b\},\ &\tau_3 := \inf\{n \geq \tau_2 : X_n \leq 1\} \\ &\vdots \\ \tau_{2k} = \inf\{n \geq \tau_{2k-1} : X_n \geq b\},\ &\tau_{2k+1} := \inf\{n \geq \tau_{2k} : X_n \leq 1\}, \\ &\vdots \end{aligned}$$

see Figure 18.2.

Let $k \geq 1$ and use the switching Lemma 18.60 repeatedly to define a new positive supermatingale $Y_n = Y_n^{(k)}$ (see Exercise 18.13 below) as follows,

$$\begin{aligned} Y_n^{(k)} &= 1_{n<\tau_1} + 1_{\tau_1 \leq n < \tau_2} X_n \\ &+ b 1_{\tau_2 \leq n < \tau_3} + b X_n 1_{\tau_3 \leq n < \tau_4} \\ &+ b^2 1_{\tau_4 \leq n < \tau_5} + b^2 X_n 1_{\tau_5 \leq n < \tau_6} \\ &\vdots \\ &+ b^{k-1} 1_{\tau_{2k-2} \leq n < \tau_{2k-1}} + b^{k-1} X_n 1_{\tau_{2k-1} \leq n < \tau_{2k}} \\ &+ b^k 1_{\tau_{2k} \leq n}. \end{aligned} \tag{18.50}$$

Since $\mathbb{E}[Y_n|\mathcal{B}_0] \le Y_0$ a.s., $Y_n \ge b^k 1_{\tau_{2k}\le n}$, and

$$Y_0 = 1_{0<\tau_1} + 1_{\tau_1=0}X_0 = 1_{X_0>1} + 1_{X_0\le 1}X_0 = 1 \wedge X_0,$$

we may infer that

$$b^k P(\tau_{2k} \le n|\mathcal{B}_0) = \mathbb{E}\left[b^k 1_{\tau_{2k}\le n}|\mathcal{B}_0\right] \le \mathbb{E}[Y_n|\mathcal{B}_0] \le 1 \wedge X_0 \text{ a.s.}$$

Using *cMCT*, we may now let $n \to \infty$ to conclude

$$P\left(U^X(1,b) \ge k|\mathcal{B}_0\right) \le P(\tau_{2k} < \infty|\mathcal{B}_0) \le \frac{1}{b^k}(1 \wedge X_0) \text{ a.s.}$$

which is Eq. (18.49). Using cDCT, we may let $k \uparrow \infty$ in this equation to discover $P\left(U_\infty^X(1,b) = \infty|\mathcal{B}_0\right) = 0$ a.s. and in particular, $U_\infty^X(1,b) < \infty$ a.s. In fact we have

$$\begin{aligned}\mathbb{E}\left[U_\infty^X(1,b)\right] &= \sum_{k=1}^\infty P\left(U_\infty^X(1,b) \ge k\right) \le \sum_{k=1}^\infty \mathbb{E}\left[\frac{1}{b^k}(1 \wedge X_0)\right] \\ &= \frac{1}{b}\frac{1}{1-1/b}\mathbb{E}[(1 \wedge X_0)] \le \frac{1}{b-1} < \infty.\end{aligned}$$

■

Exercise 18.13. In this exercise you are asked to fill in the details showing Y_n in Eq. (18.50) is still a supermartingale. To do this, define $Y_n^{(k)}$ via Eq. (18.50) and then show (making use of the switching Lemma 18.60 twice) $Y_n^{(k+1)}$ is a supermartingale under the assumption that $Y_n^{(k)}$ is a supermartingale. Finish off the induction argument by observing that the constant process, $U_n := 1$ and $V_n = 0$ are supermartingales such that $U_{\tau_1} = 1 \ge 0 = V_{\tau_1}$ on $\{\tau_1 < \infty\}$, and therefore by the switching Lemma 18.60,

$$Y_n^{(1)} = 1_{0\le n<\tau_1}U_n + 1_{\tau_1\le n}V_n = 1_{0\le n<\tau_1}$$

is also a supermartingale.

Corollary 18.63 (Positive Supermartingale convergence). *Suppose* $X = \{X_n\}_{n=0}^\infty$ *is a positive supermartingale (possibly in the extended sense), then* $X_\infty = \lim_{n\to\infty} X_n$ *exists a.s. and we have*

$$\mathbb{E}[X_\infty|\mathcal{B}_n] \le X_n \text{ for all } n \in \bar{\mathbb{N}}. \tag{18.51}$$

In particular,

$$\mathbb{E}X_\infty \le \mathbb{E}X_n \le \mathbb{E}X_0 \text{ for all } n < \infty. \tag{18.52}$$

Proof. The set,

$$\Omega_0 := \cap\left\{U_\infty^X(a,b) < \infty : a,b \in \mathbb{Q} \text{ with } a < b\right\},$$

has full measure ($P(\Omega_0) = 1$) by Dubin's upcrossing inequality in Theorem 18.62. So by Lemma 18.53, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists[7] in $[0,\infty]$. For definiteness, let $X_\infty = 0$ on Ω_0^c. Equation (18.51) is now a consequence of cFatou;

$$\mathbb{E}[X_\infty|\mathcal{B}_n] = \mathbb{E}\left[\lim_{m\to\infty} X_m|\mathcal{B}_n\right] \le \liminf_{m\to\infty} \mathbb{E}[X_m|\mathcal{B}_n] \le \liminf_{m\to\infty} X_n = X_n \text{ a.s.}$$

The supermartingale property guarantees that $\mathbb{E}X_n \le \mathbb{E}X_0$ for all $n < \infty$ while taking expectations of Eq. (18.51) implies $\mathbb{E}X_\infty \le \mathbb{E}X_n$. ■

Theorem 18.64 (Optional sampling II – Positive supermartingales). *Suppose that* $X = \{X_n\}_{n=0}^\infty$ *is a positive supermartingale,* $X_\infty := \lim_{n\to\infty} X_n$ *(which exists a.s. by Corollary 18.63), and* σ *and* τ *are* ***arbitrary*** *stopping times. Then* $X_n^\tau := X_{\tau\wedge n}$ *is a positive* $\{\mathcal{B}_n\}_{n=0}^\infty$ *– super martingale,* $X_\infty^\tau = \lim_{n\to\infty} X_{\tau\wedge n}^\tau$*, and*

$$\mathbb{E}[X_\tau|\mathcal{B}_\sigma] \le X_{\sigma\wedge\tau} \text{ a.s.} \tag{18.53}$$

Moreover, if $\mathbb{E}X_0 < \infty$*, then* $\mathbb{E}[X_\tau] = \mathbb{E}[X_\infty^\tau] < \infty$.

Proof. We already know that X^τ is a positive supermatingale by optional stopping Theorem 18.38. Hence an application of Corollary 18.63 implies that $\lim_{n\to\infty} X_n^\tau = \lim_{n\to\infty} X_{\tau\wedge n}$ is convergent and

$$\mathbb{E}\left[\lim_{n\to\infty} X_n^\tau|\mathcal{B}_m\right] \le X_m^\tau = X_{\tau\wedge m} \text{ for all } m < \infty. \tag{18.54}$$

On the set $\{\tau < \infty\}$, $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ and on the set $\{\tau = \infty\}$, $\lim_{n\to\infty} X_{\tau\wedge n} = \lim_{n\to\infty} X_n = X_\infty = X_\tau$ a.s. Therefore it follows that $\lim_{n\to\infty} X_n^\tau = X_\tau$ and Eq. (18.54) may be expressed as

$$\mathbb{E}[X_\tau|\mathcal{B}_m] \le X_{\tau\wedge m} \text{ for all } m < \infty. \tag{18.55}$$

An application of Lemma 18.30 now implies

$$\mathbb{E}[X_\tau|\mathcal{B}_\sigma] = \sum_{m\le\infty} 1_{\sigma=m}\mathbb{E}[X_\tau|\mathcal{B}_m] \le \sum_{m\le\infty} 1_{\sigma=m}X_{\tau\wedge m} = X_{\tau\wedge\sigma} \text{ a.s.}$$

■

[7] If $\mathbb{E}X_0 < \infty$, this may also be deduced by applying Corollary 18.54 to $\{-X_n\}_{n=0}^\infty$.

18.8 Martingale Closure and Regularity Results

We are now going to give a couple of theorems which have already been alluded to in Exercises 13.6, 14.8, and 14.9.

Theorem 18.65. *Let $M := \{M_n\}_{n=0}^{\infty}$ be an L^1 – bounded martingale, i.e. $C := \sup_n \mathbb{E}|M_n| < \infty$ and let $M_\infty := \lim_{n\to\infty} M_n$ which exists a.s. and satisfies, $\mathbb{E}|M_\infty| < \infty$ by Corollary 18.54. Then the following are equivalent;*

1. *There exists $X \in L^1(\Omega, \mathcal{B}, P)$ such that $M_n = \mathbb{E}[X|\mathcal{B}_n]$ for all n.*
2. *$\{M_n\}_{n=0}^{\infty}$ is uniformly integrable.*
3. *$M_n \to M_\infty$ in $L^1(\Omega, \mathcal{B}, P)$.*

Moreover, if any of the above equivalent conditions hold we may take $X = M_\infty$, i.e. $M_n = \mathbb{E}[M_\infty|\mathcal{B}_n]$.

Proof. 1. $\implies$ 2. This was already proved in Proposition 18.8.

2. $\implies$ 3. The knowledge that $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. along with the assumed uniform integrability implies L^1 – convergence by Vitali convergence Theorem 12.44.

3. $\implies$ 1. If $M_n \to M_\infty$ in $L^1(\Omega, \mathcal{B}, P)$, then by the martingale property and the $L^1(P)$ – continuity of conditional expectation we find,

$$M_n = \mathbb{E}[M_m|\mathcal{B}_n] \to \mathbb{E}[M_\infty|\mathcal{B}_n] \text{ as } m \to \infty,$$

and thus, $M_n = \mathbb{E}[M_\infty|\mathcal{B}_n]$ a.s. ∎

Definition 18.66. *A martingale satisfying any and all of the equivalent statements in Theorem 18.65 is said to be **regular.***

Theorem 18.67. *Suppose $1 < p < \infty$ and $M := \{M_n\}_{n=0}^{\infty}$ is an L^p – bounded martingale. Then $M_n \to M_\infty$ almost surely and in L^p. In particular, $\{M_n\}$ is a regular martingale.*

Proof. The almost sure convergence follows from Corollary 18.54. So, because of Corollary 12.47, to finish the proof it suffices to show $\{|M_n|^p\}_{n=0}^{\infty}$ is uniformly integrable. But by Doob's inequality, Corollary 18.47, and the MCT, we find

$$\mathbb{E}\left[\sup_k |M_k|^p\right] \le \left(\frac{p}{p-1}\right)^p \sup_k \mathbb{E}[|M_k|^p] < \infty.$$

As $|M_n|^p \le \sup_k |M_k|^p \in L^1(P)$ for all $n \in \mathbb{N}$, it follows by Example 12.39 and Exercise 12.5 that $\{|M_n|^p\}_{n=0}^{\infty}$ is uniformly integrable. ∎

Theorem 18.68 (Optional sampling III – regular martingales). *Suppose that $M = \{M_n\}_{n=0}^{\infty}$ is a regular martingale, σ and τ are **arbitrary** stopping times. Define $M_\infty := \lim_{n\to\infty} M_n$ which exists a.s.. Then $M_\infty \in L^1(P)$,*

$$M_\tau = \mathbb{E}[M_\infty|\mathcal{B}_\tau], \quad \mathbb{E}|M_\tau| \le \mathbb{E}|M_\infty| < \infty \tag{18.56}$$

and

$$\mathbb{E}[M_\tau|\mathcal{B}_\sigma] = M_{\sigma\wedge\tau} \text{ a.s.} \tag{18.57}$$

Proof. By Theorem 18.65, $M_\infty \in L^1(\Omega, \mathcal{B}, P)$ and $M_n := \mathbb{E}_{\mathcal{B}_n} M_\infty$ a.s. for all $n \le \infty$. By Lemma 18.30,

$$\mathbb{E}_{\mathcal{B}_\tau} M_\infty = \sum_{n\le\infty} 1_{\tau=n} \mathbb{E}_{\mathcal{B}_n} M_\infty = \sum_{n\le\infty} 1_{\tau=n} M_n = M_\tau.$$

Hence we have $|M_\tau| = |\mathbb{E}_{\mathcal{B}_\tau} M_\infty| \le \mathbb{E}_{\mathcal{B}_\tau}|M_\infty|$ a.s. and $\mathbb{E}|M_\tau| \le \mathbb{E}|M_\infty| < \infty$. An application of Theorem 18.31 now concludes the proof;

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = \mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau} M_\infty = \mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}} M_\infty = M_{\sigma\wedge\tau}.$$

∎

Definition 18.69. *Let $M = \{M_n\}_{n=0}^{\infty}$ be a martingale. We say that τ is a **regular stopping time for** M if M^τ is a regular martingale.*

Example 18.70. Every bounded martingale is regular. More generally if τ is a stopping time such that M^τ is bounded, then τ is a regular stopping time for M.

Remark 18.71. If τ is regular for M, then $\lim_{n\to\infty} M_n^\tau := M_\infty^\tau$ exists a.s. and in $L^1(P)$ and hence

$$\lim_{n\to\infty} M_n = M_\infty^\tau \text{ a.s. on } \{\tau = \infty\}. \tag{18.58}$$

Thus if τ is regular for M, we may define M_τ as, $M_\tau := M_\infty^\tau = \lim_{n\to\infty} M_{n\wedge\tau}$ and we will have

$$M_{\tau\wedge n} = M_n^\tau = \mathbb{E}_{\mathcal{B}_n} M_\tau \tag{18.59}$$

and

$$\mathbb{E}|M_\tau| = \lim_{n\to\infty} \mathbb{E}|M_n^\tau| \le \sup_n \mathbb{E}|M_n^\tau| < \infty.$$

Theorem 18.72. *Suppose $M = \{M_n\}_{n=0}^{\infty}$ is a martingale and σ, τ, are stopping times such that τ is a regular stopping time for M. Then*

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = M_{\tau\wedge\sigma}, \tag{18.60}$$

and if $\sigma \le \tau$ a.s. then

$$M_n^\sigma = \mathbb{E}_{\mathcal{B}_n}[\mathbb{E}_{\mathcal{B}_\sigma} M_\tau] \tag{18.61}$$

and σ is regular for M.

Proof. By assumption, $M_\tau = \lim_{n\to\infty} M_{n\wedge\tau}$ exists almost surely and in $L^1(P)$ and $M_n^\tau = \mathbb{E}[M_\tau|\mathcal{B}_n]$ for $n \le \infty$.

1. Equation (18.60) is a consequence of;

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = \sum_{n\le\infty} 1_{\sigma=n} \mathbb{E}_{\mathcal{B}_n} M_\tau = \sum_{n\le\infty} 1_{\sigma=n} M_n^\tau = M_{\sigma\wedge\tau}.$$

2. Applying $\mathbb{E}_{\mathcal{B}_\sigma}$ to Eq. (18.59) using the optional sampling Theorem 18.39 and the tower property of conditional expectation (see Theorem 18.31) shows

$$M_n^\sigma = M_{\sigma\wedge n} = \mathbb{E}_{\mathcal{B}_\sigma} M_{\tau\wedge n} = \mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_n} M_\tau = \mathbb{E}_{\mathcal{B}_n}\left[\mathbb{E}_{\mathcal{B}_\sigma} M_\tau\right].$$

The regularity of M^σ now follows by item 1. of Theorem 18.65. ■

Proposition 18.73. *Suppose that M is a martingale and τ is a stopping time. Then the τ is regular for M iff;*

1. $\mathbb{E}[|M_\tau| : \tau < \infty] < \infty$ *and*
2. $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ *is a uniformly integrable sequence of random variables.*

Moreover, condition 1. is automatically satisfied if M is L^1 – bounded, i.e. if $C := \sup_n \mathbb{E}|M_n| < \infty$.

Proof. ($\Longrightarrow$) If τ is regular for M, $M_\tau \in L^1(P)$ and $M_n^\tau = \mathbb{E}_{\mathcal{B}_n} M_\tau$ so that $M_n = \mathbb{E}_{\mathcal{B}_n} M_\tau$ a.s. on $\{n \le \tau\}$. In particular it follows that

$$\mathbb{E}[|M_\tau| : \tau < \infty] \le \mathbb{E}|M_\tau| < \infty$$

and

$$|M_n 1_{n<\tau}| = |\mathbb{E}_{\mathcal{B}_n} M_\tau 1_{n<\tau}| \le \mathbb{E}_{\mathcal{B}_n}|M_\tau| \text{ a.s.}$$

from which it follows that $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is uniformly integrable.

($\Longleftarrow$) Our goal is to show $\{M_n^\tau\}_{n=0}^\infty$ is uniformly integrable. We begin with the identity;

$$\begin{aligned}\mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a] =& \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ \tau \le n] \\ &+ \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ n < \tau].\end{aligned}$$

Since

$$\begin{aligned}\mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ \tau \le n] &= \mathbb{E}[|M_\tau| : |M_\tau| \ge a,\ \tau \le n] \\ &\le \mathbb{E}[|M_\tau 1_{\tau<\infty}| : |M_\tau 1_{\tau<\infty}| \ge a],\end{aligned}$$

if follows (by assumption 1. that $\mathbb{E}[|M_\tau 1_{\tau<\infty}|] < \infty$) that

$$\lim_{a\to\infty}\sup_n \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ \tau \le n] = 0.$$

Moreover for any $a > 0$,

$$\sup_n \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ n < \tau] = \sup_n \mathbb{E}[|M_n^\tau 1_{n<\tau}| : |M_n^\tau 1_{n<\tau}| \ge a]$$

at the latter term goes to zero as $a \to \infty$ by assumption 2. Hence we have shown,

$$\lim_{a\to\infty}\sup_n \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a] = 0$$

as desired.

Now to prove the last assertion. If $C := \sup_n \mathbb{E}|M_n| < \infty$, the (by Corollary 18.54) $M_\infty := \lim_{n\to\infty} M_n$ a.s. and $\mathbb{E}|M_\infty| < \infty$. Therefore,

$$\begin{aligned}\mathbb{E}[|M_\tau| : \tau < \infty] &\le \mathbb{E}|M_\tau| = \mathbb{E}\left[\lim_{n\to\infty}|M_{\tau\wedge n}|\right] \\ &\le \liminf_{n\to\infty}\mathbb{E}|M_{\tau\wedge n}| \le \liminf_{n\to\infty}\mathbb{E}|M_n| < \infty\end{aligned}$$

wherein we have used Fatou's lemma, the optional sampling theorem to conclude $M_{\tau\wedge n} = \mathbb{E}_{\mathcal{B}_{\tau\wedge n}} M_n$, cJensen to conclude $|M_{\tau\wedge n}| \le \mathbb{E}_{\mathcal{B}_{\tau\wedge n}}|M_n|$, and the tower property of conditional expectation to conclude $\mathbb{E}|M_{\tau\wedge n}| \le \mathbb{E}|M_n|$. ■

Corollary 18.74. *Suppose that M is an L^1 – bounded martingale and $J \in \mathcal{B}_\mathbb{R}$ is a bounded set, then $\tau = \inf\{n : M_n \notin J\}$ is a regular stopping time for M.*

Proof. According to Proposition 18.73, it suffices to show $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is a uniformly integrable sequence of random variables. However, if we choose $A < \infty$ such that $J \subset [-A, A]$, since $M_n 1_{n<\tau} \in J$ we have $|M_n 1_{n<\tau}| \le A$ which is sufficient to complete the proof. ■

18.9 Backwards (Reverse) Submartingales

In this section we will consider submartingales indexed by $\mathbb{Z}_- := \{\dots, -n, -n+1, \dots, -2, -1, 0\}$. So again we assume that we have an increasing filtration, $\{\mathcal{B}_n : n \le 0\}$, i.e. $\dots \subset \mathcal{B}_{-2} \subset \mathcal{B}_{-1} \subset \mathcal{B}_0 \subset \mathcal{B}$. As usual, we say an adapted process $\{X_n\}_{n\le 0}$ is a submartingale (martingale) provided $\mathbb{E}[X_m - X_n|\mathcal{B}_n] \ge 0\ (= 0)$ for all $m \ge n$. Observe that $\mathbb{E}X_m \ge \mathbb{E}X_n$ for $m \ge n$, so that $\mathbb{E}X_{-n}$ decreases as n increases. Also observe that $(X_{-n}, X_{-(n-1)}, \dots, X_{-1}, X_0)$ is a "finite string" submartingale relative to the filtration, $\mathcal{B}_{-n} \subset \mathcal{B}_{-(n-1)} \subset \dots \subset \mathcal{B}_{-1} \subset \mathcal{B}_0$.

It turns out that backwards submartingales are even better behaved than forward submartingales. In order to understand why, consider the case where

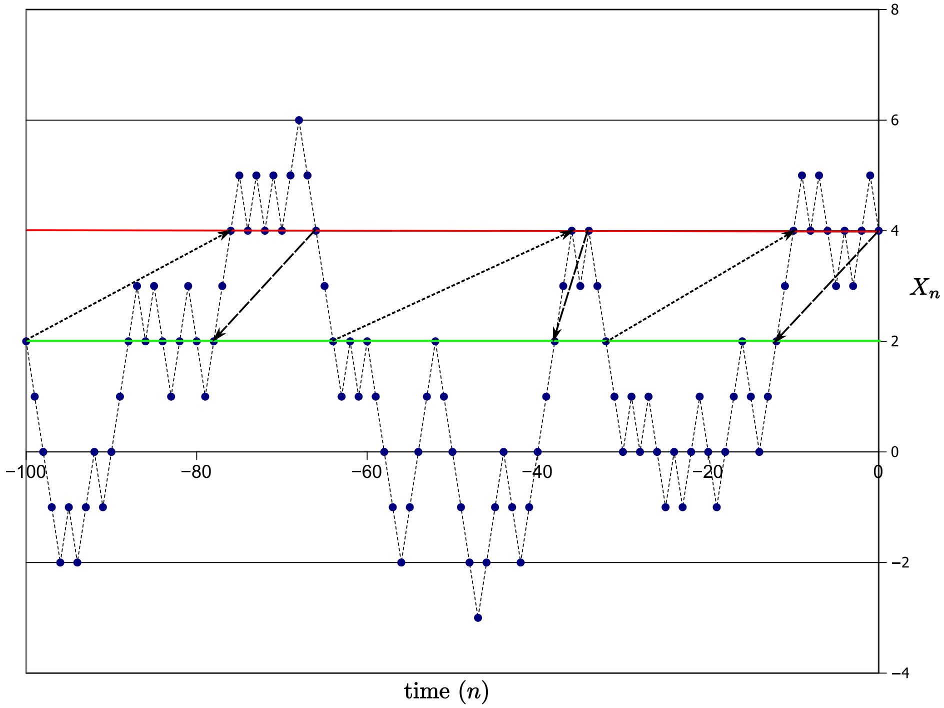


Fig. 18.5. A sample path of a backwards martingale on $[-100,0]$ indicating the down crossings of $X_0, X_{-1},\dots,X_{-100}$ and the upcrossings of $X_{-100},X_{-99},\dots,X_0$. The total number of each is the same.

$\{M_n\}_{n\le 0}$ is a backwards martingale. We then have, $\mathbb{E}[M_n|\mathcal{B}_m] = M_{m\wedge n}$ for all $m,n\le 0$. Taking $n=0$ in this equation implies that $M_m = \mathbb{E}[M_0|\mathcal{B}_m]$ and so the only backwards martingales are of the form $M_m=\mathbb{E}[M_0|\mathcal{B}_m]$ for some $M_0\in L^1(P)$. We have seen in Example 18.7 that this need not be the case for forward martingales.

Theorem 18.75 (Backwards (or reverse) submartingale convergence). *Let $\{\mathcal{B}_n : n\le 0\}$ be a reverse filtration, $\{X_n\}_{n\le 0}$ is a backwards submartingale. Then $X_{-\infty}=\lim_{n\to-\infty}X_n$ exists a.s. in $\{-\infty\}\cup\mathbb{R}$ and $X_{-\infty}^+\in L^1(\Omega,\mathcal{B},P)$. If we further assume that*

$$C:=\lim_{n\to-\infty}\mathbb{E}X_n=\inf_{n\le 0}\mathbb{E}X_n>-\infty, \tag{18.62}$$

then 1) $X_n=M_n+A_n$ where $\{M_n\}_{-\infty<n\le 0}$ is a martingale, $\{A_n\}_{-\infty<n\le 0}$ is a predictable process such that $A_{-\infty}=\lim_{n\to-\infty}A_n=0$, 2) $\{X_n\}_{n\le 0}$ is uniformly integrability, 3) $X_{-\infty}\in L^1(\Omega,\mathcal{B},P)$, and 4) $\lim_{n\to-\infty}\mathbb{E}|X_n-X_{-\infty}|=0$.

Proof. The number of downcrossings of $(X_0,X_{-1},\dots,X_{-(n-1)},X_{-n})$ across $[a,b]$, (denoted by $D_n(a,b)$) is equal to the number of upcrossings, $(X_{-n},X_{-(n-1)},\dots,X_{-1},X_0)$ across $[a,b]$, see Figure 18.5. Since $(X_{-n},X_{-(n-1)},\dots,X_{-1},X_0)$ is a $\mathcal{B}_{-n}\subset\mathcal{B}_{-(n-1)}\subset\cdots\subset\mathcal{B}_{-1}\subset\mathcal{B}_0$ submartingale, we may apply Doob's upcrossing inequality (Theorem 18.51) to find;

$$\begin{aligned}(b-a)\,\mathbb{E}[D_n(a,b)] &\le \mathbb{E}(X_0-a)_+-\mathbb{E}(X_{-n}-a)_+\\ &\le \mathbb{E}(X_0-a)_+<\infty.\end{aligned}\tag{18.63}$$

Letting $D_\infty(a,b):=\uparrow\lim_{n\to\infty}D_n(a,b)$ be the total number of downcrossing of $(X_0,X_{-1},\dots,X_{-n},\dots)$, using the MCT to pass to the limit in Eq. (18.63), we have

$$(b-a)\,\mathbb{E}[D_\infty(a,b)]\le\mathbb{E}(X_0-a)_+<\infty.$$

In particular it follows that $D_\infty(a,b)<\infty$ a.s. for all $a<b$.

As in the proof of Corollary 18.54 (making use of the obvious downcrossing analogue of Lemma 18.53), it follows that $X_{-\infty}:=\lim_{n\to-\infty}X_n$ exists in $\bar{\mathbb{R}}$ a.s. At the end of the proof, we will show that $X_{-\infty}$ takes values in $\{-\infty\}\cup\mathbb{R}$ almost surely, i.e. $X_{-\infty}<\infty$ a.s.

Now suppose that $C>-\infty$. We begin by computing the Doob decomposition of X_n as $X_n=M_n+A_n$ with A_n being predictable, increasing and satisfying, $A_{-\infty}=\lim_{n\to-\infty}A_n=0$. If such an A is to exist, following Lemma 18.16, we should define

$$A_n=\sum_{k\le n}\mathbb{E}[\Delta_kX|\mathcal{B}_{k-1}].$$

This is a well defined increasing predictable process since the submartingale property implies $\mathbb{E}[\Delta_kX|\mathcal{B}_{k-1}]\ge 0$. Moreover we have

$$\begin{aligned}\mathbb{E}A_0&=\sum_{k\le 0}\mathbb{E}[\mathbb{E}[\Delta_kX|\mathcal{B}_{k-1}]]=\sum_{k\le 0}\mathbb{E}[\Delta_kX]\\ &=\lim_{N\to\infty}(\mathbb{E}X_0-\mathbb{E}X_{-N})=\mathbb{E}X_0-\inf_{n\le 0}\mathbb{E}X_n=\mathbb{E}X_0-C<\infty.\end{aligned}$$

As $0\le A_n\le A_n^*=A_0\in L^1(P)$, it follows that $\{A_n\}_{n\le 0}$ is uniformly integrable Moreover if we define $M_n:=X_n-A_n$, then

$$\mathbb{E}[\Delta_nM|\mathcal{B}_{n-1}]=\mathbb{E}[\Delta_nX-\Delta_nA|\mathcal{B}_{n-1}]=\mathbb{E}[\Delta_nX|\mathcal{B}_{n-1}]-\Delta_nA=0\text{ a.s.}$$

Thus M is a martingale and therefore, $M_n=\mathbb{E}[M_0|\mathcal{B}_n]$ with $M_0=X_0-A_0\in L^1(P)$. An application of Proposition 18.8 implies $\{M_n\}_{n\le 0}$ is uniformly integrable and hence $X_n=M_n+A_n$ is uniformly integrable as well. (See Remark

18.76 for an alternate proof of the uniform integrability of X.) Therefore $X_{-\infty} \in L^1(\Omega,\mathcal{B},P)$ and $X_n \to X_{-\infty}$ in $L^1(\Omega,\mathcal{B},P)$ as $n\to\infty$.

To finish the proof we must show, without assumptions on $C > -\infty$, we must show $X^+_{-\infty} \in L^1(\Omega,\mathcal{B},P)$ which will also imply $P(X_{-\infty}=\infty)=0$. To prove this, notice that $X^+_{-\infty} = \lim_{n\to-\infty} X_n^+$ and that by Jensen's inequality, $\{X_n^+\}_{n=1}^\infty$ is a non-negative backwards submartingale. Since $\inf \mathbb{E}X_n^+ \geq 0 > -\infty$, it follows by what we have just proved that $X^+_{-\infty} \in L^1(\Omega,\mathcal{B},P)$. ■

*Remark 18.76 (*Not necessary to read.).* Let us give a direct proof of the fact that X is uniformly integrable if $C > -\infty$. We begin with Jensen's inequality;

$$\mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \le 2\mathbb{E}X_0^+ - \mathbb{E}X_n \le 2\mathbb{E}X_0^+ - C = K < \infty, \tag{18.64}$$

which shows that $\{X_n\}_{n=1}^\infty$ is L^1 - bounded. For uniform integrability we will use the following identity;

$$\begin{aligned}\mathbb{E}[|X| : |X|\ge\lambda] &= \mathbb{E}[X : X\ge\lambda] - \mathbb{E}[X : X\le-\lambda]\\ &= \mathbb{E}[X : X\ge\lambda] - (\mathbb{E}X - \mathbb{E}[X : X>-\lambda])\\ &= \mathbb{E}[X : X\ge\lambda] + \mathbb{E}[X : X>-\lambda] - \mathbb{E}X.\end{aligned}$$

Taking $X = X_n$ and $k\ge n$, we find

$$\begin{aligned}\mathbb{E}[|X_n| : |X_n|\ge\lambda] =& \mathbb{E}[X_n : X_n\ge\lambda] + \mathbb{E}[X_n : X_n>-\lambda] - \mathbb{E}X_n\\ \le& \mathbb{E}[X_k : X_n\ge\lambda] + \mathbb{E}[X_k : X_n>-\lambda]\\ & - \mathbb{E}X_k + (\mathbb{E}X_k - \mathbb{E}X_n)\\ =& \mathbb{E}[X_k : X_n\ge\lambda] - \mathbb{E}[X_k : X_n\le-\lambda] + (\mathbb{E}X_k - \mathbb{E}X_n)\\ =& \mathbb{E}[|X_k| : |X_n|\ge\lambda] + (\mathbb{E}X_k - \mathbb{E}X_n).\end{aligned}$$

Given $\varepsilon>0$ we may choose $k = k_\varepsilon < 0$ such that if $n\le k$, $0\le \mathbb{E}X_k - \mathbb{E}X_n \le \varepsilon$ and hence

$$\limsup_{\lambda\uparrow\infty}\sup_{n\le k}\mathbb{E}[|X_n| : |X_n|\ge\lambda] \le \limsup_{\lambda\uparrow\infty}\mathbb{E}[|X_k| : |X_n|\ge\lambda] + \varepsilon \le \varepsilon$$

wherein we have used Eq. (18.64), Chebyschev's inequality to conclude $P(|X_n|\ge\lambda)\le K/\lambda$ and then the uniform integrability of the singleton set, $\{|X_k|\}\subset L^1(\Omega,\mathcal{B},P)$. From this it now easily follows that $\{X_n\}_{n\le 0}$ is a uniformly integrable.

Corollary 18.77. *Suppose $1\le p<\infty$ and $X_n = M_n$ in Theorem 18.75, where M_n is an L^p – bounded martingale on $-\mathbb{N}\cup\{0\}$. Then $M_{-\infty} := \lim_{n\to\infty} M_n$ exists a.s. and in $L^p(P)$. Moreover $M_{-\infty} = \mathbb{E}[M_0|\mathcal{B}_{-\infty}]$, where $\mathcal{B}_{-\infty} = \cap_{n\le 0}\mathcal{B}_n$.*

Proof. Since $M_n = \mathbb{E}[M_0|\mathcal{B}_n]$ for all n, it follows by cJensen that $|M_n|^p \le \mathbb{E}[|M_0|^p|\mathcal{B}_n]$ for all n. By Proposition 18.8, $\{\mathbb{E}[|M_0|^p|\mathcal{B}_n]\}_{n\le 0}$ is uniformly integrable and so is $\{|M_n|^p\}_{n\le 0}$. By Theorem 18.75, $M_n\to M_{-\infty}$ a.s.. Hence we may now apply Corollary 12.47 to see that $M_n\to M_{-\infty}$ in $L^p(P)$. ■

Example 18.78 (Kolmogorov's SLLN). In this example we are going to give another proof of the strong law of large numbers in Theorem 16.10, also see Theorem 20.30 below for a third proof. Let $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables such that $\mathbb{E}X_n = 0$ and let $S_0 = 0$, $S_n := X_1+\dots+X_n$ and $\mathcal{B}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2},\dots)$ so that S_n is $\mathcal{B}_{-n}$ measurable for all $n\ge 0$.

1. For any permutation σ of the set $\{1,2,\dots,n\}$,
$$(X_1,\dots,X_n,S_n,S_{n+1},S_{n+2},\dots) \overset{d}{=} (X_{\sigma 1},\dots,X_{\sigma n},S_n,S_{n+1},S_{n+2},\dots)$$
and in particular
$$(X_j,S_n,S_{n+1},S_{n+2},\dots) \overset{d}{=} (X_1,S_n,S_{n+1},S_{n+2},\dots) \text{ for all } j\le n.$$
2. By Exercise 14.6 we may conclude that
$$\mathbb{E}[X_j|\mathcal{B}_{-n}] = \mathbb{E}[X_1|\mathcal{B}_{-n}] \text{ a.s. for all } j\le n. \tag{18.65}$$
To see this directly notice that if σ is any permutation of $\mathbb{N}$ leaving $\{n+1,n+2,\dots\}$ fixed, then
$$\mathbb{E}[g(X_1,\dots,X_n)\cdot f(S_n,S_{n+1},\dots)] = \mathbb{E}[g(X_{\sigma 1},\dots,X_{\sigma n})\cdot f(S_n,S_{n+1},\dots)]$$
for all bounded measurable f and g such that $g(X_1,\dots,X_n)\in L^1(P)$. From this equation it follows that
$$\mathbb{E}[g(X_1,\dots,X_n)|\mathcal{B}_{-n}] = \mathbb{E}[g(X_{\sigma 1},\dots,X_{\sigma n})|\mathcal{B}_{-n}] \text{ a.s.}$$
and then taking $g(x_1,\dots,x_n) = x_1$ give the desired result.
3. Summing Eq. (18.65) over $j=1,2,\dots,n$ gives,
$$S_n = \mathbb{E}[S_n|S_n,S_{n+1},S_{n+2},\dots] = n\mathbb{E}[X_1|S_n,S_{n+1},S_{n+2},\dots]$$
from which it follows that
$$M_{-n} := \frac{S_n}{n} := \mathbb{E}[X_1|S_n,S_{n+1},S_{n+2},\dots] \tag{18.66}$$
and hence $\left\{M_{-n} = \frac{1}{n}S_n\right\}$ is a backwards martingale.
4. By Theorem 18.75 we know;
$$\lim_{n\to\infty}\frac{S_n}{n} = \lim_{n\to-\infty} M_{-n} =: M_{-\infty} \text{ exists a.s. and in } L^1(P).$$
5. Since $M_{-\infty} = \lim_{n\to\infty}\frac{S_n}{n}$ is a $\{\sigma(X_1,\dots,X_n)\}_{n=1}^\infty$ – tail random variable it follows by Corollary 10.51 (basically by Kolmogorov's zero one law of Proposition 10.50) that $\lim_{n\to\infty}\frac{S_n}{n} = c$ a.s. for some constant c.
6. Since $\frac{S_n}{n}\to c$ in $L^1(P)$ we may conclude that
$$c = \lim_{n\to\infty}\mathbb{E}\frac{S_n}{n} = \mathbb{E}X_1.$$

Thus we have given another proof of Kolmogorov's strong law of large numbers.

18.10 Some More Martingale Exercises

(The next four problems were taken directly from http://math.nyu.edu/˜sheff/martingalenote.pdf.)

Exercise 18.14. Suppose Harriet has 7 dollars. Her plan is to make one dollar bets on fair coin tosses until her wealth reaches either 0 or 50, and then to go home. What is the expected amount of money that Harriet will have when she goes home? What is the probability that she will have 50 when she goes home?

Exercise 18.15. Consider a contract that at time N will be worth either 100 or 0. Let S_n be its price at time $0 \le n \le N$. If S_n is a martingale, and $S_0 = 47$, then what is the probability that the contract will be worth 100 at time N?

Exercise 18.16. Pedro plans to buy the contract in the previous problem at time 0 and sell it the first time T at which the price goes above 55 or below 15. What is the expected value of S_T? You may assume that the value, S_n, of the contract is bounded – there is only a finite amount of money in the world up to time N. Also note, by assumption, $T \le N$.

Exercise 18.17. Suppose S_N is with probability one either 100 or 0 and that $S_0 = 50$. Suppose further there is at least a 60% probability that the price will at some point dip to below 40 and then subsequently rise to above 60 before time N. Prove that S_n cannot be a martingale. (I don't know if this problem is correct! but if we modify the 40 to a 30 the buy low sell high strategy will show that $\{S_n\}$ is not a martingale.)

Exercise 18.18. Let $(M_n)_{n=0}^\infty$ be a martingale with $M_0 = 0$ and $E[M_n^2] < \infty$ for all n. Show that for all $\lambda > 0$,

$$P\left(\max_{1\le m\le n} M_m \ge \lambda\right) \le \frac{E[M_n^2]}{E[M_n^2]+\lambda^2}.$$

Hints: First show that for any $c > 0$ that $\left\{X_n := (M_n + c)^2\right\}_{n=0}^\infty$ is a submartingale and then observe,

$$\left\{\max_{1\le m\le n} M_m \ge \lambda\right\} \subset \left\{\max_{1\le m\le n} X_n \ge (\lambda + c)^2\right\}.$$

Now use Doob' Maximal inequality (Proposition 18.42) to estimate the probability of the last set and then choose c so as to optimize the resulting estimate you get for $P(\max_{1\le m\le n} M_m \ge \lambda)$. (Notice that this result applies to $-M_n$ as well so it also holds that;

$$P\left(\min_{1\le m\le n} M_m \le -\lambda\right) \le \frac{E[M_n^2]}{E[M_n^2]+\lambda^2} \text{ for all } \lambda > 0.$$

Exercise 18.19. Let $\{Z_n\}_{n=1}^\infty$ be independent random variables, $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$, and $f_n(\lambda) := \mathbb{E}\left[e^{i\lambda Z_n}\right]$. Suppose $\mathbb{E}e^{i\lambda S_n} = \prod_{n=1}^N f_n(\lambda)$ converges to a continuous function, $F(\lambda)$, as $N \to \infty$. Show for each $\lambda \in \mathbb{R}$ that

$$P\left(\lim_{n\to\infty} e^{i\lambda S_n} \text{ exists}\right) = 1. \tag{18.67}$$

Hints:

1. Show it is enough to find an $\varepsilon > 0$ such that Eq. (18.67) holds for $|\lambda| \le \varepsilon$.
2. Choose $\varepsilon > 0$ such that $|F(\lambda) - 1| < 1/2$ for $|\lambda| \le \varepsilon$. For $|\lambda| \le \varepsilon$, show $M_n(\lambda) := \frac{e^{i\lambda S_n}}{\mathbb{E}e^{i\lambda S_n}}$ is a bounded complex[8] martingale relative to the filtration, $\mathcal{B}_n = \sigma(Z_1, \dots, Z_n)$.

Lemma 18.79 (Protter [42, See the lemma on p. 22.]). *Let $\{x_n\}_{n=1}^\infty \subset \mathbb{R}$ such that $\left\{e^{iux_n}\right\}_{n=1}^\infty$ is convergent for Lebesgue almost every $u \in \mathbb{R}$. Then $\lim_{n\to\infty} x_n$ exists in $\mathbb{R}$.*

Proof. Let U be a uniform random variable with values in $[0,1]$. By assumption, for any $t \in \mathbb{R}$, $\lim_{n\to\infty} e^{itUx_n}$ exists a.s. Thus if n_k and m_k are any increasing sequences we have

$$\lim_{k\to\infty} e^{itUx_{n_k}} = \lim_{n\to\infty} e^{itUx_n} = \lim_{k\to\infty} e^{itUx_{m_k}} \text{ a.s.}$$

and therefore,

$$e^{it\left(Ux_{n_k} - Ux_{m_k}\right)} = \frac{e^{itUx_{n_k}}}{e^{itUx_{m_k}}} \to 1 \text{ a.s. as } k \to \infty.$$

Hence by DCT it follows that

$$\mathbb{E}\left[e^{it\left(Ux_{n_k} - Ux_{m_k}\right)}\right] \to 1 \text{ as } k \to \infty$$

and therefore

$$(x_{n_k} - x_{m_k}) \cdot U = Ux_{n_k} - Ux_{m_k} \to 0$$

in distribution and hence in probability. But his can only happen if $(x_{n_k} - x_{m_k}) \to 0$ as $k \to \infty$. As $\{n_k\}$ and $\{m_k\}$ were arbitrary, this suffices to show $\{x_n\}$ is a Cauchy sequence. ∎

Exercise 18.20 (Continuation of Exercise 18.19 – See Doob [12, Chapter VII.5]). Let $\{Z_n\}_{n=1}^\infty$ be independent random variables. Use Exercise 18.19 an Lemma 18.79 to prove the series, $\sum_{n=1}^\infty Z_n$, converges in $\mathbb{R}$ a.s. iff $\prod_{n=1}^N f_n(\lambda)$ converges to a continuous function, $F(\lambda)$ as $N \to \infty$. Conclude from this that $\sum_{n=1}^\infty Z_n$ is a.s. convergent iff $\sum_{n=1}^\infty Z_n$ is convergent in distribution.

[8] Please use the obvious generalization of a martingale for complex valued processes. It will be useful to observe that the real and imaginary parts of a complex martingales are real martingales.

18.10.1 More Random Walk Exercises

For the next four exercises, let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of Bernoulli random variables with $P(Z_n = \pm 1) = \frac{1}{2}$ and let $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$. Then S becomes a martingale relative to the filtration, $\mathcal{B}_n := \sigma(Z_1, \ldots, Z_n)$ with $\mathcal{B}_0 := \{\emptyset, \Omega\}$ – of course S_n is the (fair) simple random walk on $\mathbb{Z}$. For any $a \in \mathbb{Z}$, let

$$\sigma_a := \inf\{n : S_n = a\}.$$

Exercise 18.21. For $a < 0 < b$ with $a, b \in \mathbb{Z}$, let $\tau = \sigma_a \wedge \sigma_b$. Explain why τ is regular for S. Use this to show $P(\tau = \infty) = 0$. **Hint:** make use of Remark 18.71 and the fact that $|S_n - S_{n-1}| = |Z_n| = 1$ for all n.

Exercise 18.22. In this exercise, you are asked to use the central limit Theorem 10.35 to prove again that $P(\tau = \infty) = 0$, Exercise 18.21. **Hints:** Use the central limit theorem to show

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)\, e^{-x^2/2} dx \geq f(0)\, P(\tau = \infty) \tag{18.68}$$

for all $f \in C^3(\mathbb{R} \to [0, \infty))$ with $M := \sup_{x \in \mathbb{R}} \left|f^{(3)}(x)\right| < \infty$. Use this inequality to conclude that $P(\tau = \infty) = 0$.

Exercise 18.23. Show

$$P(\sigma_b < \sigma_a) = \frac{|a|}{b + |a|} \tag{18.69}$$

and use this to conclude $P(\sigma_b < \infty) = 1$, i.e. every $b \in \mathbb{N}$ is almost surely visited by S_n. (This last result also follows by the Hewitt-Savage Zero-One Law, see Example 10.55 where it is shown b is visited infinitely often.)

Hint: Using properties of martingales and Exercise 18.21, compute $\lim_{n\to\infty} \mathbb{E}\left[S_n^{\sigma_a \wedge \sigma_b}\right]$ in two different ways.

Exercise 18.24. Let $\tau := \sigma_a \wedge \sigma_b$. In this problem you are asked to show $\mathbb{E}[\tau] = |a|\, b$ with the aid of the following outline.

1. Use Exercise 18.4 above to conclude $N_n := S_n^2 - n$ is a martingale.
2. Now show
$$0 = \mathbb{E}N_0 = \mathbb{E}N_{\tau\wedge n} = \mathbb{E}S^2_{\tau \wedge n} - \mathbb{E}[\tau \wedge n]. \tag{18.70}$$
3. Now use DCT and MCT along with Exercise 18.23 to compute the limit as $n \to \infty$ in Eq. (18.70) to find
$$\mathbb{E}[\sigma_a \wedge \sigma_b] = \mathbb{E}[\tau] = b\,|a|. \tag{18.71}$$
4. By considering the limit, $a \to -\infty$ in Eq. (18.71), show $\mathbb{E}[\sigma_b] = \infty$.

For the next group of exercise we are now going to suppose that $P(Z_n = 1) = p > \frac{1}{2}$ and $P(Z_n = -1) = q = 1 - p < \frac{1}{2}$. As before let $\mathcal{B}_n = \sigma(Z_1, \ldots, Z_n)$, $S_0 = 0$ and $S_n = Z_1 + \cdots + Z_n$ for $n \in \mathbb{N}$. Let us review the method above and what you did in Exercise 17.10 above.

In order to follow the procedures above, we start by looking for a function, φ, such that $\varphi(S_n)$ is a martingale. Such a function must satisfy,

$$\varphi(S_n) = \mathbb{E}_{\mathcal{B}_n}\varphi(S_{n+1}) = \varphi(S_n + 1)\, p + \varphi(S_n - 1)\, q,$$

and this then leads us to try to solve the following difference equation for φ;

$$\varphi(x) = p\varphi(x+1) + q\varphi(x-1) \text{ for all } x \in \mathbb{Z}. \tag{18.72}$$

Similar to the theory of second order ODE's this equation has two linearly independent solutions which could be found by solving Eq. (18.72) with initial conditions, $\varphi(0) = 1$ and $\varphi(1) = 0$ and then with $\varphi(0) = 0$ and $\varphi(1) = 0$ for example. Rather than doing this, motivated by second order constant coefficient ODE's, let us try to find solutions of the form $\varphi(x) = \lambda^x$ with λ to be determined. Doing so leads to the equation, $\lambda^x = p\lambda^{x+1} + q\lambda^{x-1}$, or equivalently to the **characteristic equation,**

$$p\lambda^2 - \lambda + q = 0.$$

The solutions to this equation are

$$\begin{aligned}\lambda &= \frac{1 \pm \sqrt{1 - 4pq}}{2p} = \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p} \\ &= \frac{1 \pm \sqrt{4p^2 - 4p + 1}}{2p} = \frac{1 \pm \sqrt{(2p-1)^2}}{2p} = \{1, (1-p)/p\} = \{1, q/p\}.\end{aligned}$$

The most general solution to Eq. (18.72) is then given by

$$\varphi(x) = A + B\,(q/p)^x.$$

Below we will take $A = 0$ and $B = 1$. As before let $\sigma_a = \inf\{n \geq 0 : S_n = a\}$.

Exercise 18.25. Let $a < 0 < b$ and $\tau := \sigma_a \wedge \sigma_b$.

1. Apply the method in Exercise 18.21 with S_n replaced by $M_n := (q/p)^{S_n}$ to show $P(\tau = \infty) = 0$.
2. Now use the method in Exercise 18.23 to show
$$P(\sigma_a < \sigma_b) = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}. \tag{18.73}$$

3. By letting $a \to -\infty$ in Eq. (18.73), conclude $P(\sigma_b = \infty) = 0$.
4. By letting $b \to \infty$ in Eq. (18.73), conclude $P(\sigma_a < \infty) = (q/p)^{|a|}$.

Exercise 18.26. Verify,

$$M_n := S_n - n(p-q)$$

and

$$N_n := M_n^2 - \sigma^2 n$$

are martingales, where $\sigma^2 = 1 - (p-q)^2$. (This should be simple; see either Exercise 18.4 or Exercise 18.3.)

Exercise 18.27. Using exercise 18.26, show

$$\mathbb{E}(\sigma_a \wedge \sigma_b) = \left(\frac{b\left[1-(q/p)^a\right] + a\left[(q/p)^b - 1\right]}{(q/p)^b - (q/p)^a} \right)(p-q)^{-1}. \tag{18.74}$$

By considering the limit of this equation as $a \to -\infty$, show

$$\mathbb{E}[\sigma_b] = \frac{b}{p-q}$$

and by considering the limit as $b \to \infty$, show $\mathbb{E}[\sigma_a] = \infty$.

18.11 Appendix: Some Alternate Proofs

This section may be safely omitted (for now).

Proof. Alternate proof of Theorem 18.39. Let $A \in \mathcal{B}_\sigma$. Then

$$\mathbb{E}[X_\tau - X_\sigma : A] = \mathbb{E}\left[\sum_{k=0}^{N-1} 1_{\sigma \le k < \tau} \Delta_{k+1} X : A\right] = \sum_{k=1}^{N} \mathbb{E}[\Delta_k X : A \cap \{\sigma \le k < \tau\}].$$

Since $A \in \mathcal{B}_\sigma$, $A \cap \{\sigma \le k\} \in \mathcal{B}_k$ and since $\{k < \tau\} = \{\tau \le k\}^c \in \mathcal{B}_k$, it follows that $A \cap \{\sigma \le k < \tau\} \in \mathcal{B}_k$. Hence we know that

$$\mathbb{E}[\Delta_{k+1} X : A \cap \{\sigma \le k < \tau\}] \begin{matrix}\le \\ = \\ \ge\end{matrix} 0 \text{ respectively.}$$

and hence that

$$\mathbb{E}[X_\tau - X_\sigma : A] \begin{matrix}\le \\ = \\ \ge\end{matrix} 0 \text{ respectively.}$$

Since this true for all $A \in \mathcal{B}_\sigma$, Eq. (18.21) follows. ■

Lemma 18.80. *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space, $1 \le p < \infty$, and let $\mathcal{B}_\infty := \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$. Then $\cup_{n=1}^\infty L^p(\Omega, \mathcal{B}_n, P)$ is dense in $L^p(\Omega, \mathcal{B}_\infty, P)$.*

Proof. Let $M_n := L^p(\Omega, \mathcal{B}_n, P)$, then M_n is an increasing sequence of closed subspaces of $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. Further let $\mathbb{A}$ be the algebra of functions consisting of those $f \in \cup_{n=1}^\infty M_n$ such that f is bounded. As a consequence of the density Theorem 12.27, we know that $\mathbb{A}$ and hence $\cup_{n=1}^\infty M_n$ is dense in $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. This completes the proof. However for the readers convenience let us quickly review the proof of Theorem 12.27 in this context.

Let $\mathbb{H}$ denote those bounded $\mathcal{B}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{A}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded $\mathcal{B}_\infty$ –measurable $\mathbb{R}$ – valued functions on Ω, $1 \in \mathbb{H}$, $\mathbb{A} \subset \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated (or bounded) convergence theorem, $\lim_{n\to\infty} \|(f - f_n)\|_{L^p(P)} = 0$.[9] We may now choose $\varphi_n \in \mathbb{A}$ such that $\|\varphi_n - f_n\|_{L^p(P)} \le \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} \le \limsup_{n\to\infty} \|(f - f_n)\|_{L^p(P)} + \limsup_{n\to\infty} \|f_n - \varphi_n\|_{L^p(P)} = 0,$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem 8.16, now shows $\mathbb{H}$ contains all bounded $\sigma(\mathbb{A}) = \mathcal{B}_\infty$ – measurable functions on Ω. Since for any $f \in L^p(\Omega, \mathcal{B}, P)$, $f 1_{|f| \le n} \in \mathbb{H}$ there exists $\varphi_n \in \mathbb{A}$ such that $\|f_n - \varphi_n\|_p \le n^{-1}$. Using the DCT we know that $f_n \to f$ in L^p and therefore by Minikowski's inequality it follows that $\varphi_n \to f$ in L^p. ■

Theorem 18.81. *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space, $1 \le p < \infty$, and let $\mathcal{B}_\infty := \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$. Then for every $X \in L^p(\Omega, \mathcal{B}, P)$, $X_n = \mathbb{E}[X|\mathcal{B}_n]$ is a martingale and $X_n \to X_\infty := \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^p(\Omega, \mathcal{B}_\infty, P)$ as $n \to \infty$.*

Proof. We have already seen in Example 18.6 that $X_n = \mathbb{E}[X|\mathcal{B}_n]$ is always a martingale. Since conditional expectation is a contraction on L^p it follows that $\mathbb{E}|X_n|^p \le \mathbb{E}|X|^p < \infty$ for all $n \in \mathbb{N} \cup \{\infty\}$. So to finish the proof we need to show $X_n \to X_\infty$ in $L^p(\Omega, \mathcal{B}, P)$ as $n \to \infty$.

Let $M_n := L^p(\Omega, \mathcal{B}_n, P)$ and $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. If $X \in \cup_{n=1}^\infty M_n$, then $X_n = X$ for all sufficiently large n and for $n = \infty$. Now suppose that $X \in M_\infty$ and $Y \in \cup_{n=1}^\infty M_n$. Then

[9] It is at this point that the proof would break down if $p = \infty$.

$$\begin{aligned}\|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_n}X\|_p &\le \|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_\infty}Y\|_p+\|\mathbb{E}_{\mathcal{B}_\infty}Y-\mathbb{E}_{\mathcal{B}_n}Y\|_p+\|\mathbb{E}_{\mathcal{B}_n}Y-\mathbb{E}_{\mathcal{B}_n}X\|_p\\ &\le 2\|X-Y\|_p+\|\mathbb{E}_{\mathcal{B}_\infty}Y-\mathbb{E}_{\mathcal{B}_n}Y\|_p\end{aligned}$$

and hence

$$\limsup_{n\to\infty}\|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_n}X\|_p\le 2\|X-Y\|_p.$$

Using the density Lemma 18.80 we may choose $Y\in\cup_{n=1}^\infty M_n$ as close to $X\in M_\infty$ as we please and therefore it follows that $\limsup_{n\to\infty}\|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_n}X\|_p=0.$

For general $X\in L^p(\Omega,\mathcal{B},P)$ it suffices to observe that $X_\infty:=\mathbb{E}[X|\mathcal{B}_\infty]\in L^p(\Omega,\mathcal{B}_\infty,P)$ and by the tower property of conditional expectations,

$$\mathbb{E}[X_\infty|\mathcal{B}_n]=\mathbb{E}[\mathbb{E}[X|\mathcal{B}_\infty]|\mathcal{B}_n]=\mathbb{E}[X|\mathcal{B}_n]=X_n.$$

So again $X_n\to X_\infty$ in L^p as desired. ■

We are now ready to prove the converse of Theorem 18.81.

Theorem 18.82. *Suppose $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty,P)$ is a filtered probability space, $1\le p<\infty$, $\mathcal{B}_\infty:=\vee_{n=1}^\infty\mathcal{B}_n:=\sigma(\cup_{n=1}^\infty\mathcal{B}_n)$, and $\{X_n\}_{n=1}^\infty\subset L^p(\Omega,\mathcal{B},P)$ is a martingale. Further assume that $\sup_n\|X_n\|_p<\infty$ and that $\{X_n\}_{n=1}^\infty$ is uniformly integrable if $p=1$. Then there exists $X_\infty\in L^p(\Omega,\mathcal{B}_\infty,P)$ such that $X_n:=\mathbb{E}[X_\infty|\mathcal{B}_\infty]$. Moreover by Theorem 18.81 we know that $X_n\to X_\infty$ in $L^p(\Omega,\mathcal{B}_\infty,P)$ as $n\to\infty$ and hence X_∞ is uniquely determined by $\{X_n\}_{n=1}^\infty$.*

Proof. By Theorems 13.20 and 13.22 exists $X_\infty\in L^p(\Omega,\mathcal{B}_\infty,P)$ and a subsequence, $Y_k=X_{n_k}$such that

$$\lim_{k\to\infty}\mathbb{E}[Y_kh]=\mathbb{E}[X_\infty h]\text{ for all }h\in L^q(\Omega,\mathcal{B}_\infty,P)$$

where $q:=p(p-1)^{-1}$. Using the martingale property, if $h\in(\mathcal{B}_n)_b$ for some n, it follows that $\mathbb{E}[Y_kh]=\mathbb{E}[X_nh]$ for all large k and therefore that

$$\mathbb{E}[X_\infty h]=\mathbb{E}[X_nh]\text{ for all }h\in(\mathcal{B}_n)_b.$$

This implies that $X_n=\mathbb{E}[X_\infty|\mathcal{B}_n]$ as desired. ■

Theorem 18.83 (Almost sure convergence). *Suppose $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty,P)$ is a filtered probability space, $1\le p<\infty$, and let $\mathcal{B}_\infty:=\vee_{n=1}^\infty\mathcal{B}_n:=\sigma(\cup_{n=1}^\infty\mathcal{B}_n)$. Then for every $X\in L^1(\Omega,\mathcal{B},P)$, the martingale, $X_n=\mathbb{E}[X|\mathcal{B}_n]$, converges almost surely to $X_\infty:=\mathbb{E}[X|\mathcal{B}_\infty]$.*

Before starting the proof, recall from Proposition 1.5, if $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ are two bounded sequences, then

$$\begin{aligned}\limsup_{n\to\infty}&(a_n+b_n)-\liminf_{n\to\infty}(a_n+b_n)\\ &\le\limsup_{n\to\infty}a_n+\limsup_{n\to\infty}b_n-\left(\liminf_{n\to\infty}a_n+\liminf_{n\to\infty}b_n\right)\\ &=\limsup_{n\to\infty}a_n-\liminf_{n\to\infty}a_n+\limsup_{n\to\infty}b_n-\liminf_{n\to\infty}b_n.\end{aligned}\tag{18.75}$$

Proof. Since

$$X_n=\mathbb{E}[X|\mathcal{B}_n]=\mathbb{E}[\mathbb{E}[X|\mathcal{B}_\infty]|\mathcal{B}_n]=\mathbb{E}[X_\infty|\mathcal{B}_n],$$

there is no loss in generality in assuming $X=X_\infty$. If $X\in M_n:=L^1(\Omega,\mathcal{B}_n,P)$, then $X_m=X_\infty$ a.s. for all $m\ge n$ and hence $X_m\to X_\infty$ a.s. Therefore the theorem is valid for any X in the dense (by Lemma 18.80) subspace $\cup_{n=1}^\infty M_n$ of $L^1(\Omega,\mathcal{B}_\infty,P)$.

For general $X\in L^1(\Omega,\mathcal{B}_\infty,P)$, let $Y_j\in\cup M_n$ such that $Y_j\to X\in L^1(\Omega,\mathcal{B}_\infty,P)$ and let $Y_{j,n}:=\mathbb{E}[Y_j|\mathcal{B}_n]$ and $X_n:=\mathbb{E}[X|\mathcal{B}_n]$. We know that $Y_{j,n}\to Y_{j,\infty}$ a.s. for each $j\in\mathbb{N}$ and our goal is to show $X_n\to X_\infty$ a.s. By Doob's inequality in Corollary 18.47 and the L^1 - contraction property of conditional expectation we know that

$$P(X_N^*\ge a)\le\frac{1}{a}\mathbb{E}|X_N|\le\frac{1}{a}\mathbb{E}|X|$$

and so passing to the limit as $N\to\infty$ we learn that

$$P\left(\sup_n|X_n|\ge a\right)\le\frac{1}{a}\mathbb{E}|X|\text{ for all }a>0.\tag{18.76}$$

Letting $a\uparrow\infty$ then shows $P(\sup_n|X_n|=\infty)=0$ and hence $\sup_n|X_n|<\infty$ a.s. Hence we may use Eq. (18.75) with $a_n=X_n-Y_{j,n}$ and $b_n:=Y_{j,n}$ to find

$$\begin{aligned}D&=\limsup_{n\to\infty}X_n-\liminf_{n\to\infty}X_n\\ &\le\limsup_{n\to\infty}a_n-\liminf_{n\to\infty}a_n+\limsup_{n\to\infty}b_n-\liminf_{n\to\infty}b_n\\ &=\limsup_{n\to\infty}a_n-\liminf_{n\to\infty}a_n\le 2\sup_n|a_n|\\ &=2\sup_n|X_n-Y_{j,n}|,\end{aligned}$$

wherein we have used $\limsup_{n\to\infty}b_n-\liminf_{n\to\infty}b_n=0$ a.s. since $Y_{j,n}\to Y_{j,\infty}$ a.s.

We now apply Doob's inequality one more time, i.e. use Eq. (18.76) with X_n being replaced by $X_n-Y_{j,n}$ and X by $X-Y_j$, to conclude,

$$P(D\ge a)\le P\left(\sup_n|X_n-Y_{j,n}|\ge\frac{a}{2}\right)\le\frac{2}{a}\mathbb{E}|X-Y_j|\to 0\text{ as }j\to\infty.$$

Since $a > 0$ is arbitrary here, it follows that $D = 0$ a.s., i.e. $\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n$ and hence $\lim_{n\to\infty} X_n$ exists in $\mathbb{R}$ almost surely. Since we already know that $X_n \to X_\infty$ in $L^1(\Omega, \mathcal{B}, P)$, we may conclude that $\lim_{n\to\infty} X_n = X_\infty$ a.s.

Alternative proof – see Stroock [50, Corollary 5.2.7]. Let $\mathbb{H}$ denote those $X \in L^1(\Omega, \mathcal{B}_n, P)$ such that $X_n := \mathbb{E}[X|\mathcal{B}_n] \to X_\infty$ a.s. As we saw above $\mathbb{H}$ contains the dense subspace $\cup_{n=1}^\infty M_n$. It is also easy to see that $\mathbb{H}$ is a linear space. Thus it suffices to show that $\mathbb{H}$ is closed in $L^1(P)$. To prove this let $X^{(k)} \in \mathbb{H}$ with $X^{(k)} \to X$ in $L^1(P)$ and let $X_n^{(k)} := \mathbb{E}\left[X^{(k)}|\mathcal{B}_n\right]$. Then by the maximal inequality in Eq. (18.76),

$$P\left(\sup_n \left|X_n - X_n^{(k)}\right| \geq a\right) \leq \frac{1}{a}\mathbb{E}\left|X - X^{(k)}\right| \text{ for all } a > 0 \text{ and } k \in \mathbb{N}.$$

Therefore,

$$\begin{aligned}
&P\left(\sup_{n\geq N} |X - X_n| \geq 3a\right) \\
&\quad\leq P\left(\left|X - X^{(k)}\right| \geq a\right) + P\left(\sup_{n\geq N}\left|X^{(k)} - X_n^{(k)}\right| \geq a\right) \\
&\qquad + P\left(\sup_{n\geq N}\left|X_n^{(k)} - X_n\right| \geq a\right) \\
&\quad\leq \frac{2}{a}\mathbb{E}\left|X - X^{(k)}\right| + P\left(\sup_{n\geq N}\left|X^{(k)} - X_n^{(k)}\right| \geq a\right)
\end{aligned}$$

and hence

$$\limsup_{N\to\infty} P\left(\sup_{n\geq N}|X - X_n| \geq 3a\right) \leq \frac{2}{a}\mathbb{E}\left|X - X^{(k)}\right| \to 0 \text{ as } k \to \infty.$$

Thus we have shown

$$\limsup_{N\to\infty} P\left(\sup_{n\geq N}|X - X_n| \geq 3a\right) = 0 \text{ for all } a > 0.$$

Since

$$\left\{\limsup_{n\to\infty}|X - X_n| \geq 3a\right\} \subset \left\{\sup_{n\geq N}|X - X_n| \geq 3a\right\} \text{ for all } N,$$

it follows that

$$P\left(\limsup_{n\to\infty}|X - X_n| \geq 3a\right) = 0 \text{ for all } a > 0$$

and therefore $\limsup_{n\to\infty}|X - X_n| = 0$ (P a.s.) which shows that $X \in \mathbb{H}$. (This proof works equally as well in the case that X is a Banach valued random variable. One only needs to replace the absolute values in the proof by the Banach norm.) ■

19

Some Martingale Examples and Applications

Exercise 19.1. Let S_n be the total assets of an insurance company in year $n \in \mathbb{N}_0$. Assume $S_0 > 0$ is a constant and that for all $n \geq 1$ that $S_n = S_{n-1} + \xi_n$, where $\xi_n = c - Z_n$ and $\{Z_n\}_{n=1}^{\infty}$ are i.i.d. random variables having the normal distribution with mean $\mu < c$ and variance σ^2. (The number c is to be interpreted as the yearly premium.) Let $R = \{S_n \leq 0 \text{ for some } n\}$ be the event that the company eventually becomes bankrupt, i.e. is **R**uined. Show

$$P(\text{Ruin}) = P(R) \leq e^{-2(c-\mu)S_0/\sigma^2}.$$

Solution to Exercise (19.1). Let us first find λ such that $1 = \mathbb{E}\left[e^{\lambda \xi_n}\right]$. To do this let N be a standard normal random variable in which case,

$$1 \overset{\text{set}}{=} \mathbb{E}\left[e^{\lambda \xi_n}\right] = \mathbb{E}\left[e^{\lambda(c-\mu-\sigma N)}\right] = e^{\lambda(c-\mu)} e^{\left(\sigma^2\lambda^2\right)/2},$$

leads to the equation for λ;

$$\frac{\sigma^2}{2}\lambda^2 + \lambda\,(c-\mu) = 0.$$

Hence we should take $\lambda = -2\,(c-\mu)/\sigma^2$ – the other solution, $\lambda = 0$, is uninteresting. Since $\mathbb{E}\left[e^{\lambda \xi_n}\right] = 1$, we know from Example 18.11 that

$$Y_n := \exp(\lambda S_n) = e^{\lambda S_0} \prod_{j=1}^{n} e^{\lambda \xi_j}$$

is a non-negative $\mathcal{B}_n = \sigma(Z_1, \ldots, Z_n)$ – martingale. By the super-martingale or the sub-martingale convergence theorem (see Corollaries 18.63 and 18.54), it follows that $\lim_{n\to\infty} Y_n = Y_\infty$ exists a.s. Thus if τ is any stopping time we will have;

$$\mathbb{E}Y_\tau = \mathbb{E}\lim_{n\to\infty} Y_{\tau\wedge n} \leq \liminf_{n\to\infty} \mathbb{E}Y_{\tau\wedge n} = \mathbb{E}Y_0 = e^{\lambda S_0}$$

as follows from Fatou's Lemma and the optional sampling Theorem 18.39. If $\tau = \inf\{n : S_n \leq 0\}$ is the time of the companies ruin, we have $S_\tau \leq 0$ on $R = \{\tau < \infty\}$ and because $\lambda < 0$ and $S_\tau \leq 0$ on R, it follows that $Y_\tau = e^{\lambda S_\tau} \geq 1$ on R. This leads to the desired estimate;

$$P(R) \leq \mathbb{E}[Y_\tau : \tau < \infty] \leq \mathbb{E}Y_\tau \leq e^{\lambda S_0} = e^{-2(c-\mu)S_0/\sigma^2}.$$

Observe that by the strong law of large numbers that $\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}\xi_1 = c - \mu > 0$ a.s. Thus for large n we have $S_n \sim n\,(c-\mu) \to \infty$ as $n \to \infty$. The question we have addressed is what happens to the S_n for intermediate values – in particular what is the likely hood that S_n makes a sufficiently "large deviation" from the "typical" value of $n\,(c-\mu)$ in order for the company to go bankrupt.

19.1 A Polya Urn Model

In this section we are going to analyze the long run behavior of the Polya urn Markov process which was introduced in Exercise 17.4. Recall that if the urn contains r red balls and g green balls at a given time we draw one of these balls at random and replace it and add c more balls of the same color drawn. Let (r_n, g_n) be the number of red and green balls in the earn at time n. Then we have

$$P\left((r_{n+1}, g_n) = (r+c, g) \,|\, (r_n, g_n) = (r, g)\right) = \frac{r}{r+g} \text{ and}$$
$$P\left((r_{n+1}, g_n) = (r, g+c) \,|\, (r_n, g_n) = (r, g)\right) = \frac{g}{r+g}.$$

Let us observe that $r_n + g_n = r_0 + g_0 + nc$ and hence if we let X_n be the fraction of green balls in the urn at time n,

$$X_n := \frac{g_n}{r_n + g_n},$$

then

$$X_n := \frac{g_n}{r_n + g_n} = \frac{g_n}{r_0 + g_0 + nc}.$$

We now claim that $\{X_n\}_{n=0}^{\infty}$ is a martingale relative to

$$\mathcal{B}_n := \sigma\left((r_k, g_k) : k \leq n\right) = \sigma\left(X_k : k \leq n\right).$$

Indeed,

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{B}_n] &= \mathbb{E}[X_{n+1}|X_n] \\ &= \frac{r_n}{r_n+g_n}\cdot\frac{g_n}{r_n+g_n+c} + \frac{g_n}{r_n+g_n}\cdot\frac{g_n+c}{r_n+g_n+c} \\ &= \frac{g_n}{r_n+g_n}\cdot\frac{r_n+g_n+c}{r_n+g_n+c} = X_n.\end{aligned}$$

Since $X_n \ge 0$ and $\mathbb{E}X_n = \mathbb{E}X_0 < \infty$ for all n it follows by Corollary 18.54 that $X_\infty := \lim_{n\to\infty} X_n$ exists a.s. The distribution of X_∞ is described in the next theorem.

Theorem 19.1. *Let $\gamma := g/c$ and $\rho := r/c$ and $\mu := \mathrm{Law}_P(X_\infty)$. Then μ is the* ***beta distribution on*** $[0,1]$ *with parameters, γ, ρ, i.e.*

$$d\mu(x) = \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} x^{\gamma-1}(1-x)^{\rho-1}\,dx \text{ for } x\in[0,1]. \tag{19.1}$$

Proof. We will begin by computing the distribution of X_n. As an example, the probability of drawing 3 greens and then 2 reds is

$$\frac{g}{r+g}\cdot\frac{g+c}{r+g+c}\cdot\frac{g+2c}{r+g+2c}\cdot\frac{r}{r+g+3c}\cdot\frac{r+c}{r+g+4c}.$$

More generally, the probability of first drawing m greens and then $n-m$ reds is

$$\frac{g\cdot(g+c)\cdot\dots\cdot(g+(n-1)c)\cdot r\cdot(r+c)\cdot\dots\cdot(r+(n-m-1)c)}{(r+g)\cdot(r+g+c)\cdot\dots\cdot(r+g+(n-1)c)}.$$

Since this is the same probability for any of the $\binom{n}{m}$ – ways of drawing m greens and $n-m$ reds in n draws we have

$$\begin{aligned}&P(\text{Draw } m - \text{greens}) \\ &= \binom{n}{m}\frac{g\cdot(g+c)\cdot\dots\cdot(g+(m-1)c)\cdot r\cdot(r+c)\cdot\dots\cdot(r+(n-m-1)c)}{(r+g)\cdot(r+g+c)\cdot\dots\cdot(r+g+(n-1)c)} \\ &= \binom{n}{m}\frac{\gamma\cdot(\gamma+1)\cdot\dots\cdot(\gamma+(m-1))\cdot\rho\cdot(\rho+1)\cdot\dots\cdot(\rho+(n-m-1))}{(\rho+\gamma)\cdot(\rho+\gamma+1)\cdot\dots\cdot(\rho+\gamma+(n-1))}.\end{aligned} \tag{19.2}$$

Before going to the general case let us warm up with the special case, $g=r=c=1$. In this case Eq. (19.2) becomes,

$$P(\text{Draw } m - \text{greens}) = \binom{n}{m}\frac{1\cdot2\cdot\dots\cdot m\cdot1\cdot2\cdot\dots\cdot(n-m)}{2\cdot3\cdot\dots\cdot(n+1)} = \frac{1}{n+1}.$$

On the set, $\{\text{Draw } m - \text{greens}\}$, we have $X_n = \frac{1+m}{2+n}$ and hence it follows that for any $f\in C([0,1])$ that

$$\begin{aligned}\mathbb{E}[f(X_n)] &= \sum_{m=0}^n f\left(\frac{m+1}{n+2}\right)\cdot P(\text{Draw } m - \text{greens}) \\ &= \sum_{m=0}^n f\left(\frac{m+1}{n+2}\right)\frac{1}{n+1}.\end{aligned}$$

Therefore

$$\mathbb{E}[f(X)] = \lim_{n\to\infty}\mathbb{E}[f(X_n)] = \int_0^1 f(x)\,dx \tag{19.3}$$

and hence we may conclude that X_∞ has the uniform distribution on $[0,1]$.

For the general case, recall from Example 7.50 that $n! = \Gamma(n+1)$, $\Gamma(t+1) = t\Gamma(t)$, and therefore for $m\in\mathbb{N}$,

$$\Gamma(x+m) = (x+m-1)(x+m-2)\dots(x+1)x\Gamma(x). \tag{19.4}$$

Also recall Stirling's formula in Eq. (7.53) (also see Theorem 7.60) that

$$\Gamma(x) = \sqrt{2\pi}x^{x-1/2}e^{-x}[1+r(x)] \tag{19.5}$$

where $|r(x)|\to0$ as $x\to\infty$. To finish the proof we will follow the strategy of the proof of Eq. (19.3) using Stirling's formula to estimate the expression for $P(\text{Draw } m - \text{greens})$ in Eq. (19.2).

On the set, $\{\text{Draw } m - \text{greens}\}$, we have

$$X_n = \frac{g+mc}{r+g+nc} = \frac{\gamma+m}{\rho+\gamma+n} =: x_m,$$

where $\rho := r/c$ and $\gamma := g/c$. For later notice that $\Delta_m x = \frac{\gamma}{\rho+\gamma+n}$.

Using this notation we may rewrite Eq. (19.2) as

$$\begin{aligned}&P(\text{Draw } m - \text{greens}) \\ &= \binom{n}{m}\frac{\frac{\Gamma(\gamma+m)}{\Gamma(\gamma)}\cdot\frac{\Gamma(\rho+n-m)}{\Gamma(\rho)}}{\frac{\Gamma(\rho+\gamma+n)}{\Gamma(\rho+\gamma)}} \\ &= \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)}\cdot\frac{\Gamma(n+1)}{\Gamma(m+1)\Gamma(n-m+1)}\frac{\Gamma(\gamma+m)\Gamma(\rho+n-m)}{\Gamma(\rho+\gamma+n)}.\end{aligned} \tag{19.6}$$

Now by Stirling's formula,

$$\begin{aligned}\frac{\Gamma(\gamma+m)}{\Gamma(m+1)} &= \frac{(\gamma+m)^{\gamma+m-1/2}e^{-(\gamma+m)}[1+r(\gamma+m)]}{(1+m)^{m+1-1/2}e^{-(m+1)}[1+r(1+m)]} \\ &= (\gamma+m)^{\gamma-1}\cdot\left(\frac{\gamma+m}{m+1}\right)^{m+1/2}e^{-(\gamma-1)}\frac{1+r(\gamma+m)}{1+r(m+1)}. \\ &= (\gamma+m)^{\gamma-1}\cdot\left(\frac{1+\gamma/m}{1+1/m}\right)^{m+1/2}e^{-(\gamma-1)}\frac{1+r(\gamma+m)}{1+r(m+1)}\end{aligned}$$

We will keep m fairly large, so that

$$\left(\frac{1+\gamma/m}{1+1/m}\right)^{m+1/2} = \exp\left((m+1/2)\ln\left(\frac{1+\gamma/m}{1+1/m}\right)\right) \cong \exp\left((m+1/2)\left(\gamma/m-1/m\right)\right) \cong e^{\gamma-1}.$$

Hence we have

$$\frac{\Gamma(\gamma+m)}{\Gamma(m+1)} \asymp (\gamma+m)^{\gamma-1}.$$

Similarly, keeping $n-m$ fairly large, we also have

$$\frac{\Gamma(\rho+n-m)}{\Gamma(n-m+1)} \asymp (\rho+n-m)^{\rho-1} \text{ and}$$
$$\frac{\Gamma(\rho+\gamma+n)}{\Gamma(n+1)} \asymp (\rho+\gamma+n)^{\rho+\gamma-1}.$$

Combining these estimates with Eq. (19.6) gives,

$$\begin{aligned}
&P(\text{Draw } m-\text{greens}) \\
&\asymp \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot \frac{(\gamma+m)^{\gamma-1}\cdot(\rho+n-m)^{\rho-1}}{(\rho+\gamma+n)^{\rho+\gamma-1}} \\
&= \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot \frac{\left(\frac{\gamma+m}{\rho+\gamma+n}\right)^{\gamma-1}\cdot\left(\frac{\rho+n-m}{\rho+\gamma+n}\right)^{\rho-1}}{(\rho+\gamma+n)^{\rho+\gamma-1}} \\
&= \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot (x_m)^{\gamma-1}\cdot(1-x_m)^{\rho-1}\Delta_m x.
\end{aligned}$$

Therefore, for any $f \in C([0,1])$, it follows that

$$\begin{aligned}
\mathbb{E}[f(X_\infty)] &= \lim_{n\to\infty}\mathbb{E}[f(X_n)] \\
&= \lim_{n\to\infty}\sum_{m=0}^{n} f(x_m)\frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)}\cdot(x_m)^{\gamma-1}\cdot(1-x_m)^{\rho-1}\Delta_m x \\
&= \int_0^1 f(x)\frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)}x^{\gamma-1}(1-x)^{\rho-1}\, dx.
\end{aligned}$$

■

19.2 Galton Watson Branching Process

This section is taken from [14, p. 245 –249]. Let $\{\xi_i^n : i, n \geq 1\}$ be a sequence of i.i.d. non-negative integer valued random variables. Suppose that Z_n is the

number of people in the n^{th} – generation and $\xi_1^{n+1}, \dots, \xi_{Z_n}^{n+1}$ are the number of off spring of the Z_n people of generation n. Then

$$\begin{aligned}
Z_{n+1} &= \xi_1^{n+1} + \dots + \xi_{Z_n}^{n+1} \\
&= \sum_{k=1}^{\infty}\left(\xi_1^{n+1} + \dots + \xi_k^{n+1}\right) 1_{Z_n=k}.
\end{aligned} \tag{19.7}$$

represents the number of people present in generation, $n+1$. We complete the description of the process, Z_n by setting $Z_0 = 1$ and $Z_{n+1} = 0$ if $Z_n = 0$, i.e. once the population dies out it remains extinct forever after. The process $\{Z_n\}_{n\geq 0}$ is called a **Galton-Watson Branching** process, see Figure 19.1. To understand

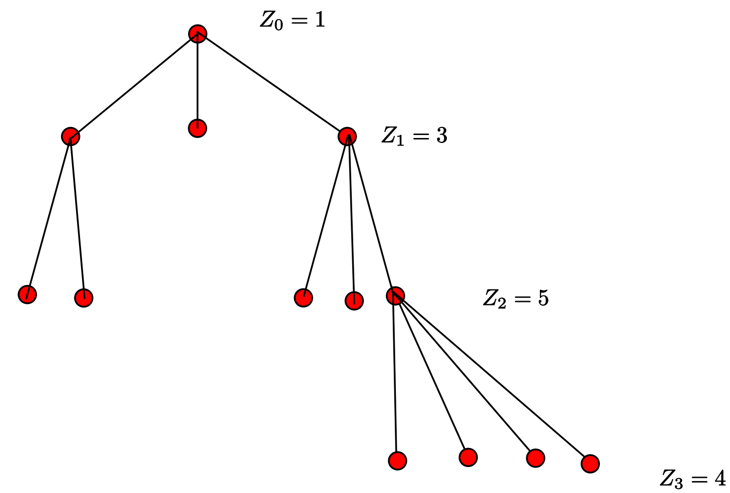


Fig. 19.1. A possible realization of a Galton Watson "tree."

Z_n a bit better observe that $Z_1 = \xi_1^1$, $Z_2 = \xi_1^2 + \dots + \xi_{\xi_1^1}^3$, $Z_3 = \xi_1^3 + \dots + \xi_{Z_2}^3$, etc. The sample path in Figure 19.1 corresponds to

$$\begin{aligned}
&\xi_1^1 = 3, \\
&\xi_1^2 = 2,\ \xi_2^2 = 0,\ \xi_3^2 = 3, \\
&\xi_1^3 = \xi_2^3 = \xi_3^3 = \xi_4^3 = 0,\ \xi_5^3 = 4, \text{ and} \\
&\xi_1^4 = \xi_2^4 = \xi_3^4 = \xi_4^4 = 0.
\end{aligned}$$

We will use later the intuitive fact that the different branches of the Galton-Watson tree evolve independently of one another – you will be asked to make this precise Exercise 19.4.

Let $\xi \overset{d}{=} \xi_i^m$, $p_k := P(\xi = k)$ be the **off-spring** distribution,

$$\mu := \mathbb{E}\xi = \sum_{k=0}^{\infty} k p_k,$$

which we assume to be finite.

Let $\mathcal{B}_0 = \{\emptyset, \Omega\}$ and

$$\mathcal{B}_n := \sigma\left(\xi_i^m : i \geq 1 \text{ and } 1 \leq m \leq n\right).$$

Making use of Eq. (19.7) and the independence of $\left\{\xi_i^{n+1}\right\}_{i=1}^{\infty}$ from $\mathcal{B}_n$ we find

$$\begin{aligned}
\mathbb{E}\left[Z_{n+1}|\mathcal{B}_n\right] &= \mathbb{E}\left[\sum_{k=1}^{\infty}\left(\xi_1^{n+1} + \dots + \xi_k^{n+1}\right) 1_{Z_n=k}|\mathcal{B}_n\right] \\
&= \sum_{k=1}^{\infty} 1_{Z_n=k}\mathbb{E}\left[\left(\xi_1^{n+1} + \dots + \xi_k^{n+1}\right)|\mathcal{B}_n\right] \\
&= \sum_{k=1}^{\infty} 1_{Z_n=k} \cdot k\mu = \mu Z_n. \qquad (19.8)
\end{aligned}$$

So we have shown, $M_n := Z_n/\mu^n$ is a positive martingale and since $M_0 = Z_0/\mu^0 = 1$ it follows that

$$1 = \mathbb{E}M_0 = \mathbb{E}M_n = \frac{\mathbb{E}Z_n}{\mu^n} \implies \mathbb{E}Z_n = \mu^n < \infty. \qquad (19.9)$$

Theorem 19.2. *If $\mu < 1$, then, almost surely, $Z_n = 0$ for a.a. n.*

Proof. When $\mu < 1$, we have

$$\mathbb{E}\sum_{n=0}^{\infty} Z_n = \sum_{n=0}^{\infty} \mu^n = \frac{1}{1-\mu} < \infty$$

and therefore $\sum_{n=0}^{\infty} Z_n < \infty$ a.s. As $Z_n \in \mathbb{N}_0$ for all n, this can only happen if $Z_n = 0$ for almost all n a.s. ∎

Theorem 19.3. *If $\mu = 1$ and $P(\xi_i^m = 1) < 1$,*[1] *then again, almost surely, $Z_n = 0$ for a.a. n.*

Proof. In this case $\{Z_n\}_{n=1}^{\infty}$ is a martingale which, being positive, is L^1 – bounded. Therefore, $\lim_{n\to\infty} Z_n =: Z_\infty$ exists. Because Z_n is integer valued, it must happen that $Z_n = Z_\infty$ a.a. If $k \in \mathbb{N}$, Since

[1] The assumption here is equivalent to $p_0 > 0$ and $\mu = 1$.

$$\{Z_\infty = k\} = \{Z_n = k \text{ a.a. } n\} = \cup_{N=1}^{\infty}\{Z_n = k \text{ for all } n \geq N\},$$

we have

$$P(Z_\infty = k) = \lim_{N\to\infty} P(Z_n = k \text{ for all } n \geq N).$$

However,

$$\begin{aligned}
P(Z_n = k \text{ for all } n \geq N) &= P\left(\xi_1^n + \dots + \xi_k^n = k \text{ for all } n \geq N\right) \\
&= \left[P\left(\xi_1^n + \dots + \xi_k^n = k\right)\right]^{\infty} = 0,
\end{aligned}$$

because, $P(\xi_1^n + \dots + \xi_k^n = k) < 1$. Indeed, since $p_1 = P(\xi = 1) < 1$ and $\mu = 1$ it follows that $p_l = P(\xi = l) > 0$ for some $l > 1$ and therefore if $P(\xi_1^n + \dots + \xi_k^n = kl) > 0$ which then implies $P(\xi_1^n + \dots + \xi_k^n = k) < 1$. Therefore we have shown $P(Z_\infty = k) = 0$ for all $k > 0$ and therefore, $Z_\infty = 0$ a.s. and hence almost surely, $Z_n = 0$ for a.a. n. ∎

Remark 19.4. By the way, the branching process, $\{Z_n\}_{n=0}^{\infty}$ with $\mu = 1$ and $P(\xi = 1) < 1$ gives a nice example of a non regular martingale. Indeed, if Z were regular, we would have

$$Z_n = \mathbb{E}\left[\lim_{m\to\infty} Z_m|\mathcal{B}_n\right] = \mathbb{E}\left[0|\mathcal{B}_n\right] = 0$$

which is clearly false.

We now wish to consider the case where $\mu := \mathbb{E}\left[\xi_i^m\right] > 1$. Let $\xi \overset{d}{=} \xi_i^m$ and for $\lambda \in \mathbb{C}$ with $|\lambda| \leq 1$ we let

$$\varphi(\lambda) := \mathbb{E}\left[\lambda^{\xi}\right] = \sum_{k\geq 0} p_k \lambda^k.$$

Notice that $\varphi(1) = 1$ and for $\lambda = s \in (-1, 1)$ we have

$$\varphi'(s) = \sum_{k\geq 0} k p_k s^{k-1} \text{ and } \varphi''(s) = \sum_{k\geq 0} k(k-1) p_k s^{k-2} \geq 0$$

with

$$\begin{aligned}
\lim_{s\uparrow 1}\varphi'(s) &= \sum_{k\geq 0} k p_k = \mathbb{E}[\xi] =: \mu \text{ and} \\
\lim_{s\uparrow 1}\varphi''(s) &= \sum_{k\geq 0} k(k-1) p_k = \mathbb{E}[\xi(\xi-1)].
\end{aligned}$$

Therefore φ is convex with $\varphi(0) = p_0$, $\varphi(1) = 1$ and $\varphi'(1) = \mu$.

Lemma 19.5. *If $\mu = \varphi'(1) > 1$, there exists a unique $\rho < 1$ so that $\varphi(\rho) = \rho$.*

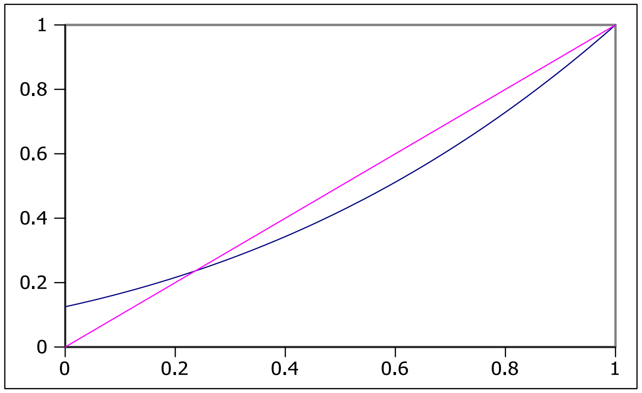


Fig. 19.2. Figure associated to $\varphi(s) = \frac{1}{8}\left(1 + 3s + 3s^2 + s^3\right)$ which is relevant for Exercise 3.13 of Durrett on p. 249. In this case $\rho \cong 0.236\,07$.

Proof. See Figure 19.2 below. ∎

Theorem 19.6 (See Durrett [14], p. 247-248.). *If $\mu > 1$, then*

$$P(\{Z_n = 0 \text{ for some } n\}) = \rho.$$

Proof. Since $\{Z_m = 0\} \subset \{Z_{m+1} = 0\}$, it follows that $\{Z_m = 0\} \uparrow \{Z_n = 0 \text{ for some } n\}$ and therefore if

$$\theta_m := P(Z_m = 0),$$

then

$$P(\{Z_n = 0 \text{ for some } n\}) = \lim_{m\to\infty} \theta_m.$$

We now show; $\theta_m = \varphi(\theta_{m-1})$. To see this, conditioned on the set $\{Z_1 = k\}$, $Z_m = 0$ iff all k – families die out in the remaining $m-1$ time units. Since each family evolves independently, the probability[2] of this event is θ_{m-1}^k. Combining this with, $P(\{Z_1 = k\}) = P\left(\xi_1^1 = k\right) = p_k$, allows us to conclude,

$$\theta_m = P(Z_m = 0) = \sum_{k=0}^{\infty} P(Z_m = 0, Z_1 = k)$$
$$= \sum_{k=0}^{\infty} P(Z_m = 0 | Z_1 = k) P(Z_1 = k) = \sum_{k=0}^{\infty} p_k \theta_{m-1}^k = \varphi(\theta_{m-1}).$$

It is now easy to see that $\theta_m \uparrow \rho$ as $m \uparrow \infty$, again see Figure 19.3.

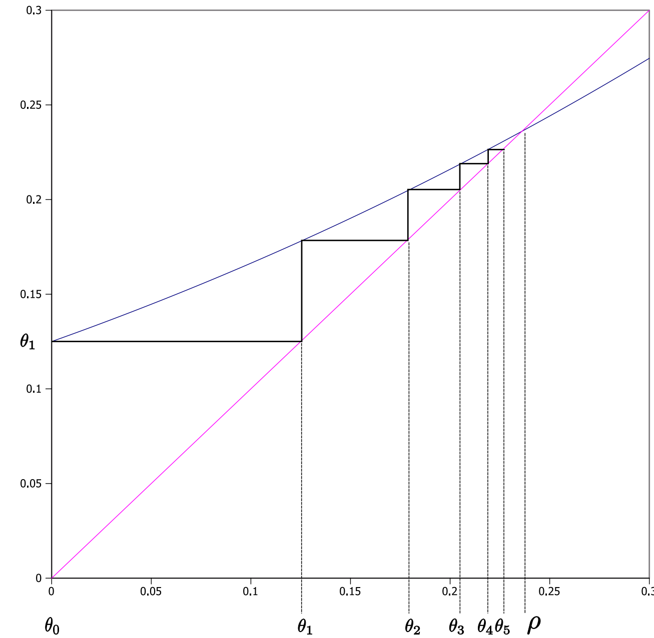


Fig. 19.3. The graphical interpretation of $\theta_m = \varphi(\theta_{m-1})$ starting with $\theta_0 = 0$.

∎

Let $S = \mathbb{N}_0$, $\{Y_i\}_{i=1}^{\infty}$ be i.i.d. S – valued random variables such that $Y \overset{d}{=} \xi_i^n$ (i.e. $P(Y_i = l) = p_l$ for $l \in \mathbb{N}_0$), and for $f : S \to \mathbb{C}$ bounded or non-negative let $Qf(0) = f(0)$ and

$$Qf(k) := \mathbb{E}[f(Y_1 + \dots + Y_k)] \text{ for all } k \geq 1.$$

Notice that

$$Qf(k) = \sum_{l \in S} f(l) p_l^{*k}$$

where

$$p_l^{*k} := P(Y_1 + \dots + Y_k = l) = \sum_{l_1 + \dots + l_k = l} p_{l_1} \dots p_{l_k}$$

with the convention that $p_n^{*0} = \delta_{0,n}$. As above, for $\lambda \in \mathbb{C}$ with $|\lambda| \leq 1$ we let

$$\varphi(\lambda) := \mathbb{E}\left[\lambda^{Y_1}\right] = \sum_{k \geq 0} p_k \lambda^k$$

be the moment generating function of for the Y_i.

[2] This argument is made precise with the aid of Exercise 19.4.

Exercise 19.2. Let $\mathcal{B}_n := \mathcal{B}_n^Z = \sigma(Z_0, \dots, Z_n)$ so that $(\Omega, \mathcal{B}, \mathcal{B}_n, P)$ is a filtered probability space. Show that $\{Z_n\}_{n=0}^\infty$ is a time homogeneous Markov process with one step transition kernel being Q. In particular verify that

$$P(Z_n = j | Z_{n-1} = k) = p_j^{*k} \text{ for all } j, k \in S \text{ and } n \geq 1$$

and

$$\mathbb{E}\left[\lambda^{Z_n} | \mathcal{B}_{n-1}\right] = \varphi(\lambda)^{Z_{n-1}} \text{ a.s.} \tag{19.10}$$

for all $\lambda \in \mathbb{C}$ with $|\lambda| \leq 1$.

Exercise 19.3. In the notation used in this section (Section 19.2), show for all $n \in \mathbb{N}$ and $\lambda_i \in \mathbb{C}$ with $|\lambda_i| \leq 1$ that

$$\mathbb{E}\left[\prod_{j=1}^n \lambda_i^{Z_i}\right] = \varphi(\lambda_1 \varphi(\dots \lambda_{n-2} \varphi(\lambda_{n-1} \varphi(\lambda_n)))).$$

For example you should show,

$$\mathbb{E}\left[\lambda_1^{Z_1} \lambda_2^{Z_2} \lambda_3^{Z_3}\right] = \varphi(\lambda_1 \varphi(\lambda_2 \varphi(\lambda_3)))$$

and

$$\mathbb{E}\left[\lambda_1^{Z_1} \lambda_2^{Z_2} \lambda_3^{Z_3} \lambda_4^{Z_4}\right] = \varphi(\lambda_1 \varphi(\lambda_2 \varphi(\lambda_3 \varphi(\lambda_4)))).$$

Exercise 19.4. Suppose that $n \geq 2$ and $f : \mathbb{N}_0^{n-1} \to \mathbb{C}$ is a bounded function or a non-negative function. Show for all $k \geq 1$ that

$$\mathbb{E}\left[f(Z_2, \dots, Z_n) | Z_1 = k\right] = \mathbb{E}\left[f\left(\sum_{l=1}^k \left(Z_1^l, \dots, Z_{n-1}^l\right)\right)\right] \tag{19.11}$$

where $\left\{Z_n^l\right\}_{n=0}^\infty$ for $1 \leq l \leq k$ are i.i.d. Galton-Watson Branching processes such that $\left\{Z_n^l\right\}_{n=0}^\infty \stackrel{d}{=} \{Z_n\}_{n=0}^\infty$ for each l.

Suggestion: it suffices to prove Eq. (19.11) for f of the form,

$$f(k_2, \dots, k_n) = \prod_{j=2}^n \lambda_i^{k_i}. \tag{19.12}$$

19.3 Kakutani's Theorem

For broad generalizations of the results in this section, see [26, Chapter IV.] or [27].

Proposition 19.7. *Suppose that μ and ν are σ – finite positive measures on $(X, \mathcal{M})$, $\nu = \nu_a + \nu_s$ is the Lebesgue decomposition of ν relative to μ, and $\rho : X \to [0, \infty)$ is a measurable function such that $d\nu_a = \rho d\mu$ so that*

$$d\nu = d\nu_a + d\nu_s = \rho d\mu + d\nu_s.$$

If $g : X \to [0, \infty)$ is another measurable function such that $g d\mu \leq d\nu$, (i.e. $\int_B g d\mu \leq \nu(B)$ for all $B \in \mathcal{M}$), then $g \leq \rho$, μ – a.e.

Proof. Let $A \in \mathcal{M}$ be chosen so that $\mu(A^c) = 0$ and $\nu_s(A) = 0$. Then, for all $B \in \mathcal{M}$,

$$\int_B g d\mu = \int_{B \cap A} g d\mu \leq \nu(B \cap A) = \int_{B \cap A} \rho d\mu = \int_B \rho d\mu.$$

So by the comparison Lemma 7.24, $g \leq \rho$. ■

Example 19.8. This example generalizes Example 18.9. Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and Q is **any** another probability measure on $(\Omega, \mathcal{B})$. By the Raydon-Nikodym Theorem 15.8, for each $n \in \bar{\mathbb{N}}$ we may write

$$dQ|_{\mathcal{B}_n} = X_n dP|_{\mathcal{B}_n} + dR_n \tag{19.13}$$

where R_n is a measure on $(\Omega, \mathcal{B}_n)$ which is singular relative to $P|_{\mathcal{B}_n}$ and $0 \leq X_n \in L^1(\Omega, \mathcal{B}_n, P)$. In this case the most we can say in general is that $X := \{X_n\}_{n \leq \infty}$ is a positive supermartingale. To verify this assertion, for $B \in \mathcal{B}_n$ and $n \leq m \leq \infty$, we have

$$Q(B) = \mathbb{E}[X_m : B] + R_m(B) \geq \mathbb{E}[X_m : B] = \mathbb{E}[\mathbb{E}_{\mathcal{B}_n}(X_m) : B]$$

from which it follows that $\mathbb{E}_{\mathcal{B}_n}(X_m) \cdot dP|_{\mathcal{B}_n} \leq dQ|_{\mathcal{B}_n}$. So according to Proposition 19.7,

$$\mathbb{E}_{\mathcal{B}_n}(X_m) \leq X_n \ (P - \text{a.s.}) \text{ for all } n \leq m \leq \infty. \tag{19.14}$$

Proposition 19.9. *Keeping the assumptions and notation used in Example 19.8, then $\lim_{n\to\infty} X_n = X_\infty$ a.s. and in particular the Lebesgue decomposition of $Q|_{\mathcal{B}_\infty}$ relative to $P|_{\mathcal{B}_\infty}$ may be written as*

$$dQ|_{\mathcal{B}_\infty} = \left(\lim_{n\to\infty} X_n\right) \cdot dP|_{\mathcal{B}_\infty} + dR_\infty. \tag{19.15}$$

Proof. By Example 19.8, we know that $\{X_n\}_{n \leq \infty}$ is a positive supermartingale and by letting $m = \infty$ in Eq. (19.14), we know

$$\mathbb{E}_{\mathcal{B}_n} X_\infty \leq X_n \text{ a.s.} \tag{19.16}$$

By the supermartingale convergence Corollary 18.63 or by the submartingale convergence Corollary 18.54 applied to $-X_n$ we know that $Y := \lim_{n\to\infty} X_n$ exists almost surely. To finish the proof it suffices to show that $Y = X_\infty$ a.s. where X_∞ is defined so that Eq. (19.13) holds for $n = \infty$.

From the regular martingale convergence Theorem 18.65 we also know that $\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}_n} X_\infty = X_\infty$ a.s. as well. So passing to the limit in Eq. (19.16) implies $X_\infty \le Y$ a.s. To prove the reverse inequality, $Y \le X_\infty$ a.s., let $B \in \mathcal{B}_m$ and $n \ge m$. Then

$$Q(B) = \mathbb{E}[X_n : B] + R_n(B) \ge \mathbb{E}[X_n : B]$$

and so by Fatou's lemma,

$$\mathbb{E}[Y : B] = \mathbb{E}\left[\liminf_{n\to\infty} X_n : B\right] \le \liminf_{n\to\infty} \mathbb{E}[X_n : B] \le Q(B). \tag{19.17}$$

Since $m \in \mathbb{N}$ was arbitrary, we have proved $\mathbb{E}[Y : B] \le Q(B)$ for all B in the algebra, $\mathcal{A} := \cup_{m\in\mathbb{N}} \mathcal{B}_m$. As a consequence of the regularity Theorem 5.44 or of the monotone class Lemma 5.52, or of Theorem[3] 5.27, it follows that $\mathbb{E}[Y : B] \le Q(B)$ for all $B \in \sigma(\mathcal{A}) = \mathcal{B}_\infty$. An application of Proposition 19.7 then implies $Y \le X_\infty$ a.s. ∎

Theorem 19.10. $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ *be a filtered probability space and Q be a probability measure on $(\Omega, \mathcal{B})$ such that $Q|_{\mathcal{B}_n} \ll P|_{\mathcal{B}_n}$ for all $n \in \mathbb{N}$. Let $M_n := \frac{dQ|_{\mathcal{B}_n}}{dP|_{\mathcal{B}_n}}$ be a version of the Raydon-Nikodym derivative of $Q|_{\mathcal{B}_n}$ relative to $P|_{\mathcal{B}_n}$, see Theorem 15.8. Recall from Example 18.9 that $\{M_n\}_{n=1}^\infty$ is a positive martingale and let $M_\infty = \lim_{n\to\infty} M_n$ which exists a.s. Then the following are equivalent;*

1. $Q|_{\mathcal{B}_\infty} \ll P|_{\mathcal{B}_\infty}$,
2. $\mathbb{E}_P M_\infty = 1$,
3. $M_n \to M_\infty$ *in* $L^1(P)$, *and*
4. $\{M_n\}_{n=1}^\infty$ *is uniformly integrable.*

Proof. Recall from Proposition 19.9 (where X_n is now M_n) that in general,

$$dQ|_{\mathcal{B}_\infty} = M_\infty \cdot dP|_{\mathcal{B}_\infty} + dR_\infty \tag{19.18}$$

where R_∞ is singular relative to $P|_{\mathcal{B}_\infty}$. Therefore, $Q|_{\mathcal{B}_\infty} \ll P|_{\mathcal{B}_\infty}$ iff $R_\infty = 0$ which happens iff $R_\infty(\Omega) = 0$, i.e. iff

$$1 = Q(\Omega) = \int_\Omega M_\infty \cdot dP|_{\mathcal{B}_\infty} = \mathbb{E}_P M_\infty.$$

This proves the equivalence of items 1. and 2. If item 2. holds, then $M_n \to M_\infty$ by the DCT, Corollary 12.9, with $g_n = f_n = M_n$ and $g = f = M_\infty$ and so item 3. holds. The implication of 3. $\Longrightarrow$ 2. is easy and the equivalence of items 3. and 4. follows from Theorem 12.44 for simply see Theorem 18.65. ∎

Remark 19.11. Recall from Exercise 10.10, that if $0 < a_n \le 1$, $\prod_{n=1}^\infty a_n > 0$ iff $\sum_{n=1}^\infty (1 - a_n) < \infty$. Indeed, $\prod_{n=1}^\infty a_n > 0$ iff

$$-\infty < \ln\left(\prod_{n=1}^\infty a_n\right) = \sum_{n=1}^\infty \ln a_n = \sum_{n=1}^\infty \ln(1 - (1 - a_n))$$

and $\sum_{n=1}^\infty \ln(1 - (1 - a_n)) > -\infty$ iff $\sum_{n=1}^\infty (1 - a_n) < \infty$. Recall that $\ln(1 - (1 - a_n)) \cong (1 - a_n)$ for a_n near 1.

Theorem 19.12 (Kakutani's Theorem). *Let $\{X_n\}_{n=1}^\infty$ be independent non-negative random variables with $\mathbb{E}X_n = 1$ for all n. Further, let $M_0 = 1$ and $M_n := X_1 \cdot X_2 \cdot \dots \cdot X_n$ – a martingale relative to the filtration, $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$ as was shown in Example 18.11. According to Corollary 18.63, $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. and $\mathbb{E}M_\infty \le 1$. The following statements are equivalent;*

1. $\mathbb{E}M_\infty = 1$,
2. $M_n \to M_\infty$ *in* $L^1(\Omega, \mathcal{B}, P)$,
3. $\{M_n\}_{n=1}^\infty$ *is uniformly integrable,*
4. $\prod_{n=1}^\infty \mathbb{E}\left(\sqrt{X_n}\right) > 0$,
5. $\sum_{n=1}^\infty \left(1 - \mathbb{E}\left(\sqrt{X_n}\right)\right) < \infty$.

*Moreover, if any one, and hence all of the above statements, **fails** to hold, then $P(M_\infty = 0) = 1$.*

Proof. If $a_n := \mathbb{E}\left(\sqrt{X_n}\right)$, then $0 < a_n$ and $a_n^2 \le \mathbb{E}X_n = 1$ with equality iff $X_n = 1$ a.s. So Remark 19.11 gives the equivalence of items 4. and 5.

The equivalence of items 1., 2. and 3. follow by the same techniques used in the proof of Theorem 19.10 above. We will now complete the proof by showing 4. $\Longrightarrow$ 3. and not(4.) $\Longrightarrow$ $P(M_\infty = 0) = 1$ which clearly implies not(1.). For both pars of the argument, let $N_0 = 1$ and N_n be the martingale (again see Example 18.11) defined by

$$N_n := \prod_{k=1}^n \frac{\sqrt{X_k}}{a_k} = \frac{\sqrt{M_n}}{\prod_{k=1}^n a_k}. \tag{19.19}$$

[3] This theorem implies that for $B \in \mathcal{B}$,

$$\mathbb{E}[X_0 : B] = \inf\{\mathbb{E}[X_0 : A] : A \in \mathcal{A}_\sigma\} \text{ and}$$
$$Q(B) = \inf\{Q(A) : A \in \mathcal{A}_\sigma\}$$

and since, by MCT, $\mathbb{E}[X_0 : A] \le Q(A)$ for all $A \in \mathcal{A}_\sigma$ it follows that Eq. (19.17) holds for all $B \in \mathcal{B}$.

Further observe that, in all cases, $N_\infty = \lim_{n\to\infty} N_n$ exists in $[0,\infty)$ μ – a.s., see Corollary 18.54 or Corollary 18.63.

4. $\implies$ 3. Since

$$N_n^2 = \prod_{k=1}^n \frac{X_k}{a_k^2} = \frac{M_n}{\left(\prod_{k=1}^n a_k\right)^2},$$

$$\mathbb{E}\left[N_n^2\right] = \frac{\mathbb{E}M_n}{\left(\prod_{k=1}^n a_k\right)^2} = \frac{1}{\left(\prod_{k=1}^n a_k\right)^2} \le \frac{1}{\left(\prod_{k=1}^\infty a_k\right)^2} < \infty,$$

and hence $\{N_n\}_{n=1}^\infty$ is bounded in L^2. Therefore, using

$$M_n = \left(\prod_{k=1}^n a_k\right)^2 N_n^2 \le N_n^2 \tag{19.20}$$

and Doob's inequality in Corollary 18.47, we find

$$\mathbb{E}\left[\sup_n M_n\right] = \mathbb{E}\left[\sup_n N_n^2\right] \le 4 \sup_n \mathbb{E}\left[N_n^2\right] < \infty. \tag{19.21}$$

Equation Eq. (19.21) certainly implies $\{M_n\}_{n=1}^\infty$ is uniformly integrable, see Proposition 12.42.

Not(4.) $\implies$ $P(M_\infty = 0) = 1$. If

$$\prod_{n=1}^\infty \mathbb{E}\left(\sqrt{X_n}\right) = \lim_{n\to\infty} \prod_{k=1}^n a_k = 0,$$

we may pass to the limit in Eq. (19.20) to find

$$M_\infty = \lim_{n\to\infty} M_n = \lim_{n\to\infty}\left[\left(\prod_{k=1}^n a_k\right)^2 \cdot N_n^2\right] = 0 \cdot \left(\lim_{n\to\infty} N_n\right)^2 = 0 \text{ a.s..}$$

■

Lemma 19.13. *Given two probability measures, μ and ν on a measurable space, $(\Omega,\mathcal{B})$, there exists a a positive measure ρ such that $d\rho := \sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda$, where λ is any other σ – finite measure on $(\Omega,\mathcal{B})$ such that $\mu \ll \lambda$ and $\nu \ll \lambda$. We will write $\sqrt{d\mu\cdot d\nu}$ for $d\rho$ in the future.*

Proof. The main point is to show that ρ is well defined. So suppose λ_1 and λ_2 are two σ – finite measures such that $\mu \ll \lambda_i$ and $\nu \ll \lambda_i$ for $i = 1,2$. Further let $\lambda := \lambda_1 + \lambda_2$ so that $\lambda_i \ll \lambda$ for $i = 1,2$. Observe that

$$\begin{aligned} d\lambda_1 &= \frac{d\lambda_1}{d\lambda}d\lambda, \\ d\mu &= \frac{d\mu}{d\lambda_1}d\lambda_1 = \frac{d\mu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}d\lambda, \text{ and} \\ d\nu &= \frac{d\nu}{d\lambda_1}d\lambda_1 = \frac{d\nu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}d\lambda. \end{aligned}$$

So

$$\begin{aligned} \sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda &= \sqrt{\frac{d\mu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}\cdot\frac{d\nu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}}d\lambda \\ &= \sqrt{\frac{d\mu}{d\lambda_1}\cdot\frac{d\nu}{d\lambda_1}}\frac{d\lambda_1}{d\lambda}d\lambda = \sqrt{\frac{d\mu}{d\lambda_1}\cdot\frac{d\nu}{d\lambda_1}}d\lambda_1 \end{aligned}$$

and by symmetry,

$$\sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda = \sqrt{\frac{d\mu}{d\lambda_2}\cdot\frac{d\nu}{d\lambda_2}}d\lambda_2.$$

This shows

$$\sqrt{\frac{d\mu}{d\lambda_2}\cdot\frac{d\nu}{d\lambda_2}}d\lambda_2 = \sqrt{\frac{d\mu}{d\lambda_1}\cdot\frac{d\nu}{d\lambda_1}}d\lambda_1$$

and hence $d\rho = \sqrt{d\mu\cdot d\nu}$ is well defined. ■

Definition 19.14. *Two probability measures, μ and ν on a measure space, $(\Omega,\mathcal{B})$ are said to be* ***equivalent*** *(written $\mu \sim \nu$) if $\mu \ll \nu$ and $\nu \ll \mu$, i.e. if μ and ν are absolutely continuous relative to one another. The* ***Hellinger integral of*** *μ* ***and*** *ν is defined as*

$$H(\mu,\nu) := \int_\Omega \sqrt{d\mu\cdot d\nu} = \int_\Omega \sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda \tag{19.22}$$

where λ is any measure (for example $\lambda = \frac{1}{2}(\mu+\nu)$ would work) on $(\Omega,\mathcal{B})$ such that there exists, $\frac{d\mu}{d\lambda}$ and $\frac{d\nu}{d\lambda}$ in $L^1(\Omega,\mathcal{B},\lambda)$ such that $d\mu = \frac{d\mu}{d\lambda}d\lambda$ and $d\nu = \frac{d\nu}{d\lambda}d\lambda$. Lemma 19.13 guarantees that $H(\mu,\nu)$ is well defined.

Proposition 19.15. *The Hellinger integral, $H(\mu,\nu)$, of two probability measures, μ and ν, is well defined. Moreover $H(\mu,\nu)$ satisfies;*

1. *$0 \le H(\mu,\nu) \le 1$,*
2. *$H(\mu,\nu) = 1$ iff $\mu = \nu$,*
3. *$H(\mu,\nu) = 0$ iff $\mu \perp \nu$, and*
4. *If $\mu \sim \nu$ or more generally if $\nu \ll \mu$, then $H(\mu,\nu) > 0$.*

Furthermore[4],

$$H(\mu,\nu)=\inf\left\{\sum_{i=1}^{n}\sqrt{\mu(A_i)\,\nu(A_i)}:\Omega=\sum_{i=1}^{n}A_i \text{ and } n\in\mathbb{N}\right\}. \tag{19.23}$$

Proof. Items 1. and 2. are both an easy consequence of the Schwarz inequality and its converse. For item 3., if $H(\mu,\nu)=0$, then $\frac{d\mu}{d\lambda}\cdot\frac{dv}{d\lambda}=0$, λ – a.e.. Therefore, if we let

$$A:=\left\{\frac{d\mu}{d\lambda}\neq 0\right\},$$

then $\frac{d\mu}{d\lambda}=1_A\frac{d\mu}{d\lambda}$ – λ –a.e. and $\frac{dv}{d\lambda}1_{A^c}=\frac{dv}{d\lambda}$ – λ – a.e. Hence it follows that $\mu(A^c)=0$ and $\nu(A)=0$ and hence $\mu\perp\nu$.

If $\nu\sim\mu$ and in particular, $v\ll\mu$, then

$$H(\mu,\nu)=\int_\Omega\sqrt{\frac{d\nu}{d\mu}\frac{d\mu}{d\mu}}d\mu=\int_\Omega\sqrt{\frac{d\nu}{d\mu}}d\mu.$$

For sake of contradiction, if $H(\mu,\nu)=0$ then $\sqrt{\frac{d\nu}{d\mu}}=0$ and hence $\frac{d\nu}{d\mu}=0$, μ – a.e. The later would imply $\nu=0$ which is impossible. Therefore, $H(\mu,\nu)>0$ if $\nu\ll\mu$. The last statement is left to the reader as Exercise 19.6. ∎

Exercise 19.5. Find a counter example to the statement that $H(\mu,\nu)>0$ implies $\nu\ll\mu$.

Exercise 19.6. Prove Eq. (19.23).

Corollary 19.16 (Kakutani [29]). *Let $\Omega=\mathbb{R}^{\mathbb{N}}$, $Y_n(\omega)=\omega_n$ for all $\omega\in\Omega$ and $n\in\mathbb{N}$, and $\mathcal{B}:=\mathcal{B}_\infty=\sigma(Y_n:n\in\mathbb{N})$ be the product σ – algebra on Ω. Further, let $\mu:=\otimes_{n=1}^\infty\mu_n$ and $\nu:=\otimes_{n=1}^\infty\nu_n$ be product measures on $(\Omega,\mathcal{B}_\infty)$ associated to two sequences of probability measures, $\{\mu_n\}_{n=1}^\infty$ and $\{\nu_n\}_{n=1}^\infty$ on $(\mathbb{R},\mathcal{B}_\mathbb{R})$, see Theorem 10.62 (take $\mu:=P\circ(Y_1,Y_2,\dots)^{-1}$). Let us further assume that $\nu_n\ll\mu_n$ for all n so that*

$$0<H(\mu_n,\nu_n)=\int_\mathbb{R}\sqrt{\frac{d\nu_n}{d\mu_m}}d\mu_n\leq 1.$$

Then precisely one of the two cases below hold;

1. $\sum_{n=1}^\infty(1-H(\mu_n,\nu_n))<\infty$ *which happens iff* $\prod_{n=1}^\infty H(\mu_n,\nu_n)>0$ *which happens iff* $\nu\ll\mu$
   or
2. $\sum_{n=1}^\infty(1-H(\mu_n,\nu_n))=\infty$ *which happens iff* $\prod_{n=1}^\infty H(\mu_n,\nu_n)=0$ *which happens iff* $\mu\perp\nu$.

In case 1. where $\nu\ll\mu$ we have

$$\frac{d\nu}{d\mu}=\prod_{n=1}^\infty\frac{d\nu_n}{d\mu_n}(Y_n)\quad \mu\text{-a.s.} \tag{19.24}$$

and in all cases we have

$$H(\mu,\nu)=\prod_{n=1}^\infty H(\mu_n,\nu_n).$$

Proof. Let $P=\mu$, $Q=\nu$, $\mathcal{B}_n:=\sigma(Y_1,\dots,Y_n)$, $X_n:=\frac{d\nu_n}{d\mu_n}(Y_n)$, and

$$M_n:=X_1\dots X_n=\frac{d\nu_1}{d\mu_1}(Y_1)\dots\frac{d\nu_n}{d\mu_n}(Y_n).$$

If $f:\mathbb{R}^n\to\mathbb{R}$ is a bounded measurable function, then

$$\begin{aligned}\mathbb{E}_\nu(f(Y_1,\dots,Y_n))&=\int_{\mathbb{R}^n}f(y_1,\dots,y_n)\,d\nu_1(y_1)\dots d\nu_n(y_n)\\&=\int_{\mathbb{R}^n}f(y_1,\dots,y_n)\frac{d\nu_1}{d\mu_1}(y_1)\dots\frac{d\nu_n}{d\mu_n}(y_n)\;d\mu_1(y_1)\dots d\mu_n(y_n)\\&=\mathbb{E}_\mu\left[f(Y_1,\dots,Y_n)\frac{d\nu_1}{d\mu_1}(Y_1)\dots\frac{d\nu_n}{d\mu_n}(Y_n)\right]\\&=\mathbb{E}_\mu[f(Y_1,\dots,Y_n)M_n]\end{aligned}$$

from which it follows that

$$d\nu|_{\mathcal{B}_n}=M_n d\mu|_{\mathcal{B}_n}.$$

Hence by Theorem 19.10, $M_\infty:=\lim_{n\to\infty}M_n$ exists a.s. and the Lebesgue decomposition of ν is given by

$$d\nu=M_\infty d\mu+dR_\infty$$

where $R_\infty\perp\mu$. Moreover $\nu\ll\mu$ iff $R_\infty=0$ which happens iff $\mathbb{E}M_\infty=1$ and $\nu\perp\mu$ iff $R_\infty=\nu$ which happens iff $M_\infty=0$. From Theorem 19.12,

$$\mathbb{E}_\mu M_\infty=1 \text{ iff } 0<\prod_{n=1}^\infty\mathbb{E}_\mu\left(\sqrt{X_n}\right)=\prod_{n=1}^\infty\int_\mathbb{R}\sqrt{\frac{d\nu_n}{d\mu_n}}d\mu_n=\prod_{n=1}^\infty H(\mu_n,\nu_n)$$

and in this case

[4] This statement and its proof may be safely omitted.

$$d\nu = M_\infty d\mu = \left(\prod_{k=1}^{\infty} X_k\right) \cdot d\mu = \left(\prod_{n=1}^{\infty} \frac{d\nu_n}{d\mu_n}(Y_n)\right) \cdot d\mu.$$

On the other hand, if

$$\prod_{n=1}^{\infty} \mathbb{E}_\mu\left(\sqrt{X_n}\right) = \prod_{n=1}^{\infty} H(\mu_n, \nu_n) = 0,$$

Theorem 19.12 implies $M_\infty = 0$, μ – a.s. in which case Theorem 19.10 implies $\nu = R_\infty$ and so $\nu \perp \mu$.

(The rest of the argument may be safely omitted.) For the last assertion, if $\prod_{n=1}^{\infty} H(\mu_n, \nu_n) = 0$ then $\mu \perp \nu$ and hence $H(\mu, \nu) = 0$. Conversely if $\prod_{n=1}^{\infty} H(\mu_n, \nu_n) > 0$, then $M_n \to M_\infty$ in $L^1(\mu)$ and therefore

$$\begin{aligned}\mathbb{E}_\mu\left[\left|\sqrt{M_n} - \sqrt{M_\infty}\right|^2\right] &\le \mathbb{E}_\mu\left[\left|\sqrt{M_n} - \sqrt{M_\infty}\right| \cdot \left|\sqrt{M_n} + \sqrt{M_\infty}\right|\right] \\ &= \mathbb{E}_\mu\left[|M_n - M_\infty|\right] \to 0 \text{ as } n \to \infty.\end{aligned}$$

Since $d\nu = M_\infty d\mu$ in this case, it follows that

$$H(\mu, \nu) = \mathbb{E}_\mu\left[\sqrt{M_\infty}\right] = \lim_{n\to\infty} \mathbb{E}_\mu\left[\sqrt{M_n}\right] = \lim_{n\to\infty} \prod_{k=1}^{n} H(\mu_k, \nu_k) = \prod_{k=1}^{\infty} H(\mu_k, \nu_k).$$

■

Example 19.17. Suppose that $\nu_n = \delta_1$ for all n and $\mu_n = \left(1 - p_n^2\right)\delta_0 + p_n^2 \delta_1$ with $p_n \in (0, 1)$. Then $\nu_n \ll \mu_n$ with

$$\frac{d\nu_n}{d\mu_n} = 1_{\{1\}} p_n^{-2}$$

and

$$H(\mu_n, \nu_n) = \int_{\mathbb{R}} \sqrt{1_{\{1\}} p_n^{-2}} d\mu_n = \sqrt{p_n^{-2}} \cdot p_n^2 = p_n.$$

So in this case $\nu \ll \mu$ iff $\sum_{n=1}^{\infty} (1 - p_n) < \infty$. Observe that μ is never absolutely continuous relative to ν.

On the other hand; if we further assume in Corollary 19.16 that $\mu_n \sim \nu_n$, then either; $\mu \sim \nu$ or $\mu \perp \nu$ depending on whether $\prod_{n=1}^{\infty} H(\mu_n, \nu_n) > 0$ or $\prod_{n=1}^{\infty} H(\mu_n, \nu_n) = 0$ respectively.

In the next group of problems you will be given probability measures, μ_n and ν_n on $\mathbb{R}$ and you will be asked to decide if $\mu := \otimes_{n=1}^{\infty} \mu_n$ and $\nu := \otimes_{n=1}^{\infty} \nu_n$ are equivalent. For the solutions of these problems you will want to make use of the following Gaussian integral formula;

$$\begin{aligned}\int_{\mathbb{R}} \exp\left(-\frac{a}{2}x^2 + bx\right) dx &= \int_{\mathbb{R}} \exp\left(-\frac{a}{2}\left(x - \frac{b}{a}\right)^2 + \frac{b^2}{2a}\right) dx \\ &= e^{\frac{b^2}{2a}} \int_{\mathbb{R}} \exp\left(-\frac{a}{2}x^2\right) dx = \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a}}\end{aligned}$$

which is valid for all $a > 0$ and $b \in \mathbb{R}$.

Exercise 19.7 (A Discrete Cameron-Martin Theorem). Suppose $t > 0$, $\{a_n\} \subset \mathbb{R}$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$ and $d\nu_n(x) = \frac{1}{\sqrt{2\pi t}} e^{-(x+a_n)^2/2t} dx$. Show $\mu \sim \nu$ iff $\sum_{k=1}^{\infty} a_k^2 < \infty$.

Exercise 19.8. Suppose $s, t > 0$, $\{a_n\} \subset \mathbb{R}$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$ and $d\nu_n(x) = \frac{1}{\sqrt{2\pi s}} e^{-(x+a_n)^2/2s} dx$. Show $\mu \perp \nu$ if $s \neq t$.

Exercise 19.9. Suppose $\{t_n\} \subset (0, \infty)$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ and $d\nu_n(x) = \frac{1}{\sqrt{2\pi t_n}} e^{-x^2/2t_n} dx$. If $\sum_{n=1}^{\infty} (t_n - 1)^2 < \infty$ then $\mu \sim \nu$.

Part IV

(Weak) Convergence of Random Sums

20

Random Sums

As usual let $(\Omega, \mathcal{B}, P)$ be a probability space. The general theme of this chapter is to consider arrays of random variables, $\{X_k^n\}_{k=1}^n$, for each $n \in \mathbb{N}$. We are going to look for conditions under which $\lim_{n\to\infty} \sum_{k=1}^n X_k^n$ exists almost surely or in L^p for some $0 \le p < \infty$. Typically we will start with a sequence of random variables, $\{X_k\}_{k=1}^\infty$ and consider the convergence of

$$S_n = \frac{X_1 + \cdots + X_n}{b_n} - a_n$$

for appropriate choices of sequence of numbers, $\{a_n\}$ and $\{b_n\}$. This fits into our general scheme by taking $X_k^n = X_k^n / b_n - a_n/n$.

20.1 Weak Laws of Large Numbers

Theorem 20.1 (An L^2 – Weak Law of Large Numbers). *Let $\{X_n\}_{n=1}^\infty$ be a sequence of uncorrelated square integrable random variables, $\mu_n = \mathbb{E}X_n$ and $\sigma_n^2 = \mathrm{Var}(X_n)$. If there exists an increasing positive sequence, $\{a_n\}$ and $\mu \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{j=1}^n \mu_j = \mu \text{ and } \lim_{n\to\infty} \frac{1}{a_n^2} \sum_{j=1}^n \sigma_j^2 = 0,$$

then $\frac{S_n}{a_n} \to \mu$ in $L^2(P)$ (and hence also in probability).

Exercise 20.1. Prove Theorem 20.1.

Example 20.2. Suppose that $\{X_k\}_{k=1}^\infty \subset L^2(P)$ are uncorrelated identically distributed random variables. Then

$$\frac{S_n}{n} \overset{L^2(P)}{\to} \mu = \mathbb{E}X_1 \text{ as } n \to \infty.$$

To see this, simply apply Theorem 20.1 with $a_n = n$. More generally if $b_n \uparrow \infty$ such that $\lim_{n\to\infty} (n/b_n^2) = 0$, then

$$\mathrm{Var}\left(\frac{S_n}{b_n}\right) = \frac{1}{b_n^2} \cdot n \,\mathrm{Var}(X_1) \to 0 \text{ as } n \to \infty$$

and therefore

$$(S_n - n\mu)/b_n \to 0 \text{ in } L^2(P).$$

Note well: since $L^2(P)$ convergence implies $L^p(P)$ – convergence for $0 \le p \le 2$, where by $L^0(P)$ – **convergence** we mean convergence in probability. The remainder of this chapter is mostly devoted to proving *a.s.* convergence for the quantities in Theorem 12.25 and Proposition 20.10 under various assumptions. These results will be described in the next section.

Theorem 20.3 (Weak Law of Large Numbers). *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of independent random variables. Let and*

$$S_n := \sum_{j=1}^n X_j \text{ and } a_n := \sum_{k=1}^n \mathbb{E}(X_k : |X_k| \le n).$$

If

$$\lim_{n\to\infty} \sum_{k=1}^n P(|X_k| > n) = 0 \text{ and} \tag{20.1}$$

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k| \le n\right) = 0, \tag{20.2}$$

then

$$\frac{S_n - a_n}{n} \overset{P}{\to} 0.$$

Proof. A key ingredient in this proof and proofs of other versions of the law of large numbers is to introduce truncations of the $\{X_k\}$. In this case we consider

$$S_n' := \sum_{k=1}^n X_k 1_{|X_k| \le n}.$$

Since $\{S_n \ne S_{n'}\} \subset \cup_{k=1}^n \{|X_k| > n\}$,

$$P\left(\left|\frac{S_n - a_n}{n} - \frac{S_n' - a_n}{n}\right| > \varepsilon\right) = P\left(\left|\frac{S_n - S_n'}{n}\right| > \varepsilon\right)$$
$$\le P(S_n \ne S_{n'}) \le \sum_{k=1}^n P(|X_k| > n) \to 0 \text{ as } n \to \infty.$$

Hence it suffices to show $\frac{S_n'-a_n}{n} \xrightarrow{P} 0$ as $n\to\infty$ and for this it suffices to show, $\frac{S_n'-a_n}{n} \overset{L^2(P)}{\to} 0$ as $n\to\infty$.

Observe that $\mathbb{E}S_n' = a_n$ and therefore,

$$\mathbb{E}\left(\left[\frac{S_n'-a_n}{n}\right]^2\right) = \frac{1}{n^2}\operatorname{Var}(S_n') = \frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}\left(X_k 1_{|X_k|\le n}\right)$$
$$\le \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 1_{|X_k|\le n}\right) \to 0 \text{ as } n\to\infty,$$

wherein we have used $\operatorname{Var}(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \le \mathbb{E}Y^2$ in the last inequality. ∎

We are now going to use this result to prove Feller's weak law of large numbers which will be valid with an assumption which is weaker than first moments existing.

Remark 20.4. If $X \in L^1(P)$, Chebyschev's inequality along with the dominated convergence theorem implies

$$\tau(x) := xP(|X|\ge x) \le \mathbb{E}[|X| : |X|\ge x] \to 0 \text{ as } x\to\infty.$$

If X is a random variable such that $\tau(x) = xP(|X|\ge x)\to 0$ as $x\to\infty$, we say that X is in "weak L^1."

Exercise 20.2. Let $\Omega = (0,1]$, $\mathcal{B} = \mathcal{B}_{(0,1]}$ be the Borel σ – algebra, $P = m$ be Lebesgue measure on $(\Omega,\mathcal{B})$, and $X(y) := (y|\ln y|)^{-1}\cdot 1_{y\le 1/2}$ for $y\in\Omega$. Show that $X\notin L^1(P)$ yet $\lim_{x\to\infty} xP(|X|\ge x) = 0$.

Lemma 20.5. *Let X be a random variable such that $\tau(x) := xP(|X|\ge x)\to 0$ as $x\to\infty$, then*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[|X|^2 : |X|\le n\right] = 0. \tag{20.3}$$

Proof. To prove this we observe that

$$\mathbb{E}\left[|X|^2 : |X|\le n\right] = \mathbb{E}\left[2\int 1_{0\le x\le |X|\le n} x dx\right] = 2\int P(0\le x\le |X|\le n)\, x dx$$
$$\le 2\int_0^n xP(|X|\ge x)\,dx = 2\int_0^n \tau(x)\,dx$$

so that

$$\frac{1}{n}\mathbb{E}\left[|X|^2 : |X|\le n\right] = \frac{2}{n}\int_0^n \tau(x)\,dx.$$

It is now easy to check (we leave it to the reader) that

$$\lim_{n\to\infty}\frac{1}{n}\int_0^n \tau(x)\,dx = 0.$$

∎

Corollary 20.6 (Feller's WLLN). *If $\{X_n\}_{n=1}^\infty$ are i.i.d. and $\tau(x) := xP(|X_1|>x)\to 0$ as $x\to\infty$, then the hypothesis of Theorem 20.3 are satisfied so that*

$$\frac{S_n}{n} - \mathbb{E}(X_1 : |X_1|\le n) \xrightarrow{P} 0.$$

Proof. Since

$$\sum_{k=1}^n P(|X_k|>n) = nP(|X_1|>n) = \tau(n)\to 0 \text{ as } n\to\infty,$$

Eq. (20.1) is satisfied. Equation (20.2) follows from Lemma 20.5 and the identity,

$$\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k|\le n\right) = \frac{1}{n}\mathbb{E}\left[|X_1|^2 : |X_1|\le n\right].$$

∎

As a direct corollary of Feller's WLLN and Remark 20.4 we get Khintchin's weak law of large numbers.

Corollary 20.7 (Khintchin's WLLN). *If $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^1(P)$ – random variables, then $\frac{1}{n}S_n \xrightarrow{P} \mu = \mathbb{E}X_1$. This convergence holds in $L^1(P)$ as well since $\left\{\frac{1}{n}S_n\right\}_{n=1}^\infty$ is uniformly integrable under these hypothesis.*

This result is also clearly a consequence of Komogorov's strong law of large numbers.

20.1.1 A WLLN Example

Theorem 20.8 (Shannon's Theorem). *Let $\{X_i\}_{i=1}^\infty$ be a sequence of i.i.d. random variables with values in $\{1,2,\dots,r\}\subset\mathbb{N}$, $p(k) := P(X_i = k) > 0$ for $1\le k\le r$, and*

$$H(p) := -\mathbb{E}[\ln p(X_1)] = -\sum_{k=1}^r p(k)\ln p(k)$$

be the entropy of $p = \{p_k\}_{k=1}^r$. If we define $\pi_n(\omega) := p(X_1(\omega))\dots p(X_n(\omega))$ to be the "probability of the realization" $(X_1(\omega),\dots,X_n(\omega))$, then for all $\varepsilon>0$,

$$P\left(e^{-n(H(p)+\varepsilon)}\le \pi_n \le e^{-n(H(p)-\varepsilon)}\right)\to 1 \text{ as } n\to\infty.$$

Thus the probability, π_n, that the random sample $\{X_1,\dots,X_n\}$ should occur is approximately $e^{-nH(p)}$ with high probability. The number $H(p)$ is called the entropy of the distribution, $\{p(k)\}_{k=1}^r$.

Proof. Since $\{\ln p(X_i)\}_{i=1}^{\infty}$ are i.i.d. it follows by the Weal law of large numbers that

$$-\frac{1}{n}\ln\pi_n = -\frac{1}{n}\sum_{i=1}^{n}\ln p(X_i) \xrightarrow{P} -\mathbb{E}[\ln p(X_1)] = -\sum_{k=1}^{r} p(k)\ln p(k) =: H(p),$$

i.e. for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|H(p) - \frac{1}{n}\ln\pi_n\right| > \varepsilon\right) = 0.$$

Since

$$\begin{aligned}\left\{\left|H(p)+\frac{1}{n}\ln\pi_n\right| > \varepsilon\right\} &= \left\{H(p)+\frac{1}{n}\ln\pi_n > \varepsilon\right\} \cup \left\{H(p)+\frac{1}{n}\ln\pi_n < -\varepsilon\right\} \\ &= \left\{\frac{1}{n}\ln\pi_n > -H(p)+\varepsilon\right\} \cup \left\{\frac{1}{n}\ln\pi_n < -H(p)-\varepsilon\right\} \\ &= \left\{\pi_n > e^{n(-H(p)+\varepsilon)}\right\} \cup \left\{\pi_n < e^{n(-H(p)-\varepsilon)}\right\}\end{aligned}$$

it follows that

$$\begin{aligned}\left\{\left|H(p)-\frac{1}{n}\ln\pi_n\right| > \varepsilon\right\}^c &= \left\{\pi_n \le e^{n(-H(p)+\varepsilon)}\right\} \cap \left\{\pi_n \ge e^{n(-H(p)-\varepsilon)}\right\} \\ &= \left\{e^{-n(H(p)+\varepsilon)} \le \pi_n \le e^{-n(H(p)-\varepsilon)}\right\},\end{aligned}$$

and therefore

$$P\left(e^{-n(H(p)+\varepsilon)} \le \pi_n \le e^{-n(H(p)-\varepsilon)}\right) \to 1 \text{ as } n\to\infty.$$

■

For our next example, let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. random variables with common distribution function, $F(x) := P(X_n \le x)$. For $x \in \mathbb{R}$ let $F_n(x)$ be the **empirical distribution function** defined by,

$$F_n(x) := \frac{1}{n}\sum_{j=1}^{n} 1_{X_j\le x} = \left(\frac{1}{n}\sum_{j=1}^{n}\delta_{X_j}\right)((-\infty, x]).$$

Since $\mathbb{E}1_{X_j\le x} = F(x)$ and $\{1_{X_j\le x}\}_{j=1}^{\infty}$ are Bernoulli random variables, the weak law of large numbers implies $F_n(x) \xrightarrow{P} F(x)$ as $n\to\infty$. As usual, for $p\in(0,1)$ let

$$F^{\leftarrow}(p) := \inf\{x : F(x) \ge p\}$$

and recall that $F^{\leftarrow}(p) \le x$ iff $F(x) \ge p$. Let us notice that

$$\begin{aligned}F_n^{\leftarrow}(p) &= \inf\{x : F_n(x) \ge p\} = \inf\left\{x : \sum_{j=1}^{n} 1_{X_j\le x} \ge np\right\} \\ &= \inf\{x : \#\{j \le n : X_j \le x\} \ge np\}.\end{aligned}$$

Recall from Definition 11.10 that the **order statistic of** $(X_1,\dots,X_n)$ is the finite sequence, $\left(X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}\right)$, where $\left(X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}\right)$ denotes $(X_1,\dots,X_n)$ arranged in increasing order with possible repetitions. It follows from the formula in Definition 11.10 that $X_k^{(n)}$ are all random variables for $k \le n$ but it will be useful to give another proof. Indeed, $X_k^{(n)} \le x$ iff $\#\{j \le n : X_j \le x\} \ge k$ iff $\sum_{j=1}^{n} 1_{X_j\le x} \ge k$, i.e.

$$\left\{X_k^{(n)} \le x\right\} = \left\{\sum_{j=1}^{n} 1_{X_j\le x} \ge k\right\} \in \mathcal{B}.$$

Moreover, if we let $\lceil x\rceil = \min\{n\in\mathbb{Z} : n \ge x\}$, the reader may easily check that $F_n^{\leftarrow}(p) = X_{\lceil np\rceil}^{(n)}$.

Proposition 20.9. *Keeping the notation above. Suppose that $p\in(0,1)$ is a point where*

$$F(F^{\leftarrow}(p)-\varepsilon) < p < F(F^{\leftarrow}(p)+\varepsilon) \text{ for all } \varepsilon > 0$$

then $X_{\lceil np\rceil}^{(n)} = F_n^{\leftarrow}(p) \xrightarrow{P} F^{\leftarrow}(p)$ as $n\to\infty$. Thus we can recover, with high probability, the p^{th} – quantile of the distribution F by observing $\{X_i\}_{i=1}^{n}$.

Proof. Let $\varepsilon > 0$. Then

$$\begin{aligned}\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\}^c &= \{F_n^{\leftarrow}(p) \le \varepsilon + F^{\leftarrow}(p)\} = \{F_n^{\leftarrow}(p) \le \varepsilon + F^{\leftarrow}(p)\} \\ &= \{F_n(\varepsilon + F^{\leftarrow}(p)) \ge p\}\end{aligned}$$

so that

$$\begin{aligned}\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\} &= \{F_n(F^{\leftarrow}(p)+\varepsilon) < p\} \\ &= \{F_n(\varepsilon + F^{\leftarrow}(p)) - F(\varepsilon + F^{\leftarrow}(p)) < p - F(F^{\leftarrow}(p)+\varepsilon)\}.\end{aligned}$$

Letting $\delta_\varepsilon := F(F^{\leftarrow}(p)+\varepsilon) - p > 0$, we have, as $n\to\infty$, that

$$P(\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\}) = P(F_n(\varepsilon + F^{\leftarrow}(p)) - F(\varepsilon + F^{\leftarrow}(p)) < -\delta_\varepsilon) \to 0.$$

Similarly, let $\delta_\varepsilon := p - F(F^{\leftarrow}(p)-\varepsilon) > 0$ and observe that

$$\{F^{\leftarrow}(p) - F_n^{\leftarrow}(p) \ge \varepsilon\} = \{F_n^{\leftarrow}(p) \le F^{\leftarrow}(p) - \varepsilon\} = \{F_n(F^{\leftarrow}(p)-\varepsilon) \ge p\}$$

and hence,

$$\begin{aligned}
&P\left(F^{\leftarrow}(p)-F_n^{\leftarrow}(p)\geq\varepsilon\right)\\
&\quad=P\left(F_n\left(F^{\leftarrow}(p)-\varepsilon\right)-F\left(F^{\leftarrow}(p)-\varepsilon\right)\geq p-F\left(F^{\leftarrow}(p)-\varepsilon\right)\right)\\
&\quad=P\left(F_n\left(F^{\leftarrow}(p)-\varepsilon\right)-F\left(F^{\leftarrow}(p)-\varepsilon\right)\geq\delta_\varepsilon\right)\to 0 \text{ as } n\to\infty.
\end{aligned}$$

Thus we have shown that $X_{\lceil np\rceil}^{(n)}\xrightarrow{P}F^{\leftarrow}(p)$ as $n\to\infty$. ∎

20.2 Kolmogorov's Convergence Criteria

Proposition 20.10 (L^2 - Convergence of Random Sums). *Suppose that $\{Y_k\}_{k=1}^\infty\subset L^2(P)$ are uncorrelated. If $\sum_{k=1}^\infty \operatorname{Var}(Y_k)<\infty$ then*

$$\sum_{k=1}^\infty (Y_k-\mu_k) \text{ converges in } L^2(P).$$

where $\mu_k:=\mathbb{E}Y_k$.

Proof. Letting $S_n:=\sum_{k=1}^n (Y_k-\mu_k)$, it suffices by the completeness of $L^2(P)$ (see Theorem 12.25) to show $\|S_n-S_m\|_2\to 0$ as $m,n\to\infty$. Supposing $n>m$, we have

$$\begin{aligned}
\|S_n-S_m\|_2^2&=\mathbb{E}\left(\sum_{k=m+1}^n (Y_k-\mu_k)\right)^2\\
&=\sum_{k=m+1}^n \operatorname{Var}(Y_k)=\sum_{k=m+1}^n \sigma_k^2\to 0 \text{ as } m,n\to\infty.
\end{aligned}$$

∎

Theorem 20.11 (Kolmogorov's Convergence Criteria). *Suppose that $\{Y_n\}_{n=1}^\infty$ are independent square integrable random variables. If $\sum_{j=1}^\infty \operatorname{Var}(Y_j)<\infty$, then $\sum_{j=1}^\infty (Y_j-\mathbb{E}Y_j)$ converges a.s. In particular if $\sum_{j=1}^\infty \operatorname{Var}(Y_j)<\infty$ and $\sum_{j=1}^\infty \mathbb{E}Y_j$ is convergent, then $\sum_{j=1}^\infty Y_j$ converges a.s. and in $L^2(P)$.*

Proof. This is a special case of Theorem 18.67. Indeed, let $S_n:=\sum_{j=1}^n (Y_j-\mathbb{E}Y_j)$ with $S_0=0$. Then $\{S_n\}_{n=0}^\infty$ is a martingale relative to the filtration, $\mathcal{B}_n=\sigma(S_0,\dots S_n)$. By assumption we have

$$\mathbb{E}S_n^2=\sum_{j=1}^n \operatorname{Var}(Y_j)\leq\sum_{j=1}^\infty \operatorname{Var}(Y_j)<\infty$$

so that $\{S_n\}_{n=0}^\infty$ is bounded in $L^2(P)$. Therefore by Theorem 18.67, $\sum_{j=1}^\infty (Y_j-\mathbb{E}Y_j)=\lim_{n\to\infty}S_n$ exists a.s. and in $L^2(P)$.

Another way to prove this is to appeal Proposition 20.10 above and Lévy's Theorem 20.46 below. As second method is to make use of Kolmogorov's inequality and we will give this proof below. ∎

Example 20.12 (Brownian Motion). Let $\{N_n\}_{n=1}^\infty$ be i.i.d. standard normal random variable, i.e.

$$P(N_n\in A)=\int_A \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \text{ for all } A\in\mathcal{B}_\mathbb{R}.$$

Let $\{\omega_n\}_{n=1}^\infty\subset\mathbb{R}$, $\{a_n\}_{n=1}^\infty\subset\mathbb{R}$, and $t\in\mathbb{R}$, then

$$\sum_{n=1}^\infty a_nN_n\sin\omega_nt \text{ converges a.s.}$$

provided $\sum_{n=1}^\infty a_n^2<\infty$. This is a simple consequence of Kolmogorov's convergence criteria, Theorem 20.11, and the facts that $\mathbb{E}[a_nN_n\sin\omega_nt]=0$ and

$$\operatorname{Var}(a_nN_n\sin\omega_nt)=a_n^2\sin^2\omega_nt\leq a_n^2.$$

As a special case, if we take $\omega_n=(2n-1)\frac{\pi}{2}$ and $a_n=\frac{\sqrt{2}}{\pi(2n-1)}$, then it follows that

$$B_t:=\frac{2\sqrt{2}}{\pi}\sum_{k=1,3,5,\dots}\frac{N_k}{k}\sin\left(k\frac{\pi}{2}t\right) \tag{20.4}$$

is a.s. convergent for all $t\in\mathbb{R}$. The factor $\frac{2\sqrt{2}}{\pi k}$ has been determined by requiring,

$$\int_0^1\left[\frac{d}{dt}\frac{2\sqrt{2}}{\pi k}\sin(k\pi t)\right]^2dt=1$$

as seen by,

$$\begin{aligned}
\int_0^1\left[\frac{d}{dt}\sin\left(\frac{k\pi}{2}t\right)\right]^2dt&=\frac{k^2\pi^2}{2^2}\int_0^1\left[\cos\left(\frac{k\pi}{2}t\right)\right]^2dt\\
&=\frac{k^2\pi^2}{2^2}\frac{2}{k\pi}\left[\frac{k\pi}{4}t+\frac{1}{4}\sin k\pi t\right]_0^1=\frac{k^2\pi^2}{2^3}.
\end{aligned}$$

Fact: Wiener in 1923 showed the series in Eq. (20.4) is in fact almost surely uniformly convergent. Given this, the process, $t\to B_t$ is almost surely continuous. The process $\{B_t:0\leq t\leq 1\}$ is **Brownian Motion.**

Kolmogorov's convergence criteria becomes a powerful tool when combined with the following real variable lemma.

Lemma 20.13 (Kronecker's Lemma). *Suppose that* $\{x_k\} \subset \mathbb{R}$ *and* $\{a_k\} \subset (0,\infty)$ *are sequences such that* $a_k \uparrow \infty$ *and* $\sum_{k=1}^{\infty} \frac{x_k}{a_k}$ *is convergent in* $\mathbb{R}$. *Then*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} x_k = 0.$$

BRUCE – see Kallenberg for a better proof of this lemma.

Proof. Before going to the proof, let us warm-up by proving the following continuous version of the lemma. Let $a(s) \in (0,\infty)$ and $x(s) \in \mathbb{R}$ be continuous functions such that $a(s) \uparrow \infty$ as $s \to \infty$ and $\int_1^\infty \frac{x(s)}{a(s)} ds$ exists. We are going to show

$$\lim_{n\to\infty} \frac{1}{a(n)} \int_1^n x(s)\, ds = 0.$$

Let $X(s) := \int_0^s x(u)\, du$ and

$$r(s) := \int_s^\infty \frac{X'(u)}{a(u)} du = \int_s^\infty \frac{x(u)}{a(u)} du.$$

Then by assumption, $r(s) \to 0$ as $s \to 0$ and $X'(s) = -a(s) r'(s)$. Integrating this equation shows

$$X(s) - X(s_0) = -\int_{s_0}^{s} a(u) r'(u)\, du = -a(u) r(u)|_{u=s_0}^{s} + \int_{s_0}^{s} r(u) a'(u)\, du.$$

Dividing this equation by $a(s)$ and then letting $s \to \infty$ gives

$$\begin{aligned}\limsup_{s\to\infty} \frac{|X(s)|}{a(s)} &= \limsup_{s\to\infty} \left[\frac{a(s_0) r(s_0) - a(s) r(s)}{a(s)} + \frac{1}{a(s)} \int_{s_0}^{s} r(u) a'(u)\, du\right] \\ &\le \limsup_{s\to\infty} \left[-r(s) + \frac{1}{a(s)} \int_{s_0}^{s} |r(u)|\, a'(u)\, du\right] \\ &\le \limsup_{s\to\infty} \left[\frac{a(s) - a(s_0)}{a(s)} \sup_{u\ge s_0} |r(u)|\right] = \sup_{u\ge s_0} |r(u)| \to 0 \text{ as } s_0 \to \infty.\end{aligned}$$

With this as warm-up, we go to the discrete case.

Let

$$S_k := \sum_{j=1}^{k} x_j \text{ and } r_k := \sum_{j=k}^{\infty} \frac{x_j}{a_j},$$

so that $r_k \to 0$ as $k \to \infty$ by assumption. Since $x_k = a_k (r_k - r_{k+1})$, we find

$$\begin{aligned}\frac{S_n}{a_n} &= \frac{1}{a_n} \sum_{k=1}^{n} a_k (r_k - r_{k+1}) = \frac{1}{a_n} \left[\sum_{k=1}^{n} a_k r_k - \sum_{k=2}^{n+1} a_{k-1} r_k\right] \\ &= \frac{1}{a_n} \left[a_1 r_1 - a_n r_{n+1} + \sum_{k=2}^{n} (a_k - a_{k-1}) r_k\right]. \text{ (summation by parts)}\end{aligned}$$

Using the fact that $a_k - a_{k-1} \ge 0$ for all $k \ge 2$, and

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=2}^{m} (a_k - a_{k-1}) |r_k| = 0$$

for any $m \in \mathbb{N}$; we may conclude

$$\begin{aligned}\limsup_{n\to\infty} \left|\frac{S_n}{a_n}\right| &\le \limsup_{n\to\infty} \frac{1}{a_n} \left[\sum_{k=2}^{n} (a_k - a_{k-1}) |r_k|\right] \\ &= \limsup_{n\to\infty} \frac{1}{a_n} \left[\sum_{k=m}^{n} (a_k - a_{k-1}) |r_k|\right] \\ &\le \sup_{k\ge m} |r_k| \cdot \limsup_{n\to\infty} \frac{1}{a_n} \left[\sum_{k=m}^{n} (a_k - a_{k-1})\right] \\ &= \sup_{k\ge m} |r_k| \cdot \limsup_{n\to\infty} \frac{1}{a_n} [a_n - a_{m-1}] = \sup_{k\ge m} |r_k|.\end{aligned}$$

This completes the proof since $\sup_{k\ge m} |r_k| \to 0$ as $m \to \infty$.

(See Kallenberg for a better proof.) ■

Corollary 20.14. *Let* $\{X_n\}$ *be a sequence of independent square integrable random variables and* b_n *be a sequence such that* $b_n \uparrow \infty$. *If*

$$\sum_{k=1}^{\infty} \frac{\mathrm{Var}(X_k)}{b_k^2} < \infty$$

then

$$\frac{S_n - \mathbb{E}S_n}{b_n} \to 0 \textit{ a.s. and in } L^2(P).$$

Proof. By Kolmogorov's convergence criteria, Theorem 20.11,

$$\sum_{k=1}^{\infty} \frac{X_k - \mathbb{E}X_k}{b_k} \text{ is convergent a.s. and in } L^2(P).$$

Therefore an application of Kronecker's Lemma 20.13 implies

$$0=\lim_{n\to\infty}\frac{1}{b_n}\sum_{k=1}^n\left(X_k-\mathbb{E}X_k\right)=\lim_{n\to\infty}\frac{S_n-\mathbb{E}S_n}{b_n}\text{ a.s.}$$

Similarly by Kronecker's Lemma 20.13 we know that

$$0=\lim_{n\to\infty}\frac{1}{b_n^2}\sum_{k=1}^n\operatorname{Var}\left(X_k\right)=\lim_{n\to\infty}\mathbb{E}\left(\frac{S_n-\mathbb{E}S_n}{b_n}\right)^2$$

which gives the $L^2(P)$ – convergence statement as well. ■

As an immediate corollary we have the following corollary.

Corollary 20.15 (L^2 – SSLN). *Let $\{X_n\}$ be a sequence of independent random variables such that $\sigma^2=\mathbb{E}X_n^2<\infty$ and $\mu=\mathbb{E}X_n$ are independent of n. As above let $S_n=\sum_{k=1}^n X_k$. If $\{b_n\}_{n=1}^\infty\subset(0,\infty)$ is a sequence such that $b_n\uparrow\infty$ and $\sum_{n=1}^\infty\frac{1}{b_n^2}<\infty$, then*

$$\frac{1}{b_n}\left(S_n-n\mu\right)\to 0 \text{ a.s. and in } L^2(P) \tag{20.5}$$

We may rewrite Eq. (20.5) as

$$S_n=n\mu+o(1)\,b_n \text{ or } \frac{S_n}{n}=\mu+o(1)\,\frac{b_n}{n}.$$

Example 20.16. For example, we could take $b_n=n$ or $b_n=n^p$ for an $p>1/2$, or $b_n=n^{1/2}(\ln n)^{1/2+\varepsilon}$ for any $\varepsilon>0$. The idea here is that

$$\sum_{n=2}^\infty\frac{1}{\left(n^{1/2}(\ln n)^{1/2+\varepsilon}\right)^2}=\sum_{n=2}^\infty\frac{1}{n(\ln n)^{1+2\varepsilon}}$$

which may be analyzed by comparison with the integral

$$\int_2^\infty\frac{1}{x\ln^{1+2\varepsilon}x}dx=\int_{\ln 2}^\infty\frac{1}{e^y y^{1+2\varepsilon}}e^y dy=\int_{\ln 2}^\infty\frac{1}{y^{1+2\varepsilon}}dy<\infty,$$

wherein we have made the change of variables, $y=\ln x$. When $b_n=n^{1/2}(\ln n)^{1/2+\varepsilon}$ we may conclude that

$$\frac{S_n}{n}=\mu+o(1)\frac{(\ln n)^{1/2+\varepsilon}}{n^{1/2}},$$

i.e. the fluctuations of $\frac{S_n}{n}$ about the mean, μ, have order smaller than $n^{-1/2}(\ln n)^{1/2+\varepsilon}$.

Fact 20.17 (Missing Reference) *Under the hypothesis in Corollary 20.15,*

$$\lim_{n\to\infty}\frac{S_n-n\mu}{n^{1/2}(\ln\ln n)^{1/2}}=\sqrt{2}\sigma \text{ a.s.}$$

We end this section with another example of using Kolmogorov's convergence criteria in conjunction with Kronecker's Lemma 20.13.

Lemma 20.18. *Let $\{X_n\}_{n=1}^\infty$ be independent square integrable random variables such that $\mathbb{E}S_n\uparrow\infty$ as $n\to\infty$. Then*

$$\sum_{n=1}^\infty\operatorname{Var}\left(\frac{X_n}{\mathbb{E}S_n}\right)=\sum_{n=1}^\infty\frac{\operatorname{Var}(X_n)}{(\mathbb{E}S_n)^2}<\infty\implies\frac{S_n}{\mathbb{E}S_n}\to 1 \text{ a.s.}$$

Proof. Kolmogorov's convergence criteria, Theorem 20.11 we know that

$$\sum_{n=1}^\infty\frac{X_n-\mathbb{E}X_n}{\mathbb{E}S_n}\text{ is a.s. convergent.}$$

It then follows by Kronecker's Lemma 20.13 that

$$0=\lim_{n\to\infty}\frac{1}{\mathbb{E}S_n}\sum_{i=1}^n\left(X_n-\mathbb{E}X_n\right)=\lim_{n\to\infty}\frac{S_n}{\mathbb{E}S_n}-1\text{ a.s.}$$

■

Example 20.19. Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. square integrable random variables with $\mu:=\mathbb{E}X_n>0$ and $\sigma^2:=\operatorname{Var}(X_n)<\infty$. Since $\mathbb{E}S_n=\mu n\uparrow\infty$ and

$$\sum_{n=1}^\infty\frac{\operatorname{Var}(X_n)}{(\mathbb{E}S_n)^2}=\sum_{n=1}^\infty\frac{\sigma^2}{\mu^2n^2}<\infty,$$

we may conclude that $\lim_{n\to\infty}\frac{S_n}{\mu n}=1$ a.s., i.e. $S_n/n\to\mu$ a.s. as we already know.

We now assume that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables with a continuous distribution function and let A_j denote the event when X_j is a record, i.e.

$$A_j:=\left\{X_j>\max\left\{X_1,X_2,\dots,X_{k-1}\right\}\right\}.$$

Recall from Renyi Theorem 10.23 that $\{A_j\}_{j=1}^\infty$ are independent and $P(A_j)=\frac{1}{j}$ for all j.

Proposition 20.20. *Keeping the preceding notation and let $S_n:=\sum_{j=1}^n 1_{A_j}$ denote the number of records in the first n observations. Then $\lim_{n\to\infty}\frac{S_n}{\ln n}=1$ a.s.*

Proof. In this case

$$\mathbb{E}S_n = \sum_{j=1}^{n} \mathbb{E}1_{A_j} = \sum_{j=1}^{n} \frac{1}{j} \sim \int_1^n \frac{1}{x} dx = \ln n \uparrow \infty$$

and

$$\operatorname{Var}(1_{A_n}) = \mathbb{E}1_{A_n}^2 - (\mathbb{E}1_{A_n})^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}$$

so by that

$$\begin{aligned}\sum_{n=1}^{\infty} \operatorname{Var}\left(\frac{1_{A_n}}{\mathbb{E}S_n}\right) &= \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n^2}\right) \frac{1}{\left(\sum_{j=1}^{n} \frac{1}{j}\right)^2} \\ &\le \sum_{n=1}^{\infty} \frac{1}{\left(\sum_{j=1}^{n} \frac{1}{j}\right)^2} \frac{1}{n} \\ &\lessapprox 1 + \int_2^{\infty} \frac{1}{\ln^2 x} \frac{1}{x} dx = 1 + \int_{\ln 2}^{\infty} \frac{1}{y^2} dy < \infty.\end{aligned}$$

Therefore by Lemma 20.18 we may conclude that $\lim_{n\to\infty} \frac{S_n}{\mathbb{E}S_n} = 1$ a.s.

So to finish the proof it only remains to show

$$\lim_{n\to\infty} \frac{\mathbb{E}S_n}{\ln n} \lim_{n\to\infty} \frac{\sum_{j=1}^{n} \frac{1}{j}}{\ln n} = 1. \tag{20.6}$$

To see this write

$$\begin{aligned}\ln(n+1) &= \int_1^{n+1} \frac{1}{x} dx = \sum_{j=1}^{n} \int_j^{j+1} \frac{1}{x} dx \\ &= \sum_{j=1}^{n} \int_j^{j+1} \left(\frac{1}{x} - \frac{1}{j}\right) dx + \sum_{j=1}^{n} \frac{1}{j} \\ &= \rho_n + \sum_{j=1}^{n} \frac{1}{j}\end{aligned} \tag{20.7}$$

where

$$|\rho_n| = \sum_{j=1}^{n} \left|\ln \frac{j+1}{j} - \frac{1}{j}\right| = \sum_{j=1}^{n} \left|\ln(1 + 1/j) - \frac{1}{j}\right| \sim \sum_{j=1}^{n} \frac{1}{j^2}$$

and hence we conclude that $\lim_{n\to\infty} \rho_n < \infty$. So dividing Eq. (20.7) by $\ln n$ and letting $n \to \infty$ gives the desired limit in Eq. (20.6). ■

20.3 The Strong Law of Large Numbers Revisited

Definition 20.21. *Two sequences, $\{X_n\}$ and $\{X_n'\}$, of random variables are **tail equivalent** if*

$$\mathbb{E}\left[\sum_{n=1}^{\infty} 1_{X_n \ne X_n'}\right] = \sum_{n=1}^{\infty} P(X_n \ne X_n') < \infty.$$

Proposition 20.22. *Suppose $\{X_n\}$ and $\{X_n'\}$ are tail equivalent. Then*

1. *$\sum (X_n - X_n')$ converges a.s.*
2. *The sum $\sum X_n$ is convergent a.s. iff the sum $\sum X_n'$ is convergent a.s. More generally we have*

$$P\left(\left\{\sum X_n \text{ is convergent}\right\} \triangle \left\{\sum X_n' \text{ is convergent}\right\}\right) = 1$$

3. *If there exists a random variable, X, and a sequence $a_n \uparrow \infty$ such that*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} X_k = X \text{ a.s}$$

then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} X_k' = X \text{ a.s}$$

Proof. If $\{X_n\}$ and $\{X_n'\}$ are tail equivalent, we know by the first Borel - Cantelli Lemma 7.14 that $P(X_n = X_n' \text{ for a.a. } n) = 1$. The proposition is an easy consequence of this observation. ■

Remark 20.23. In what follows we will typically have a sequence, $\{X_n\}_{n=1}^{\infty}$, of independent random variables and $X_n' = f_n(X_n)$ for some "cutoff" functions, $f_n : \mathbb{R} \to \mathbb{R}$. In this case the collection of sets, $\{A_n := \{X_n \ne X_n'\}\}_{n=1}^{\infty}$ are independent and so by the Borel zero one law (Lemma 10.41) we will have

$$P(X_n \ne X_n' \text{ i.o. } n) = 0 \iff \sum_{n=1}^{\infty} P(X_n \ne X_n') < \infty.$$

So in this case $\{X_n\}$ and $\{X_n'\}$ are tail equivalent iff $P(X_n \ne X_n' \text{ a.a. } n) = 1$. For example if $\{k_n\}_{n=1}^{\infty} \subset (0, \infty)$ and $X_n' := X_n \cdot 1_{|X_n| \le k_n}$ then the following are equivalent;

1. $P(|X_n| \le k_n \text{ a.a. } n) = 1$,
2. $P(|X_n| > k_n \text{ i.o. } n) = 0$,
3. $\sum_{n=1}^{\infty} P(X_n \ne X_n') = \sum_{n=1}^{\infty} P(|X_n| > k_n) < \infty$,

4. $\{X_n\}$ and $\{X_n'\}$ are tail equivalent.

Lemma 20.24. *Suppose that $X : \Omega \to \mathbb{R}$ is a random variable, then*

$$\mathbb{E}\,|X|^p = \int_0^\infty ps^{p-1} P\left(|X| \geq s\right) ds = \int_0^\infty ps^{p-1} P\left(|X| > s\right) ds.$$

Proof. By the fundamental theorem of calculus,

$$|X|^p = \int_0^{|X|} ps^{p-1} ds = p\int_0^\infty 1_{s\leq |X|} \cdot s^{p-1} ds = p\int_0^\infty 1_{s<|X|} \cdot s^{p-1} ds.$$

Taking expectations of this identity along with an application of Tonelli's theorem completes the proof. ∎

Lemma 20.25. *If X is a random variable and $\varepsilon > 0$, then*

$$\sum_{n=1}^\infty P\left(|X| \geq n\varepsilon\right) \leq \frac{1}{\varepsilon}\mathbb{E}\,|X| \leq \sum_{n=0}^\infty P\left(|X| \geq n\varepsilon\right). \tag{20.8}$$

Proof. First observe that for all $y \geq 0$ we have,

$$\sum_{n=1}^\infty 1_{n\leq y} \leq y \leq \sum_{n=1}^\infty 1_{n\leq y} + 1 = \sum_{n=0}^\infty 1_{n\leq y}. \tag{20.9}$$

Taking $y = |X|/\varepsilon$ in Eq. (20.9) and then take expectations gives the estimate in Eq. (20.8). ∎

Proposition 20.26. *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables, then the following are equivalent:*

1. $\mathbb{E}\,|X_1| < \infty$.
2. *There exists $\varepsilon > 0$ such that $\sum_{n=1}^\infty P\left(|X_1| \geq \varepsilon n\right) < \infty$.*
3. *For all $\varepsilon > 0$, $\sum_{n=1}^\infty P\left(|X_1| \geq \varepsilon n\right) < \infty$.*
4. $\lim_{n\to\infty} \frac{|X_n|}{n} = 0$ *a.s.*

Proof. The equivalence of items 1., 2., and 3. easily follows from Lemma 20.25. So to finish the proof it suffices to show 3. is equivalent to 4. To this end we start by noting that $\lim_{n\to\infty} \frac{|X_n|}{n} = 0$ a.s. iff

$$0 = P\left(\frac{|X_n|}{n} \geq \varepsilon \text{ i.o.}\right) = P\left(|X_n| \geq n\varepsilon \text{ i.o.}\right) \text{ for all } \varepsilon > 0. \tag{20.10}$$

Because $\{|X_n| \geq n\varepsilon\}_{n=1}^\infty$ are independent sets, the Borel zero-one law (Lemma 10.41) shows the statement in Eq. (20.10) is equivalent to $\sum_{n=1}^\infty P\left(|X_n| \geq n\varepsilon\right) < \infty$ for all $\varepsilon > 0$. ∎

Corollary 20.27. *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables such that $\frac{1}{n}S_n \to c \in \mathbb{R}$ a.s., then $X_n \in L^1(P)$ and $\mu := \mathbb{E}X_n = c$.*

Proof. If $\frac{1}{n}S_n \to c$ a.s. then $\varepsilon_n := \frac{S_{n+1}}{n+1} - \frac{S_n}{n} \to 0$ a.s. and therefore,

$$\begin{aligned}\frac{X_{n+1}}{n+1} &= \frac{S_{n+1}}{n+1} - \frac{S_n}{n+1} = \varepsilon_n + S_n\left[\frac{1}{n} - \frac{1}{n+1}\right] \\ &= \varepsilon_n + \frac{1}{(n+1)}\frac{S_n}{n} \to 0 + 0\cdot c = 0.\end{aligned}$$

Hence an application of Proposition 20.26 shows $X_n \in L^1(P)$. Moreover by Exercise 12.6, $\left\{\frac{1}{n}S_n\right\}_{n=1}^\infty$ is a uniformly integrable sequenced and therefore,

$$\mu = \mathbb{E}\left[\frac{1}{n}S_n\right] \to \mathbb{E}\left[\lim_{n\to\infty}\frac{1}{n}S_n\right] = \mathbb{E}\left[c\right] = c.$$

∎

Lemma 20.28. *For all $x \geq 0$,*

$$\varphi(x) := \sum_{n=1}^\infty \frac{1}{n^2} 1_{x\leq n} = \sum_{n\geq x} \frac{1}{n^2} \leq 2\cdot \min\left(\frac{1}{x}, 1\right).$$

Proof. The proof will be by comparison with the integral, $\int_a^\infty \frac{1}{t^2}dt = 1/a$. For example,

$$\sum_{n=1}^\infty \frac{1}{n^2} \leq 1 + \int_1^\infty \frac{1}{t^2}dt = 1 + 1 = 2$$

and so

$$\sum_{n\geq x} \frac{1}{n^2} \leq \sum_{n=1}^\infty \frac{1}{n^2} \leq 2 \leq \frac{2}{x} \text{ for } 0 < x \leq 1.$$

Similarly, for $x > 1$,

$$\sum_{n\geq x} \frac{1}{n^2} \leq \frac{1}{x^2} + \int_x^\infty \frac{1}{t^2}dt = \frac{1}{x^2} + \frac{1}{x} = \frac{1}{x}\left(1 + \frac{1}{x}\right) \leq \frac{2}{x},$$

see Figure 20.1 below. ∎

Lemma 20.29. *Suppose that $X : \Omega \to \mathbb{R}$ is a random variable, then*

$$\sum_{n=1}^\infty \frac{1}{n^2}\mathbb{E}\left[|X|^2 : 1_{|X|\leq n}\right] \leq 2\mathbb{E}\,|X|.$$

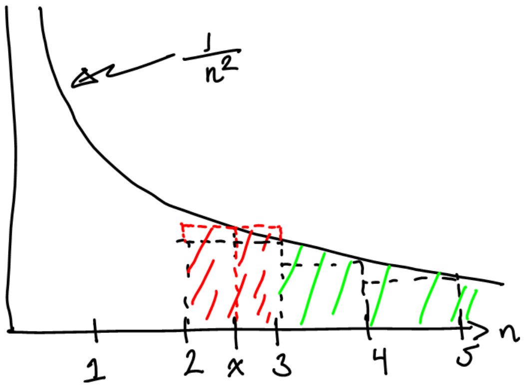


Fig. 20.1. Estimating $\sum_{n\geq x} 1/n^2$ with an integral.

Proof. This is a simple application of Lemma 20.28;

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\mathbb{E}\left[|X|^2 : 1_{|X|\leq n}\right] = \mathbb{E}\left[|X|^2\sum_{n=1}^{\infty}\frac{1}{n^2}1_{|X|\leq n}\right] = \mathbb{E}\left[|X|^2\varphi(|X|)\right]$$
$$\leq 2\mathbb{E}\left[|X|^2\left(\frac{1}{|X|}\wedge 1\right)\right] \leq 2\mathbb{E}|X|.$$

■

With this as preparation we are now in a position to give another proof of the Kolmogorov's strong law of large numbers which has already appeared in Theorem 16.10 and Example 18.78.

Theorem 20.30 (Kolmogorov's Strong Law of Large Numbers). *Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables and let $S_n := X_1 + \cdots + X_n$. Then there exists $\mu \in \mathbb{R}$ such that $\frac{1}{n}S_n \to \mu$ a.s. iff X_n is integrable and in which case $\mathbb{E}X_n = \mu$.*

Proof. The implication, $\frac{1}{n}S_n \to \mu$ a.s. implies $X_n \in L^1(P)$ and $\mathbb{E}X_n = \mu$ has already been proved in Corollary 20.27. So let us now assume $X_n \in L^1(P)$ and let $\mu := \mathbb{E}X_n$.

Let $X_n' := X_n 1_{|X_n|\leq n}$. By Lemma 20.25,

$$\sum_{n=1}^{\infty} P(X_n' \neq X_n) = \sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X_1| > n) \leq \mathbb{E}|X_1| < \infty,$$

and hence $\{X_n\}$ and $\{X_n'\}$ are tail equivalent. Therefore, by Proposition 20.22, it suffices to show $\lim_{n\to\infty}\frac{1}{n}S_n' = \mu$ a.s. where $S_n' := X_1' + \cdots + X_n'$. But by Lemma 20.29,

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}(X_n')}{n^2} \leq \sum_{n=1}^{\infty}\frac{\mathbb{E}|X_n'|^2}{n^2} = \sum_{n=1}^{\infty}\frac{\mathbb{E}\left[|X_n|^2 1_{|X_n|\leq n}\right]}{n^2}$$
$$= \sum_{n=1}^{\infty}\frac{\mathbb{E}\left[|X_1|^2 1_{|X_1|\leq n}\right]}{n^2} \leq 2\mathbb{E}|X_1| < \infty. \quad (20.11)$$

Therefore by Kolmogorov's convergence criteria, Theorem 20.11,

$$\sum_{n=1}^{\infty}\frac{X_n' - \mathbb{E}X_n'}{n} \text{ is almost surely convergent.}$$

Kronecker's Lemma 20.13 then implies

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(X_k' - \mathbb{E}X_k') = 0 \text{ a.s.}$$

So to finish the proof, it only remains to observe

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}X_k' = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[X_k 1_{|X_k|\leq k}\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[X_1 1_{|X_1|\leq k}\right] = \mu.$$

Here we have used the dominated convergence theorem to see that $a_k := \mathbb{E}\left[X_1 1_{|X_1|\leq k}\right] \to \mu$ as $k\to\infty$ from which it is easy (and standard) to check that $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} a_k = \mu$. ■

Remark 20.31. If $\mathbb{E}|X_1| = \infty$ but $\mathbb{E}X_1^- < \infty$, then $\frac{1}{n}S_n \to \infty$ a.s. To prove this, for $M > 0$ let $X_n^M := X_n \wedge M$ and $S_n^M := \sum_{i=1}^{n} X_i^M$. It follows from Theorem 20.30 that $\frac{1}{n}S_n^M \to \mu^M := \mathbb{E}X_1^M$ a.s.. Since $S_n \geq S_n^M$, we may conclude that

$$\liminf_{n\to\infty}\frac{S_n}{n} \geq \liminf_{n\to\infty}\frac{1}{n}S_n^M = \mu^M \text{ a.s.}$$

Since $\mu^M \to \infty$ as $M \to \infty$, it follows that $\liminf_{n\to\infty}\frac{S_n}{n} = \infty$ a.s. and hence that $\lim_{n\to\infty}\frac{S_n}{n} = \infty$ a.s.

Exercise 20.3 (Resnik 7.9). Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 = 0$. Following the ideas in the proof of Theorem 20.30, show for any bounded sequence $\{c_n\}_{n=1}^{\infty}$ of real numbers that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} c_k X_k = 0 \text{ a.s.}$$

20.3.1 Strong Law of Large Number Examples

Example 20.32 (Renewal Theory). Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. non-negative integrable random variables such that $P(X_i>0)>0$. Think of the X_i as the life time of bulb number i, $\mu:=\mathbb{E}X_i$ is the mean life time of each bulb, and $T_n:=X_1+\cdots+X_n$ is the time that the n^{th} – bulb burns out. (We assume the bulbs are replaced immediately on burning out.) By convention, we set $T_0=0$.

Let
$$N_t:=\sup\{n\geq 0: T_n\leq t\}$$
denote the number of bulbs which have burned out up to time t. Since $\mathbb{E}X_i<\infty$, $X_i<\infty$ a.s. and therefore $T_n<\infty$ a.s. for all n. From this observation it follows that $N_t\uparrow\infty$ on the set, $\Omega_1:=\cap_{i=1}^{\infty}\{X_i<\infty\}$ – a subset of Ω with full measure.

It is reasonable to guess that $N_t\sim t/\mu$ and indeed we will show;
$$\lim_{t\uparrow\infty}\frac{1}{t}N_t=\frac{1}{\mu}\text{ a.s.} \tag{20.12}$$

To prove Eq. (20.12), by the SSLN, if $\Omega_0:=\left\{\lim_{n\to\infty}\frac{1}{n}T_n=\mu\right\}$ then $P(\Omega_0)=1$. From the definition of N_t, $T_{N_t}\leq t<T_{N_t+1}$ and so
$$\frac{T_{N_t}}{N_t}\leq\frac{t}{N_t}<\frac{T_{N_t+1}}{N_t}.$$

For $\omega\in\Omega_0\cap\Omega_1$ we have
$$\begin{aligned}\mu=\lim_{t\to\infty}\frac{T_{N_t(\omega)}(\omega)}{N_t(\omega)}&\leq\liminf_{t\to\infty}\frac{t}{N_t(\omega)}\\&\leq\limsup_{t\to\infty}\frac{t}{N_t(\omega)}\leq\lim_{t\to\infty}\left[\frac{T_{N_t(\omega)+1}(\omega)}{N_t(\omega)+1}\frac{N_t(\omega)+1}{N_t(\omega)}\right]=\mu.\end{aligned}$$

Example 20.33 (Renewal Theory II). Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. and $\{Y_i\}_{i=1}^{\infty}$ be i.i.d. non-negative integrable random variables with $\{X_i\}_{i=1}^{\infty}$ being independent of the $\{Y_i\}_{i=1}^{\infty}$ and let $\mu=\mathbb{E}X_1$ and $\nu=\mathbb{E}Y_1$. Again assume that $P(X_i>0)>0$. We will interpret Y_i to be the amount of time the i^{th} – bulb remains out after burning out before it is replaced by bulb number $i+1$. Let R_t be the amount of time that we have a working bulb in the time interval $[0,t]$. We are now going to show
$$\lim_{t\uparrow\infty}\frac{1}{t}R_t=\frac{\mathbb{E}X_1}{\mathbb{E}X_1+\mathbb{E}Y_1}=\frac{\mu}{\mu+\nu}.$$

To prove this, let $T_n:=\sum_{i=1}^{n}(X_i+Y_i)$ be the time that the n^{th} – bulb is replaced and
$$N_t:=\sup\{n\geq 0: T_n\leq t\}$$
denote the number of bulbs which have burned out up to time n. By Example 20.32 we know that
$$\lim_{t\uparrow\infty}\frac{1}{t}N_t=\frac{1}{\mu+\nu}\text{ a.s., i.e. }N_t=\frac{1}{\mu+\nu}t+o(t)\text{ a.s.}$$

Let us now set $\tilde{R}_t=\sum_{i=1}^{N_t}X_i$ and observe that
$$\tilde{R}_t\leq R_t\leq\tilde{R}_t+X_{N_t+1}.$$

By Proposition 20.26 we know that $X_n/n\to 0$ a.s. and therefore,
$$\lim_{t\uparrow\infty}\frac{X_{N_t+1}}{t}=\lim_{t\uparrow\infty}\left[\frac{X_{N_t+1}}{N_t+1}\cdot\frac{N_t+1}{t}\right]=0\cdot\frac{1}{\mu+\nu}=0\text{ a.s.}$$

Thus it follows that $\lim_{t\uparrow\infty}\frac{1}{t}R_t=\lim_{t\uparrow\infty}\frac{1}{t}\tilde{R}_t$ a.s. and the latter limit may be computed using the strong law of large numbers;
$$\frac{1}{t}\tilde{R}_t=\frac{1}{t}\sum_{i=1}^{N_t}X_i=\frac{N_t}{t}\cdot\frac{1}{N_t}\sum_{i=1}^{N_t}X_i\to\frac{1}{\mu+\nu}\cdot\mu\text{ a.s.}$$

Theorem 20.34 (Glivenko-Cantelli Theorem). *Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables and $F(x):=P(X_i\leq x)$. Further let $\mu_n:=\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ be the* ***empirical distribution*** *with* ***empirical distribution function,***
$$F_n(x):=\mu_n((-\infty,x])=\frac{1}{n}\sum_{i=1}^{n}1_{X_i\leq x}.$$

Then
$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|=0\text{ a.s.}$$

Proof. Since $\{1_{X_i\leq x}\}_{i=1}^{\infty}$ are i.i.d random variables with $\mathbb{E}1_{X_i\leq x}=P(X_i\leq x)=F(x)$, it follows by the strong law of large numbers that
$$\lim_{n\to\infty}F_n(x)=F(x)\text{ a.s. for all }x\in\mathbb{R}. \tag{20.13}$$

Our goal is to now show that this convergence is uniform.[1] To do this we will use another application of the strong law of large numbers applied to $\{1_{X_i<x}\}$ in order to conclude that, for all $x\in\mathbb{R}$,

[1] Observation. If F is continuous then, by what we have just shown, there is a set $\Omega_0\subset\Omega$ such that $P(\Omega_0)=1$ and on Ω_0, $F_n(r)\to F(r)$ for all $r\in\mathbb{Q}$. Moreover on Ω_0, if $x\in\mathbb{R}$ and $r\leq x\leq s$ with $r,s\in\mathbb{Q}$, we have
$$F(r)=\lim_{n\to\infty}F_n(r)\leq\liminf_{n\to\infty}F_n(x)\leq\limsup_{n\to\infty}F_n(x)\leq\lim_{n\to\infty}F_n(s)=F(s).$$

$$\lim_{n\to\infty} F_n(x-) = F(x-) \text{ a.s. for all } x \in \mathbb{R}. \tag{20.14}$$

Keep in mind that the exceptional set of probability zero depend on x.

Given $k \in \mathbb{N}$, let $\Lambda_k := \left\{\frac{i}{k} : i = 1, 2, \ldots, k-1\right\}$ and let $x_i := \inf\{x : F(x) \ge i/k\}$ for $i = 1, 2, \ldots, k-1$, see Figure 20.2. Let us further set $x_k = \infty$ and $x_0 = -\infty$ and let Ω_k denote the subset of Ω of full measure where Eqs. (20.13) and (20.14) hold for $x \in \{x_i : 1 \le i \le k-1\}$. For $\omega \in \Omega_k$ we may find $N(\omega) \in \mathbb{N}$ (N is random) so that

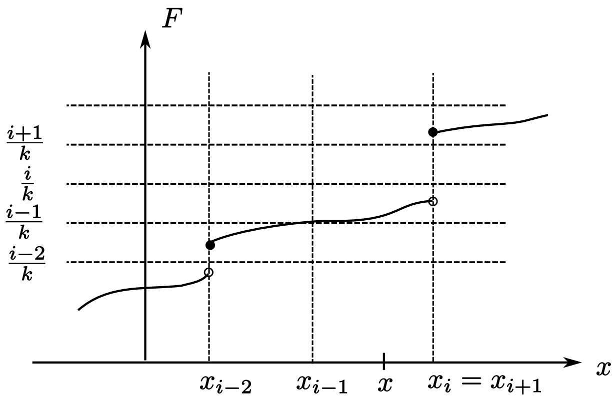


Fig. 20.2. Constructing the sequence of points $\{x_i\}_{i=0}^k$.

$$|F_n(x_i) - F(x_i)| < 1/k \text{ and } |F_n(x_i-) - F(x_i-)| < 1/k$$

for $n \ge N(\omega)$, $1 \le i \le k-1$, and $\omega \in \Omega_k$ with $P(\Omega_k) = 1$.

Observe that it is possible that $x_i = x_{i+1}$ for some of the i. This can occur when F has jumps of size greater than $1/k$,[2] see Figure 20.2. Now suppose i has been chosen so that $x_{i-1} < x_i$ and let $x \in (x_{i-1}, x_i)$. We then have for $\omega \in \Omega_k$ and $n \ge N(\omega)$ that

$$F_n(x) \le F_n(x_i-) \le F(x_i-) + 1/k \le F(x) + 2/k$$

We may now let $s \downarrow x$ and $r \uparrow x$ to conclude, on Ω_0, on

$$F(x) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x) \text{ for all } x \in \mathbb{R},$$

i.e. on Ω_0, $\lim_{n\to\infty} F_n(x) = F(x)$. Thus, in this special case we have shown that off a fixed null set independent of x that $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$.

[2] In fact if $F(x) = \delta_0((-\infty, x]) = 1_{x\ge 0}$, then $x_1 = \cdots = x_{k-1} = 0$ for all k.

and

$$F_n(x) \ge F_n(x_{i-1}) \ge F(x_{i-1}) - 1/k \ge F(x_i-) - 2/k \ge F(x) - 2/k.$$

From this it follows on Ω_k that $|F(x) - F_n(x)| \le 2/k$ for $n \ge N$ and therefore,

$$\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| \le 2/k.$$

Hence it follows on $\Omega_0 := \cap_{k=1}^\infty \Omega_k$ (a set with $P(\Omega_0) = 1$) that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0.$$

■

20.4 Kolmogorov's Three Series Theorem

The next theorem generalizes Theorem 20.11 by giving necessary and sufficient conditions for a random series of independent random variables to converge.

Theorem 20.35 (Kolmogorov's Three Series Theorem). *Suppose that $\{X_n\}_{n=1}^\infty$ are independent random variables. Then the random series, $\sum_{n=1}^\infty X_n$, is almost surely convergent in $\mathbb{R}$ iff there exists $c > 0$ such that*

1. $\sum_{n=1}^\infty P(|X_n| > c) < \infty$,
2. $\sum_{n=1}^\infty \operatorname{Var}\left(X_n 1_{|X_n|\le c}\right) < \infty$, *and*
3. $\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n|\le c}\right)$ *converges.*

Moreover, if the three series above converge for some $c > 0$ then they converge for all values of $c > 0$.

Remark 20.36. We have seen another necessary and sufficient condition in Exercise 18.20, namely $\sum_{n=1}^\infty X_n$, is almost surely convergent in $\mathbb{R}$ iff $\sum_{n=1}^\infty X_n$ is convergent in distribution. We will also see below that $\sum_{n=1}^\infty X_n$, is almost surely convergent in $\mathbb{R}$ iff $\sum_{n=1}^\infty X_n$, is convergent in probability, see Lévy's Theorem 20.46 below.

Proof. Proof of sufficiency. Suppose the three series converge for some $c > 0$. If we let $X_n' := X_n 1_{|X_n|\le c}$, then

$$\sum_{n=1}^\infty P(X_n' \ne X_n) = \sum_{n=1}^\infty P(|X_n| > c) < \infty.$$

Hence $\{X_n\}$ and $\{X_n'\}$ are tail equivalent and so it suffices to show $\sum_{n=1}^\infty X_n'$ is almost surely convergent. However, by the convergence of the second series we learn

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n'\right) = \sum_{n=1}^{\infty} \operatorname{Var}\left(X_n 1_{|X_n| \le c}\right) < \infty$$

and so by Kolmogorov's convergence criteria, Theorem 20.11,

$$\sum_{n=1}^{\infty} \left(X_n' - \mathbb{E}X_n'\right) \text{ is almost surely convergent.}$$

Finally, the third series guarantees that $\sum_{n=1}^{\infty} \mathbb{E}X_n' = \sum_{n=1}^{\infty} \mathbb{E}\left(X_n 1_{|X_n| \le c}\right)$ is convergent, therefore we may conclude $\sum_{n=1}^{\infty} X_n'$ is convergent.

The necessity proof will be completed after the next two lemmas. Another proof of necessity may be found in Chapter 23, see Theorem 23.10. ■

Lemma 20.37. *Suppose that $\{Y_n\}_{n=1}^{\infty}$ are independent random variables such that there exists $c < \infty$ such that $|Y_n| \le c < \infty$ a.s. and further assume $\mathbb{E}Y_n = 0$. If $\sum_{n=1}^{\infty} Y_n$ is almost surely convergent then $\sum_{n=1}^{\infty} \mathbb{E}Y_n^2 < \infty$. More precisely the following estimate holds,*

$$\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 \le \frac{(\lambda + c)^2}{P\left(\sup_n |S_n| \le \lambda\right)} \text{ for all } \lambda > 0, \tag{20.15}$$

where as usual, $S_n := \sum_{j=1}^{n} Y_j$.

Remark 20.38. It follows from Eq. (20.15) that if $P\left(\sup_n |S_n| < \infty\right) > 0$, then $\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 < \infty$ and hence by Kolmogorov's convergence criteria (Theorem 20.11), $\sum_{j=1}^{\infty} Y_j = \lim_{n\to\infty} S_n$ exists a.s. and in particular, $P\left(\sup_n |S_n| < \infty\right) = 1$. This also follows from the fact that $\sup_n |S_n| < \infty$ is a tail event and hence $P\left(\sup_n |S_n| < \infty\right)$ is either 0 or 1 and as $P\left(\sup_n |S_n| < \infty\right) > 0$ we must have $P\left(\sup_n |S_n| < \infty\right) = 1$.

Proof. We will begin by proving that for every $N \in \mathbb{N}$ and $\lambda > 0$ we have

$$\mathbb{E}\left[S_N^2\right] \le \frac{(\lambda + c)^2}{P\left(\sup_{n \le N} |S_n| \le \lambda\right)} \le \frac{(\lambda + c)^2}{P\left(\sup_n |S_n| \le \lambda\right)}. \tag{20.16}$$

To prove Eq. (20.16), let τ be the stopping time,

$$\tau = \tau_\lambda := \inf\{n \ge 1 : |S_n| > \lambda\}$$

where $\inf \emptyset = \infty$. Then

$$\begin{aligned}
\mathbb{E}\left[S_N^2\right] &= \mathbb{E}\left[S_N^2 : \tau \le N\right] + \mathbb{E}\left[S_N^2 : \tau > N\right] \\
&\le \mathbb{E}\left[S_N^2 : \tau \le N\right] + \lambda^2 P[\tau > N]
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[S_N^2 : \tau \le N\right] &= \sum_{j=1}^{N} \mathbb{E}\left[S_N^2 : \tau = j\right] = \sum_{j=1}^{N} \mathbb{E}\left[|S_j + S_N - S_j|^2 : \tau = j\right] \\
&= \sum_{j=1}^{N} \mathbb{E}\left[S_j^2 + 2S_j\left(S_N - S_j\right) + \left(S_N - S_j\right)^2 : \tau = j\right] \\
&= \sum_{j=1}^{N} \mathbb{E}\left[S_j^2 : \tau = j\right] + \sum_{j=1}^{N} \mathbb{E}\left[\left(S_N - S_j\right)^2\right] P[\tau = j] \\
&\le \sum_{j=1}^{N} \mathbb{E}\left[\left(S_{j-1} + Y_j\right)^2 : \tau = j\right] + \mathbb{E}\left[S_N^2\right] \sum_{j=1}^{N} P[\tau = j] \\
&\le \sum_{j=1}^{N} \mathbb{E}\left[(\lambda + c)^2 : \tau = j\right] + \mathbb{E}\left[S_N^2\right] P[\tau \le N] \\
&= \left[(\lambda + c)^2 + \mathbb{E}\left[S_N^2\right]\right] P[\tau \le N].
\end{aligned}$$

Combining the previous two estimates gives;

$$\begin{aligned}
\mathbb{E}\left[S_N^2\right] &\le \left[(\lambda + c)^2 + \mathbb{E}\left[S_N^2\right]\right] P[\tau \le N] + \lambda^2 P[\tau > N] \\
&\le \left[(\lambda + c)^2 + \mathbb{E}\left[S_N^2\right]\right] P[\tau \le N] + (\lambda + c)^2 P[\tau > N] \\
&= (\lambda + c)^2 + P[\tau \le N] \cdot \mathbb{E}\left[S_N^2\right],
\end{aligned}$$

form which Eq. (20.16) follows upon noting that

$$\{\tau \le N\} = \left\{\sup_{n \le N} |S_n| > \lambda\right\} = \left\{\sup_{n \le N} |S_n| \le \lambda\right\}^c.$$

Since S_n is convergent a.s., it follows that $P\left(\sup_n |S_n| < \infty\right) = 1$ and therefore,

$$\lim_{\lambda \uparrow \infty} P\left(\sup_n |S_n| \le \lambda\right) = 1.$$

Hence for λ sufficiently large, $P\left(\sup_n |S_n| \le \lambda\right) > 0$ and we learn from Eq. (20.16) that

$$\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 = \lim_{N\to\infty} \mathbb{E}\left[S_N^2\right] \le \frac{(\lambda + c)^2}{P\left(\sup_n |S_n| \le \lambda\right)} < \infty.$$

■

Lemma 20.39. *Suppose that $\{Y_n\}_{n=1}^\infty$ are independent random variables such that there exists $c<\infty$ such that $|Y_n|\le c$ a.s. for all n. If $\sum_{n=1}^\infty Y_n$ converges in $\mathbb{R}$ a.s. then $\sum_{n=1}^\infty \mathbb{E}Y_n$ converges as well.*

Proof. Let $(\Omega_0,\mathcal{B}_0,P_0)$ be the probability space that $\{Y_n\}_{n=1}^\infty$ is defined on and let

$$\Omega := \Omega_0\times\Omega_0,\ \mathcal{B}:=\mathcal{B}_0\otimes\mathcal{B}_0, \text{ and } P:=P_0\otimes P_0.$$

Further let $Y_n'(\omega_1,\omega_2):=Y_n(\omega_1)$ and $Y_n''(\omega_1,\omega_2):=Y_n(\omega_2)$ and

$$Z_n(\omega_1,\omega_2):=Y_n'(\omega_1,\omega_2)-Y_n''(\omega_1,\omega_2)=Y_n(\omega_1)-Y_n(\omega_2).$$

Then $|Z_n|\le 2c$ a.s., $\mathbb{E}Z_n=0$, and

$$\sum_{n=1}^\infty Z_n(\omega_1,\omega_2)=\sum_{n=1}^\infty Y_n(\omega_1)-\sum_{n=1}^\infty Y_n(\omega_2) \text{ exists}$$

for P a.e. (ω_1,ω_2). Hence it follows from Lemma 20.37 that

$$\begin{aligned}\infty>\sum_{n=1}^\infty \mathbb{E}Z_n^2&=\sum_{n=1}^\infty \operatorname{Var}(Z_n)=\sum_{n=1}^\infty \operatorname{Var}(Y_n'-Y_n'')\\&=\sum_{n=1}^\infty[\operatorname{Var}(Y_n')+\operatorname{Var}(Y_n'')]=2\sum_{n=1}^\infty \operatorname{Var}(Y_n).\end{aligned}$$

Thus by Kolmogorov's convergence theorem, it follows that $\sum_{n=1}^\infty (Y_n-\mathbb{E}Y_n)$ is convergent. Since $\sum_{n=1}^\infty Y_n$ is a.s. convergent, we may conclude that $\sum_{n=1}^\infty \mathbb{E}Y_n$ is also convergent. ■

We are now ready to complete the proof of Theorem 20.35.

Proof of Theorem 20.35. Our goal is to show if $\{X_n\}_{n=1}^\infty$ are independent random variables such that $\sum_{n=1}^\infty X_n$, is almost surely convergent then for all $c>0$ the following three series converge;

1. $\sum_{n=1}^\infty P(|X_n|>c)<\infty$,
2. $\sum_{n=1}^\infty \operatorname{Var}\left(X_n 1_{|X_n|\le c}\right)<\infty$, and
3. $\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n|\le c}\right)$ converges.

Since $\sum_{n=1}^\infty X_n$ is almost surely convergent, it follows that $\lim_{n\to\infty}X_n=0$ a.s. and hence for every $c>0$, $P(\{|X_n|\ge c \text{ i.o.}\})=0$. According the Borel zero one law (Lemma 10.41) this implies for every $c>0$ that $\sum_{n=1}^\infty P(|X_n|>c)<\infty$. Given this, we now know that $\{X_n\}$ and $\left\{X_n^c:=X_n 1_{|X_n|\le c}\right\}$ are tail equivalent for all $c>0$ and in particular $\sum_{n=1}^\infty X_n^c$ is almost surely convergent for all $c>0$. So according to Lemma 20.39 (with $Y_n=X_n^c$),

$$\sum_{n=1}^\infty \mathbb{E}X_n^c=\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n|\le c}\right) \text{ converges.}$$

Letting $Y_n:=X_n^c-\mathbb{E}X_n^c$, we may now conclude that $\sum_{n=1}^\infty Y_n$ is almost surely convergent. Since $\{Y_n\}$ is uniformly bounded and $\mathbb{E}Y_n=0$ for all n, an application of Lemma 20.37 allows us to conclude

$$\sum_{n=1}^\infty \operatorname{Var}\left(X_n 1_{|X_n|\le c}\right)=\sum_{n=1}^\infty \mathbb{E}Y_n^2<\infty.$$

■

Exercise 20.4 (Two Series Theorem – Resnik 7.15). Prove that the three series theorem reduces to a two series theorem when the random variables are positive. That is, if $X_n\ge 0$ are independent, then $\sum_n X_n<\infty$ a.s. iff for any $c>0$ we have

$$\sum_n P(X_n>c)<\infty \text{ and} \tag{20.17}$$
$$\sum_n \mathbb{E}[X_n 1_{X_n\le c}]<\infty, \tag{20.18}$$

that is it is unnecessary to verify the convergence of the second series in Theorem 20.35 involving the variances.

20.4.1 Examples

Lemma 20.40. *Suppose that $\{Y_n\}_{n=1}^\infty$ are independent square integrable random variables such that $Y_n\overset{d}{=}N\left(\mu_n,\sigma_n^2\right)$. Then $\sum_{j=1}^\infty Y_j$ converges a.s. iff $\sum_{j=1}^\infty \sigma_j^2<\infty$ and $\sum_{j=1}^\infty \mu_j$ converges.*

Proof. The implication "$\Longleftarrow$" is true without the assumption that the Y_n are normal random variables as pointed out in Theorem 20.11. To prove the converse directions we will make use of the Kolmogorov's three series Theorem 20.35. Namely, if $\sum_{j=1}^\infty Y_j$ converges a.s. then the three series in Theorem 20.35 converge for all $c>0$.

1. Since $Y_n\overset{d}{=}\sigma_n N+\mu_n$, we have for any $c>0$ that

$$\infty>\sum_{n=1}^\infty P(|\sigma_n N+\mu_n|>c). \tag{20.19}$$

If $\lim_{n\to\infty}\mu_n\ne 0$ then there is a $c>0$ such that either $\mu_n\ge 2c$ for infinitely many n or $\mu_n\le -2c$ for infinitely many n. It then follows that either $\{N>0\}\subset\{|\sigma_n N+\mu_n|>c\}$ n i.o. or $\{N<0\}\subset\{|\sigma_n N+\mu_n|>c\}$ n i.o. In either case we would have $P(|\sigma_n N+\mu_n|>c)\ge 1/2$ i.o. which would violate Eq. (20.19) and so we may concluded that $\lim_{n\to\infty}\mu_n=0$. Similarly if $\lim_{n\to\infty}\sigma_n\ne 0$, then there exists $\alpha<\infty$ such that

$$\{N \geq \alpha\} \subset \{|\sigma_n N + \mu_n| > 1\} \ n \text{ i.o.}$$

which would imply $P(|\sigma_n N + \mu_n| > 1) \geq P(N \geq \alpha) > 0$ for infinitely many n. This again violate Eq. (20.19) and thus we may conclude that $\lim_{n\to\infty} \mu_n = \lim_{n\to\infty} \sigma_n = 0$.

2. Let $\chi_n := 1_{|\sigma_n N + \mu_n| \leq c} \in \{0,1\}$. The convergence of the second series for all $c > 0$ implies

$$\infty > \sum_{n=1}^{\infty} \operatorname{Var}\left(Y_n 1_{|Y_n| \leq c}\right) = \sum_{n=1}^{\infty} \operatorname{Var}\left([\sigma_n N + \mu_n] \chi_n\right). \tag{20.20}$$

If we can show

$$\operatorname{Var}\left([\sigma_n N + \mu_n] \chi_n\right) \geq \frac{1}{2}\sigma_n^2 \text{ for large } n, \tag{20.21}$$

it would then follow from Eq. (20.20) that $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. We may now use Kolmogorov's convergence criteria (Theorem 20.11) to infer that $\sum_{n=1}^{\infty}(Y_n - \mu_n)$ is almost surely convergent which then implies that $\sum_{n=1}^{\infty} \mu_n$ is convergent as $\mu_n = Y_n - (Y_n - \mu_n)$ and $\sum_{n=1}^{\infty} Y_n$ and $\sum_{n=1}^{\infty}(Y_n - \mu_n)$ are both convergent a.s. So to finish the proof we need to prove the estimate in Eq. (20.21).

Let $\alpha_n := \operatorname{Var}(N\chi_n)$ and $\beta_n := P(\chi_n = 1)$ so that $\operatorname{Var}(\chi_n) = \beta_n(1 - \beta_n)$ and

$$\varepsilon_n := \operatorname{Cov}(N\chi_n, \chi_n) = \mathbb{E}[N\chi_n \cdot \chi_n] - \mathbb{E}[N\chi_n]\,\mathbb{E}[\chi_n] = \mathbb{E}[N\chi_n](1 - \beta_n).$$

Therefore, using $\operatorname{Var}(\sigma X + \mu Y) = \sigma^2 \operatorname{Var}(X) + \mu^2 \operatorname{Var}(Y) + 2\sigma\mu \operatorname{Cov}(X,Y)$, we find

$$\begin{aligned}\operatorname{Var}\left([\sigma_n N + \mu_n]\chi_n\right) &= \operatorname{Var}(\sigma_n N\chi_n + \mu_n \chi_n) \\ &= \sigma_n^2 \alpha_n + \mu_n^2 \beta_n (1 - \beta_n) + 2\sigma_n \mu_n \varepsilon_n.\end{aligned}$$

Making use of the estimate, $2ab \leq a^2 + b^2$ valid for all $a, b \geq 0$, it follows that

$$\begin{aligned}\operatorname{Var}\left([\sigma_n N + \mu_n]\chi_n\right) &\geq \sigma_n^2 \alpha_n + \mu_n^2 \beta_n (1 - \beta_n) - 2|\varepsilon_n|\,\sigma_n\,|\mu_n| \\ &\geq \sigma_n^2(\alpha_n - |\varepsilon_n|) + \mu_n^2(\beta_n(1 - \beta_n) - |\varepsilon_n|) \\ &= \sigma_n^2(\alpha_n - |\varepsilon_n|) + (1 - \beta_n)(\beta_n - |\mathbb{E}[N\chi_n]|)\,\mu_n^2.\end{aligned}$$

This estimate along with the observations that $1 - \beta_n \geq 0$, $\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \beta_n = 1$, $\lim_{n\to\infty} \mathbb{E}[N\chi_n] = 0$ (use DCT) and $\lim_{n\to\infty} \varepsilon_n = 0$ easily implies Eq. (20.21).

An alternative proof that $\sum_{n=1}^{\infty} \mu_n$ is convergent using the the third series in Theorem 20.35. For all $c > 0$ the third series implies

$$\sum_{n=1}^{\infty} \mathbb{E}\left([\sigma_n N + \mu_n]\, 1_{|\sigma_n N + \mu_n| \leq c}\right) \text{ is convergent, i.e.}$$

$$\sum_{n=1}^{\infty} [\sigma_n \delta_n + \mu_n \beta_n] \text{ is convergent.}$$

where $\delta_n := \mathbb{E}\left(N \cdot 1_{|\sigma_n N + \mu_n| \leq c}\right)$ and $\beta_n := \mathbb{E}\left(1_{|\sigma_n N + \mu_n| \leq c}\right)$. With a little effort one can show,

$$\delta_n \sim e^{-k/\sigma_n^2} \text{ and } 1 - \beta_n \sim e^{-k/\sigma_n^2} \text{ for large } n.$$

Since $e^{-k/\sigma_n^2} \leq C\sigma_n^2$ for large n, it follows that $\sum_{n=1}^{\infty} |\sigma_n \delta_n| \leq C \sum_{n=1}^{\infty} \sigma_n^3 < \infty$ so that $\sum_{n=1}^{\infty} \mu_n \beta_n$ is convergent. Moreover,

$$\sum_{n=1}^{\infty} |\mu_n(\beta_n - 1)| \leq C \sum_{n=1}^{\infty} |\mu_n|\,\sigma_n^2 < \infty$$

and hence

$$\sum_{n=1}^{\infty} \mu_n = \sum_{n=1}^{\infty} \mu_n \beta_n - \sum_{n=1}^{\infty} \mu_n(\beta_n - 1)$$

must also be convergent. ■

Example 20.41. As another simple application of Theorem 20.35, let us use it to give a proof of Theorem 20.11. We will apply Theorem 20.35 with $X_n := Y_n - \mathbb{E}Y_n$. We need to then check the three series in the statement of Theorem 20.35 converge. For the first series we have by the Markov inequality,

$$\sum_{n=1}^{\infty} P(|X_n| > c) \leq \sum_{n=1}^{\infty} \frac{1}{c^2}\mathbb{E}|X_n|^2 = \frac{1}{c^2}\sum_{n=1}^{\infty} \operatorname{Var}(Y_n) < \infty.$$

For the second series, observe that

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n 1_{|X_n| \leq c}\right) \leq \sum_{n=1}^{\infty} \mathbb{E}\left[\left(X_n 1_{|X_n| \leq c}\right)^2\right] \leq \sum_{n=1}^{\infty} \mathbb{E}\left[X_n^2\right] = \sum_{n=1}^{\infty} \operatorname{Var}(Y_n) < \infty$$

and we estimate the third series as;

$$\sum_{n=1}^{\infty} \left|\mathbb{E}\left(X_n 1_{|X_n| \leq c}\right)\right| \leq \sum_{n=1}^{\infty} \mathbb{E}\left(\frac{1}{c}|X_n|^2 1_{|X_n| \leq c}\right) \leq \frac{1}{c}\sum_{n=1}^{\infty} \operatorname{Var}(Y_n) < \infty.$$

20.5 Maximal Inequalities

Theorem 20.42 (Kolmogorov's Inequality). *Let $\{X_n\}$ be a sequence of independent random variables with mean zero, $S_n := X_1 + \cdots + X_n$, and $S_n^* = \max_{j\le n} |S_j|$. Then for any $\alpha > 0$ we have*

$$P(S_N^* \ge \alpha) \le \frac{1}{\alpha^2}\mathbb{E}\left[S_N^2 : S_N^* \ge \alpha\right]. \tag{20.22}$$

Proof. First proof. As $\{S_n\}_{n=1}^\infty$ is a martingale relative to the filtration, $\mathcal{B}_n = \sigma(S_1, \dots, S_n)$, the inequality in Eq. (20.22) is a special case of Proposition 18.42 with $X_n = S_n^2$, also see Example 18.48.

*__Second direct proof.__ Let $\tau = \inf\{j : |S_j| \ge \alpha\}$ with the infimum of the empty set being taken to be equal to ∞. Observe that

$$\{\tau = j\} = \{|S_1| < \alpha, \dots, |S_{j-1}| < \alpha, |S_j| \ge \alpha\} \in \sigma(X_1, \dots, X_j).$$

Now

$$\begin{aligned}\mathbb{E}\left[S_N^2 : |S_N^*| > \alpha\right] &= \mathbb{E}\left[S_N^2 : \tau \le N\right] = \sum_{j=1}^N \mathbb{E}\left[S_N^2 : \tau = j\right] \\ &= \sum_{j=1}^N \mathbb{E}\left[(S_j + S_N - S_j)^2 : \tau = j\right] \\ &= \sum_{j=1}^N \mathbb{E}\left[S_j^2 + (S_N - S_j)^2 + 2S_j(S_N - S_j) : \tau = j\right] \\ &\overset{(*)}{=} \sum_{j=1}^N \mathbb{E}\left[S_j^2 + (S_N - S_j)^2 : \tau = j\right] \\ &\ge \sum_{j=1}^N \mathbb{E}\left[S_j^2 : \tau = j\right] \ge \alpha^2 \sum_{j=1}^N P[\tau = j] = \alpha^2 P(|S_N^*| > \alpha).\end{aligned}$$

The equality, $(*)$, is a consequence of the observations: 1) $1_{\tau=j}S_j$ is $\sigma(X_1, \dots, X_j)$ – measurable, 2) $(S_n - S_j)$ is $\sigma(X_{j+1}, \dots, X_n)$ – measurable and hence $1_{\tau=j}S_j$ and $(S_n - S_j)$ are independent, and so 3)

$$\begin{aligned}\mathbb{E}[S_j(S_N - S_j) : \tau = j] &= \mathbb{E}[S_j 1_{\tau=j}(S_N - S_j)] \\ &= \mathbb{E}[S_j 1_{\tau=j}] \cdot \mathbb{E}[S_N - S_j] = \mathbb{E}[S_j 1_{\tau=j}] \cdot 0 = 0.\end{aligned}$$

■

Remark 20.43 (Another proof of Theorem 20.11). Suppose that $\{Y_j\}_{j=1}^\infty$ are independent random variables such that $\sum_{j=1}^\infty \operatorname{Var}(Y_j) < \infty$ and let $S_n := \sum_{j=1}^n X_j$ where $X_j := Y_j - \mathbb{E}Y_j$. According to Kolmogorov's inequality, Theorem 20.42, for all $M < N$,

$$\begin{aligned}P\left(\max_{M\le j\le N} |S_j - S_M| \ge \alpha\right) &\le \frac{1}{\alpha^2}\mathbb{E}\left[(S_N - S_M)^2\right] = \frac{1}{\alpha^2}\sum_{j=M+1}^N \mathbb{E}\left[X_j^2\right] \\ &= \frac{1}{\alpha^2}\sum_{j=M+1}^N \operatorname{Var}(X_j).\end{aligned}$$

Letting $N \to \infty$ in this inequality shows, with $Q_M := \sup_{j\ge M} |S_j - S_M|$,

$$P(Q_M \ge \alpha) \le \frac{1}{\alpha^2}\sum_{j=M+1}^\infty \operatorname{Var}(X_j).$$

Since

$$\delta_M := \sup_{j,k\ge M} |S_j - S_k| \le \sup_{j,k\ge M} [|S_j - S_M| + |S_M - S_k|] \le 2Q_M$$

we may further conclude,

$$P(\delta_M \ge 2\alpha) \le \frac{1}{\alpha^2}\sum_{j=M+1}^\infty \operatorname{Var}(X_j) \to 0 \text{ as } M \to \infty,$$

i.e. $\delta_M \overset{P}{\to} 0$ as $M \to \infty$. Since δ_M is decreasing in M, it follows that $\lim_{M\to\infty} \delta_M =: \delta$ exists and because $\delta_M \overset{P}{\to} 0$ we may concluded that $\delta = 0$ a.s. Thus we have shown

$$\lim_{m,n\to\infty} |S_n - S_m| = 0 \text{ a.s.}$$

and therefore $\{S_n\}_{n=1}^\infty$ is almost surely Cauchy and hence almost surely convergent. This gives a second proof of Kolmogorov's convergence criteria in Theorem 20.11.

Corollary 20.44 (L^2 – SSLN). *Let $\{X_n\}$ be a sequence of independent random variables with mean zero, and $\sigma^2 = \mathbb{E}X_n^2 < \infty$. Letting $S_n = \sum_{k=1}^n X_k$ and $p > 1/2$, we have*

$$\frac{1}{n^p}S_n \to 0 \text{ a.s.}$$

If $\{Y_n\}$ is a sequence of independent random variables $\mathbb{E}Y_n = \mu$ and $\sigma^2 = \operatorname{Var}(X_n) < \infty$, then for any $\beta \in (0, 1/2)$,

$$\frac{1}{n}\sum_{k=1}^{n} Y_k - \mu = O\left(\frac{1}{n^\beta}\right).$$

Proof. (The proof of this Corollary may be skipped as it has already been proved, see Corollary 20.15.) From Theorem 20.42, we have for every $\varepsilon > 0$ that

$$P\left(\frac{S_N^*}{N^p} \ge \varepsilon\right) = P\left(S_N^* \ge \varepsilon N^p\right) \le \frac{1}{\varepsilon^2 N^{2p}}\mathbb{E}\left[S_N^2\right] = \frac{1}{\varepsilon^2 N^{2p}} CN = \frac{C}{\varepsilon^2 N^{(2p-1)}}.$$

Hence if we suppose that $N_n = n^\alpha$ with $\alpha(2p-1) > 1$, then we have

$$\sum_{n=1}^{\infty} P\left(\frac{S_{N_n}^*}{N_n^p} \ge \varepsilon\right) \le \sum_{n=1}^{\infty} \frac{C}{\varepsilon^2 n^{\alpha(2p-1)}} < \infty$$

and so by the first Borel – Cantelli lemma we have

$$P\left(\left\{\frac{S_{N_n}^*}{N_n^p} \ge \varepsilon \text{ for } n \text{ i.o.}\right\}\right) = 0.$$

From this it follows that $\lim_{n\to\infty} \frac{S_{N_n}^*}{N_n^p} = 0$ a.s.

To finish the proof, for $m \in \mathbb{N}$, we may choose $n = n(m)$ such that

$$n^\alpha = N_n \le m < N_{n+1} = (n+1)^\alpha.$$

Since

$$\frac{S_{N_{n(m)}}^*}{N_{n(m)+1}^p} \le \frac{S_m^*}{m^p} \le \frac{S_{N_{n(m)+1}}^*}{N_{n(m)}^p}$$

and

$$N_{n+1}/N_n \to 1 \text{ as } n \to \infty,$$

it follows that

$$\begin{aligned} 0 = \lim_{m\to\infty} \frac{S_{N_{n(m)}}^*}{N_{n(m)}^p} &= \lim_{m\to\infty} \frac{S_{N_{n(m)}}^*}{N_{n(m)+1}^p} \le \lim_{m\to\infty} \frac{S_m^*}{m^p} \\ &\le \lim_{m\to\infty} \frac{S_{N_{n(m)+1}}^*}{N_{n(m)}^p} = \lim_{m\to\infty} \frac{S_{N_{n(m)+1}}^*}{N_{n(m)+1}^p} = 0 \text{ a.s.} \end{aligned}$$

That is $\lim_{m\to\infty} \frac{S_m^*}{m^p} = 0$ a.s. ∎

We are going to give three more maximal inequalities before ending this section. In all case we will start with $\{X_n\}_{n=1}^\infty$ a sequence of (possibly with values in a separable Banach space, Y) random variables and we will let $S_n := \sum_{k\le n} X_k$ and $S_n^* := \max_{k\le n}\|S_k\|$. If τ is any $\mathcal{B}_n^X$ – stopping time and $f \ge 0$, then

$$\mathbb{E}\left[f(S_n - S_\tau) : \tau \le n\right] = \sum_{k=1}^{n} \mathbb{E}\left[f(S_n - S_k) : \tau = k\right] = \sum_{k=1}^{n} \mathbb{E}\left[f(S_n - S_k)\right] \cdot P(\tau = k). \tag{20.23}$$

Theorem 20.45 (Skorohod's Inequality). *Suppose that $\{X_n\}_{n=1}^\infty$ are independent real or Banach valued random variables. Then for all $\alpha > 0$ we have*

$$P(\|S_N\| \ge \alpha) \ge (1 - c_N(\alpha)) P(S_N^* \ge 2\alpha) \text{ and} \tag{20.24}$$
$$P(\|S_N\| > \alpha) \ge (1 - c_N(\alpha)) P(S_N^* > 2\alpha) \tag{20.25}$$

where

$$c_N(\alpha) := \max_{1\le k\le N} P(\|S_N - S_k\| > \alpha).$$

Proof. We only prove Eq. (20.24) since the proof of Eq. (20.25) is similar and in fact can be deduced from Eq. (20.24) by a simple limiting argument. If $\tau = \inf\{n : \|S_n\| \ge 2\alpha\}$, then $\{\tau \le N\} = \{S_N^* \ge 2\alpha\}$ we have

$$\begin{aligned} \|S_N\| &= \|S_\tau + S_N - S_\tau\| \ge \|S_\tau\| - \|S_N - S_\tau\| \\ &\ge 2\alpha - \|S_N - S_\tau\|. \end{aligned}$$

From this it follows that

$$\{\tau \le N \ \& \ \|S_N - S_\tau\| \le \alpha\} \subset \{\|S_N\| \ge \alpha\}$$

and therefore,

$$\begin{aligned} P(\|S_N\| \ge \alpha) &\ge P(\tau \le N \ \& \ \|S_N - S_\tau\| \le \alpha) \\ &= \sum_{k=1}^{N} P(\tau = k) \cdot P(\|S_N - S_k\| \le \alpha) \\ &\ge \min_{1\le k\le N} P(\|S_N - S_k\| \le \alpha) \cdot \sum_{k=1}^{N} P(\tau = k) \\ &= c_N(\alpha) \cdot P(S_N^* \ge 2\alpha). \end{aligned}$$

∎

As an application of Theorem 20.45 we have the following convergence result.

Theorem 20.46 (Lévy's Theorem). *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables then $\sum_{n=1}^\infty X_n$ converges in probability iff $\sum_{n=1}^\infty X_n$ converges a.s.*

Proof. Let $S_n := \sum_{k=1}^n X_k$. Since almost sure convergence implies convergence in probability, it suffices to show; if S_n is convergent in probability then S_n is almost surely convergent. Given $M \in \mathbb{M}$, let $Q_M := \sup_{n\ge M}|S_n - S_M|$ and for $M < N$, let $Q_{M,N} := \sup_{M \le n \le N} |S_n - S_M|$. Given $\varepsilon \in (0,1)$, by assumption, there exists $M = M(\varepsilon) \in \mathbb{N}$ such that $\max_{M \le j \le N} P(|S_N - S_j| > \varepsilon) < \varepsilon$ for all $N \ge M$. An application of Skorohod's inequality (Theorem 20.45), then shows

$$P(Q_{M,N} \ge 2\varepsilon) \le \frac{P(|S_N - S_M| > \varepsilon)}{(1 - \max_{M \le j \le N} P(|S_N - S_j| > \varepsilon))} \le \frac{\varepsilon}{1-\varepsilon}.$$

Since $Q_{M,N} \uparrow Q_M$ as $N \to \infty$, we may conclude

$$P(Q_M \ge 2\varepsilon) \le \frac{\varepsilon}{1-\varepsilon}.$$

Since,

$$\delta_M := \sup_{m,n \ge M} |S_n - S_m| \le \sup_{m,n \ge M} [|S_n - S_M| + |S_M - S_m|] = 2Q_M$$

we may further conclude, $P(\delta_M > 4\varepsilon) \le \frac{\varepsilon}{1-\varepsilon}$ and since $\varepsilon > 0$ is arbitrary, it follows that $\delta_M \xrightarrow{P} 0$ as $M \to \infty$. Moreover, since δ_M is decreasing in M, it follows that $\lim_{M\to\infty} \delta_M =: \delta$ exists and because $\delta_M \xrightarrow{P} 0$ we may concluded that $\delta = 0$ a.s. Thus we have shown

$$\lim_{m,n\to\infty} |S_n - S_m| = 0 \text{ a.s.}$$

and therefore $\{S_n\}_{n=1}^\infty$ is almost surely Cauchy and hence almost surely convergent. ■

Remark 20.47 (Yet another proof of Theorem 20.11). Suppose that $\{Y_j\}_{j=1}^\infty$ are independent random variables such that $\sum_{j=1}^\infty \operatorname{Var}(Y_j) < \infty$. By Proposition 20.10, the sum, $\sum_{j=1}^\infty (Y_j - \mathbb{E}Y_j)$, is $L^2(P)$ convergent and hence convergent in probability. An application of Lévy's Theorem 20.46 then shows $\sum_{j=1}^\infty (Y_j - \mathbb{E}Y_j)$ is almost surely convergent which gives another proof of Kolmogorov's convergence criteria in Theorem 20.11.

The next maximal inequality will be useful later in proving the "functional central limit theorem." It is actually a simple corollary of Skorohod's inequality (Theorem 20.45) along with Chebyshev's inequality.

Corollary 20.48 (Ottaviani's maximal ineqaulity). *Suppose that $\{X_n\}_{n=1}^\infty$ are independent real or Banach valued square integrable random variables. Then for all $\alpha > 0$ we have*

$$P(\|S_N\| \ge \alpha) \ge \left(1 - \frac{1}{\alpha^2} \max_{1 \le k \le N} \mathbb{E}\|S_N - S_k\|^2\right) P(S_N^* \ge 2\alpha)$$

and in particular if $\alpha^2 > \max_{1 \le k \le N} \mathbb{E}\|S_N - S_k\|^2$, then

$$P(S_N^* \ge 2\alpha) \le \left(1 - \frac{1}{\alpha^2} \max_{1 \le k \le N} \mathbb{E}\|S_N - S_k\|^2\right)^{-1} P(\|S_N\| \ge \alpha).$$

If we further assume that $\{X_n\}$ are real (or Hilbert valued) mean zero random variables, then

$$P(S_N^* \ge 2\alpha) \le \left(1 - \frac{1}{\alpha^2} \mathbb{E}\|S_N - X_1\|^2\right)^{-1} P(\|S_N\| \ge \alpha). \tag{20.26}$$

Proof. The first and second inequalities follow by Chebyshev's inequality and Skorohod's Theorem 20.45. When the $\{X_n\}$ are real or Hilbert valued mean zero square integrable random variables, we have

$$\max_{1 \le k \le N} \mathbb{E}\|S_N - S_k\|^2 = \max_{1 \le k \le N} \sum_{j=k+1}^N \mathbb{E}\|X_j\|^2 = \sum_{j=2}^N \mathbb{E}\|X_j\|^2 = \|S_N - X_1\|^2.$$

■

Corollary 20.49. *Suppose $\lambda > 1$ and $\{X_n\}_{n=1}^\infty$ are independent real square integrable random variables with $\mathbb{E}X_n = 0$ and $\operatorname{Var}(X_n) = 1$ for all n. Then*

$$P\left(S_n^* \ge 2\lambda\sqrt{n}\right) \le \left(1 - \frac{1}{\lambda^2}\right)^{-1} \cdot P\left(|S_n| \ge \lambda\sqrt{n}\right)$$

and if we further assume that $\{X_n\}_{n=1}^\infty$ are i.i.d., then

$$\lim_{n\to\infty} P\left(S_n^* \ge 2\lambda\sqrt{n}\right) \le \sqrt{\frac{2}{\pi}} \left(1 - \frac{1}{\lambda^2}\right)^{-1} \frac{1}{\lambda} e^{-\lambda^2/2}.$$

Proof. The first inequality follows from Eq. (20.26) of Corollary 20.48 with $\alpha = \lambda\sqrt{n}$. For the second inequality we use the central limit theorem to conclude that

$$P\left(|S_n| \ge \lambda\sqrt{n}\right) = P\left(\frac{|S_n|}{\sqrt{n}} \ge \lambda\right) \to P(|Z| \ge \lambda)$$

where Z is a standard normal random variable. We then estimate $P(|Z| \ge \lambda)$ using the Gaussian tail estimates in Lemma 7.59. ■

We can significantly improve on Corollary 20.48 if we further assume that X_n is symmetric for n in which case the following reflection principle holds.

Theorem 20.50. *Suppose that $\{X_n\}_{n=1}^\infty$ are independent real or Banach valued random variables such that $X_n \overset{d}{=} -X_n$ for all n and τ is any $\left\{\mathcal{B}_n^X = \sigma(X_1, \dots, X_n)\right\}_{n=1}^\infty$ stopping time. If we set*

$$\begin{aligned} S_n^\tau &:= 1_{n\le\tau} S_n - 1_{n>\tau}\left(S_n - S_\tau\right) \\ &= 1_{n\le\tau} S_n - 1_{n>\tau} \sum_{\tau<k\le n} X_k, \end{aligned}$$

then $(\{S_n\}_{n=1}^\infty, \tau)$ *and* $(\{S_n^\tau\}_{n=1}^\infty, \tau)$ *have the same distribution. (Notice* $S_n^\tau = S_n$ *for* $n \le \tau$ *and* S_n^τ *is* S_n *"reflected about* S_τ*" for* $n > \tau$*.)*

Proof. Let $N \in \mathbb{N}$ be given and $f : S^N \to \mathbb{R}$ be a bounded measurable function. Then, for all $k \le N$ we have,

$$\begin{aligned} &\mathbb{E}\left[f\left(S_1^\tau, \ldots, S_N^\tau\right) : \tau = k\right] \\ &= \mathbb{E}\left[f\left(S_1, \ldots, S_k, \left(S_k - X_{k+1}\right), \ldots, \left(S_k - X_{k+1} - \cdots - X_N\right)\right) : \tau = k\right] \\ &= \mathbb{E}\left[f\left(S_1, \ldots, S_k, \left(S_k + X_{k+1}\right), \ldots, \left(S_k + X_{k+1} + \cdots + X_N\right)\right) : \tau = k\right] \\ &= \mathbb{E}\left[f\left(S_1, \ldots, S_N\right) : \tau = k\right], \\ &= \end{aligned}$$

wherein we have used $(X_{k+1}, \ldots, X_N) \overset{d}{=} -(X_{k+1}, \ldots, X_N)$, $(X_{k+1}, \ldots, X_N)$ is independent of $\mathcal{B}_k^X$, and $\{\tau = k\}$ and $(S_1, \ldots, S_k)$ are $\mathcal{B}_k^X$ – measurable. This completes the proof since on $\{\tau = \infty\}$, $\{S_n\}_{n=1}^\infty = \{S_n^\tau\}_{n=1}^\infty$. ∎

In order to exploit this principle we will need to combine it with the following simple **geometric reflection property** for Banach spaces; if $r > 0$ and $x, y \in Y$ (Y is a normed space) such that $\|x\| \ge r$ while $\|x - y\| < r$, then $\|x + y\| > r$. This is easy to believe (draw the picture for $Y = \mathbb{R}^2$) and it is also easy to prove;

$$\begin{aligned} \|x + y\| &= \|2x - (x - y)\| \\ &\ge \|2x\| - \|x - y\| \\ &\ge 2r - \|x - y\| > 2r - r = r. \end{aligned}$$

Proposition 20.51 (Reflection Principle). *Suppose that* $\{X_n\}_{n=1}^\infty$ *are independent real or Banach valued random variables such that* $X_n \overset{d}{=} -X_n$ *for all* n*. Then*

$$P(S_N^* \ge r) \le P(\|S_N\| \ge r) + P(\|S_N\| > r) \le 2P(\|S_N\| \ge r). \tag{20.27}$$

Proof. Let $\tau := \inf\{n : \|S_n\| \ge r\}$ (a $\left\{\mathcal{B}_n^X\right\}_{n=1}^\infty$ – stopping time), then

$$\begin{aligned} P(S_N^* \ge r) &= P(S_N^* \ge r, \|S_N\| \ge r) + P(S_N^* \ge r, \|S_N\| < r) \\ &= P(\|S_N\| \ge r) + P(\tau \le N, \|S_N\| < r). \end{aligned}$$

Moreover by the reflection principle (Theorem 20.50),

$$\begin{aligned} P(\tau \le N, \|S_N\| < r) &= P(\tau \le N, \|S_N^\tau\| < r) \\ &= P\left(\tau \le N, \|S_\tau - (S_N - S_\tau)\| < r\right). \end{aligned}$$

If $\|S_\tau\| \ge r$ and $\|S_\tau - (S_N - S_\tau)\| < r$, then by the geometric reflection property, $\|S_N\| = \|S_\tau + (S_N - S_\tau)\| > r$ and therefore

$$P\left(\tau \le N, \|S_\tau - (S_N - S_\tau)\| < r\right) \ge P\left(\tau \le N, \|S_N\| > r\right) = P\left(\|S_N\| > r\right).$$

Combining this inequality with the first displayed inequality in the proof easily gives the result.

∎

Exercise 20.5 (Simple Random Walk Reflection principle). Let $\{X_n\}_{n=1}^\infty$ be i.i.d Bernoulli random variables with $P(X_n = \pm 1) = \frac{1}{2}$ for all n and let $S_n := \sum_{k\le n} X_k$ be the standard simple random walk on $\mathbb{Z}$. Show for every $r \in \mathbb{N}$ that

$$P\left(\max_{k\le n} S_k \ge r\right) = P(S_n \ge r) + P(S_n > r).$$

21

Weak Convergence Results

In this chapter we will discuss a couple of different ways to decide weather two probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ are "close" to one another. This will arise later as follows. Suppose $\{Y_n\}_{n=1}^\infty$ is a sequence of random variables and Y is another random variable (possibly defined on a different probability space). We would like to understand when, for large n, Y_n and Y have nearly the "same" distribution, i.e. when is $\mu_n := \mathrm{Law}(Y_n)$ close to $\mu = \mathrm{Law}(Y)$ for large n.

We will often be the case that $Y_n = X_1 + \cdots + X_n$ where $\{X_i\}_{i=1}^n$ are independent random variables. For this reason it will be useful to record the procedure for computing the law of Y_n in terms of the laws of the $\{X_i\}_{i=1}^n$. So before going to the main theme of this chapter let us pause to introduce the relevant notion of the convolution of probability measures on $\mathbb{R}^n$.

21.1 Convolutions

Definition 21.1. *Let μ and ν be two probability measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. The* ***convolution*** *of μ and ν, denoted $\mu * \nu$, is the measure, $P \circ (X+Y)^{-1}$ where $\{X, Y\}$ are two independent random vectors such that $P \circ X^{-1} = \mu$ and $P \circ Y^{-1} = \nu$.*

Of course we may give a more direct definition of the convolution of μ and ν by observing for $A \in \mathcal{B}_{\mathbb{R}^n}$ that

$$\begin{aligned}
(\mu * \nu)(A) &= P(X + Y \in A) \\
&= \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} d\nu(y) 1_A(x+y) \qquad (21.1)\\
&= \int_{\mathbb{R}^n} \nu(A - x)\, d\mu(x) \qquad (21.2)\\
&= \int_{\mathbb{R}^n} \mu(A - x)\, d\nu(x). \qquad (21.3)
\end{aligned}$$

This may also be expressed as,

$$(\mu * \nu)(A) = \int_{\mathbb{R}^n \times \mathbb{R}^n} 1_A(x+y)\, d\mu(x)\, d\nu(y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} 1_A(x+y)\, d(\mu \otimes \nu)(x, y). \qquad (21.4)$$

Exercise 21.1. Let μ, ν, and γ be three probability measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. Show;

1. $\mu * \nu = \nu * \mu$.
2. $\mu * (\nu * \gamma) = (\mu * \nu) * \gamma$. (So it is now safe to write $\mu * \nu * \gamma$ for either side of this equation.)
3. $(\mu * \delta_x)(A) = \mu(A - x)$ for all $x \in \mathbb{R}^n$ where $\delta_x(A) := 1_A(x)$ for all $A \in \mathcal{B}_{\mathbb{R}^n}$ and in particular $\mu * \delta_0 = \mu$.

As a consequence of item 2. of this exercise, if $\{Y_i\}_{i=1}^n$ are independent random vectors in $\mathbb{R}^n$ with $\mu_i = \mathrm{Law}(Y_i)$, then

$$\mathrm{Law}(Y_1 + \cdots + Y_n) = \mu_1 * \mu_2 * \cdots * \mu_n. \qquad (21.5)$$

Solution to Exercise (21.1). 1. The first item follows from Eq. (21.4).

2.For the second we have,

$$\begin{aligned}
[\mu * (\nu * \gamma)](A) &= \int_{\mathbb{R}^n} (\nu * \gamma)(A - x)\, d\mu(x) \\
&= \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} d\nu(y)\, \gamma(A - x - y) \\
&= \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} d\nu(y) \int_{\mathbb{R}^n} d\gamma(z)\, 1_{A-x-y}(z) \\
&= \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} d\nu(y) \int_{\mathbb{R}^n} d\gamma(z)\, 1_A(x+y+z) \\
&= \int_{[\mathbb{R}^n]^3} 1_A(x+y+z)\, d(\mu \otimes \nu \otimes \gamma)(x, y, z).
\end{aligned}$$

Similarly one shows $[(\mu * \nu) * \gamma](A)$ satisfies the same formula.

3. For the last item,

$$\begin{aligned}
(\mu * \delta_x)(A) &= \int_{\mathbb{R}^n} \delta_x(A - y)\, d\mu(y) = \int_{\mathbb{R}^n} 1_{x \in A - y} d\mu(y) \\
&= \int_{\mathbb{R}^n} 1_{x+y \in A} d\mu(y) = \int_{\mathbb{R}^n} 1_{y \in A - x} d\mu(y) = \mu(A - x).
\end{aligned}$$

Remark 21.2. Suppose that $d\mu(x) = u(x)\,dx$ where $u(x) \geq 0$ and $\int_{\mathbb{R}^n} u(x)\,dx = 1$. Then using the translation invariance of Lebesgue measure and Tonelli's theorem, we have

$$\mu * \nu(f) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x+y)\,u(x)\,dx d\nu(y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x)\,u(x-y)\,dx d\nu(y)$$

from which it follows that

$$d(\mu * \nu)(x) = \left[\int_{\mathbb{R}^n} u(x-y)\,d\nu(y)\right] dx.$$

If we further assume that $d\nu(x) = v(x)\,dx$, then we have

$$d(\mu * \nu)(x) = \left[\int_{\mathbb{R}^n} u(x-y)\,v(y)\,dy\right] dx.$$

To simplify notation we write,

$$u * v(x) = \int_{\mathbb{R}^n} u(x-y)\,v(y)\,dy = \int_{\mathbb{R}^n} v(x-y)\,u(y)\,dy.$$

Example 21.3. Suppose that $n = 1$, $d\mu(x) = 1_{[0,1]}(x)\,dx$ and $d\nu(x) = 1_{[-1,0]}(x)\,dx$ so that $\nu(A) = \mu(-A)$. In this case

$$d(\mu * \nu)(x) = \left(1_{[0,1]} * 1_{[-1,0]}\right)(x)\,dx$$

where

$$\begin{aligned}\left(1_{[0,1]} * 1_{[-1,0]}\right)(x) &= \int_{\mathbb{R}} 1_{[-1,0]}(x-y)\,1_{[0,1]}(y)\,dy \\ &= \int_{\mathbb{R}} 1_{[0,1]}(y-x)\,1_{[0,1]}(y)\,dy \\ &= \int_{\mathbb{R}} 1_{[0,1]+x}(y)\,1_{[0,1]}(y)\,dy \\ &= m([0,1] \cap (x+[0,1])) = (1-|x|)_+ .\end{aligned}$$

21.2 Total Variation Distance

Definition 21.4. *Let μ and ν be two probability measure on a measurable space, $(\Omega, \mathcal{B})$. The total variation distance, $d_{TV}(\mu,\nu)$, is defined as*

$$d_{TV}(\mu,\nu) := \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|, \tag{21.6}$$

i.e. $d_{TV}(\mu,\nu)$ is simply the supremum norm of $\mu - \nu$ as a function on $\mathcal{B}$.

Notation 21.5 *Suppose that X and Y are random variables, let*

$$d_{TV}(X,Y) := d_{TV}(\mu_X, \mu_Y) = \sup_{A \in \mathcal{B}_{\mathbb{R}}} |P(X \in A) - P(Y \in A)|,$$

where $\mu_X = P \circ X^{-1}$ and $\mu_Y = P \circ Y^{-1}$.

Example 21.6. For $x \in \mathbb{R}^n$, let $\delta_x(A) := 1_A(x)$ for all $A \in \mathcal{B}_{\mathbb{R}^n}$. Then one easily shows that $d_{TV}(\delta_x, \delta_y) = 1_{x \neq y}$. Thus if $x \neq y$, in this metric δ_x and δ_y are one unit apart no matter how close x and y are in $\mathbb{R}^n$. (This is not always a desirable feature and because of this will introduce shortly another notion of closeness for measures.)

Exercise 21.2. Let $\mathcal{P}_1$ denote the set of probability measures on $(\Omega, \mathcal{B})$. Show d_{TV} is a complete metric on $\mathcal{P}_1$.

Exercise 21.3. Suppose that μ, ν, and γ are probability measures on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. Show $d_{TV}(\mu * \nu, \mu * \gamma) \leq d_{TV}(\nu, \gamma)$. Use this fact along with Exercise 21.2 to show,

$$d_{TV}(\mu_1 * \mu_2 * \cdots * \mu_n, \nu_1 * \nu_2 * \cdots * \nu_n) \leq \sum_{i=1}^n d_{TV}(\mu_i, \nu_i)$$

for all choices probability measures, μ_i and ν_i on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

Remark 21.7. The function, $\lambda : \mathcal{B} \to \mathbb{R}$ defined by, $\lambda(A) := \mu(A) - \nu(A)$ for all $A \in \mathcal{B}$, is an example of a "signed measure." For signed measures, one usually defines

$$\|\lambda\|_{TV} := \sup\left\{\sum_{i=1}^n |\lambda(A_i)| : n \in \mathbb{N} \text{ and partitions, } \{A_i\}_{i=1}^n \subset \mathcal{B} \text{ of } \Omega\right\}.$$

You are asked to show in Exercise 21.4 below, that when $\lambda = \mu - \nu$, $d_{TV}(\mu,\nu) = \frac{1}{2}\|\mu - \nu\|_{TV}$.

Lemma 21.8 (Scheffé's Lemma). *Suppose that m is another positive measure on $(\Omega, \mathcal{B})$ such that there exists measurable functions, $f, g : \Omega \to [0,\infty)$, such that $d\mu = f dm$ and $d\nu = g dm$.*[1] *Then*

$$d_{TV}(\mu,\nu) = \frac{1}{2}\int_\Omega |f - g|\,dm.$$

Moreover, if $\{\mu_n\}_{n=1}^\infty$ is a sequence of probability measure of the form, $d\mu_n = f_n dm$ with $f_n : \Omega \to [0,\infty)$, and $f_n \to g$, m - a.e., then $d_{TV}(\mu_n, \nu) \to 0$ as $n \to \infty$.

[1] Fact: it is always possible to do this by taking $m = \mu + \nu$ for example.

Proof. Let $\lambda = \mu - \nu$ and $h := f - g : \Omega \to \mathbb{R}$ so that $d\lambda = hdm$. Since

$$\lambda(\Omega) = \mu(\Omega) - \nu(\Omega) = 1 - 1 = 0,$$

if $A \in \mathcal{B}$ we have

$$\lambda(A) + \lambda(A^c) = \lambda(\Omega) = 0.$$

In particular this shows $|\lambda(A)| = |\lambda(A^c)|$ and therefore,

$$\begin{aligned}|\lambda(A)| &= \frac{1}{2}\left[|\lambda(A)| + |\lambda(A^c)|\right] = \frac{1}{2}\left[\left|\int_A hdm\right| + \left|\int_{A^c} hdm\right|\right] \qquad (21.7)\\ &\leq \frac{1}{2}\left[\int_A |h|\,dm + \int_{A^c}|h|\,dm\right] = \frac{1}{2}\int_\Omega |h|\,dm.\end{aligned}$$

This shows

$$d_{TV}(\mu,\nu) = \sup_{A\in\mathcal{B}} |\lambda(A)| \leq \frac{1}{2}\int_\Omega |h|\,dm.$$

To prove the converse inequality, simply take $A = \{h > 0\}$ (note $A^c = \{h \leq 0\}$) in Eq. (21.7) to find

$$\begin{aligned}|\lambda(A)| &= \frac{1}{2}\left[\int_A hdm - \int_{A^c} hdm\right]\\ &= \frac{1}{2}\left[\int_A |h|\,dm + \int_{A^c}|h|\,dm\right] = \frac{1}{2}\int_\Omega |h|\,dm.\end{aligned}$$

For the second assertion, observe that $|f_n - g| \to 0$ m – a.e., $|f_n - g| \leq G_n := f_n + g \in L^1(m)$, $G_n \to G := 2g$ a.e. and $\int_\Omega G_n dm = 2 \to 2 = \int_\Omega Gdm$ and $n \to \infty$. Therefore, by the dominated convergence theorem 7.27,

$$\lim_{n\to\infty} d_{TV}(\mu_n,\nu) = \frac{1}{2}\lim_{n\to\infty}\int_\Omega |f_n - g|\,dm = 0.$$

■

For a concrete application of Scheffé's Lemma 21.8, see Proposition 21.49 below.

Corollary 21.9. *Let $\|h\|_\infty := \sup_{\omega\in\Omega}|h(\omega)|$ when $h : \Omega \to \mathbb{R}$ is a bounded random variable. Continuing the notation in Scheffé's lemma above, we have*

$$d_{TV}(\mu,\nu) = \frac{1}{2}\sup\left\{\left|\int_\Omega hd\mu - \int_\Omega hd\nu\right| : \|h\|_\infty \leq 1\right\}. \qquad (21.8)$$

Consequently,

$$\left|\int_\Omega hd\mu - \int_\Omega hd\nu\right| \leq 2d_{TV}(\mu,\nu)\cdot\|h\|_\infty \qquad (21.9)$$

and in particular, for all bounded and measurable functions, $h : \Omega \to \mathbb{R}$,

$$\lim_{n\to\infty} d_{TV}(\mu_n,\nu) = 0 \implies \lim_{n\to\infty}\int_\Omega hd\mu_n = \int_\Omega hd\nu. \qquad (21.10)$$

Proof. *We begin by observing that*

$$\begin{aligned}\left|\int_\Omega hd\mu - \int_\Omega hd\nu\right| &= \left|\int_\Omega h(f-g)\,dm\right| \leq \int_\Omega |h|\,|f-g|\,dm\\ &\leq \|h\|_\infty \int_\Omega |f-g|\,dm = 2d_{TV}(\mu,\nu)\,\|h\|_\infty.\end{aligned}$$

Moreover, from the proof of Scheffé's Lemma 21.8, we have

$$d_{TV}(\mu,\nu) = \frac{1}{2}\left|\int_\Omega hd\mu - \int_\Omega hd\nu\right|$$

when $h := 1_{f>g} - 1_{f\leq g}$. These two equations prove Eqs. (21.8) and (21.9) and the latter implies Eq. (21.10). ■

Exercise 21.4. Under the hypothesis of Scheffé's Lemma 21.8, show

$$\|\mu - \nu\|_{TV} = \int_\Omega |f - g|\,dm = 2d_{TV}(\mu,\nu).$$

Exercise 21.5. Suppose that Ω is a (at most) countable set, $\mathcal{B} := 2^\Omega$, and $\{\mu_n\}_{n=0}^\infty$ are probability measures on $(\Omega,\mathcal{B})$. Show

$$d_{TV}(\mu_n,\mu_0) = \frac{1}{2}\sum_{\omega\in\Omega}|\mu_n(\{\omega\}) - \mu_0(\{\omega\})|$$

and $\lim_{n\to\infty} d_{TV}(\mu_n,\mu_0) = 0$ iff $\lim_{n\to\infty}\mu_n(\{\omega\}) = \mu_0(\{\omega\})$ for all $\omega \in \Omega$.

Exercise 21.6. Let $\mu_p(\{1\}) = p$ and $\mu_p(\{0\}) = 1 - p$ and $\nu_\lambda(\{n\}) := e^{-\lambda}\frac{\lambda^n}{n!}$ for all $n \in \mathbb{N}_0$.

1. Find $d_{TV}(\mu_p,\mu_q)$ for all $0 \leq p, q \leq 1$.
2. Show $d_{TV}(\mu_p,\nu_p) = p(1 - e^{-p})$ for all $0 \leq p \leq 1$. From this estimate and the estimate,
$$1 - e^{-p} = \int_0^p e^{-x}dx \leq \int_0^p 1dx = p, \qquad (21.11)$$
it follows that $d_{TV}(\mu_p,\nu_p) \leq p^2$ for all $0 \leq p \leq 1$.
3. Show
$$d_{TV}(\nu_\lambda,\nu_\gamma) \leq |\lambda - \gamma|\,e^{|\lambda-\gamma|} \text{ for all } \lambda,\gamma \in \mathbb{R}_+.$$
I got this estimate with the aid of the fundamental theorem of calculus along with crude estimates on each term in the infinite series for $d_{TV}(\nu_\lambda,\nu_\gamma)$. Andy Parrish got a much better estimate, namely that
$$d_{TV}(\nu_\lambda,\nu_\gamma) \leq |\lambda - \gamma| \text{ for all } \lambda,\gamma \in \mathbb{R}_+! \qquad (21.12)$$

The next theorem should be compared with Exercise 7.6 which may be stated as follows. If $\{Z_i\}_{i=1}^n$ are i.i.d. Bernoulli random variables with $P(Z_i=1) = p = O(1/n)$ and $S = Z_1 + \dots + Z_n$, then $P(S=k) \cong P(\mathrm{Poisson}(pn) = k)$ which is valid for $k \ll n$.

Theorem 21.10 (Law of rare events). *Let $\{Z_i\}_{i=1}^n$ be independent Bernoulli random variables with $P(Z_i=1) = p_i \in (0,1)$ and $P(Z_i=0) = 1-p_i$, $S := Z_1 + \dots + Z_n$, $a := p_1 + \dots + p_n$, and $X \overset{d}{=} \mathrm{Poi}(a)$. Then for any[2] $A \in \mathcal{B}_{\mathbb{R}}$ we have*

$$|P(S \in A) - P(X \in A)| \le \sum_{i=1}^n p_i^2, \tag{21.13}$$

or in short,

$$d_{TV}\left(\sum_{i=1}^n Z_i, X\right) \le \sum_{i=1}^n p_i^2.$$

(Of course this estimate has no content unless $\sum_{i=1}^n p_i^2 < 1$.)

Proof. Let $\{X_i\}_{i=1}^n$ be independent random variables with $X_i \overset{d}{=} Pois(p_i)$ for each i. It then follows from Exercises 21.3 and 21.6 that,

$$d_{TV}\left(\sum_{i=1}^n Z_i, \sum_{i=1}^n X_i\right) \le \sum_{i=1}^n d_{TV}(Z_i, X_i) = \sum_{i=1}^n p_i\left(1-e^{-p_i}\right) \le \sum_{i=1}^n p_i^2.$$

■

The reader should compare the proof of this theorem with the proof of the central limit theorem in Theorem 10.37. For another less quantitative Poisson limit theorem, see Theorem 23.16.

For the next result we will suppose that $(Y, \mathcal{M}, \mu)$ is a finite measure space with the following properties;

1. $\{y\} \in \mathcal{M}$ and $\mu(\{y\}) = 0$ for all $y \in Y$,
2. to any $A \in \mathcal{M}$ and $\varepsilon > 0$, there exists a finite partition $\{A_n\}_{n=1}^{N=N(\varepsilon)} \subset \mathcal{M}$ of A such that $\mu(A_n) \le \varepsilon$ for all n.

In what follows below we will write $F(A) = o(\mu(A))$ provided there exits an increasing function, $\delta : \mathbb{R}_+ \to \mathbb{R}_+$, such that $\delta(x) \to 0$ as $x \to 0$ and $|F(A)| \le \mu(A)\,\delta(\mu(A))$ for all $A \in \mathcal{M}$.

Proposition 21.11 (Why Poisson). *Suppose $(Y, \mathcal{M}, \mu)$ is finite measure space with the properties given above and $\{N(A) : A \in \mathcal{M}\}$ is a collection of $\mathbb{N}_0$ – valued random variables with the following properties;*

1. *If $\{A_j\}_{j=1}^n \subset \mathcal{M}$ are disjoint, then $\{N(A_i)\}_{i=1}^n$ are independent random variables and*
$$N\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n N(A_i) \text{ a.s.}$$
2. *$P(N(A) \ge 2) = o(\mu(A))$.*
3. *$|P(N(A) \ge 1) - \mu(A)| = o(\mu(A))$.*

Then $N(A) \overset{d}{=} \mathrm{Poi}(\mu(A))$ for all $A \in \mathcal{M}$ and in particular $\mathbb{E}N(A) = \mu(A)$ for all $A \in \mathcal{M}$.

Proof. Let $A \in \mathcal{M}$ and $\varepsilon > 0$ be given. Choose a partition $\{A_i^\varepsilon\}_{i=1}^N \subset \mathcal{M}$ of A such that $\mu(A_i^\varepsilon) \le \varepsilon$ for all i. Let $Z_i := 1_{N(A_i^\varepsilon) \ge 1}$ and $S := \sum_{i=1}^N Z_i$. Using

$$N(A) = \sum_{i=1}^N N(A_i^\varepsilon)$$

and Lemma 21.12, we have

$$|P(N(A) = k) - P(S=k)| \le P(N(A) \ne S) \le \sum_{i=1}^N P(Z_i \ne N(A_i^\varepsilon)).$$

Since $\{Z_i \ne N(A_i^\varepsilon)\} = \{N(A_i^\varepsilon) \ge 2\}$ and $P(N(A_i^\varepsilon) \ge 2) = o(\mu(A_i^\varepsilon))$, it follows that

$$\begin{aligned} |P(N(A) = k) - P(S=k)| &\le \sum_{i=1}^N \mu(A_i^\varepsilon)\,\delta(\mu(A_i^\varepsilon)) \\ &\le \sum_{i=1}^N \mu(A_i^\varepsilon)\,\delta(\varepsilon) = \delta(\varepsilon)\,\mu(A). \end{aligned} \tag{21.14}$$

On the other hand, $\{Z_i\}_{i=1}^N$ are independent Bernoulli random variables with

$$P(Z_i = 1) = P(N(A_i^\varepsilon) \ge 1),$$

and $a_\varepsilon = \sum_{i=1}^N P(N(A_i^\varepsilon) \ge 1)$. Then by the Law of rare events Theorem 21.10,

$$\begin{aligned} \left|P(S=k) - \frac{a_\varepsilon^k}{k!}e^{-a_\varepsilon}\right| &\le \sum_{i=1}^N \left[P(N(A_i^\varepsilon) \ge 1)\right]^2 \le \sum_{i=1}^N \left[\mu(A_i^\varepsilon) + o(\mu(A_i^\varepsilon))\right]^2 \\ &\le \sum_{i=1}^N \mu(A_i^\varepsilon)^2 (1+\delta'(\varepsilon))^2 = (1+\delta'(\varepsilon))^2\,\varepsilon\mu(A). \end{aligned} \tag{21.15}$$

[2] Actually, since S and X are $\mathbb{N}_0$ – valued, we may as well assume that $A \subset \mathbb{N}_0$.

Combining Eqs. (21.14) and (21.15) shows

$$\left|P(N(A)=k)-\frac{a_\varepsilon^k}{k!}e^{-a_\varepsilon}\right|\le\left[\delta(\varepsilon)+(1+\delta'(\varepsilon))^2\varepsilon\right]\mu(A) \qquad (21.16)$$

where a_ε satisfies

$$\begin{aligned}|a_\varepsilon-\mu(A)|&=\left|\sum_{i=1}^N\left[P(N(A_i^\varepsilon)\ge1)-\mu(A_i^\varepsilon)\right]\right|\\&\le\sum_{i=1}^N\left|\left[P(N(A_i^\varepsilon)\ge1)-\mu(A_i^\varepsilon)\right]\right|\le\sum_{i=1}^N o(\mu(A_i^\varepsilon))\\&\le\sum_{i=1}^N\mu(A_i^\varepsilon)\left|\delta'(\mu(A_i^\varepsilon))\right|\le\mu(A)\,\delta'(\varepsilon).\end{aligned}$$

Hence we may let $\varepsilon\downarrow0$ in Eq. (21.16) to find

$$P(N(A)=k)=\frac{(\mu(A))^k}{k!}e^{-\mu(A)}.$$

■

See [42, p. 13-16.] for another variant of this theorem in the case that $\Omega=\mathbb{R}_+$. See Theorem 11.11 and Exercises 11.6 – 11.8 for concrete constructions of Poisson processes.

21.3 A Coupling Estimate

Lemma 21.12 (Coupling Estimates). *Suppose X and Y are any random variables on a probability space, $(\Omega,\mathcal{B},P)$ and $A\in\mathcal{B}_\mathbb{R}$. Then*

$$|P(X\in A)-P(Y\in A)|\le P(\{X\in A\}\triangle\{Y\in A\})\le P(X\ne Y). \quad (21.17)$$

Proof. The proof is simply;

$$\begin{aligned}|P(X\in A)-P(Y\in A)|&=|\mathbb{E}[1_A(X)-1_A(Y)]|\\&\le\mathbb{E}|1_A(X)-1_A(Y)|=\mathbb{E}1_{\{X\in A\}\triangle\{Y\in A\}}\\&\le\mathbb{E}1_{X\ne Y}=P(X\ne Y).\end{aligned}$$

■

Pushing the above proof a little more we have, if $\{A_i\}$ is a partition of Ω, then

$$\begin{aligned}\sum_i|P(X\in A_i)-P(Y\in A_i)|&=\sum_i|\mathbb{E}[1_{A_i}(X)-1_{A_i}(Y)]|\\&\le\sum_i\mathbb{E}|1_{A_i}(X)-1_{A_i}(Y)|\\&\le\mathbb{E}[1_{X\ne Y}:X\in A_i\text{ or }Y\in A_i]\\&\le\mathbb{E}[1_{X\ne Y}:X\in A_i]+\mathbb{E}[1_{X\ne Y}:Y\in A_i]\\&=2P(X\ne Y).\end{aligned}$$

This shows

$$\|X_*P-Y_*P\|_{TV}\le2P(X\ne Y).$$

This is really not more general than Eq. (21.17) since the Hahn decomposition theorem we know that in fact the signed measure, $\mu:=X_*P-Y_*P$, has total variation given by

$$\|\mu\|_{TV}=\mu(\Omega_+)-\mu(\Omega_-)$$

where $\Omega=\Omega_+\sum\Omega_-$ with Ω_+ being a positive set and Ω_- being a negative set. Moreover, since $\mu(\Omega)=0$ we must in fact have $\mu(\Omega_+)=-\mu(\Omega_-)$ so that

$$\begin{aligned}\|X_*P-Y_*P\|_{TV}&=\|\mu\|_{TV}=2\mu(\Omega_+)=2|P(X\in\Omega_+)-P(Y\in\Omega_+)|\\&\le2P(X\ne Y).\end{aligned}$$

Here is perhaps a better way to view the above lemma. Suppose that we are given two probability measures, μ,ν, on $(\mathbb{R},\mathcal{B}_\mathbb{R})$ (or any other measurable space, $(S,\mathcal{B}_S)$). We would like to estimate $\|\mu-\nu\|_{TV}$. The lemma states that if X,Y are random variables (vectors) on some probability space such that $\mathrm{Law}_P(X)=\mu$ and $\mathrm{Law}_P(Y)=\nu$, then

$$\|\mu-\nu\|_{TV}\le2P(X\ne Y).$$

Suppose that we let $\rho:=\mathrm{Law}_P(X,Y)$ on $(S\times S,\mathcal{B}_S\otimes\mathcal{B}_S)$, $\pi_i:S\times S\to S$ be the projection maps for $i=1,2$, then $(\pi_1)_*\rho=\mu$, $(\pi_2)_*\rho=v$, and

$$\|\mu-\nu\|_{TV}\le2P(X\ne Y)=2\rho(\pi_1\ne\pi_2)=2\rho\left(S^2\setminus\Delta\right)$$

where $\Delta=\{(s,s):s\in S\}$ is the diagonal in S^2. Thus finding a coupling amounts to fining a probability measure, ρ, on $\left(S^2,\mathcal{B}_S\otimes\mathcal{B}_S\right)$ whose marginals are μ and ν respectively. Then we will have the coupling estimate,

$$\|\mu-\nu\|_{TV}\le2\rho\left(S^2\setminus\Delta\right).$$

As an example of how to use Lemma 21.12 let us give a coupling proof of Theorem 21.10.

Proof. (A coupling proof of Theorem 21.10.) We are going to construct a coupling for S_*P and X_*P. Finding such a coupling amounts to representing

X and S on the same probability space. We are going to do this by building all random variables in site out of $\{U_i\}_{i=1}^n$, where the $\{U_i\}_{i=1}^n$ are i.i.d. random variables distributed uniformly on $[0,1]$.

If we define,

$$Z_i := 1_{(1-p_i,1]}(U_i) = 1_{1-p_i<U_i\le 1},$$

then $\{Z_i\}_{i=1}^n$ are independent Bernoulli random variables with $P(Z_i=1)=p_i$. We are now also going to construct out of the $\{U_i\}_{i=1}^n$, a sequence of independent Poisson random variables, $\{X_i\}_{i=1}^n$ with $X_i = \mathrm{Poi}(p_i)$. To do this define

$$\alpha_i(k) := P(\mathrm{Poi}(p_i)\le k) = e^{-p_i}\sum_{j=0}^k \frac{p_i^j}{j!}$$

with the convention that $\alpha_i(-1)=0$. Notice that

$$e^{-p_i} \le \alpha_i(k) \le \alpha_i(k+1) \le 1 \text{ for all } k\in\mathbb{N}_0$$

and for p_i small, we have

$$\alpha_i(0) = e^{-p_i} \cong 1-p_i$$
$$\alpha_i(1) = e^{-p_i}(1+p_i) \cong 1-p_i^2.$$

If we define (see Figure 21.1) by

$$X_i := \sum_{k=0}^\infty k 1_{\alpha_i(k-1)<U_i\le\alpha_i(k)},$$

then $X_i = \mathrm{Poi}(p_i)$ since

$$\begin{aligned} P(X_i=k) &= P(\alpha_i(k-1)<U_i\le\alpha_i(k)) \\ &= \alpha_i(k)-\alpha_i(k-1) = e^{-p_i}\frac{p_i^k}{k!}. \end{aligned}$$

It is also clear that $\{X_i\}_{i=1}^n$ are independent and hence by Lemma 11.1, it follows that $X := \sum_{i=1}^n X_i \overset{d}{=} \mathrm{Poi}(a)$.

An application of Lemma 21.12 now shows

$$|P(S\in A) - P(X\in A)| \le P(S\ne X)$$

and since $\{S\ne X\}\subset \cup_{i=1}^n\{X_i\ne Z_i\}$, we may conclude

$$|P(S\in A)-P(X\in A)| \le \sum_{i=1}^n P(X_i\ne Z_i).$$

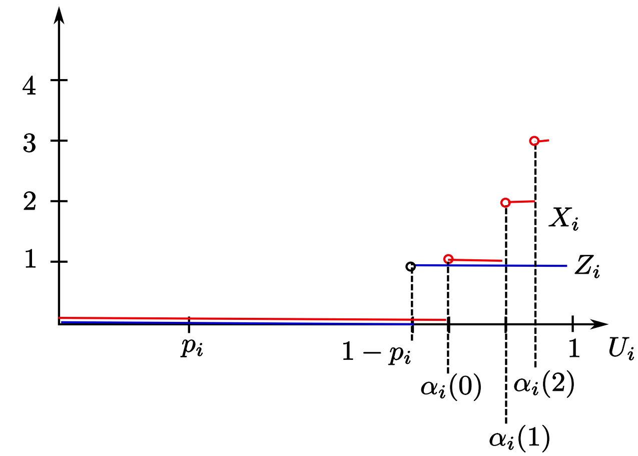


Fig. 21.1. Plots of X_i and Z_i as functions of U_i.

As is easily seen from Figure 21.1,

$$\begin{aligned} P(X_i\ne Z_i) &= [\alpha_i(0)-(1-p_i)] + 1-\alpha_i(1) \\ &= \left[e^{-p_i}-(1-p_i)\right] + 1 - e^{-p_i}(1+p_i) \\ &= p_i\left(1-e^{-p_i}\right) \le p_i^2 \end{aligned}$$

where we have used the estimate in Eq. (21.11) for the last inequality. ■

21.4 Weak Convergence

Recall that to each right continuous increasing function, $F:\mathbb{R}\to\mathbb{R}$ there is a unique measure, μ_F, on $\mathcal{B}_\mathbb{R}$ such that $\mu_F((a,b]) = F(b)-F(a)$ for all $-\infty<a\le b<\infty$. To simplify notation in this section we will now write $F(A)$ for $\mu_F(A)$ for all $A\in\mathcal{B}_\mathbb{R}$ and in particular $F((a,b]) := F(b)-F(a)$ for all $-\infty<a\le b<\infty$.

Example 21.13. Suppose that $P\left(X_n=\frac{i}{n}\right) = \frac{1}{n}$ for $i\in\{1,2,\dots,n\}$ so that X_n is a discrete "approximation" to the uniform distribution, i.e. to U where $P(U\in A) = m(A\cap[0,1])$ for all $A\in\mathcal{B}_\mathbb{R}$. If we let $A_n = \left\{\frac{i}{n} : i=1,2,\dots,n\right\}$, then $P(X_n\in A_n) = 1$ while $P(U\in A_n)=0$. Therefore, it follows that $d_{TV}(X_n,U)=1$ for all n.[3]

[3] More generally, if μ and ν are two probability measure on $(\mathbb{R},\mathcal{B}_\mathbb{R})$ such that $\mu(\{x\})=0$ for all $x\in\mathbb{R}$ while ν concentrates on a countable set, then $d_{TF}(\mu,\nu)=1$.

Nevertheless we would like X_n to be close to U in distribution. Let us observe that if we let $F_n(y) := P(X_n \le y)$ and $F(y) := P(U \le y)$, then

$$F_n(y) = P(X_n \le y) = \frac{1}{n}\#\left\{i \in \{1,2,\dots,n\} : \frac{i}{n} \le y\right\}$$

and

$$F(y) := P(U \le y) = (y \wedge 1) \vee 0.$$

From these formula, it easily follows that $F(y) = \lim_{n\to\infty} F_n(y)$ for all $y \in \mathbb{R}$, see Figure 21.2. This suggest that we should say that X_n converges in distribu-

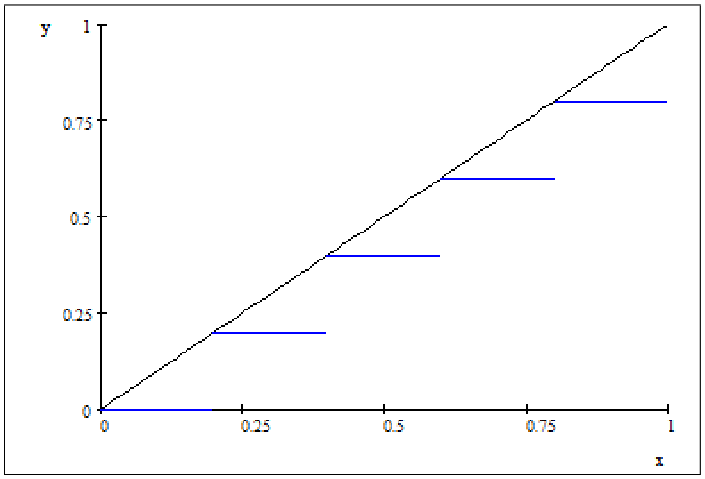


Fig. 21.2. The plot of F_5 in blue and F is black on $[0,1]$.

tion to X iff $P(X_n \le y) \to P(X \le y)$ for all $y \in \mathbb{R}$. However, the next simple example shows this definition is also too restrictive.

Example 21.14. Suppose that $P(X_n = 1/n) = 1$ for all n and $P(X_0 = 0) = 1$. Then it is reasonable to insist that X_n converges of X_0 in distribution. However, $F_n(y) = 1_{y \ge 1/n} \to 1_{y\ge 0} = F_0(y)$ for all $y \in \mathbb{R}$ **except** for $y = 0$. Observe that y is the only point of discontinuity of F_0.

Notation 21.15 *Let (X,d) be a metric space, $f : X \to \mathbb{R}$ be a function. The set of $x \in X$ where f is continuous (discontinuous) at x will be denoted by $\mathcal{C}(f)$ ($\mathcal{D}(f)$).*

Remark 21.16. If $F : \mathbb{R} \to [0,1]$ is a non-decreasing function, then $\mathcal{C}(F)$ is at most countable. To see this, suppose that $\varepsilon > 0$ is given and let $\mathcal{C}_\varepsilon := \{y \in \mathbb{R} : F(y+) - F(y-) \ge \varepsilon\}$. If $y < y'$ with $y, y' \in \mathcal{C}_\varepsilon$, then $F(y+) < F(y'-)$ and $(F(y-), F(y+))$ and $(F(y'-), F(y'+))$ are disjoint intervals of length greater that ε. Hence it follows that

$$1 = m([0,1]) \ge \sum_{y \in \mathcal{C}_\varepsilon} m((F(y-), F(y+))) \ge \varepsilon \cdot \#(\mathcal{C}_\varepsilon)$$

and hence that $\#(\mathcal{C}_\varepsilon) \le \varepsilon^{-1} < \infty$. Therefore $\mathcal{C} := \cup_{k=1}^\infty \mathcal{C}_{1/k}$ is at most countable.

Definition 21.17. *Let $\{F, F_n : n = 1,2,\dots\}$ be a collection of right continuous non-increasing functions from $\mathbb{R}$ to $[0,1]$. Then*

1. *F_n converges to F **vaguely** and write, $F_n \xrightarrow{v} F$, iff $F_n((a,b]) \to F((a,b])$ for all $a, b \in \mathcal{C}(F)$.*
2. *F_n converges to F **weakly** and write, $F_n \xrightarrow{w} F$, iff $F_n(x) \to F(x)$ for all $x \in \mathcal{C}(F)$.*
3. *We say F is **proper**, if F is a distribution function of a probability measure, i.e. if $F(\infty) = 1$ and $F(-\infty) = 0$.*

Example 21.18. If X_n and U are as in Example 21.13 and $F_n(y) := P(X_n \le y)$ and $F(y) := P(Y \le y)$, then $F_n \xrightarrow{v} F$ and $F_n \xrightarrow{w} F$.

Lemma 21.19. *Let $\{F, F_n : n = 1,2,\dots\}$ be a collection of proper distribution functions. Then $F_n \xrightarrow{v} F$ iff $F_n \xrightarrow{w} F$. In the case where F_n and F are proper and $F_n \xrightarrow{w} F$, we will write $F_n \implies F$.*

Proof. If $F_n \xrightarrow{w} F$, then $F_n((a,b]) = F_n(b) - F_n(a) \to F(b) - F(a) = F((a,b])$ for all $a, b \in \mathcal{C}(F)$ and therefore $F_n \xrightarrow{v} F$. So now suppose $F_n \xrightarrow{v} F$ and let $a < x$ with $a, x \in \mathcal{C}(F)$. Then

$$F(x) = F(a) + \lim_{n\to\infty}[F_n(x) - F_n(a)] \le F(a) + \liminf_{n\to\infty} F_n(x).$$

Letting $a \downarrow -\infty$, using the fact that F is proper, implies

$$F(x) \le \liminf_{n\to\infty} F_n(x).$$

Likewise,

$$F(x) - F(a) = \lim_{n\to\infty}[F_n(x) - F_n(a)] \ge \limsup_{n\to\infty}[F_n(x) - 1] = \limsup_{n\to\infty} F_n(x) - 1$$

which upon letting $a \uparrow \infty$, (so $F(a) \uparrow 1$) allows us to conclude,

$$F(x) \ge \limsup_{n\to\infty} F_n(x).$$

■

Definition 21.20. *A sequence of random variables, $\{X_n\}_{n=1}^\infty$ is said to* ***converge weakly*** *or to* ***converge in distribution*** *to a random variable X (written $X_n \implies X$) iff $F_n(y) := P(X_n \le y) \implies F(y) := P(X \le y)$.*

Example 21.21 (Central Limit Theorem). The central limit theorem (see Theorems 7.62, Corollary 10.39, and Theorem 22.26) states; if $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^2(P)$ random variables with $\mu := \mathbb{E}X_1$ and $\sigma^2 = \mathrm{Var}(X_1)$, then

$$\frac{S_n - n\mu}{\sqrt{n}} \implies N(0,\sigma) \stackrel{d}{=} \sigma N(0,1).$$

Written out explicitly we find

$$\begin{aligned}\lim_{n\to\infty} P\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) &= P(a < N(0,1) \le b) \\ &= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx\end{aligned}$$

or equivalently put

$$\lim_{n\to\infty} P\left(n\mu + \sigma\sqrt{n}a < S_n \le n\mu + \sigma\sqrt{n}b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx.$$

More intuitively, we have

$$S_n \stackrel{d}{\cong} n\mu + \sqrt{n}\sigma N(0,1) \stackrel{d}{=} N\left(n\mu, n\sigma^2\right).$$

Example 21.22. Suppose that $P(X_n = n) = 1$ for all n, then $F_n(y) = 1_{y \ge n} \to 0 = F(y)$ as $n \to \infty$. Notice that F is not a distribution function because all of the mass went off to $+\infty$. Similarly, if we suppose, $P(X_n = \pm n) = \frac{1}{2}$ for all n, then $F_n = \frac{1}{2}1_{[-n,n)} + 1_{[n,\infty)} \to \frac{1}{2} = F(y)$ as $n \to \infty$. Again, F is not a distribution function on $\mathbb{R}$ since half the mass went to $-\infty$ while the other half went to $+\infty$.

Example 21.23. Suppose X is a non-zero random variables such that $X \stackrel{d}{=} -X$, then $X_n := (-1)^n X \stackrel{d}{=} X$ for all n and therefore, $X_n \implies X$ as $n \to \infty$. On the other hand, X_n does not converge to X almost surely or in probability.

Lemma 21.24. *Suppose X is a random variable, $\{c_n\}_{n=1}^\infty \subset \mathbb{R}$, and $X_n = X + c_n$. If $c := \lim_{n\to\infty} c_n$ exists, then $X_n \implies X + c$.*

Proof. Let $F(x) := P(X \le x)$ and

$$F_n(x) := P(X_n \le x) = P(X + c_n \le x) = F(x - c_n).$$

Clearly, if $c_n \to c$ as $n \to \infty$, then for all $x \in \mathcal{C}(F(\cdot - c))$ we have $F_n(x) \to F(x-c)$. Since $F(x-c) = P(X + c \le x)$, we see that $X_n \implies X + c$. Observe that $F_n(x) \to F(x-c)$ only for $x \in \mathcal{C}(F(\cdot - c))$ but this is sufficient to assert $X_n \implies X + c$. ∎

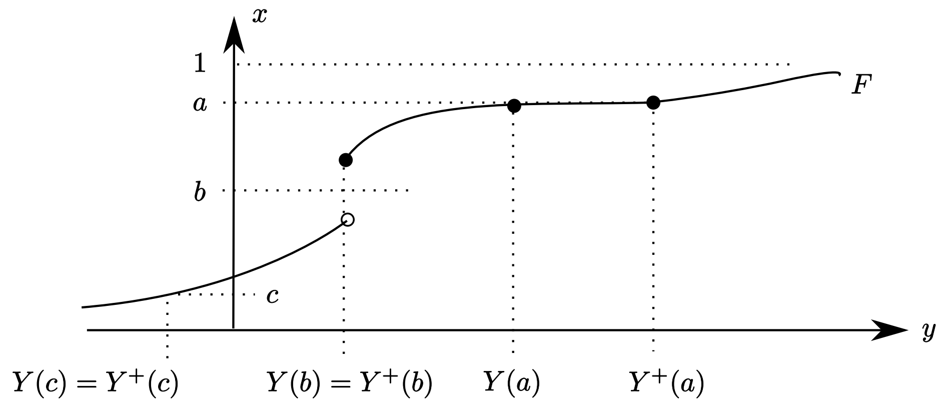


Fig. 21.3. The functions Y and Y^+ associated to F.

Lemma 21.25. *Suppose $\{X_n\}_{n=1}^\infty$ is a sequence of random variables on a common probability space and $c \in \mathbb{R}$. Then $X_n \implies c$ iff $X_n \xrightarrow{P} c$.*

Proof. Recall that $X_n \xrightarrow{P} c$ iff for all $\varepsilon > 0$, $P(|X_n - c| > \varepsilon) \to 0$. Since

$$\{|X_n - c| > \varepsilon\} = \{X_n > c + \varepsilon\} \cup \{X_n < c - \varepsilon\}$$

it follows $X_n \xrightarrow{P} c$ iff $P(X_n > x) \to 0$ for all $x > c$ and $P(X_n < x) \to 0$ for all $x < c$. These conditions are also equivalent to $P(X_n \le x) \to 1$ for all $x > c$ and $P(X_n \le x) \le P(X_n < x') \to 0$ for all $x < c$ (where $x < x' < c$). So $X_n \xrightarrow{P} c$ iff

$$\lim_{n\to\infty} P(X_n \le x) = \begin{cases} 0 \text{ if } x < c \\ 1 \text{ if } x > c \end{cases} = F(x)$$

where $F(x) = P(c \le x) = 1_{x \ge c}$. Since $\mathcal{C}(F) = \mathbb{R}\setminus\{c\}$, we have shown $X_n \xrightarrow{P} c$ iff $X_n \implies c$. ∎

Notation 21.26 *Given a proper distribution function, $F : \mathbb{R} \to [0,1]$, let $Y = F^{\leftarrow} : (0,1) \to \mathbb{R}$ be the function defined by*

$$Y(x) = F^{\leftarrow}(x) = \sup\{y \in \mathbb{R} : F(y) < x\}.$$

Similarly, let

$$Y^+(x) := \inf\{y \in \mathbb{R} : F(y) > x\}.$$

We will need the following simple observations about Y and Y^+ which are easily understood from Figure 21.3.

1. $Y(x) \le Y^+(x)$ and $Y(x) < Y^+(x)$ iff x is the height of a "flat spot" of F.
2. The set,
$$E := \left\{x \in (0,1) : Y(x) < Y^+(x)\right\}, \tag{21.18}$$
of flat spot heights is at most countable. This is because, $\left\{(Y(x), Y^+(x))\right\}_{x\in E}$ is a collection of pairwise disjoint intervals which is necessarily countable. (Each such interval contains a rational number.)
3. The following inequality holds,
$$F(Y(x)-) \le x \le F(Y(x)) \text{ for all } x \in (0,1). \tag{21.19}$$
Indeed, if $y > Y(x)$, then $F(y) \ge x$ and by right continuity of F it follows that $F(Y(x)) \ge x$. Similarly, if $y < Y(x)$, then $F(y) < x$ and hence $F(Y(x)-) \le x$.
4. $\{x \in (0,1) : Y(x) \le y_0\} = (0, F(y_0)] \cap (0,1)$. To prove this assertion first suppose that $Y(x) \le y_0$, then according to Eq. (21.19) we have $x \le F(Y(x)) \le F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely, if $x \in (0,1)$ and $x \le F(y_0)$, then $Y(x) \le y_0$ by definition of Y.
5. As a consequence of item 4. we see that Y is $\mathcal{B}_{(0,1)}/\mathcal{B}_{\mathbb{R}}$ – measurable and $m \circ Y^{-1} = F$, where m is Lebesgue measure on $\left((0,1), \mathcal{B}_{(0,1)}\right)$.

Theorem 21.27 (Baby Skorohod Theorem). *Suppose that $\{F_n\}_{n=0}^\infty$ is a collection of distribution functions such that $F_n \implies F_0$. Then there exists a probability space, $(\Omega, \mathcal{B}, P)$ and random variables, $\{Y_n\}_{n=1}^\infty$ such that $P(Y_n \le y) = F_n(y)$ for all $n \in \mathbb{N} \cup \{\infty\}$ and $\lim_{n\to\infty} Y_n = Y$ a.s..*

Proof. We will take $\Omega := (0,1)$, $\mathcal{B} = \mathcal{B}_{(0,1)}$, and $P = m$ – Lebesgue measure on Ω and let $Y_n := F_n^{\leftarrow}$ and $Y := F_0^{\leftarrow}$ as in Notation 21.26. Because of the above comments, $P(Y_n \le y) = F_n(y)$ and $P(Y \le y) = F_0(y)$ for all $y \in \mathbb{R}$. So in order to finish the proof it suffices to show, $Y_n(x) \to Y(x)$ for all $x \notin E$, where E is the countable null set defined as in Eq. (21.18).

We now suppose $x \notin E$. If $y \in \mathcal{C}(F_0)$ with $y < Y(x)$, we have $\lim_{n\to\infty} F_n(y) = F_0(y) < x$ and in particular, $F_n(y) < x$ for almost all n. This implies that $Y_n(x) \ge y$ for a.a. n and hence that $\liminf_{n\to\infty} Y_n(x) \ge y$. Letting $y \uparrow Y(x)$ with $y \in \mathcal{C}(F_0)$ then implies
$$\liminf_{n\to\infty} Y_n(x) \ge Y(x).$$

Similarly, for $x \notin E$ and $y \in \mathcal{C}(F_0)$ with $Y(x) = Y^+(x) < y$, we have $\lim_{n\to\infty} F_n(y) = F_0(y) > x$ and in particular, $F_n(y) > x$ for almost all n. This implies that $Y_n(x) \le y$ for a.a. n and hence that $\limsup_{n\to\infty} Y_n(x) \le y$. Letting $y \downarrow Y(x)$ with $y \in \mathcal{C}(F_0)$ then implies
$$\limsup_{n\to\infty} Y_n(x) \le Y(x).$$

Hence we have shown, for $x \notin E$, that
$$\limsup_{n\to\infty} Y_n(x) \le Y(x) \le \liminf_{n\to\infty} Y_n(x)$$
which shows
$$\lim_{n\to\infty} F_n^{\leftarrow}(x) = \lim_{n\to\infty} Y_n(x) = Y(x) = F^{\leftarrow}(x) \text{ for all } x \notin E. \tag{21.20}$$
■

In preparation for the full version of Skorohod's Theorem 21.58 it will be useful to record a special case of Theorem 21.27 which has both a stronger hypothesis and a stronger conclusion.

Theorem 21.28 (Prenatal Skorohod Theorem). *Suppose $S = \{1, 2, \dots, m\} \subset \mathbb{R}$ and $\{\mu_n\}_{n=1}^\infty$ is a sequence of probabilities on S such that $\mu_n \implies \mu$ for some probability μ on S. Let $P := \mu \otimes m$ on $\Omega := S \times (0,1]$, $Y(i,\theta) = i$ for all $(i,\theta) \in \Omega$. Then there exists $Y_n : \Omega \to S$ such that $\mathrm{Law}_P(Y_n) = \mu_n$ for all n and $Y_n(i,\theta) = i$ if $\theta \le \mu_n(i)/\mu(i)$ where we take $0/0 = 1$ in this expression. In particular, $\lim_{n\to\infty} Y_n(i,\theta) = Y(i,\theta)$ a.s.*

Proof. The main point is to show for any probability measure, ν, on S there exists $Y_\nu : \Omega \to S$ such that $Y_\nu(i,\theta) = i$ when $\theta \le \nu(i)/\mu(i)$ and $\mathrm{Law}_P(Y_\nu) = \nu$. If we can do this then we need only take $Y_n = Y_{\mu_n}$ for all n to complete the proof.

In the proof to follow we will use the simple observation that for any $a \in (0,1)$ and $\alpha_i \ge 0$ with $\sum_{i=1}^m \alpha_i = 1$, then there exists a partition, $\{J_i\}_{i=1}^m$ of $(a,1]$ such that $m(J_i) = \alpha_i m((a,1]) = \alpha_i(1-a)$ – simply take $J_i = (a_{i-1}, a_i]$ where $a_0 = a$ and $a_i = \left(\sum_{j\le i} \alpha_i\right) a$ for $1 \le i \le m$.

Let ν be any probability on S and let
$$A_i := \{i\} \times \left(0, \frac{\nu(i)}{\mu(i)} \wedge 1\right]$$
and
$$C = \Omega \setminus \left(\sum_{i=1}^m A_i\right) = \sum_{i=1}^m \{i\} \times \left(\frac{\nu(i)}{\mu(i)} \wedge 1, 1\right]$$
and observe that
$$P(A_i) = \mu(i) \cdot \left(\frac{\nu(i)}{\mu(i)} \wedge 1\right) = \nu(i) \wedge \mu(i).$$

Using the observation in the previous paragraph we may write $\{k\} \times \left(\frac{\nu(k)}{\mu(k)} \wedge 1, 1\right] = \sum_{i=1}^m C_{k,i}$ with

$$P(C_{k,i}) = \alpha_i \cdot P\left(\{k\} \times \left(\frac{\nu(k)}{\mu(k)} \wedge 1, 1\right]\right).$$

The sets $C_i := \sum_{k=1}^m C_{k,i}$ then form a partition of C such that $P(C_i) = \alpha_i P(C)$ for all i.

We now define

$$Y_\nu(i,\theta) := \sum_{i=1}^m i 1_{A_i \cup C_i}$$

so that $Y_\nu = i$ on A_i and in particular $Y_\nu(i,\theta) = i$ when $\theta \le v(i)/\mu(i)$.

To finish the proof we need only choose the $\{\alpha_i\}_{i=1}^m$ so that $P(Y_\nu = i) = \nu(i)$ for all i, i.e. we must require,

$$\begin{aligned}\nu(i) = P(Y_\nu = i) &= P(A_i \cup C_i) = P(A_i) + \alpha_i P(C)\\ &= \nu(i) \wedge \mu(i) + \alpha_i P(C)\end{aligned} \tag{21.21}$$

and therefore we must define

$$\alpha_i = (\nu(i) - \nu(i) \wedge \mu(i))/P(C) \ge 0.$$

To see this is an admissible choice (i.e. $\sum_{i=1}^m \alpha_i = 1$) notice that

$$\begin{aligned}P(C) &= \sum_i [\mu(i) - \nu(i) \wedge \mu(i)]\\ &= \sum_{\nu(i)<\mu(i)} (\mu(i) - \nu(i)) = \sum_{\mu(i)\le\nu(i)} (\nu(i) - \mu(i)),\end{aligned} \tag{21.22}$$

wherein we have used the fact that

$$\sum_{i\in S} (\mu(i) - \nu(i)) = 1 - 1 = 0.$$

Making use of these identities we find,

$$\sum_{i\in S} \alpha_i = \frac{1}{P(C)} \sum_{\mu(i)\le\nu(i)} (\nu(i) - \mu(i)) = 1.$$

■

The next theorem summarizes a number of useful equivalent characterizations of weak convergence. (The reader should compare Theorem 21.29 with Corollary 21.9.) In this theorem we will write $BC(\mathbb{R})$ for the bounded continuous functions, $f : \mathbb{R} \to \mathbb{R}$ (or $f : \mathbb{R} \to \mathbb{C}$) and $C_c(\mathbb{R})$ for those $f \in C(\mathbb{R})$ which have compact support, i.e. $f(x) \equiv 0$ if $|x|$ is sufficiently large.

Theorem 21.29. *Suppose that $\{\mu_n\}_{n=0}^\infty$ is a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and for each n, let $F_n(y) := \mu_n((-\infty, y])$ be the (proper) distribution function associated to μ_n. Then the following are equivalent.*

1. *For all $f \in BC(\mathbb{R})$,*
$$\int_\mathbb{R} f d\mu_n \to \int_\mathbb{R} f d\mu_0 \text{ as } n \to \infty. \tag{21.23}$$
2. *Eq. (21.23) holds for all $f \in BC(\mathbb{R})$ which are uniformly continuous.*
3. *Eq. (21.23) holds for all $f \in C_c(\mathbb{R})$.*
4. *$F_n \implies F$.*
5. *There exists a probability space $(\Omega, \mathcal{B}, P)$ and random variables, Y_n, on this space such that $P \circ Y_n^{-1} = \mu_n$ for all n and $Y_n \to Y_0$ a.s.*

Proof. Clearly 1. $\implies$ 2. $\implies$ 3. and 5. $\implies$ 1. by the dominated convergence theorem. Indeed, we have

$$\int_\mathbb{R} f d\mu_n = \mathbb{E}[f(Y_n)] \overset{\text{D.C.T.}}{\to} \mathbb{E}[f(Y)] = \int_\mathbb{R} f d\mu_0$$

for all $f \in BC(\mathbb{R})$. Therefore it suffices to prove 3. $\implies$ 4. and 4. $\implies$ 5. The proof of 4. $\implies$ 5. will be the content of Skorohod's Theorem 21.27 below. Given Skorohod's Theorem, we will now complete the proof.

(3. $\implies$ 4.) Let $-\infty < a < b < \infty$ with $a, b \in \mathcal{C}(F_0)$ and for $\varepsilon > 0$, let $f_\varepsilon(x) \ge 1_{(a,b]}$ and $g_\varepsilon(x) \le 1_{(a,b]}$ be the functions in $C_c(\mathbb{R})$ pictured in Figure 21.4. Then

$$\limsup_{n\to\infty} \mu_n((a,b]) \le \limsup_{n\to\infty} \int_\mathbb{R} f_\varepsilon d\mu_n = \int_\mathbb{R} f_\varepsilon d\mu_0 \tag{21.24}$$

and

$$\liminf_{n\to\infty} \mu_n((a,b]) \ge \liminf_{n\to\infty} \int_\mathbb{R} g_\varepsilon d\mu_n = \int_\mathbb{R} g_\varepsilon d\mu_0. \tag{21.25}$$

Since $f_\varepsilon \to 1_{[a,b]}$ and $g_\varepsilon \to 1_{(a,b)}$ as $\varepsilon \downarrow 0$, we may use the dominated convergence theorem to pass to the limit as $\varepsilon \downarrow 0$ in Eqs. (21.24) and (21.25) to conclude,

$$\limsup_{n\to\infty} \mu_n((a,b]) \le \mu_0([a,b]) = \mu_0((a,b])$$

and

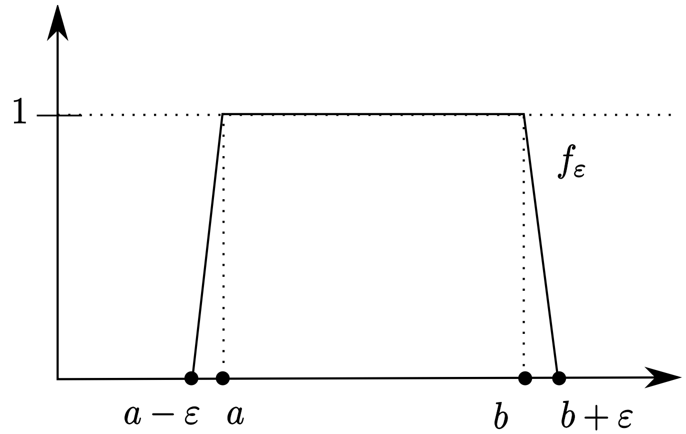
$$\liminf_{n\to\infty} \mu_n((a,b]) \ge \mu_0((a,b)) = \mu_0((a,b]),$$

where the second equality in each of the equations holds because a and b are points of continuity of F_0. Hence we have shown that $\lim_{n\to\infty} \mu_n((a,b])$ exists and is equal to $\mu_0((a,b])$.

■

Example 21.30. Suppose that $\{\mu_n\}_{n=1}^\infty$ and μ are measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\lim_{n\to\infty} d_{TV}(\mu_n, \mu) = 0$, then $\mu_n \implies \mu$. To prove this simply observe that for $f \in BC(\mathbb{R})$ we have by Corollary 21.9 that

$$|\mu(f) - \mu_n(f)| \le 2\|f\|_u d_{TV}(\mu_n, \mu) \to 0 \text{ as } n \to \infty.$$



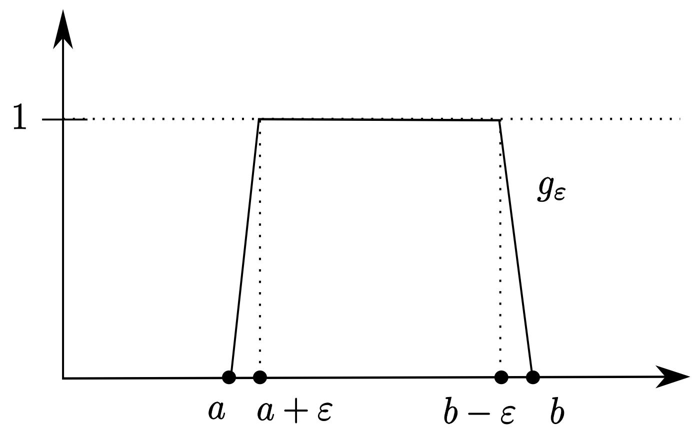


Fig. 21.4. The picture definition of the trapezoidal functions, f_ε and g_ε.

Corollary 21.31. *Suppose that $\{X_n\}_{n=0}^\infty$ is a sequence of random variables, such that $X_n \xrightarrow{P} X_0$, then $X_n \implies X_0$. (Recall that Example 21.23 shows the converse is in general false.)*

Proof. Let $g \in BC(\mathbb{R})$, then by Corollary 12.12, $g(X_n) \xrightarrow{P} g(X_0)$ and since g is bounded, we may apply the dominated convergence theorem (see Corollary 12.9) to conclude that $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X_0)]$. ■

We end this section with a few more equivalent characterizations of weak convergence. The combination of Theorem 21.29 and 21.32 is often called the Portmanteau[4] Theorem. A review of the notions of closure, interior, and boundary of a set A which are used in the next theorem may be bound in Subsection 21.9.1 below.

Theorem 21.32 (The Baby Portmanteau Theorem). *Suppose $\{F_n\}_{n=0}^\infty$ are proper distribution functions. (Recall that we are denoting $\mu_{F_n}(A)$ simply by $F_n(A)$ for all $A \in \mathcal{B}_\mathbb{R}$.) Then the following are equivalent.*

1. *$F_n \implies F_0$.*
2. *$\liminf_{n\to\infty} F_n(U) \ge F_0(U)$ for open subsets, $U \subset \mathbb{R}$.*
3. *$\limsup_{n\to\infty} F_n(C) \le F_0(C)$ for all closed subsets, $C \subset \mathbb{R}$.*
4. *$\lim_{n\to\infty} F_n(A) = F_0(A)$ for all $A \in \mathcal{B}_\mathbb{R}$ such that $F_0(\mathrm{bd}(A)) = 0$.*

Proof. (1. $\implies$ 2.) By Skorohod's Theorem 21.27 we may choose random variables, Y_n, such that $P(Y_n \le y) = F_n(y)$ for all $y \in \mathbb{R}$ and $n \in \mathbb{N}$ and $Y_n \to Y_0$ a.s. as $n \to \infty$. Since U is open, it follows that

$$1_U(Y) \le \liminf_{n\to\infty} 1_U(Y_n) \text{ a.s.}$$

and so by Fatou's lemma,

$$\begin{aligned} F(U) &= P(Y \in U) = \mathbb{E}[1_U(Y)] \\ &\le \liminf_{n\to\infty} \mathbb{E}[1_U(Y_n)] = \liminf_{n\to\infty} P(Y_n \in U) = \liminf_{n\to\infty} F_n(U). \end{aligned}$$

(2. $\iff$ 3.) This follows from the observations: 1) $C \subset \mathbb{R}$ is closed iff $U := C^c$ is open, 2) $F(U) = 1 - F(C)$, and 3) $\liminf_{n\to\infty}(-F_n(C)) = -\limsup_{n\to\infty} F_n(C)$.

(2. and 3. $\iff$ 4.) If $F_0(\mathrm{bd}(A)) = 0$, then $A^o \subset A \subset \bar{A}$ with $F_0(\bar{A} \setminus A^o) = F_0(\mathrm{bd}(A)) = 0$. Therefore

$$F_0(A) = F_0(A^o) \le \liminf_{n\to\infty} F_n(A^o) \le \limsup_{n\to\infty} F_n(\bar{A}) \le F_0(\bar{A}) = F_0(A).$$

(4. $\implies$ 1.) Let $a, b \in \mathcal{C}(F_0)$ and take $A := (a, b]$. Then $F_0(\mathrm{bd}(A)) = F_0(\{a, b\}) = 0$ and therefore, $\lim_{n\to\infty} F_n((a, b]) = F_0((a, b])$, i.e. $F_n \implies F_0$. ■

Exercise 21.7. Suppose that F is a continuous proper distribution function. Show,

1. $F : \mathbb{R} \to [0, 1]$ is uniformly continuous.

[4] Portmanteua: 1) A new word formed by joining two others and combining their meanings, or 2) A large travelling bag made of stiff leather.

2. If $\{F_n\}_{n=1}^\infty$ is a sequence of distribution functions converging weakly to F, then F_n converges to F uniformly on $\mathbb{R}$, i.e.

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F(x)-F_n(x)|=0.$$

In particular, it follows that

$$\begin{aligned}\sup_{a<b}|\mu_F((a,b])-\mu_{F_n}((a,b])| &= \sup_{a<b}|F(b)-F(a)-(F_n(b)-F_n(a))|\\ &\le \sup_b|F(b)-F_n(b)|+\sup_a|F_n(a)-F_n(a)|\\ &\to 0 \text{ as } n\to\infty.\end{aligned}$$

Hints for part 2. Given $\varepsilon>0$, show that there exists, $-\infty=\alpha_0<\alpha_1<\dots<\alpha_n=\infty$, such that $|F(\alpha_{i+1})-F(\alpha_i)|\le\varepsilon$ for all i. Now show, for $x\in[\alpha_i,\alpha_{i+1})$, that

$$|F(x)-F_n(x)|\le(F(\alpha_{i+1})-F(\alpha_i))+|F(\alpha_i)-F_n(\alpha_i)|+(F_n(\alpha_{i+1})-F_n(\alpha_i)).$$

Most of the results above generalize to the case where $\mathbb{R}$ is replaced by a complete separable metric space as described in Section 21.9 below. The definition of weak convergence in this generality is as follows.

Definition 21.33 (Weak convergence). *Let (S,ρ) be a metric space. A sequence of probability measures $\{\mu_n\}_{n=1}^\infty$ is said to converge weakly to a probability μ if $\lim_{n\to\infty}\mu_n(f)=\mu(f)$ for every $f\in BC(S)$.*[5] *We will write this convergence as $\mu_n\implies\mu$ or $\mu_n\overset{w}{\to}\mu$ as $n\to\infty$.*

As a warm up to these general results and compactness results to come, let us consider in more detail the case where $S=\mathbb{R}^d$.

Proposition 21.34. *Suppose that $\{\mu_n\}_{n=1}^\infty\cup\{\mu\}$ are probability measures on $(S:=\mathbb{R}^d,\mathcal{B}=\mathcal{B}_{\mathbb{R}^d})$ such that $\mu(f)=\lim_{n\to\infty}\mu_n(f)$ for all $f\in C_c^\infty(S)$ then $\lim_{n\to\infty}\mu_n(f)=\mu(f)$ for all $f\in C_c(S)$.*

Proof. Let $\rho\in C_c^\infty(S)$ such that $0\le\rho\le 1_{C_1}$ and $\int_S\rho(z)\,dz=1$. For $f\in C_c(S)$ and $\varepsilon>0$, let

$$f_\varepsilon(x):=\int_S f(x+\varepsilon z)\rho(z)\,dz. \tag{21.26}$$

It then follows that

$$\begin{aligned}M_\varepsilon &:= \max_x|f(x)-f_\varepsilon(x)|=\max_x\left|\int_S[f(x)-f(x+\varepsilon z)]\rho(z)\,dz\right|\\ &\le \max_x\int_S|f(x)-f(x+\varepsilon z)|\rho(z)\,dz\\ &\le \max_x\max_{|z|\le\varepsilon}|f(x)-f(x+z)|\end{aligned}$$

where the latter expression goes to zero as $\varepsilon\downarrow 0$ by the uniform continuity of f. Thus we have shown that $f_\varepsilon\to f$ uniformly in x as $\varepsilon\downarrow 0$. Making the change of variables $y=x+\varepsilon z$ in Eq. (21.26) shows

$$f_\varepsilon(x):=\frac{1}{\varepsilon^d}\int_S f(y)\rho\left(\frac{y-x}{\varepsilon}\right)dy$$

from which it follows that f_ε is smooth. Using this information we find,

$$\begin{aligned}&\limsup_{n\to\infty}|\mu(f)-\mu_n(f)|\\ &\quad\le\limsup_{n\to\infty}[|\mu(f)-\mu(f_\varepsilon)|+|\mu(f_\varepsilon)-\mu_n(f_\varepsilon)|+|\mu_n(f_\varepsilon)-\mu_n(f)|]\\ &\quad\le 2M_\varepsilon\to 0 \text{ as } \varepsilon\downarrow 0.\end{aligned}$$

■

Theorem 21.35. *Suppose that $\{\mu_n\}_{n=1}^\infty\cup\{\mu\}$ are probability measures on $(S:=\mathbb{R}^d,\mathcal{B}=\mathcal{B}_{\mathbb{R}^d})$ (or some other locally compact Hausdorff space) such that $\mu(f)=\lim_{n\to\infty}\mu_n(f)$ for all $f\in C_c(S)$, then;*

1. *For all $\varepsilon>0$ there exists a compact set $K_\varepsilon\subset S$ such that $\mu(K_\varepsilon)\ge 1-\varepsilon$ and $\mu_n(K_\varepsilon)\ge 1-\varepsilon$ for all $n\in\mathbb{N}$.*
2. *If $f\in BC(S)$, then $\lim_{n\to\infty}\mu_n(f)=\mu(f)$.*

Proof. For all $R>0$ let $C_R:=\{x\in S:|x|\le R\}$ and then choose $\varphi_R\in C_c(S)$ such that $\varphi_R=1$ on $C_{R/2}$ and $0\le\varphi_R\le 1_{C_R}$.

1. With this notation it follows that
$$\mu_n(C_R)\ge\mu_n(\varphi_R)\to\mu(\varphi_R)\ge\mu(C_{R/2}).$$
Choose R so large that $\mu_n(C_R)\ge\mu(C_{R/2})\ge 1-\varepsilon/2$. Then for $n\ge N_\varepsilon$ we will have $\mu_n(C_R)\ge 1-\varepsilon$ for all $n\ge N_\varepsilon$. By increasing R more if necessary we may also assume that $\mu_n(C_R)\ge 1-\varepsilon$ for all $n<N_\varepsilon$. Taking $K_\varepsilon:=C_R$ for this R completes the proof of item 1.
2. Let $f\in BC(S)$ and for $R>0$ let $f_R:=\varphi_R\cdot f\in C_c(S)$. Then

[5] This is actually "**weak-* convergence**" when viewing $\mu_n\in BC(S)^*$.

$$\limsup_{n\to\infty} |\mu(f) - \mu_n(f)|$$
$$\leq \limsup_{n\to\infty} \left[|\mu(f) - \mu(f_R)| + |\mu(f) - \mu_n(f_R)| + |\mu_n(f_R) - \mu_n(f)| \right]$$
$$= |\mu(f) - \mu(f_R)| + \limsup_{n\to\infty} |\mu_n(f_R) - \mu_n(f)|. \tag{21.27}$$

By the dominated convergence theorem, $\lim_{R\to\infty} |\mu(f) - \mu(f_R)| = 0$. For the second term if $M = \max_{x\in S} |f(x)|$ we will have

$$\sup_n |\mu_n(f_R) - \mu_n(f)| \leq \sup_n \mu_n(|f_R - f|)$$
$$\leq M \cdot \sup_n \mu_n(\varphi_R \neq 1) \leq M \cdot \sup_n \mu_n\left(S \setminus C_{R/2}\right).$$

However, by item 1. it follows that $\lim_{R\to\infty} \sup_n \mu_n\left(S \setminus C_{R/2}\right) = 0$. Therefore letting $R \to \infty$ in Eq. (21.27) show that $\limsup_{n\to\infty} |\mu(f) - \mu_n(f)| = 0$. ■

21.5 "Derived" Weak Convergence

Lemma 21.36. *Let (X,d) be a metric space, $f : X \to \mathbb{R}$ be a function, and $\mathcal{D}(f)$ be the set of $x \in X$ where f is discontinuous at x. Then $\mathcal{D}(f)$ is a Borel measurable subset of X.*

Proof. For $x \in X$ and $\delta > 0$, let $B_x(\delta) = \{y \in X : d(x,y) < \delta\}$. Given $\delta > 0$, let $f^\delta : X \to \mathbb{R}\cup\{\infty\}$ be defined by,

$$f^\delta(x) := \sup_{y\in B_x(\delta)} f(y).$$

We will begin by showing f^δ is **lower semi-continuous**, i.e. $\left\{f^\delta \leq a\right\}$ is closed (or equivalently $\left\{f^\delta > a\right\}$ is open) for all $a \in \mathbb{R}$. Indeed, if $f^\delta(x) > a$, then there exists $y \in B_x(\delta)$ such that $f(y) > a$. Since this y is in $B_{x'}(\delta)$ whenever $d(x,x') < \delta - d(x,y)$ (because then, $d(x',y) \leq d(x,y) + d(x,x') < \delta$) it follows that $f^\delta(x') > a$ for all $x' \in B_x(\delta - d(x,y))$. This shows $\left\{f^\delta > a\right\}$ is open in X.

We similarly define $f_\delta : X \to \mathbb{R}\cup\{-\infty\}$ by

$$f_\delta(x) := \inf_{y\in B_x(\delta)} f(y).$$

Since $f_\delta = -(-f)^\delta$, it follows that

$$\{f_\delta \geq a\} = \left\{(-f)^\delta \leq -a\right\}$$

is closed for all $a \in \mathbb{R}$, i.e. f_δ is **upper semi-continuous**. Moreover, $f_\delta \leq f \leq f^\delta$ for all $\delta > 0$ and $f^\delta \downarrow f^0$ and $f_\delta \uparrow f_0$ as $\delta \downarrow 0$, where $f_0 \leq f \leq f^0$ and $f_0 : X \to \mathbb{R}\cup\{-\infty\}$ and $f^0 : X \to \mathbb{R}\cup\{\infty\}$ are measurable functions. The proof is now complete since it is easy to see that

$$\mathcal{D}(f) = \left\{f^0 > f_0\right\} = \left\{f^0 - f_0 \neq 0\right\} \in \mathcal{B}_X.$$

■

Remark 21.37. Suppose that $x_n \to x$ with $x \in \mathcal{C}(f) := \mathcal{D}(f)^c$. Then $f(x_n) \to f(x)$ as $n \to \infty$.

Theorem 21.38 (Continuous Mapping Theorem). *Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. If $X_n \implies X_0$ and $P(X_0 \in \mathcal{D}(f)) = 0$, then $f(X_n) \implies f(X_0)$. If in addition, f is bounded, $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X_0)$. (This result generalizes easily to the case where $f : S \to T$ is a Borel measurable function between metric spaces and X_n, X_0 are not S – valued random functions.)*

Proof. Let $\{Y_n\}_{n=0}^\infty$ be random variables on some probability space as in Theorem 21.27. For $g \in BC(\mathbb{R})$ we observe that $\mathcal{D}(g\circ f) \subset \mathcal{D}(f)$ and therefore,

$$P(Y_0 \in \mathcal{D}(g\circ f)) \leq P(Y_0 \in \mathcal{D}(f)) = P(X_0 \in \mathcal{D}(f)) = 0.$$

Hence it follows that $g\circ f\circ Y_n \to g\circ f\circ Y_0$ a.s. So an application of the dominated convergence theorem (see Corollary 12.9) implies

$$\mathbb{E}[g(f(X_n))] = \mathbb{E}[g(f(Y_n))] \to \mathbb{E}[g(f(Y_0))] = \mathbb{E}[g(f(X_0))]. \tag{21.28}$$

This proves the first assertion. For the second assertion we take $g(x) = (x \wedge M) \vee (-M)$ in Eq. (21.28) where M is a bound on $|f|$. ■

Theorem 21.39 (Slutzky's Theorem). *Suppose that $X_n \implies X \in \mathbb{R}^m$ and $Y_n \xrightarrow{P} c \in \mathbb{R}^n$ where $c \in \mathbb{R}^n$ is constant. Assuming all random vectors are on the same probability space we will have $(X_n, Y_n) \implies (X, c)$ – see Definition 21.33. In particular if $m = n$, by taking $f(x,y) = g(x+y)$ and $f(x,y) = h(x\cdot y)$ with $g \in BC(\mathbb{R}^n)$ and $h \in BC(\mathbb{R})$, we learn $X_n + Y_n \implies X + c$ and $X_n \cdot Y_n \implies X \cdot c$ respectively. (The first part of this theorem generalizes to metric spaces as well.)*

Proof. According to Theorem 21.35 it suffices to show for

$$\lim_{n\to\infty} \mathbb{E}[f(X_n, Y_n)] = \mathbb{E}[f(X,c)] \tag{21.29}$$

for all $f \in BC(\mathbb{R}^{m\times n})$ which are uniformly continuous or even only $f \in C(_c\mathbb{R}^{m\times n})$. For a uniformly continuous function we have for every $\varepsilon > 0$ a $\delta := \delta(\varepsilon) > 0$ such that

$$|f(x,y) - f(x',y')| \le \varepsilon \text{ if } \|(x,y) - (x',y')\| \le \delta.$$

Then

$$\begin{aligned}|\mathbb{E}[f(X_n,Y_n) - f(X_n,c)]| &\le \mathbb{E}[|f(X_n,Y_n) - f(X_n,c)| : \|Y_n - c\| \le \delta] \\ &\quad + \mathbb{E}[|f(X_n,Y_n) - f(X_n,c)| : \|Y_n - c\| > \delta] \\ &\le \varepsilon + 2MP(\|Y_n - c\| > \delta) \to \varepsilon \text{ as } n\to\infty,\end{aligned}$$

where $M = \sup|f|$. Since, $X_n \implies X$, we know $\mathbb{E}[f(X_n,c)] \to \mathbb{E}[f(X,c)]$ and hence we have shown,

$$\begin{aligned}&\limsup_{n\to\infty} |\mathbb{E}[f(X_n,Y_n) - f(X,c)]| \\ &\quad\le \limsup_{n\to\infty}|\mathbb{E}[f(X_n,Y_n) - f(X_n,c)]| + \limsup_{n\to\infty}|\mathbb{E}[f(X_n,c) - f(X,c)]| \le \varepsilon.\end{aligned}$$

As $\varepsilon > 0$ was arbitrary this proves Eq. (21.29).

■

Theorem 21.40 (δ – method). *Suppose that $\{X_n\}_{n=1}^\infty$ are random variables, $b \in \mathbb{R}$, $a_n \in \mathbb{R}\setminus\{0\}$ with $\lim_{n\to\infty} a_n = 0$, and*

$$Y_n := \frac{X_n - b}{a_n} \implies Z.$$

If $g : \mathbb{R} \to \mathbb{R}$ be a measurable function which is differentiable at b, then

$$\frac{g(X_n) - g(b)}{a_n} \implies g'(b)\, Z. \tag{21.30}$$

Put more informally, if

$$X_n \overset{d}{\simeq} b + a_n Z \text{ then } g(X_n) \overset{d}{\simeq} g(b) + g'(b)\, a_n Z.$$

Proof. Informally we have $X_n = a_nY_n + b \overset{d}{\cong} a_nZ + b$ and therefore

$$\frac{g(X_n) - g(b)}{a_n} \overset{d}{\cong} \frac{g(a_nZ + b) - g(b)}{a_nZ} Z \to g'(b)\, Z \text{ as } n \to \infty.$$

We now make the proof rigorous.

By Skorohod's Theorem 21.27 we may assume that $\{Y_n\}_{n=1}^\infty$ and Z are on the same probability space and that $Y_n \to Z$ a.s. and we may take $X_n := a_nY_n + b$. By the definition of the derivative of g at b, we have

$$g(b + \Delta) - g(b) = g'(b)\,\Delta + \varepsilon(\Delta)\,\Delta$$

where $\varepsilon(\Delta) \to 0$ as $\Delta \to 0$. Taking $\Delta = a_nY_n$ in this equation shows

$$\begin{aligned}\frac{g(X_n) - g(b)}{a_n} &= \frac{g(a_nY_n + b) - g(b)}{a_n} \\ &= \frac{g'(b)\,a_nY_n + \varepsilon(a_nY_n)\,a_nY_n}{a_n} \to g'(b)\, Z \text{ a.s.}\end{aligned}$$

which implies Eq. (21.30) because of Corollary 21.31.

■

Example 21.41. Suppose that $\{U_n\}_{n=1}^\infty$ are i.i.d. random variables which are uniformly distributed on $[0,1]$ and let $Y_n := \prod_{j=1}^n U_j^{\frac{1}{n}}$. Our goal is to find a_n and b_n such that $\frac{Y_n - b_n}{a_n}$ is weakly convergent to a non-constant random variable. To this end, let

$$X_n := \ln Y_n = \frac{1}{n}\sum_{j=1}^n \ln U_j.$$

Since

$$\mathbb{E}[\ln U_1] = \int_0^1 \ln x dx = -1,$$
$$\mathbb{E}[\ln U_1]^2 = \int_0^1 \ln^2 x dx = 2,$$

$\mathrm{Var}(\ln U_1) = 1$ and so by the central limit theorem

$$\sqrt{n}[X_n - (-1)] = \frac{\sum_{j=1}^n [\ln U_j + 1]}{\sqrt{n}} \implies Z \overset{d}{=} N(0,1).$$

In other words, $X_n \overset{d}{\simeq} -1 + \frac{1}{\sqrt{n}}Z$ and so by the δ – method if $g'(-1)$ eixts, then

$$g(X_n) \overset{d}{\simeq} g(-1) + g'(-1)\frac{1}{\sqrt{n}}Z.$$

Taking $g(x) = e^x$ then implies $Y_n \overset{d}{\simeq} e^{-1} + e^{-1}\frac{1}{\sqrt{n}}Z$ or more precisely,

$$\sqrt{n}\left[\prod_{j=1}^n U_j^{\frac{1}{n}} - e^{-1}\right] = \sqrt{n}\left(Y_n - e^{-1}\right) \implies e^{-1}Z = N\left(0, e^{-2}\right).$$

Exercise 21.8. Given a function, $f : X \to \mathbb{R}$ and a point $x \in X$, let

$$\liminf_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \inf_{y\in B'_x(\delta)} f(y) \text{ and} \tag{21.31}$$

$$\limsup_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \sup_{y\in B'_x(\delta)} f(y), \tag{21.32}$$

where

$$B'_x(\delta) := \{y \in X : 0 < d(x,y) < \delta\}.$$

Show f is lower (upper) semi-continuous iff $\liminf_{y\to x} f(y) \geq f(x)$ $\left(\limsup_{y\to x} f(y) \leq f(x)\right)$ for all $x \in X$.

Solution to Exercise (21.8). Suppose Eq. (21.31) holds, $a \in \mathbb{R}$, and $x \in X$ such that $f(x) > a$. Since,

$$\lim_{\varepsilon\downarrow 0} \inf_{y\in B'_x(\delta)} f(y) = \liminf_{y\to x} f(y) \geq f(x) > a,$$

it follows that $\inf_{y\in B'_x(\delta)} f(y) > a$ for some $\delta > 0$. Hence we may conclude that $B_x(\delta) \subset \{f > a\}$ which shows $\{f > a\}$ is open.

Conversely, suppose now that $\{f > a\}$ is open for all $a \in \mathbb{R}$. Given $x \in X$ and $a < f(x)$, there exists $\delta > 0$ such that $B_x(\delta) \subset \{f > a\}$. Hence it follows that $\liminf_{y\to x} f(y) \geq a$ and then letting $a \uparrow f(x)$ then implies $\liminf_{y\to x} f(y) \geq f(x)$.

21.6 Convergence of Types

Given a sequence of random variables $\{X_n\}_{n=1}^\infty$ we often look for centerings $\{b_n\}_{n=1}^\infty \subset \mathbb{R}$ and scalings $\{a_n > 0\}_{n=1}^\infty$ such that there exists a non-constant random variable Y such that

$$\frac{X_n - b_n}{a_n} \implies Y. \tag{21.33}$$

Assuming this can be done it is reasonable to ask how unique are the centering, scaling parameters, and the limiting distribution Y. To answer this question let us suppose there exists another collection of centerings $\{\beta_n\}_{n=1}^\infty \subset \mathbb{R}$ and scalings $\{\alpha_n > 0\}_{n=1}^\infty$ along with a non-constant random variable Z such that Thus if

$$\frac{X_n - \beta_n}{\alpha_n} \implies Z. \tag{21.34}$$

Working informally we expect that

$$X_n \overset{d}{\cong} \alpha_n Z + \beta_n$$

and putting this expression back into Eq. (21.34) leads us to expect;

$$\frac{\alpha_n}{a_n} Z + \frac{\beta_n - b_n}{a_n} = \frac{\alpha_n Z + \beta_n - b_n}{a_n} \implies Y.$$

It is reasonable to expect that this can only happen if the limits

$$A = \lim_{n\to\infty} \frac{\alpha_n}{a_n} \in (0,\infty) \text{ and } B := \lim_{n\to\infty} \frac{\beta_n - b_n}{a_n} \tag{21.35}$$

exist and

$$Y \overset{d}{=} AZ + B. \tag{21.36}$$

Notice that $A > 0$ as both Y and Z are assumed to be non-constant. That these results are correct is the content of Theorem 21.45 below.

Let us now explain how to choose the $\{a_n\}$ and the $\{b_n\}$. Let $F_n(x) := P(X_n \leq x)$, then Eq. (21.33) states,

$$F_n(a_n y + b_n) = P(X_n \leq a_n y + b_n) = P\left(\frac{X_n - b_n}{a_n} \leq y\right) \implies P(Y \leq y).$$

Taking $y = 0$ and $y = 1$ in this equation leads us to expect,

$$\lim_{n\to\infty} F_n(b_n) = P(Y \leq 0) = \gamma_1 \in (0,1) \text{ and}$$
$$\lim_{n\to\infty} F_n(a_n + b_n) = P(Y \leq 1) = \gamma_2 \in (0,1).$$

In fact there is nothing so special about 0 and 1 in these equation for if $Y \overset{d}{=} AZ + B$ we will have $Z = A^{-1}(Y - B)$ and so

$$P(Y \leq 0) = P(AZ + B \leq 0) = P(Z \leq -B/A) \text{ and}$$
$$P(Y \leq 1) = P(AZ + B \leq 1) = P(Z \leq (1 - B)/A).$$

Definition 21.42. *Two random variables, Y and Z, are said to be of the* ***same type*** *if there exists constants, $A > 0$ and $B \in \mathbb{R}$ such that Eq. (21.36) holds. Alternatively put, if $U(y) := P(Y \leq y)$ and $V(z) := P(Z \leq z)$, then U and V should satisfy,*

$$V(z) = P(Z \leq z) = P(Y \leq Az + B) = U(Az + B)$$

for all $z \in \mathbb{R}$.

Remark 21.43. Suppose that $Y \overset{d}{=} AZ + B$ and Y and Z are square integrable random variables. Then

$$\mathbb{E}Y = A \cdot \mathbb{E}Z + B \text{ and } \operatorname{Var}(Y) = A^2 \operatorname{Var}(Z)$$

from which it follows that $A^2 = \mathrm{Var}(Y)/\mathrm{Var}(Z)$ and $B = \mathbb{E}Y - A \cdot \mathbb{E}Z$. In particular, given $Y \in L^2(P)$ there is a unique Z of the same type such that $\mathbb{E}Z = 0$ and $\mathrm{Var}(Z) = 1$. On these grounds it is often reasonable to try to choose $\{b_n\}$ and $\{a_n > 0\}$ so that $\bar{X}_n := a_n^{-1}(X_n - b_n)$ has mean zero and variance one.

We will need the following elementary observation for the proof of Theorem 21.45.

Lemma 21.44. *If Y is non-constant (a.s.) random variable and $U(y) := P(Y \le y)$, then $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$ for all γ_1 sufficiently close to 0 and γ_2 sufficiently close to 1 – see Notation 21.26 for the meaning of $U^{\leftarrow}$.*

Proof. Observe that Y is constant iff $U(y) = 1_{y \ge c}$ for some $c \in \mathbb{R}$, i.e. iff U only takes on the values, $\{0, 1\}$. So since Y is not constant, there exists $y \in \mathbb{R}$ such that $0 < U(y) < 1$. Hence if $\gamma_2 > U(y)$ then $U^{\leftarrow}(\gamma_2) \ge y$ and if $\gamma_1 < U(y)$ then $U^{\leftarrow}(\gamma_1) \le y$. Moreover, if we suppose that γ_1 is not the height of a flat spot of U, then in fact, $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$. This inequality then remains valid as γ_1 decreases and γ_2 increases. ∎

Theorem 21.45 (Convergence of Types). *Suppose $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables and $a_n, \alpha_n \in (0, \infty)$, $b_n, \beta_n \in \mathbb{R}$ are constants and Y and Z are non-constant random variables. Then*

1. *if both Eq. (21.33) and Eq. (21.34) hold then the limits, in Eq. (21.35) exists and $Y \overset{d}{=} AZ + B$ and in particular Y and Z are of the same type.*
2. *If the limits in Eq. (21.35) hold then either of the convergences in Eqs. (21.33) or (21.34) implies the others with Z and Y related by Eq. (21.36).*
3. *If there are some constants, $a_n > 0$ and $b_n \in \mathbb{R}$ and a non-constant random variable Y, such that Eq. (21.33) holds, then Eq. (21.34) holds using α_n and β_n of the form,*
$$\alpha_n := F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1) \text{ and } \beta_n := F_n^{\leftarrow}(\gamma_1) \tag{21.37}$$
for some $0 < \gamma_1 < \gamma_2 < 1$. If the F_n are invertible functions, Eq. (21.37) may be written as
$$F_n(\beta_n) = \gamma_1 \text{ and } F_n(\alpha_n + \beta_n) = \gamma_2. \tag{21.38}$$

Proof. (2) Assume the limits in Eq. (21.35) hold. If Eq. (21.33) is satisfied, then by Slutsky's Theorem 13.22,
$$\begin{aligned}\frac{X_n - \beta_n}{\alpha_n} &= \frac{X_n - b_n + b_n - \beta_n}{a_n}\frac{a_n}{\alpha_n} \\ &= \frac{X_n - b_n}{a_n}\frac{a_n}{\alpha_n} - \frac{\beta_n - b_n}{a_n}\frac{a_n}{\alpha_n} \\ &\implies A^{-1}(Y - B) =: Z\end{aligned}$$

Similarly, if Eq. (21.34) is satisfied, then
$$\frac{X_n - b_n}{a_n} = \frac{X_n - \beta_n}{\alpha_n}\frac{\alpha_n}{a_n} + \frac{\beta_n - b_n}{a_n} \implies AZ + B =: Y.$$

(1) If $F_n(y) := P(X_n \le y)$, then
$$P\left(\frac{X_n - b_n}{a_n} \le y\right) = F_n(a_n y + b_n) \text{ and } P\left(\frac{X_n - \beta_n}{\alpha_n} \le y\right) = F_n(\alpha_n y + \beta_n).$$
By assumption we have
$$F_n(a_n y + b_n) \implies U(y) \text{ and } F_n(\alpha_n y + \beta_n) \implies V(y).$$
If $w := \sup\{y : F_n(a_n y + b_n) < x\}$, then $a_n w + b_n = F_n^{\leftarrow}(x)$ and hence
$$\sup\{y : F_n(a_n y + b_n) < x\} = \frac{F_n^{\leftarrow}(x) - b_n}{a_n}.$$
Similarly,
$$\sup\{y : F_n(\alpha_n y + \beta_n) < x\} = \frac{F_n^{\leftarrow}(x) - \beta_n}{\alpha_n}.$$
With these identities, it now follows from the proof of Skorohod's Theorem 21.27 (see Eq. (21.20)) that there exists an at most countable subset, Λ, of $(0, 1)$ such that,
$$\begin{aligned}\frac{F_n^{\leftarrow}(x) - b_n}{a_n} &= \sup\{y : F_n(a_n y + b_n) < x\} \to U^{\leftarrow}(x) \text{ and} \\ \frac{F_n^{\leftarrow}(x) - \beta_n}{\alpha_n} &= \sup\{y : F_n(\alpha_n y + \beta_n) < x\} \to V^{\leftarrow}(x)\end{aligned}$$
for all $x \notin \Lambda$. Since Y and Z are not constants a.s., we can choose, by Lemma 21.44, $\gamma_1 < \gamma_2$ not in Λ such that $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$ and $V^{\leftarrow}(\gamma_1) < V^{\leftarrow}(\gamma_2)$. In particular it follows that
$$\begin{aligned}\frac{F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1)}{a_n} &= \frac{F_n^{\leftarrow}(\gamma_2) - b_n}{a_n} - \frac{F_n^{\leftarrow}(\gamma_1) - b_n}{a_n} \\ &\to U^{\leftarrow}(\gamma_2) - U^{\leftarrow}(\gamma_1) > 0\end{aligned} \tag{21.39}$$
and similarly
$$\frac{F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1)}{\alpha_n} \to V^{\leftarrow}(\gamma_2) - V^{\leftarrow}(\gamma_1) > 0.$$
Taking ratios of the last two displayed equations shows,

$$\frac{\alpha_n}{a_n} \to A := \frac{U^{\leftarrow}(\gamma_2) - U^{\leftarrow}(\gamma_1)}{V^{\leftarrow}(\gamma_2) - V^{\leftarrow}(\gamma_1)} \in (0,\infty).$$

Moreover,

$$\frac{F_n^{\leftarrow}(\gamma_1) - b_n}{a_n} \to U^{\leftarrow}(\gamma_1) \text{ and} \tag{21.40}$$

$$\frac{F_n^{\leftarrow}(\gamma_1) - \beta_n}{a_n} = \frac{F_n^{\leftarrow}(\gamma_1) - \beta_n}{\alpha_n}\frac{\alpha_n}{a_n} \to AV^{\leftarrow}(\gamma_1)$$

and therefore,

$$\frac{\beta_n - b_n}{a_n} = \frac{F_n^{\leftarrow}(\gamma_1) - \beta_n}{a_n} - \frac{F_n^{\leftarrow}(\gamma_1) - b_n}{a_n} \to AV^{\leftarrow}(\gamma_1) - U^{\leftarrow}(\gamma_1) := B.$$

(3) Now suppose that we define $\alpha_n := F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1)$ and $\beta_n := F_n^{\leftarrow}(\gamma_1)$, then according to Eqs. (21.39) and (21.40)we have

$$\alpha_n/a_n \to U^{\leftarrow}(\gamma_2) - U^{\leftarrow}(\gamma_1) \in (0,1) \text{ and}$$
$$\frac{\beta_n - b_n}{a_n} \to U^{\leftarrow}(\gamma_1) \text{ as } n \to \infty.$$

Thus we may always center and scale the $\{X_n\}$ using α_n and β_n of the form described in Eq. (21.37). ■

21.7 Weak Convergence Examples

Example 21.46. Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. $\exp(\lambda)$ – random variables, i.e. $X_n \ge 0$ a.s. and $P(X_n \ge x) = e^{-\lambda x}$ for all $x \ge 0$. In this case

$$F(x) := P(X_1 \le x) = 1 - e^{-\lambda(x\vee 0)} = \left(1 - e^{-\lambda x}\right)_+.$$

Consider $M_n := \max(X_1, \dots, X_n)$. We have, for $x \ge 0$ and $c_n \in (0,\infty)$ that

$$F_n(x) := P(M_n \le x) = P\left(\cap_{j=1}^n \{X_j \le x\}\right)$$
$$= \prod_{j=1}^n P(X_j \le x) = [F(x)]^n = \left(1 - e^{-\lambda x}\right)^n.$$

We now wish to find $a_n > 0$ and $b_n \in \mathbb{R}$ such that $\frac{M_n - b_n}{a_n} \implies Y$.

1. To this end we note that

$$P\left(\frac{M_n - b_n}{a_n} \le x\right) = P(M_n \le a_n x + b_n)$$
$$= F_n(a_n x + b_n) = [F(a_n x + b_n)]^n.$$

If we demand (c.f. Eq. (21.38) above)

$$P\left(\frac{M_n - b_n}{a_n} \le 0\right) = F_n(b_n) = [F(b_n)]^n \to \gamma_1 \in (0,1),$$

then $b_n \to \infty$ and we find

$$\ln\gamma_1 \sim n\ln F(b_n) = n\ln\left(1 - e^{-\lambda b_n}\right) \sim -ne^{-\lambda b_n}.$$

From this it follows that $b_n \sim \lambda^{-1}\ln n$. Given this, we now try to find a_n by requiring,

$$P\left(\frac{M_n - b_n}{a_n} \le 1\right) = F_n(a_n + b_n) = [F(a_n + b_n)]^n \to \gamma_2 \in (0,1).$$

However, by what we have done above, this requires $a_n + b_n \sim \lambda^{-1}\ln n$. Hence we may as well take a_n to be constant and for simplicity we take $a_n = 1$.

2. We now compute

$$\lim_{n\to\infty} P\left(M_n - \lambda^{-1}\ln n \le x\right) = \lim_{n\to\infty}\left(1 - e^{-\lambda\left(x + \lambda^{-1}\ln n\right)}\right)^n$$
$$= \lim_{n\to\infty}\left(1 - \frac{e^{-\lambda x}}{n}\right)^n = \exp\left(-e^{-\lambda x}\right).$$

The function $F(x) = \exp\left(-e^{-\lambda x}\right)$ is the CDF for a "Gumbel distribution," see Figure 21.5. Thus letting Y be a random variable with this distribution (i.e. $P(Y \le x) = \exp\left(-e^{-\lambda x}\right)$) we have shown $M_n - \frac{1}{\lambda}\ln n \implies Y$, i.e.

$$\max(X_1, \dots, X_n) - \frac{1}{\lambda}\ln n \implies Y.$$

Example 21.47. For $p \in (0,1)$, let X_p denote the number of trials to get success in a sequence of independent trials with success probability p. Then $P(X_p > n) = (1-p)^n$ and therefore for $x > 0$,

$$P(pX_p > x) = P\left(X_p > \frac{x}{p}\right) = (1-p)^{\left[\frac{x}{p}\right]} = e^{\left[\frac{x}{p}\right]\ln(1-p)}$$
$$\sim e^{-p\left[\frac{x}{p}\right]} \to e^{-x} \text{ as } p \to 0.$$

Therefore $pX_p \implies T$ where $T \overset{d}{=} \exp(1)$, i.e. $P(T > x) = e^{-x}$ for $x \ge 0$ or alternatively, $P(T \le y) = 1 - e^{-y\vee 0}$.

Remarks on this example. Let us see in a couple of ways where the appropriate centering and scaling of the X_p come from in this example. For this let $q = 1 - p$, then $P(X_p = n) = (1-p)^{n-1}p = q^{n-1}p$ for $n \in \mathbb{N}$. Also let

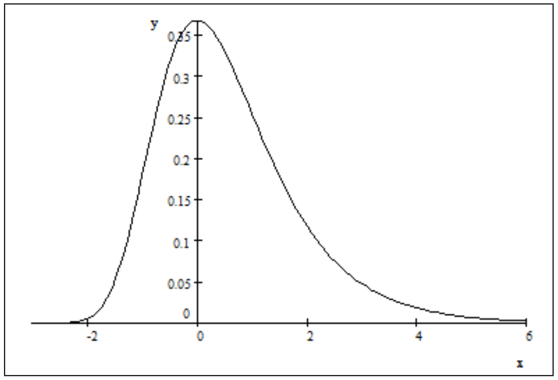


Fig. 21.5. Here is a plot of the density function for Y when $\lambda = 1$.

$$F_p(x) = P(X_p \le x) = P(X_p \le [x]) = 1 - q^{[x]}$$

where $[x] := \sum_{n=1}^{\infty} n \cdot 1_{[n,n+1)}$.

Method 1. Our goal is to choose $a_p > 0$ and $b_p \in \mathbb{R}$ such that $\lim_{p \downarrow 0} F_p(a_p x + b_p)$ exists. As above, we first demand (taking $x = 0$) that

$$\lim_{p \downarrow 0} F_p(b_p) = \gamma_1 \in (0,1).$$

Since, $\gamma_1 \sim F_p(b_p) \sim 1 - q^{b_p}$ we require, $q^{b_p} \sim 1 - \gamma_1$ and hence, $c \sim b_p \ln q = b_p \ln(1-p) \sim -b_p p$. This suggests that we take $b_p = 1/p$ say. Having done this, we would like to choose a_p such that

$$F_0(x) := \lim_{p \downarrow 0} F_p(a_p x + b_p) \text{ exists.}$$

Since,

$$F_0(x) \sim F_p(a_p x + b_p) \sim 1 - q^{a_p x + b_p}$$

this requires that

$$(1-p)^{a_p x + b_p} = q^{a_p x + b_p} \sim 1 - F_0(x)$$

and hence that

$$\ln(1 - F_0(x)) = (a_p x + b_p)\ln q \sim (a_p x + b_p)(-p) = -p a_p x - 1.$$

From this (setting $x = 1$) we see that $p a_p \sim c > 0$. Hence we might take $a_p = 1/p$ as well. We then have

$$F_p(a_p x + b_p) = F_p\left(p^{-1}x + p^{-1}\right) = 1 - (1-p)^{\left[p^{-1}(x+1)\right]}$$

which is equal to 0 if $x \le -1$, and for $x > -1$ we find

$$(1-p)^{\left[p^{-1}(x+1)\right]} = \exp\left(\left[p^{-1}(x+1)\right]\ln(1-p)\right) \to \exp(-(x+1)).$$

Hence we have shown,

$$\lim_{p \downarrow 0} F_p(a_p x + b_p) = [1 - \exp(-(x+1))]\, 1_{x \ge -1}$$

$$\frac{X_p - 1/p}{1/p} = pX_p - 1 \implies T - 1$$

or again that $pX_p \implies T$.

Method 2. (Center and scale using the first moment and the variance of X_p.) The generating function is given by

$$f(z) := \mathbb{E}\left[z^{X_p}\right] = \sum_{n=1}^{\infty} z^n q^{n-1} p = \frac{pz}{1-qz}.$$

Observe that $f(z)$ is well defined for $|z| < \frac{1}{q}$ and that $f(1) = 1$, reflecting the fact that $P(X_p \in \mathbb{N}) = 1$, i.e. a success must occur almost surely. Moreover, we have

$$f'(z) = \mathbb{E}\left[X_p z^{X_p - 1}\right],\ f''(z) = \mathbb{E}\left[X_p(X_p-1) z^{X_p-2}\right], \dots$$
$$f^{(k)}(z) = \mathbb{E}\left[X_p(X_p-1)\dots(X_p-k+1) z^{X_p-k}\right]$$

and in particular,

$$\mathbb{E}\left[X_p(X_p-1)\dots(X_p-k+1)\right] = f^{(k)}(1) = \left(\frac{d}{dz}\right)^k |_{z=1} \frac{pz}{1-qz}.$$

Since

$$\frac{d}{dz}\frac{pz}{1-qz} = \frac{p(1-qz)+qpz}{(1-qz)^2} = \frac{p}{(1-qz)^2}$$

and

$$\frac{d^2}{dz^2}\frac{pz}{1-qz} = 2\frac{pq}{(1-qz)^3}$$

it follows that

$$\mu_p := \mathbb{E}X_p = \frac{p}{(1-q)^2} = \frac{1}{p} \text{ and}$$

$$\mathbb{E}\left[X_p(X_p-1)\right] = 2\frac{pq}{(1-q)^3} = \frac{2q}{p^2}.$$

Therefore,

$$\sigma_p^2 = \operatorname{Var}(X_p) = \mathbb{E}X_p^2 - (\mathbb{E}X_p)^2 = \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2$$
$$= \frac{2q+p-1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

Thus, if we had used μ_p and σ_p to center and scale X_p we would have considered,

$$\frac{X_p - \frac{1}{p}}{\frac{\sqrt{1-p}}{p}} = \frac{pX_p - 1}{\sqrt{1-p}} \implies T - 1$$

instead.

Theorem 21.48 (This is already done in Theorem 7.62). *Let $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables such that $P(X_n = \pm 1) = 1/2$ and let $S_n := X_1 + \dots + X_n$ – the position of a drunk after n steps. Observe that $|S_n|$ is an odd integer if n is odd and an even integer if n is even. Then $\frac{S_m}{\sqrt{m}} \implies N(0,1)$ as $m \to \infty$.*

Proof. (Sketch of the proof.) We start by observing that $S_{2n} = 2k$ iff

$$\#\{i \le 2n : X_i = 1\} = n + k \text{ while}$$
$$\#\{i \le 2n : X_i = -1\} = 2n - (n+k) = n - k$$

and therefore,

$$P(S_{2n} = 2k) = \binom{2n}{n+k}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n+k)! \cdot (n-k)!}\left(\frac{1}{2}\right)^{2n}.$$

Recall Stirling's formula states,

$$n! \sim n^n e^{-n}\sqrt{2\pi n} \text{ as } n \to \infty$$

and therefore,

$$P(S_{2n} = 2k)$$
$$\sim \frac{(2n)^{2n} e^{-2n}\sqrt{4\pi n}}{(n+k)^{n+k} e^{-(n+k)}\sqrt{2\pi(n+k)} \cdot (n-k)^{n-k} e^{-(n-k)}\sqrt{2\pi(n-k)}}\left(\frac{1}{2}\right)^{2n}$$
$$= \sqrt{\frac{n}{\pi(n+k)(n-k)}}\left(1 + \frac{k}{n}\right)^{-(n+k)} \cdot \left(1 - \frac{k}{n}\right)^{-(n-k)}$$
$$= \frac{1}{\sqrt{\pi n}}\sqrt{\frac{1}{\left(1+\frac{k}{n}\right)\left(1-\frac{k}{n}\right)}}\left(1 - \frac{k^2}{n^2}\right)^{-n} \cdot \left(1 + \frac{k}{n}\right)^{-k} \cdot \left(1 - \frac{k}{n}\right)^{k}$$
$$= \frac{1}{\sqrt{\pi n}}\left(1 - \frac{k^2}{n^2}\right)^{-n} \cdot \left(1 + \frac{k}{n}\right)^{-k-1/2} \cdot \left(1 - \frac{k}{n}\right)^{k-1/2}.$$

So if we let $x := 2k/\sqrt{2n}$, i.e. $k = x\sqrt{n/2}$ and $k/n = \frac{x}{\sqrt{2n}}$, we have

$$P\left(\frac{S_{2n}}{\sqrt{2n}} = x\right)$$
$$\sim \frac{1}{\sqrt{\pi n}}\left(1 - \frac{x^2}{2n}\right)^{-n} \cdot \left(1 + \frac{x}{\sqrt{2n}}\right)^{-x\sqrt{n/2}-1/2} \cdot \left(1 - \frac{x}{\sqrt{2n}}\right)^{x\sqrt{n/2}-1/2}$$
$$\sim \frac{1}{\sqrt{\pi n}} e^{x^2/2} \cdot e^{\frac{x}{\sqrt{2n}}\left(-x\sqrt{n/2}-1/2\right)} \cdot e^{-\frac{x}{\sqrt{2n}}\left(x\sqrt{n/2}-1/2\right)}$$
$$\sim \frac{1}{\sqrt{\pi n}} e^{-x^2/2},$$

wherein we have repeatedly used

$$(1 + a_n)^{b_n} = e^{b_n \ln(1+a_n)} \sim e^{b_n a_n} \text{ when } a_n \to 0.$$

We now compute

$$P\left(a \le \frac{S_{2n}}{\sqrt{2n}} \le b\right) = \sum_{a \le x \le b} P\left(\frac{S_{2n}}{\sqrt{2n}} = x\right)$$
$$= \frac{1}{\sqrt{2\pi}} \sum_{a \le x \le b} e^{-x^2/2} \frac{2}{\sqrt{2n}} \tag{21.41}$$

where the sum is over x of the form, $x = \frac{2k}{\sqrt{2n}}$ with $k \in \{0, \pm 1, \dots, \pm n\}$. Since $\frac{2}{\sqrt{2n}}$ is the increment of x as k increases by 1, we see the latter expression in Eq. (21.41) is the Riemann sum approximation to

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

This proves $\frac{S_{2n}}{\sqrt{2n}} \implies N(0,1)$. Since

$$\frac{S_{2n+1}}{\sqrt{2n+1}} = \frac{S_{2n} + X_{2n+1}}{\sqrt{2n+1}} = \frac{S_{2n}}{\sqrt{2n}} \frac{1}{\sqrt{1 + \frac{1}{2n}}} + \frac{X_{2n+1}}{\sqrt{2n+1}},$$

it follows directly (or see Slutsky's Theorem 21.32) that $\frac{S_{2n+1}}{\sqrt{2n+1}} \implies N(0,1)$ as well. ■

Proposition 21.49. *Suppose that $\{U_n\}_{n=1}^\infty$ are i.i.d. random variables which are uniformly distributed in $(0,1)$. Let $U_{(k,n)}$ denote the position of the k^{th} – largest number from the list, $\{U_1, U_2, \dots, U_n\}$. Further let $k(n)$ be chosen so that $\lim_{n\to\infty} k(n) = \infty$ while $\lim_{n\to\infty} \frac{k(n)}{n} = 0$ and let*

$$X_n := \frac{U_{(k(n),n)} - k(n)/n}{\frac{\sqrt{k(n)}}{n}}.$$

Then $d_{TV}(X_n, N(0,1)) \to 0$ *as* $n \to \infty$.

Proof. (Sketch only. See Resnick, Proposition 8.2.1 for more details.) Observe that, for $x \in (0,1)$, that

$$P\left(U_{(k,n)} \le x\right) = P\left(\sum_{i=1}^n X_i \ge k\right) = \sum_{l=k}^n \binom{n}{l} x^l (1-x)^{n-l}.$$

From this it follows that $\rho_n(x) := 1_{(0,1)}(x) \frac{d}{dx} P\left(U_{(k,n)} \le x\right)$ is the probability density for $U_{(k,n)}$. It now turns out that $\rho_n(x)$ is a Beta distribution,

$$\rho_n(x) = \binom{n}{k} k \cdot x^{k-1} (1-x)^{n-k}.$$

Giving a direct computation of this result is not so illuminating. So let us go another route. To do this we are going to estimate, $P\left(U_{(k,n)} \in (x, x+\Delta]\right)$, for $\Delta \in (0,1)$. Observe that if $U_{(k,n)} \in (x, x+\Delta]$, then there must be at least one $U_i \in (x, x+\Delta]$, for otherwise, $U_{(k,n)} \le x + \Delta$ would imply $U_{(k,n)} \le x$ as well and hence $U_{(k,n)} \notin (x, x+\Delta]$. Let

$$\Omega_i := \{U_i \in (x, x+\Delta] \text{ and } U_j \notin (x, x+\Delta] \text{ for } j \ne i\}.$$

Since

$$\begin{aligned} P(U_i, U_j \in (x, x+\Delta] \text{ for some } i \ne j \text{ with } i, j \le n) &\le \sum_{i<j\le n} P(U_i, U_j \in (x, x+\Delta]) \\ &\le \frac{n^2 - n}{2} \Delta^2, \end{aligned}$$

we see that

$$\begin{aligned} P\left(U_{(k,n)} \in (x, x+\Delta]\right) &= \sum_{i=1}^n P\left(U_{(k,n)} \in (x, x+\Delta], \Omega_i\right) + O\left(\Delta^2\right) \\ &= nP\left(U_{(k,n)} \in (x, x+\Delta], \Omega_1\right) + O\left(\Delta^2\right). \end{aligned}$$

Now on the set, Ω_1; $U_{(k,n)} \in (x, x+\Delta]$ iff there are exactly $k-1$ of $U_2, \dots, U_n$ in $[0, x]$ and $n-k$ of these in $[x+\Delta, 1]$. This leads to the conclusion that

$$P\left(U_{(k,n)} \in (x, x+\Delta]\right) = n\binom{n-1}{k-1} x^{k-1} (1-(x+\Delta))^{n-k} \Delta + O\left(\Delta^2\right)$$

and therefore,

$$\rho_n(x) = \lim_{\Delta \downarrow 0} \frac{P\left(U_{(k,n)} \in (x, x+\Delta]\right)}{\Delta} = \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} (1-x)^{n-k}.$$

By Stirling's formula,

$$\begin{aligned} &\frac{n!}{(k-1)! \cdot (n-k)!} \\ &\sim \frac{n^n e^{-n} \sqrt{2\pi n}}{(k-1)^{(k-1)} e^{-(k-1)} \sqrt{2\pi (k-1)} (n-k)^{(n-k)} e^{-(n-k)} \sqrt{2\pi (n-k)}} \\ &= \frac{\sqrt{n} e^{-1}}{\sqrt{2\pi}} \frac{1}{\left(\frac{k-1}{n}\right)^{(k-1)} \sqrt{\frac{k-1}{n}} \left(\frac{n-k}{n}\right)^{(n-k)} \sqrt{\frac{n-k}{n}}} \\ &= \frac{\sqrt{n} e^{-1}}{\sqrt{2\pi}} \frac{1}{\left(\frac{k-1}{n}\right)^{(k-1/2)} \left(1 - \frac{k}{n}\right)^{(n-k+1/2)}}. \end{aligned}$$

Since

$$\begin{aligned} \left(\frac{k-1}{n}\right)^{(k-1/2)} &= \left(\frac{k}{n}\right)^{(k-1/2)} \cdot \left(\frac{k-1}{k}\right)^{(k-1/2)} \\ &= \left(\frac{k}{n}\right)^{(k-1/2)} \cdot \left(1 - \frac{1}{k}\right)^{(k-1/2)} \\ &\sim e^{-1} \left(\frac{k}{n}\right)^{(k-1/2)} \end{aligned}$$

we arrive at

$$\frac{n!}{(k-1)! \cdot (n-k)!} \sim \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{1}{\left(\frac{k}{n}\right)^{(k-1/2)} \left(1 - \frac{k}{n}\right)^{(n-k+1/2)}}.$$

By the change of variables formula, with

$$x = \frac{u - k(n)/n}{\frac{\sqrt{k(n)}}{n}}$$

on noting the $du = \frac{\sqrt{k(n)}}{n} dx$, $x = -\sqrt{k(n)}$ at $u = 0$, and

$$\begin{aligned} x &= \frac{1 - k(n)/n}{\frac{\sqrt{k(n)}}{n}} = \frac{n - k(n)}{\sqrt{k(n)}} \\ &= \frac{n}{\sqrt{k(n)}} \left(1 - \frac{k(n)}{n}\right) = \sqrt{n} \sqrt{\frac{n}{k(n)}} \left(1 - \frac{k(n)}{n}\right) =: b_n, \end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[F\left(X_n\right)\right] &= \int_0^1 \rho_n\left(u\right) F\left(\frac{u - k\left(n\right)/n}{\frac{\sqrt{k(n)}}{n}}\right) du \\ &= \int_{-\sqrt{k(n)}}^{b_n} \frac{\sqrt{k\left(n\right)}}{n} \rho_n\left(\frac{\sqrt{k\left(n\right)}}{n} x + k\left(n\right)/n\right) F\left(x\right) du.\end{aligned}$$

Using this information, it is then shown in Resnick that

$$\frac{\sqrt{k\left(n\right)}}{n}\rho_n\left(\frac{\sqrt{k\left(n\right)}}{n}x + k\left(n\right)/n\right) \to \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

which upon an application of Scheffé's Lemma 21.8 completes the proof. ■

Remark 21.50. It is possible to understand the normalization constants in the definition of X_n by computing the mean and the variance of $U_{(n,k)}$. After some computations (see Chapter **??**), one arrives at

$$\begin{aligned}\mathbb{E}U_{(k,n)} &= \int_0^1 \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} \left(1-x\right)^{n-k} x dx \\ &= \frac{k}{n+1} \sim \frac{k}{n}, \\ \mathbb{E}U_{(k,n)}^2 &= \int_0^1 \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} \left(1-x\right)^{n-k} x^2 dx \\ &= \frac{\left(k+1\right)k}{\left(n+2\right)\left(n+1\right)} \text{ and} \\ \operatorname{Var}\left(U_{(k,n)}\right) &= \frac{\left(k+1\right)k}{\left(n+2\right)\left(n+1\right)} - \frac{k^2}{\left(n+1\right)^2} \\ &= \frac{k}{n+1}\left[\frac{k+1}{n+2} - \frac{k}{n+1}\right] \\ &= \frac{k}{n+1}\left[\frac{n-k+1}{\left(n+2\right)\left(n+1\right)}\right] \sim \frac{k}{n^2}.\end{aligned}$$

21.8 Compactness and tightness of measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$

Suppose that $\Lambda \subset \mathbb{R}$ is a dense set and F and $\tilde{F}$ are two right continuous functions. If $F = \tilde{F}$ on Λ, then $F = \tilde{F}$ on $\mathbb{R}$. Indeed, for $x \in \mathbb{R}$ we have

$$F\left(x\right) = \lim_{\Lambda \ni \lambda \downarrow x} F\left(\lambda\right) = \lim_{\Lambda \ni \lambda \downarrow x} \tilde{F}\left(\lambda\right) = \tilde{F}\left(x\right).$$

Lemma 21.51. *If $G : \Lambda \to \mathbb{R}$ is a non-decreasing function, then*

$$F\left(x\right) := G_+\left(x\right) := \inf\left\{G\left(\lambda\right) : x < \lambda \in \Lambda\right\} \tag{21.42}$$

is a non-decreasing right continuous function.

Proof. To show F is right continuous, let $x \in \mathbb{R}$ and $\lambda \in \Lambda$ such that $\lambda > x$. Then for any $y \in (x, \lambda)$,

$$F\left(x\right) \leq F\left(y\right) = G_+\left(y\right) \leq G\left(\lambda\right)$$

and therefore,

$$F\left(x\right) \leq F\left(x+\right) := \lim_{y \downarrow x} F\left(y\right) \leq G\left(\lambda\right).$$

Since $\lambda > x$ with $\lambda \in \Lambda$ is arbitrary, we may conclude, $F\left(x\right) \leq F\left(x+\right) \leq G_+\left(x\right) = F\left(x\right)$, i.e. $F\left(x+\right) = F\left(x\right)$. ■

Proposition 21.52. *Suppose that $\{F_n\}_{n=1}^\infty$ is a sequence of distribution functions and $\Lambda \subset \mathbb{R}$ is a dense set such that $G\left(\lambda\right) := \lim_{n\to\infty} F_n\left(\lambda\right) \in [0,1]$ exists for all $\lambda \in \Lambda$. If, for all $x \in \mathbb{R}$, we define $F = G_+$ as in Eq. (21.42), then $F_n\left(x\right) \to F\left(x\right)$ for all $x \in \mathcal{C}\left(F\right)$. (Note well; as we have already seen, it is possible that $F\left(\infty\right) < 1$ and $F\left(-\infty\right) > 0$ so that F need not be a distribution function for a measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.)*

Proof. Suppose that $x, y \in \mathbb{R}$ with $x < y$ and and $s, t \in \Lambda$ are chosen so that $x < s < y < t$. Then passing to the limit in the inequality,

$$F_n\left(s\right) \leq F_n\left(y\right) \leq F_n\left(t\right)$$

implies

$$F\left(x\right) = G_+\left(x\right) \leq G\left(s\right) \leq \liminf_{n\to\infty} F_n\left(y\right) \leq \limsup_{n\to\infty} F_n\left(y\right) \leq G\left(t\right).$$

Taking the infimum over $t \in \Lambda \cap (y, \infty)$ and then letting $x \in \mathbb{R}$ tend up to y, we may conclude

$$F\left(y-\right) \leq \liminf_{n\to\infty} F_n\left(y\right) \leq \limsup_{n\to\infty} F_n\left(y\right) \leq F\left(y\right) \text{ for all } y \in \mathbb{R}.$$

This completes the proof, since $F\left(y-\right) = F\left(y\right)$ for $y \in \mathcal{C}\left(F\right)$. ■

The next theorem deals with weak convergence of measures on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$. So as not have to introduce any new machinery, the reader should identify $\bar{\mathbb{R}}$ with $[-1,1] \subset \mathbb{R}$ via the map,

$$[-1,1] \ni x \to \tan\left(\frac{\pi}{2}x\right) \in \bar{\mathbb{R}}.$$

Hence a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ may be identified with a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ which is supported on $[-1,1]$. Using this identification, we see that a $-\infty$ should only be considered a point of continuity of a distribution function, $F : \bar{\mathbb{R}} \to [0,1]$ iff and only if $F(-\infty) = 0$. On the other hand, ∞ is always a point of continuity.

Theorem 21.53 (Helly's Selection Theorem). *Every sequence of probability measures, $\{\mu_n\}_{n=1}^\infty$, on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ has a sub-sequence which is weakly convergent to a probability measure, μ_0 on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$.*

Proof. Using the identification described above, rather than viewing μ_n as probability measures on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$, we may view them as probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ which are supported on $[-1,1]$, i.e. $\mu_n([-1,1]) = 1$. As usual, let

$$F_n(x) := \mu_n((-\infty, x]) = \mu_n((-\infty, x] \cap [-1,1]).$$

Since $\{F_n(x)\}_{n=1}^\infty \subset [0,1]$ and $[0,1]$ is compact, for each $x \in \mathbb{R}$ we may find a convergence subsequence of $\{F_n(x)\}_{n=1}^\infty$. Hence by Cantor's diagonalization argument we may find a subsequence, $\{G_k := F_{n_k}\}_{k=1}^\infty$ of the $\{F_n\}_{n=1}^\infty$ such that $G(x) := \lim_{k\to\infty} G_k(x)$ exists for all $x \in \Lambda := \mathbb{Q}$.

Letting $F(x) := G(x+)$ as in Eq. (21.42), it follows from Lemma 21.51 and Proposition 21.52 that $G_k = F_{n_k} \implies F_0$. Moreover, since $G_k(x) = 0$ for all $x \in \mathbb{Q} \cap (-\infty, -1)$ and $G_k(x) = 1$ for all $x \in \mathbb{Q} \cap [1, \infty)$. Therefore, $F_0(x) = 1$ for all $x \geq 1$ and $F_0(x) = 0$ for all $x < -1$ and the corresponding measure, μ_0 is supported on $[-1,1]$. Hence μ_0 may now be transferred back to a measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$. ∎

Example 21.54. Here are there simple examples showing that probabilities may indeed transfer to the points at $\pm\infty$; 1) $\delta_{-n} \implies \delta_{-\infty}$, 2) $\delta_n \implies \delta_\infty$ and 3) $\frac{1}{2}(\delta_n + \delta_{-n}) \implies \frac{1}{2}(\delta_\infty + \delta_{-\infty})$.

The next question we would like to address is when is the limiting measure, μ_0 on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ concentrated on $\mathbb{R}$. The following notion of tightness is the key to answering this question.

Definition 21.55. *A collection of probability measures, Γ, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is **tight** iff for every $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that*

$$\inf_{\mu \in \Gamma} \mu([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon. \tag{21.43}$$

*We further say that a collection of random variables, $\{X_\lambda : \lambda \in \Lambda\}$ is **tight** iff the collection probability measures, $\left\{P \circ X_\lambda^{-1} : \lambda \in \Lambda\right\}$ is tight. Equivalently put, $\{X_\lambda : \lambda \in \Lambda\}$ is tight iff*

$$\lim_{M\to\infty} \sup_{\lambda \in \Lambda} P(|X_\lambda| \geq M) = 0. \tag{21.44}$$

Observe that the definition of uniform integrability (see Definition 12.38) is considerably stronger than the notion of tightness. It is also worth observing that if $\alpha > 0$ and $C := \sup_{\lambda \in \Lambda} \mathbb{E}|X_\lambda|^\alpha < \infty$, then by Chebyschev's inequality,

$$\sup_\lambda P(|X_\lambda| \geq M) \leq \sup_\lambda \left[\frac{1}{M^\alpha} \mathbb{E}|X_\lambda|^\alpha\right] \leq \frac{C}{M^\alpha} \to 0 \text{ as } M \to \infty$$

and therefore $\{X_\lambda : \lambda \in \Lambda\}$ is tight.

Theorem 21.56. *Let $\Gamma := \{\mu_n\}_{n=1}^\infty$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then Γ is tight, iff every subsequently limit measure, μ_0, on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ is supported on $\mathbb{R}$. In particular if Γ is tight, there is a weakly convergent subsequence of Γ converging to a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. (This is greatly generalized in Prokhorov's Theorem 21.61 below.)*

Proof. Suppose that $\mu_{n_k} \implies \mu_0$ with μ_0 being a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$. As usual, let $F_0(x) := \mu_0([-\infty, x])$. If Γ is tight and $\varepsilon > 0$ is given, we may find $M_\varepsilon < \infty$ such that $M_\varepsilon, -M_\varepsilon \in \mathcal{C}(F_0)$ and $\mu_n([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon$ for all n. Hence it follows that

$$\mu_0([-M_\varepsilon, M_\varepsilon]) = \lim_{k\to\infty} \mu_{n_k}([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon$$

and by letting $\varepsilon \downarrow 0$ we conclude that $\mu_0(\mathbb{R}) = \lim_{\varepsilon \downarrow 0} \mu_0([-M_\varepsilon, M_\varepsilon]) = 1$.

Conversely, suppose there is a subsequence $\{\mu_{n_k}\}_{k=1}^\infty$ such that $\mu_{n_k} \implies \mu_0$ with μ_0 being a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ such that $\mu_0(\mathbb{R}) < 1$. In this case $\varepsilon_0 := \mu_0(\{-\infty, \infty\}) > 0$ and hence for all $M < \infty$ we have

$$\mu_0([-M, M]) \leq \mu_0\left(\bar{\mathbb{R}}\right) - \mu_0(\{-\infty, \infty\}) = 1 - \varepsilon_0.$$

By choosing M so that $-M$ and M are points of continuity of F_0, it then follows that

$$\lim_{k\to\infty} \mu_{n_k}([-M, M]) = \mu_0([-M, M]) \leq 1 - \varepsilon_0.$$

Therefore,

$$\inf_{n \in \mathbb{N}} \mu_n(([-M, M])) \leq 1 - \varepsilon_0 \text{ for all } M < \infty$$

and $\{\mu_n\}_{n=1}^\infty$ is not tight. ∎

21.9 Metric Space Extensions

The goal of this section is to extend the notions of weak convergence when $\mathbb{R}$ is replace by a metric space (S, ρ). Standard references for the material here are [5] and [40] – also see [17] and [33]. Throughout this section, (S, ρ) will be a metric space and $\mathcal{B}_S$ will be the Borel σ – algebra on S, i.e. the σ – algebra generated by the open subsets of S. Recall that $V \subset S$ is open if it is the union of open balls of the form

$$B(x,r) := \{y \in S : \rho(x,y) < r\}$$

where $x \in S$ and $r \geq 0$. It should be noted that if S is separable (i.e. contains a countable dense set, $\Lambda \subset S$), then every open set may be written as a union of balls with $x \in \Lambda$ and $r \in \mathbb{Q}$ and so in the separable case

$$\mathcal{B}_S = \sigma(B(x,r) : x \in \Lambda,\ r \in \mathbb{Q}) = \sigma(B(x,r) : x \in S,\ r \geq 0).$$

Let us now state the theorems of this section.

Definition 21.57. *Let (S,τ) be a topological space, $\mathcal{B} := \sigma(\tau)$ be the Borel σ – algebra, and μ be a probability measure on $(S,\mathcal{B})$. We say that $A \in \mathcal{B}$ is a* ***continuity set*** *for μ provided $\mu(\mathrm{bd}(A)) = 0$. Notice that this is equivalent to saying that $\mu(A^\circ) = \mu(A) = \mu(\bar{A})$.*

Theorem 21.58 (Skorohod Theorem). *Let (S,ρ) be a separable metric space and $\{\mu_n\}_{n=0}^\infty$ be probability measures on $(S,\mathcal{B}_S)$ such that $\lim_{n\to\infty}\mu_n(A) = \mu(A)$ for all $A \in \mathcal{B}$ such that $\mu(\mathrm{bd}(A)) = 0$.*[6] *Then there exists a probability space, $(\Omega,\mathcal{B},P)$ and measurable functions, $Y_n : \Omega \to S$, such that $\mu_n = P \circ Y_n^{-1}$ for all $n \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and $\lim_{n\to\infty} Y_n = Y$ a.s.*

Proposition 21.59 (The Portmanteau Theorem). *Suppose that S is a complete separable metric space and $\{\mu_n\} \cup \{\mu\}$ are probability measure on $(S, \mathcal{B} := \mathcal{B}_S)$. Then the following are equivalent:*

1. *$\mu_n \implies \mu$ as $n \to \infty$, i.e. $\mu_n(f) \to \mu(f)$ for all $f \in BC(S)$.*
2. *$\mu_n(f) \to \mu(f)$ for every $f \in BC(S)$ which is uniformly continuous.*
3. *$\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ for all $F \sqsubset S$.*
4. *$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ for all $G \subset_o S$.*
5. *$\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all $A \in \mathcal{B}$ such that $\mu(\mathrm{bd}(A)) = 0$.*

Definition 21.60. *Let S be a topological space. A collection of probability measures Λ on $(S,\mathcal{B}_S)$ is said to be* ***tight*** *if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \in \mathcal{B}_S$ such that $\mu(K_\varepsilon) \geq 1 - \varepsilon$ for all $\mu \in \Lambda$.*

Theorem 21.61 (Prokhorov's Theorem). *Suppose S is a separable metrizable space and $\Lambda = \{\mu_n\}_{n=1}^\infty$ is a tight sequence of probability measures on $\mathcal{B}_{S_1}$. Then there exists a subsequence $\{\mu_{n_k}\}_{k=1}^\infty$ which is weakly convergent to a probability measure μ on $\mathcal{B}_S$.*

Conversely, if we further assume that (S,ρ) is a complete and Λ is a sequentially compact subset of the probability measures on $(S,\mathcal{B}_S)$ with the weak topology, then Λ is tight. (The converse direction is not so important for us.)

For the next few exercises, let (S_1,ρ_1) and (S_2,ρ_2) be separable metric spaces and $\mathcal{B}_{S_1}$ and $\mathcal{B}_{S_2}$ be the Borel σ – algebras on S_1 and S_2 respectively. Further define a metric, ρ, on $S := S_1 \times S_2$ by

$$\rho((x_1,x_2),(y_1,y_2)) = \rho_1(x_1,y_1) \vee \rho_2(x_2,y_2)$$

and let $\mathcal{B}_{S_1\times S_2}$ be the Borel σ – algebra on $S_1 \times S_2$. For $i = 1,2$, let $\pi_i : S_1 \times S_2 \to S_i$ be the projection maps and recall that

$$\mathcal{B}_{S_1} \otimes \mathcal{B}_{S_2} = \sigma(\pi_1,\pi_2) = \sigma\left(\pi_1^{-1}(\mathcal{B}_{S_1}) \cup \pi_2^{-1}(\mathcal{B}_{S_2})\right).$$

Exercise 21.9 (Continuous Mapping Theorem). Let (S_1,ρ_1) and (S_2,ρ_2) be separable metric spaces and $\mathcal{B}_{S_1}$ and $\mathcal{B}_{S_2}$ be the Borel σ – algebras on S_1 and S_2 respectively. Let Further suppose that $\{\mu_n\} \cup \{\mu\}$ are probability measures on $(S_1,\mathcal{B}_{S_1})$ such that $\mu_n \implies \mu$. If $f : S_1 \to S_2$ is a Borel measurable function such that $\mu(\mathcal{D}(f)) = 0$ (see Notation 21.15), then $f_*\mu_n \implies f_*\mu$ where $f_*\mu := \mu \circ f^{-1}$.

Exercise 21.10. Prove the analogue of Lemma 6.25, namely show $\mathcal{B}_{S_1\times S_2} = \mathcal{B}_{S_1} \otimes \mathcal{B}_{S_2}$. Hint: you may find Exercise 6.6 helpful.

Exercise 21.11. Let (S_1,ρ_1) and (S_2,ρ_2) be separable metric spaces and $\mathcal{B}_{S_1}$ and $\mathcal{B}_{S_2}$ be the Borel σ – algebras on S_1 and S_2 respectively. Further suppose that $\{\mu_n\} \cup \{\mu\}$ and $\{\nu_n\} \cup \{\nu\}$ are probability measures on $(S_1,\mathcal{B}_{S_1})$ and $(S_2,\mathcal{B}_{S_2})$ respectively. If $\mu_n \implies \mu$ and $\nu_n \implies \nu$, then $\mu_n \otimes \nu_n \implies \mu \otimes \nu$.

Exercise 21.10 and 21.11 have obvious generalizations to finite product spaces. In particular, if $\left\{X_n^{(i)}\right\}_{n=0}^\infty$ are sequences of random variables for $1 \leq i \leq K$ such that for each n, $\left\{X_n^{(i)}\right\}_{i=1}^K$ are independent random variables with $X_n^{(i)} \implies X_0^{(i)}$ as $n \to \infty$ for each $1 \leq i \leq K$, then

$$\left(X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(K)}\right) \implies \left(X_0^{(1)}, X_0^{(2)}, \dots, X_0^{(K)}\right) \text{ as } n \to \infty.$$

These comments will be useful for Exercise 21.12 below.

Definition 21.62 (Convergence of finite dimensional distributions). *Let $\{X_n(t) : t \geq 0\}_{n=0}^\infty$ be a collection of random processes, $X_n(t) : \Omega \to \mathbb{R}$. We say that X_n converges to X_0 in finite dimensional distributions and write $X_n \overset{f.d.}{\implies} X_0$ provided for every finite subset $\Lambda := \{0 = t_0 < t_1 < t_2 < \dots < t_K\}$ of $\mathbb{R}_+$ we have*

$$(X_n(t_0), \dots, X_n(t_K)) \implies (X_0(t_0), \dots, X_0(t_K)) \text{ as } n \to \infty.$$

[6] In Proposition 21.59 below we will see that this assumption is equivalent to assuming $\mu_n \implies \mu$.

Exercise 21.12. Let $\{X_n\}_{n=1}^\infty$ be an i.i.d. sequence of random variables with zero mean and $\mathrm{Var}(X_n) = 1$. For $t \geq 0$, let $B_n(t) := \frac{1}{\sqrt{n}} S_{[nt]}$ where $[nt]$ is the nearest integer to nt less than or equal to nt and $S_m := \sum_{k \leq m} X_k$ where $S_0 = 0$ by definition. Show that $B_n \overset{\text{f.d.}}{\Longrightarrow} B$ where $\{B(t) : t \geq 0\}$ is a Brownian motion as defined in definition 17.21. You might use the following outline.

1. For any $0 \leq s < t < \infty$, explain why $B_n(t) - B_n(s) \implies N(0, (t-s))$.
2. Given $\Lambda := \{0 = t_0 < t_1 < t_2 < \dots < t_K\} \subset \mathbb{R}_+$ argue that $\{B_n(t_i) - B_n(t_{i-1})\}_{i=1}^K$ are independent and then show
$$\{B_n(t_i) - B_n(t_{i-1})\}_{i=1}^K \implies \{B(t_i) - B(t_{i-1})\}_{i=1}^K \text{ as } n \to \infty.$$
3. Now show that $\{B_n(t_i)\}_{i=1}^K \implies \{B(t_i)\}_{i=1}^K$ as $n \to \infty$.

The rest of this section is devoted to the proofs of these results. (These proofs may safely be skipped on first reading.)

21.9.1 A point set topology review

Before getting down to business let me recall a few basic point set topology results which we will need. Recall that if (S, τ) is a topological space that $\bar{A} \subset S$, the **closure of** A, is defined by

$$\bar{A} := \cap \{C : A \subset C \sqsubset S\} \text{ and } A^\circ := \cup \{V : \tau \ni V \subset A\}$$

and the **interior of** A is defined by

$$A^\circ = \cup \{V : \tau \ni V \subset A\}.$$

Thus $\bar{A}$ is the smallest closed set containing A and A° is the largest open set contained in A. The relationship between the interior and closure operations is;

$$\begin{aligned}(A^\circ)^c &= \cap \{V^c : \tau \ni V \subset A\} \\ &= \cap \{C : A^c \subset C \sqsubset S\} = \overline{A^c}.\end{aligned}$$

Finally recall that the **topological boundary** of a set $A \subset S$ is defined by $\mathrm{bd}(A) := \bar{A} \setminus A^\circ$ which may also be expressed as

$$\mathrm{bd}(A) = \bar{A} \cap (A^\circ)^c = \bar{A} \cap \overline{A^c} \qquad (= \mathrm{bd}(A^c)).$$

In the case of a metric space we may describe $\bar{A}$ and $\mathrm{bd}(A)$ as

$$\begin{aligned}\bar{A} &= \{x \in S : \exists \{x_n\} \subset A \ni x = \lim_{n\to\infty} x_n\} \text{ and} \\ \mathrm{bd}(A) &= \{x \in S : \exists \{x_n\} \subset A \text{ and } \{y_n\} \subset A^c \ni \lim_{n\to\infty} y_n = x = \lim_{n\to\infty} x_n\}.\end{aligned}$$

So the boundary of A consists of those points in S which are arbitrarily close to points inside of A and outside of A. In the metric space case of most interest, the next lemma is easily proved using this characterization.

Lemma 21.63. *For any subsets, A and B, of S we have* $\mathrm{bd}(A \cap B) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B)$, $\mathrm{bd}(A \setminus B) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B)$, *and* $\mathrm{bd}(A \cup B) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B)$.

Proof. We begin by observing that $A^\circ \cap B^\circ \subset A \cap B \subset \bar{A} \cap \bar{B}$ from which it follows that

$$A^\circ \cap B^\circ \subset [A \cap B]^\circ \subset A \cap B \subset \overline{A \cap B} \subset \bar{A} \cap \bar{B}$$

and hence,

$$\mathrm{bd}(A \cap B) \subset [\bar{A} \cap \bar{B}] \setminus [A^\circ \cap B^\circ].$$

Combining this inclusion with

$$\begin{aligned}[\bar{A} \cap \bar{B}] \setminus [A^\circ \cap B^\circ] &= [\bar{A} \cap \bar{B}] \cap [A^\circ \cap B^\circ]^c = [\bar{A} \cap \bar{B}] \cap [(A^\circ)^c \cup (B^\circ)^c] \\ &= [\bar{A} \cap \bar{B} \cap (A^\circ)^c] \cup [\bar{A} \cap \bar{B} \cap (B^\circ)^c] \\ &\subset [\bar{A} \cap (A^\circ)^c] \cup [\bar{B} \cap (B^\circ)^c] = \mathrm{bd}(A) \cup \mathrm{bd}(B)\end{aligned}$$

completes the proof of the first assertion. The second and third assertions are easy consequence of the first because;

$$\mathrm{bd}(A \setminus B) = \mathrm{bd}(A \cap B^c) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B^c) = \mathrm{bd}(A) \cup \mathrm{bd}(B)$$

and

$$\begin{aligned}\mathrm{bd}(A \cup B) &= \mathrm{bd}([A \cup B]^c) = \mathrm{bd}(A^c \cap B^c) \\ &\subset \mathrm{bd}(A^c) \cup \mathrm{bd}(B^c) = \mathrm{bd}(A) \cup \mathrm{bd}(B).\end{aligned}$$

■

21.9.2 Proof of Skorohod's Theorem 21.58

Lemma 21.64. *Let (S, ρ) be a separable metric space, $\mathcal{B}$ be the Borel σ – algebra on S, and μ be a probability measure on $\mathcal{B}$. Then for every $\varepsilon > 0$ there exists a countable partition, $\{B_n\}_{n=1}^\infty$, of S such that $B_n \in \mathcal{B}$, $\mathrm{diam}(B_n) \leq \varepsilon$ and B_n is a μ – continuity set (i.e. $\mu(\mathrm{bd}(B_n)) = 0$) for all n.*

Proof. For $x \in S$ and $r \geq 0$ let $S(x,r) := \{y \in S : \rho(x,y) = r\}$. For any finite subset, $\Gamma \subset [0, \infty)$, we have $\sum_{r \in \Gamma} S(x,r) \subset S$ and therefore,

$$\sum_{r \in \Gamma} \mu(S(x,r)) \leq \mu(S) = 1.$$

As $\Gamma \subset_f [0, \infty)$ was arbitrary we may conclude that $\sum_{r \geq 0} \mu(S(x,r)) \leq 1 < \infty$ and therefore the set $Q_x := \{r \geq 0 : \mu(S(x,r)) > 0\}$ is at most countable.

If $B(x,r) := \{y \in S : \rho(x,y) < r\}$ and $C(x,r) := \{y \in S : \rho(x,y) \le r\}$ are the open and closed r – balls about x respectively, we have $S(x,r) = C(x,r) \setminus B(x,r)$. As

$$\mathrm{bd}(B(x,r)) = \overline{B(x,r)} \setminus B(x,r) \subset C(x,r) \setminus B(x,r) = S(x,r),$$

it follows that $B(x,r)$ is a μ – continuity set for all $r \notin Q_x$. With these preparations in hand we are now ready to complete the proof.

Let $\{x_n\}_{n=1}^\infty$ be a countable dense subset of S and let $Q := \cup_{n=1}^\infty Q_{x_n}$ – a countable subset of $[0,\infty)$. Choose $r \in [0,\infty) \setminus Q$ such that $r \le \varepsilon/2$ and then define

$$B_n := B(x_n, r) \setminus [B(x_1, r) \cup \dots \cup B(x_{n-1}, r)].$$

It is clear that $\{B_n\}_{n=1}^\infty \subset \mathcal{B}$ is a partition of S with $\mathrm{diam}(B_n) \le 2r \le \varepsilon$. Moreover, we know that

$$\begin{aligned}\mathrm{bd}(B_n) &\subset \mathrm{bd}(B(x_n,r)) \cup \mathrm{bd}(B(x_1,r) \cup \dots \cup B(x_{n-1},r)) \\ &\subset \cup_{k=1}^n \mathrm{bd}(B(x_k,r)) \subset \cup_{k=1}^n S(x_k,r)\end{aligned}$$

and therefore as $r \notin Q$ we have

$$\mu(\mathrm{bd}(B_n)) \le \sum_{k=1}^n \mu(S(x_k,r)) = 0$$

so that B_n is a μ – continuity set for each $n \in \mathbb{N}$. ■

We are now ready to prove Skorohod's Theorem 21.58.

Proof. (of Skorohod's Theorem 21.58) We will be following the proof in Kallenberg [30, Theorem 4.30 on page 79.]. In this proof we will be using an auxiliary probability space $(\Omega_0, \mathcal{B}_0, P_0)$ which is sufficiently rich so as to support the collection of independent random variables needed in the proof.[7] The final probability space will then be given by $(\Omega, \mathcal{B}, P) = (\Omega_0 \times S, \mathcal{B}_0 \otimes \mathcal{B}_S, P_0 \otimes \mu)$ and the random variable Y will be defined by $Y(\omega, x) := x$ for all $(\omega, x) \in \Omega$. Let us now start the proof.

Given $p \in \mathbb{N}$, use Lemma 21.64 to construct a partition, $\{B_n\}_{n=1}^\infty$, of S such that $\mathrm{diam}(B_n) < 2^{-p}$ and $\mu(\mathrm{bd}(B_n)) = 0$ for all n. Choose m sufficiently large so that $\mu\left(\sum_{n=m+1}^\infty B_n\right) < 2^{-p}$ and let $B_0 := \sum_{n=m+1}^\infty B_n$ so that $\{B_k\}_{k=0}^m$ is a partition of S. Now define

$$\kappa := \sum_{k=0}^m k 1_{B_k}(Y) = \sum_{k=0}^m k 1_{Y \in B_k}$$

[7] An examination of the proof will show that Ω_0 can be taken to be $(0,1) \times S^{\mathbb{N}}$ equipped with a well chosen infinite product measure.

and let Θ be a random variable on Ω which is independent of Y and has the uniform distribution on $[0,1]$. For each $n \in \mathbb{N}$, the Prenatal Skorohod Theorem 21.28 implies there exists $\tilde{\kappa}_n : (0,1) \times \{0, \dots, m\} \to \{0, \dots, m\}$ such that $\tilde{\kappa}_n(\theta, k) = k$ when $\theta \le \mu_n(B_k)/\mu(B_k)$ and $\mathrm{Law}_{m \times \{\mu(B_k)\}_{k=0}^m}(\kappa_n) = \{\mu_n(B_k)\}_{k=0}^m$. Now let $\kappa_n := \tilde{\kappa}_n(\Theta, \kappa)$ so that $P(\kappa_n = k) = \mu_n(B_k)$ for all $n \in \mathbb{N}$ and $0 \le k \le m$ and $\kappa_n = k$ when $\Theta \le \mu_n(B_\kappa)/\mu(B_\kappa)$. Since $\mu(\mathrm{bd}(B_k)) = 0$ for all k it follows $\mu_n(B_k) \to \mu(B_k)$ for all $0 \le k \le m$ and therefore $\lim_{n\to\infty} \kappa_n = \kappa$, P – a.s.

Now choose ξ_n^k independent of everything such that $P\left(\xi_n^k \in A\right) = \mu_n(A|B_k)$ for all n and $0 \le k \le n$. Then define

$$Y_n^p := \xi_n^{\kappa_n(\theta,\kappa)} = \sum_{k=0}^m 1_{\kappa_n(\theta,\kappa)=k} \cdot \xi_n^k.$$

Notice that

$$\begin{aligned}P(Y_n^p \in A) &= \sum_{k=0}^m P\left(\xi_n^k \in A \;\&\; \kappa_n(\theta,\kappa) = k\right) \\ &= \sum_{k=0}^m \mu_n(A|B_k)\,\mu_n(B_k) = \sum_{k=0}^m \mu_n(A \cap B_k) = \mu_n(A),\end{aligned}$$

and

$$\left\{\rho(Y_n^p, Y) > 2^{-p}\right\} \subset \{Y \in B_0\} \cup \{\kappa \ne \kappa_n\}$$

so that

$$\begin{aligned}P\left(\cup_{n\ge N}\left\{\rho(Y_n^p, Y) > 2^{-p}\right\}\right) &\le P(Y \in B_0) + P(\cup_{n\ge N}\{\kappa_n \ne \kappa\}) \\ &< 2^{-p} + P(\cup_{n\ge N}\{\kappa_n \ne \kappa\}).\end{aligned}$$

Since $\kappa_n \to \kappa$ a.s. it follows that

$$0 = P(\kappa_n \ne \kappa \text{ i.o. } n) = \lim_{N\to\infty} P(\cup_{n\ge N}\{\kappa_n \ne \kappa\})$$

and so there exists $n_p < \infty$ such that

$$P\left(\cup_{n\ge n_p}\left\{\rho(Y_n^p, Y) > 2^{-p}\right\}\right) < 2^{-p}.$$

To finish the proof, construct $\{Y_n^p\}_{n=1}^\infty$ and $n_p \in \mathbb{N}$ as above for each $p \in \mathbb{N}$. By replacing n_p by $\sum_{i=1}^p n_i$ if necessary, we may assume that $n_1 < n_2 < n_3 < \dots$. As

$$\sum_p P\left(\cup_{n\ge n_p}\left\{\rho(Y_n^p, Y) > 2^{-p}\right\}\right) < \sum_p 2^{-p} < \infty$$

it follows from the first Borel Cantelli lemma that $P(N) = 0$ where

$$N := \left\{ \left[\cup_{n \geq n_p} \left\{ \rho\left(Y_n^p, Y\right) > 2^{-p} \right\} \right] \text{ i.o. } p \right\}.$$

So off the null set N we have $\rho\left(Y_n^p, Y\right) \leq 2^{-p}$ for all $n \geq n_p$ and a.a. p. We now define $\{Y_n\}_{n=1}^\infty$ by

$$Y_n := Y_n^p \text{ for } n_p \leq n < n_{p+1} \text{ and } p \in \mathbb{N}.$$

Then by construction we have $\text{Law}\,(Y_n) = \mu_n$ for all n and $\rho\left(Y_n, Y\right) \to 0$ a.s. ■

21.9.3 Proof of Proposition – The Portmanteau Theorem 21.59

Proof. (of Proposition 21.59.) 1. $\implies$ 2. is obvious.

For 2. $\implies$ 3., let

$$\varphi(t) := \begin{cases} 1 & \text{if} \quad t \leq 0 \\ 1-t & \text{if } 0 \leq t \leq 1 \\ 0 & \text{if} \quad t \geq 1 \end{cases} \tag{21.45}$$

and let $f_n(x) := \varphi(n\rho(x, F))$. Then $f_n \in BC(S, [0,1])$ is uniformly continuous, $0 \leq 1_F \leq f_n$ for all n and $f_n \downarrow 1_F$ as $n \to \infty$. Passing to the limit $n \to \infty$ in the equation

$$0 \leq \mu_n(F) \leq \mu_n(f_m)$$

gives

$$0 \leq \limsup_{n\to\infty} \mu_n(F) \leq \mu(f_m)$$

and then letting $m \to \infty$ in this inequality implies item 3.

3. $\iff$ 4. Assuming item 3., let $F = G^c$, then

$$\begin{aligned} 1 - \liminf_{n\to\infty} \mu_n(G) &= \limsup_{n\to\infty}(1 - \mu_n(G)) = \limsup_{n\to\infty} \mu_n(G^c) \\ &\leq \mu(G^c) = 1 - \mu(G) \end{aligned}$$

which implies 4. Similarly 4. $\implies$ 3.

3. $\iff$ 5. Recall that $\text{bd}\,(A) = \bar{A} \setminus A^o$, so if $\mu(\text{bd}\,(A)) = 0$ and 3. (and hence also 4. holds) we have

$$\begin{aligned} \limsup_{n\to\infty} \mu_n(A) &\leq \limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(A) \text{ and} \\ \liminf_{n\to\infty} \mu_n(A) &\geq \liminf_{n\to\infty} \mu_n(A^o) \geq \mu(A^o) = \mu(A) \end{aligned}$$

from which it follows that $\lim_{n\to\infty} \mu_n(A) = \mu(A)$. Conversely, let $F \sqsubset S$ and set $F_\delta := \{x \in S : \rho(x, F) \leq \delta\}$.[8] Then

$$\text{bd}\,(F_\delta) \subset F_\delta \setminus \{x \in S : \rho(x, F) < \delta\} = A_\delta$$

where $A_\delta := \{x \in S : \rho(x, F) = \delta\}$. Since $\{A_\delta\}_{\delta > 0}$ are all disjoint, we must have

$$\sum_{\delta > 0} \mu(A_\delta) \leq \mu(S) \leq 1$$

and in particular the set $\Lambda := \{\delta > 0 : \mu(A_\delta) > 0\}$ is at most countable. Let $\delta_n \notin \Lambda$ be chosen so that $\delta_n \downarrow 0$ as $n \to \infty$, then

$$\mu(F_{\delta_m}) = \lim_{n\to\infty} \mu_n(F_{\delta_m}) \geq \limsup_{n\to\infty} \mu_n(F).$$

Let $m \to \infty$ in this equation to conclude $\mu(F) \geq \limsup_{n\to\infty} \mu_n(F)$ as desired.

To finish the proof it suffices to show 5. $\implies$ 1. which is easily done using Skorohod's Theorem 21.58 just as was done in the proof of Theorem 21.29. For those not wanting to use Skorohod's theorem we also provide a direct proof that 3. $\implies$ 1.

Alternate finish to the proof (3. $\implies$ 1.). By an affine change of variables it suffices to consider $f \in C(S, (0,1))$ in which case we have

$$\sum_{i=1}^k \frac{(i-1)}{k} 1_{\left\{\frac{(i-1)}{k} \leq f < \frac{i}{k}\right\}} \leq f \leq \sum_{i=1}^k \frac{i}{k} 1_{\left\{\frac{(i-1)}{k} \leq f < \frac{i}{k}\right\}}. \tag{21.46}$$

Let $F_i := \left\{\frac{i}{k} \leq f\right\}$ and notice that $F_k = \emptyset$. Then for any probability μ,

$$\sum_{i=1}^k \frac{(i-1)}{k} \left[\mu(F_{i-1}) - \mu(F_i)\right] \leq \mu(f) \leq \sum_{i=1}^k \frac{i}{k} \left[\mu(F_{i-1}) - \mu(F_i)\right]. \tag{21.47}$$

Since

$$\begin{aligned} &\sum_{i=1}^k \frac{(i-1)}{k} \left[\mu(F_{i-1}) - \mu(F_i)\right] \\ &= \sum_{i=1}^k \frac{(i-1)}{k} \mu(F_{i-1}) - \sum_{i=1}^k \frac{(i-1)}{k} \mu(F_i) \\ &= \sum_{i=1}^{k-1} \frac{i}{k} \mu(F_i) - \sum_{i=1}^k \frac{i-1}{k} \mu(F_i) = \frac{1}{k} \sum_{i=1}^{k-1} \mu(F_i) \end{aligned}$$

and

[8] We let $\rho(x, F) := \inf\{\rho(x, y) : y \in F\}$ so that $\rho(x, F)$ is the distance of x from F. Recall that $\rho(\cdot, F) : S \to [0, \infty)$ is a continuous map.

$$\begin{aligned}\sum_{i=1}^{k} \frac{i}{k}\left[\mu(F_{i-1})-\mu(F_i)\right] &= \sum_{i=1}^{k} \frac{i-1}{k}\left[\mu(F_{i-1})-\mu(F_i)\right] + \sum_{i=1}^{k} \frac{1}{k}\left[\mu(F_{i-1})-\mu(F_i)\right] \\ &= \sum_{i=1}^{k-1} \mu(F_i) + \frac{1}{k},\end{aligned}$$

Eq. (21.47) becomes,

$$\frac{1}{k}\sum_{i=1}^{k-1} \mu(F_i) \le \mu(f) \le \frac{1}{k}\sum_{i=1}^{k-1} \mu(F_i) + 1/k.$$

Using this equation with $\mu = \mu_n$ and then with $\mu = \mu$ we find

$$\begin{aligned}\limsup_{n\to\infty} \mu_n(f) &\le \limsup_{n\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k-1} \mu_n(F_i) + 1/k\right] \\ &\le \frac{1}{k}\sum_{i=1}^{k-1} \mu(F_i) + 1/k \le \mu(f) + 1/k.\end{aligned}$$

Since k is arbitrary, $\limsup_{n\to\infty} \mu_n(f) \le \mu(f)$. Replacing f by $1-f$ in this inequality also gives $\liminf_{n\to\infty} \mu_n(f) \ge \mu(f)$ and hence we have shown $\lim_{n\to\infty} \mu_n(f) = \mu(f)$ as claimed. ■

21.9.4 Proof of Prokhorov's compactness Theorem 21.61

The following proof relies on results not proved in these notes up to this point. The missing results may be found by searching for "Riesz-Markov Theorem" in the notes at

http://www.math.ucsd.edu/~bdriver/240A-C-03-04/240_lecture_notes.htm.

Proof. (of Prokhorov's compactness Theorem 21.61) First suppose that S is compact. In this case $C(S)$ is a Banach space which is separable by the Stone – Weirstrass theorem, see Exercise **??** in the analysis notes. By the Riesz theorem, Corollary **??** of the analysis notes, we know that $C(S)^*$ is in one to one correspondence with the complex measures on $(S, \mathcal{B}_S)$. We have also seen that $C(S)^*$ is metrizable and the unit ball in $C(S)^*$ is weak - * compact, see Theorem **??** of the analysis notes. Hence there exists a subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$ which is weak -* convergent to a probability measure μ on S. Alternatively, use the cantor's diagonalization procedure on a countable dense set $\Gamma \subset C(S)$ so find $\{\mu_{n_k}\}_{k=1}^{\infty}$ such that $\Lambda(f) := \lim_{k\to\infty} \mu_{n_k}(f)$ exists for all $f \in \Gamma$. Then for $g \in C(S)$ and $f \in \Gamma$, we have

$$\begin{aligned}\left|\mu_{n_k}(g) - \mu_{n_l}(g)\right| &\le \left|\mu_{n_k}(g) - \mu_{n_k}(f)\right| + \left|\mu_{n_k}(f) - \mu_{n_l}(f)\right| \\ &\quad + \left|\mu_{n_l}(f) - \mu_{n_l}(g)\right| \\ &\le 2\left\|g-f\right\|_\infty + \left|\mu_{n_k}(f) - \mu_{n_l}(f)\right|\end{aligned}$$

which shows

$$\limsup_{n\to\infty} \left|\mu_{n_k}(g) - \mu_{n_l}(g)\right| \le 2\left\|g-f\right\|_\infty.$$

Letting $f \in \Lambda$ tend to g in $C(S)$ shows $\limsup_{n\to\infty} \left|\mu_{n_k}(g) - \mu_{n_l}(g)\right| = 0$ and hence $\Lambda(g) := \lim_{k\to\infty} \mu_{n_k}(g)$ for all $g \in C(S)$. It is now clear that $\Lambda(g) \ge 0$ for all $g \ge 0$ so that Λ is a positive linear functional on S and thus there is a probability measure μ such that $\Lambda(g) = \mu(g)$.

General case. By Theorem 9.64 we may assume that S is a subset of a compact metric space which we will denote by $\bar{S}$. We now extend μ_n to $\bar{S}$ by setting $\bar{\mu}_n(A) := \bar{\mu}_n(A \cap S)$ for all $A \in \mathcal{B}_{\bar{S}}$. By what we have just proved, there is a subsequence $\{\bar{\mu}'_k := \bar{\mu}_{n_k}\}_{k=1}^{\infty}$ such that $\bar{\mu}'_k$ converges weakly to a probability measure $\bar{\mu}$ on $\bar{S}$. The main thing we now have to prove is that "$\bar{\mu}(S) = 1$," this is where the tightness assumption is going to be used. Given $\varepsilon > 0$, let $K_\varepsilon \subset S$ be a compact set such that $\bar{\mu}_n(K_\varepsilon) \ge 1-\varepsilon$ for all n. Since K_ε is compact in S it is compact in $\bar{S}$ as well and in particular a closed subset of $\bar{S}$. Therefore by Proposition 21.59

$$\bar{\mu}(K_\varepsilon) \ge \limsup_{k\to\infty} \bar{\mu}'_k(K_\varepsilon) = 1-\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows with $S_0 := \cup_{n=1}^{\infty} K_{1/n}$ satisfies $\bar{\mu}(S_0) = 1$. Because $S_0 \in \mathcal{B}_S \cap \mathcal{B}_{\bar{S}}$, we may view $\bar{\mu}$ as a measure on $\mathcal{B}_S$ by letting $\mu(A) := \bar{\mu}(A \cap S_0)$ for all $A \in \mathcal{B}_S$. Given a closed subset $F \subset S$, choose $\tilde{F} \sqsubset \bar{S}$ such that $F = \tilde{F} \cap S$. Then

$$\limsup_{k\to\infty} \mu'_k(F) = \limsup_{k\to\infty} \bar{\mu}'_k(\tilde{F}) \le \bar{\mu}(\tilde{F}) = \bar{\mu}(\tilde{F} \cap S_0) = \mu(F),$$

which shows $\mu'_k \implies \mu$.

Converse direction. Suppose now that (S, ρ) is complete and Λ is a sequentially compact subset of the probability measures on $(S, \mathcal{B}_S)$. We first will prove if $\{G_n\}_{n=1}^{\infty}$ is a sequence of open subsets of S such that $G_n \uparrow S$, then

$$c := \sup_n \inf_{\mu\in\Lambda} \mu(G_n) = \lim_{n\to\infty} \inf_{\mu\in\Lambda} \mu(G_n) = 1.$$

Suppose for sake of contradiction that $c < 1$ and let $c' \in (c, 1)$. By our assumption we have $\inf_{\mu\in\Lambda} \mu(G_n) \le c$ for all n therefore there exists $\mu_n \in \Lambda$ such that $\mu_n(G_n) \le c'$ for all $n \in \mathbb{N}$. By passing to a subsequence of $\{n\}$ and corresponding subsequence $\{G'_n\}$ of the $\{G_n\}$, we may assume that $\nu_n := \mu_{k_n} \implies \mu$ for some probability measure μ on S and $\nu_n(G'_n) \leq c'$ for all n where $G'_n \uparrow S$ as $n \uparrow \infty$. For fixed $N \in \mathbb{N}$ we have $\nu_n(G'_N) \leq \nu_n(G'_n) \leq c'$ for $n \geq N$. Passing to the limit as $n \to \infty$ in these inequalities then implies

$$\mu(G'_N) \leq \liminf_{n\to\infty} \nu_n(G'_N) \leq c' < 1.$$

However this is absurd since $\mu(G'_N) \uparrow 1$ as $N \to \infty$ since μ is a probability measure on S and $G'_N \uparrow S$ as $N \uparrow \infty$.

We may now finish the proof as follows. Let $\varepsilon > 0$ be given and let $\{x_k\}_{k=1}^{\infty}$ be a countable dense subset of S. For each $m \in \mathbb{N}$ the open sets $G_n := \cup_{k=1}^{n} B\left(x_k, \frac{1}{m}\right) \uparrow S$ and so by the above claim there exists n_m such $V_m := G_{n_m}$ satisfies $\inf_k \mu_k(V_m) \geq 1 - \varepsilon 2^{-m}$. We now let $A := \cap_m V_m$ so that $\mu_k(A) \geq 1 - \varepsilon$ for all k. As A is totally bounded and S is complete, $K_\varepsilon := \bar{A}$ is the desired compact subset of S such that $\mu_k(K_\varepsilon) \geq 1 - \varepsilon$ for all k. ■

Characteristic Functions (Fourier Transform)

Notation 22.1 *Given a measure μ on a measurable space, $(\Omega, \mathcal{B})$ and a function, $f \in L^1(\mu)$, we will often write $\mu(f)$ for $\int_\Omega f d\mu$.*

Let us recall Definition 8.10 here.

Definition 22.2. *Given a probability measure, μ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, let*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^n} e^{i\lambda\cdot x} d\mu(x)$$

*be the **Fourier transform or characteristic function of** μ. If $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ is a random vector on some probability space $(\Omega, \mathcal{B}, P)$, then we let $f(\lambda) := f_X(\lambda) := \mathbb{E}\left[e^{i\lambda\cdot X}\right]$. Of course, if $\mu := P \circ X^{-1}$, then $f_X(\lambda) = \hat{\mu}(\lambda)$.*

From Corollary 8.11 that we know if μ and ν are two probability measures on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ such that $\hat{\mu} = \hat{\nu}$ then $\mu = \nu$ – i.e. the Fourier transform map is injective. In this chapter we are going to, among other things, characterize those functions which are characteristic functions and we will also construct an inversion formula.

22.1 Basic Properties of the Characteristic Function

Definition 22.3. *A function $f : \mathbb{R}^n \to \mathbb{C}$ is said to be **positive definite**, iff $f(-\lambda) = \overline{f(\lambda)}$ for all $\lambda \in \mathbb{R}^n$ and for all $m \in \mathbb{N}$, $\{\lambda_j\}_{j=1}^m \subset \mathbb{R}^n$ the matrix, $\left(\{f(\lambda_j - \lambda_k)\}_{j,.k=1}^m\right)$ is non-negative. More explicitly we require,*

$$\sum_{j,k=1}^m f(\lambda_j - \lambda_k)\xi_j\bar{\xi}_k \geq 0 \text{ for all } (\xi_1, \ldots, \xi_m) \in \mathbb{C}^m.$$

Notation 22.4 *For $l \in \mathbb{N}\cup\{0\}$, let $C^l(\mathbb{R}^n, \mathbb{C})$ denote the vector space of functions, $f : \mathbb{R}^n \to \mathbb{C}$ which are l - time continuously differentiable. More explicitly, if $\partial_j := \frac{\partial}{\partial x_j}$, then $f \in C^l(\mathbb{R}^n, \mathbb{C})$ iff the partial derivatives, $\partial_{j_1} \ldots \partial_{j_k} f$, exist and are continuous for $k = 1, 2, \ldots, l$ and all $j_1, \ldots, j_k \in \{1, 2, \ldots, n\}$.*

Proposition 22.5 (Basic Properties of $\hat{\mu}$). *Let μ and ν be two probability measures on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, then;*

1. *$\hat{\mu}(0) = 1$, and $|\hat{\mu}(\lambda)| \leq 1$ for all λ.*
2. *$\hat{\mu}(\lambda)$ is continuous.*
3. *$\overline{\hat{\mu}(\lambda)} = \hat{\mu}(-\lambda)$ for all $\lambda \in \mathbb{R}^n$ and in particular, $\hat{\mu}$ is real valued iff μ is symmetric, i.e. iff $\mu(-A) = \mu(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^n}$. (If $\mu = P \circ X^{-1}$ for some random vector X, then μ is symmetric iff $X \overset{d}{=} -X$.)*
4. *$\hat{\mu}$ is a positive definite function.*
   (Bochner's Theorem 22.43 below asserts that if f is a function satisfying properties of $\hat{\mu}$ in itmes 1 – 4 above, then $f = \hat{\mu}$ for some probability measure μ.)
5. *If $\int_{\mathbb{R}^n} \|x\|^l d\mu(x) < \infty$, then $\hat{\mu} \in C^l(\mathbb{R}^n, \mathbb{C})$ and*
   $$\partial_{j_1} \ldots \partial_{j_m}\hat{\mu}(\lambda) = \int_{\mathbb{R}^n} (ix_{j_1} \ldots ix_{j_m}) e^{i\lambda\cdot x} d\mu(x) \text{ for all } m \leq l.$$
6. *If X and Y are independent random vectors then*
   $$f_{X+Y}(\lambda) = f_X(\lambda) f_Y(\lambda) \text{ for all } \lambda \in \mathbb{R}^n.$$
   This may be alternatively expressed as
   $$\widehat{\mu * \nu}(\lambda) = \hat{\mu}(\lambda)\hat{\nu}(\lambda) \text{ for all } \lambda \in \mathbb{R}^n.$$
7. *If $a \in \mathbb{R}$, $b \in \mathbb{R}^n$, and $X : \Omega \to \mathbb{R}^n$ is a random vector, then*
   $$f_{aX+b}(\lambda) = e^{i\lambda\cdot b} f_X(a\lambda).$$

Proof. The proof of items 1., 2., 6., and 7. are elementary and will be left to the reader. It also easy to see that $\overline{\hat{\mu}(\lambda)} = \hat{\mu}(-\lambda)$ and $\hat{\mu}(\lambda) = \hat{\mu}(-\lambda)$ if μ is symmetric. Therefore if μ is symmetric, then $\hat{\mu}(\lambda)$ is real. Conversely if $\hat{\mu}(\lambda)$ is real then

$$\hat{\mu}(\lambda) = \hat{\mu}(-\lambda) = \int_{\mathbb{R}^n} e^{i\lambda\cdot x} d\nu(x) = \hat{\nu}(\lambda)$$

where $\nu(A) := \mu(-A)$. The uniqueness Corollary 8.11 then implies $\mu = \nu$, i.e. μ is symmetric. This proves item 3.

Item 5. follows by induction using Corollary 7.30. For item 4. let $m \in \mathbb{N}$, $\{\lambda_j\}_{j=1}^m \subset \mathbb{R}^n$ and $(\xi_1, \dots, \xi_m) \in \mathbb{C}^m$. Then

$$\begin{aligned}\sum_{j,k=1}^m \hat{\mu}(\lambda_j - \lambda_k)\xi_j\bar{\xi}_k &= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i(\lambda_j-\lambda_k)\cdot x}\xi_j\bar{\xi}_k d\mu(x) \\ &= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i\lambda_j\cdot x}\xi_j\overline{e^{i\lambda_k\cdot x}\xi_k}d\mu(x) \\ &= \int_{\mathbb{R}^n} \left|\sum_{j=1}^m e^{i\lambda_j\cdot x}\xi_j\right|^2 d\mu(x) \geq 0.\end{aligned}$$

■

Example 22.6 (Example 21.3 continued.). Let $d\mu(x) = 1_{[0,1]}(x)\,dx$ and $\nu(A) = \mu(-A)$. Then

$$\begin{aligned}\hat{\mu}(\lambda) &= \int_0^1 e^{i\lambda x}dx = \frac{e^{i\lambda}-1}{i\lambda}, \\ \hat{\nu}(\lambda) &= \hat{\mu}(-\lambda) = \overline{\hat{\mu}(\lambda)} = \frac{e^{-i\lambda}-1}{-i\lambda}, \text{ and} \\ \widehat{\mu * \nu}(\lambda) &= \hat{\mu}(\lambda)\hat{\nu}(\lambda) = |\hat{\mu}(\lambda)|^2 = \left|\frac{e^{i\lambda}-1}{i\lambda}\right|^2 = \frac{2}{\lambda^2}[1-\cos\lambda].\end{aligned}$$

According to example 21.3 we also have $d(\mu * \nu)(x) = (1-|x|)_+\,dx$ and so directly we find

$$\begin{aligned}\widehat{\mu * \nu}(\lambda) &= \int_{\mathbb{R}} e^{i\lambda x}(1-|x|)_+\,dx = \int_{\mathbb{R}} \cos(\lambda x)(1-|x|)_+\,dx \\ &= 2\int_0^1 (1-x)\cos\lambda x\ dx = 2\int_0^1 (1-x)\,d\frac{\sin\lambda x}{\lambda} \\ &= -2\int_0^1 d(1-x)\frac{\sin\lambda x}{\lambda} = 2\int_0^1 \frac{\sin\lambda x}{\lambda}dx = 2\frac{-\cos\lambda x}{\lambda^2}\Big|_{x=0}^{x=1} \\ &= 2\frac{1-\cos\lambda}{\lambda^2}.\end{aligned}$$

For the most part we are now going to stick to the one dimensional case, i.e. X will be a random variable and μ will be a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The following Lemma is a special case of item 4. of Proposition 22.5.

Lemma 22.7. *Suppose $n \in \mathbb{N}$ and X is random variables such that $\mathbb{E}[|X|^n] < \infty$. If $\mu = P \circ X^{-1}$ is the distribution of X, then $\hat{\mu}(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ is C^n – differentiable and*

$$\hat{\mu}^{(l)}(\lambda) = \mathbb{E}\left[(iX)^l e^{i\lambda X}\right] = \int_{\mathbb{R}} (ix)^l e^{i\lambda x}d\mu(x) \text{ for } l = 0, 1, 2, \dots, n.$$

In particular it follows that

$$\mathbb{E}\left[X^l\right] = \frac{\hat{\mu}^{(l)}(0)}{i^l}.$$

The following theorem is a partial converse to this lemma. Hence the combination of Lemma 22.7 and Theorem 22.8 (see also Corollary 22.35 below) shows that there is a correspondence between the number of moments of X and the differentiability of f_X.

Theorem 22.8. *Let X be a random variable, $m \in \{0, 1, 2, \dots\}$, $f(\lambda) = \mathbb{E}\left[e^{i\lambda X}\right]$. If $f \in C^{2m}(\mathbb{R}, \mathbb{C})$ such that $g := f^{(2m)}$ is differentiable in a neighborhood of 0 and $g''(0) = f^{(2m+2)}(0)$ exists. Then $\mathbb{E}\left[X^{2m+2}\right] < \infty$ and $f \in C^{2m+2}(\mathbb{R}, \mathbb{C})$.*

Proof. This will be proved by induction on m. Let $m \in \mathbb{N}_0$ be given and suppose that

$$u(\lambda) = \mathbb{E}\left[X^{2m}\cos(\lambda X)\right] = \operatorname{Re}\mathbb{E}\left[X^{2m}e^{i\lambda X}\right]$$

is differentiable in a neighborhood of 0 and further suppose that $u''(0)$ exists. Since u is an even function of λ, u' is an odd function of λ near 0 and therefore $u'(0) = 0$. By the mean value theorem, to each $\lambda > 0$ with λ near 0, there exists $0 < c_\lambda < \lambda$ such that

$$\frac{u(\lambda) - u(0)}{\lambda} = u'(c_\lambda) = u'(c_\lambda) - u'(0)$$

and so

$$\frac{u(0) - u(\lambda)}{\lambda c_\lambda} = -\frac{u'(c_\lambda) - u'(0)}{c_\lambda} \to -u''(0) \text{ as } \lambda \downarrow 0. \tag{22.1}$$

Using Eq. (22.1) along with Fatou's lemma, we may pass to the limit as $\lambda \downarrow 0$ in the inequality,

$$\mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda^2}\right] \leq \mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda c_\lambda}\right] = \frac{u(0) - u(\lambda)}{\lambda c_\lambda},$$

to find

$$\frac{1}{2}\mathbb{E}\left[X^{2m+2}\right] = \liminf_{\lambda\downarrow 0}\mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda^2}\right] \leq \liminf_{\lambda\downarrow 0}\frac{u(0) - u(\lambda)}{\lambda c_\lambda} = -u''(0) < \infty.$$

With this result in hand the theorem is now easily proved by induction. We start with $m = 0$ and recall from Proposition 22.5 that $f \in C(\mathbb{R}, \mathbb{C})$). Assuming

f is differentiable in a neighborhood of 0 and $f''(0)$ exists we may apply the above result with $u = \operatorname{Re} f$ in order to learn $\mathbb{E}\left[X^2\right] < \infty$. An application of Lemma 22.7 then implies that $f \in C^2(\mathbb{R}, \mathbb{C})$. The induction step is handled in much the same way upon noting,

$$f^{(2m)}(\lambda) = (-1)^m \mathbb{E}\left[X^{2m} e^{i\lambda X}\right]$$

so that

$$u(\lambda) := (-1)^m \operatorname{Re} f^{(2m)}(\lambda) = \mathbb{E}\left[X^{2m} \cos(\lambda X)\right].$$

■

Corollary 22.9. *Suppose that X is a $\mathbb{R}^d$ – valued random vector such that for all $\lambda \in \mathbb{R}^d$ the function*

$$f_\lambda(t) := f_X(t\lambda) = \mathbb{E}\left[e^{it\lambda \cdot X}\right]$$

is $2m$ – times differentiable in a neighborhood of $t = 0$, then $\mathbb{E}\|X\|^{2m} < \infty$ and $f_X \in C^{2m}\left(\mathbb{R}^d, \mathbb{C}\right)$.

Proof. Applying Theorem 22.8 with X replaced by $\lambda \cdot X$ shows that $\mathbb{E}\left[|\lambda \cdot X|^{2m}\right] < \infty$ for all $\lambda \in \mathbb{R}^d$. In particular, taking $\lambda = e_i$ (the i^{th} – standard basis vector) implies that $\mathbb{E}\left[|X_i|^{2m}\right] < \infty$ for $1 \le i \le d$. So by Minikowski's inequality;

$$\|X\|_{L^{2m}(P)} = \left\|\sum_{i=1}^d X_i e_i\right\|_{L^{2m}(P)} \le \left\|\sum_{i=1}^d |X_i| \, \|e_i\|\right\|_{L^{2m}(P)} \le \sum_{i=1}^d \|e_i\| \cdot \|X_i\|_{L^{2m}(P)} < \infty,$$

i.e. $\mathbb{E}\|X\|^{2m} < \infty$. The fact that $f_X \in C^{2m}\left(\mathbb{R}^d, \mathbb{C}\right)$ now follows from Proposition 22.5. ■

22.2 Examples

Example 22.10. If $-\infty < a < b < \infty$ and $d\mu(x) = \frac{1}{b-a} 1_{[a,b]}(x)\, dx$ then

$$\hat{\mu}(\lambda) = \frac{1}{b-a} \int_a^b e^{i\lambda x} dx = \frac{e^{i\lambda b} - e^{i\lambda a}}{i\lambda(b-a)}.$$

If $a = -c$ and $b = c$ with $c > 0$, then

$$\hat{\mu}(\lambda) = \frac{\sin \lambda c}{\lambda c}.$$

Observe that

$$\hat{\mu}(\lambda) = 1 - \frac{1}{3!}\lambda^2 c^2 + \dots$$

and therefore, $\hat{\mu}'(0) = 0$ and $\hat{\mu}''(0) = -\frac{1}{3}c^2$ and hence it follows that

$$\int_{\mathbb{R}} x d\mu(x) = 0 \text{ and } \int_{\mathbb{R}} x^2 d\mu(x) = \frac{1}{3}c^2.$$

Example 22.11. Suppose Z is a Poisson random variable with mean $a > 0$, i.e. $P(Z = n) = e^{-a}\frac{a^n}{n!}$. Then

$$f_Z(\lambda) = \mathbb{E}\left[e^{i\lambda Z}\right] = e^{-a} \sum_{n=0}^\infty e^{i\lambda n} \frac{a^n}{n!} = e^{-a} \sum_{n=0}^\infty \frac{\left(ae^{i\lambda}\right)^n}{n!} = \exp\left(a\left(e^{i\lambda} - 1\right)\right).$$

Differentiating this result gives,

$$f_Z'(\lambda) = iae^{i\lambda} \exp\left(a\left(e^{i\lambda} - 1\right)\right) \text{ and}$$
$$f_Z''(\lambda) = \left(-a^2 e^{i2\lambda} - ae^{i\lambda}\right) \exp\left(a\left(e^{i\lambda} - 1\right)\right)$$

from which we conclude,

$$\mathbb{E}Z = \frac{1}{i} f_Z'(0) = a \text{ and } \mathbb{E}Z^2 = -f_Z''(0) = a^2 + a.$$

Therefore, $\mathbb{E}Z = a = \operatorname{Var}(Z)$.

Example 22.12. Suppose $T \overset{d}{=} \exp(a)$, i.e. $T \ge 0$ a.s. and $P(T \ge t) = e^{-at}$ for all $t \ge 0$. Recall that $\mu = \operatorname{Law}(T)$ is given by

$$d\mu(t) = F_T'(t)\, dt = ae^{-at} 1_{t \ge 0} dt.$$

Therefore,

$$\mathbb{E}\left[e^{iaT}\right] = \int_0^\infty ae^{-at} e^{i\lambda t} dt = \frac{a}{a - i\lambda} = \hat{\mu}(\lambda).$$

Since

$$\hat{\mu}'(\lambda) = i\frac{a}{(a - i\lambda)^2} \text{ and } \hat{\mu}''(\lambda) = -2\frac{a}{(a - i\lambda)^3}$$

it follows that

$$\mathbb{E}T = \frac{\hat{\mu}'(0)}{i} = a^{-1} \text{ and } \mathbb{E}T^2 = \frac{\hat{\mu}''(0)}{i^2} = \frac{2}{a^2}$$

and hence $\operatorname{Var}(T) = \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 = a^{-2}$.

Example 22.13. From Exercise 7.15, if $d\mu(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, then $\hat{\mu}(\lambda) = e^{-\lambda^2/2}$ and we may deduce

$$\int_{\mathbb{R}} x d\mu(x) = 0 \text{ and } \int_{\mathbb{R}} x^2 d\mu(x) = 1.$$

Recall from Section 9.8 that we have defined a random vector, $X \in \mathbb{R}^d$, to be Gaussian iff

$$\mathbb{E}\left[e^{i\lambda \cdot X}\right] = \exp\left(-\frac{1}{2}\operatorname{Var}(\lambda \cdot X) + i\mathbb{E}(\lambda \cdot X)\right).$$

We define a probability measure, μ, on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$ to be **Gaussian** iff there is a Gaussian random vector, X, such that $\operatorname{Law}(X) = \mu$. This can be expressed directly in terms of μ as; μ is **Gaussian** iff

$$\hat{\mu}(\lambda) = \exp\left(-\frac{1}{2} q(\lambda, \lambda) + i\lambda \cdot m\right) \text{ for all } \lambda \in \mathbb{R}^d$$

where

$$m := \int_{\mathbb{R}^d} x d\mu(x) \text{ and } q(\lambda, \lambda) := \int_{\mathbb{R}^d} (\lambda \cdot x)^2 d\mu(x) - (\lambda \cdot m)^2.$$

Example 22.14. If μ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $n \in \mathbb{N}$, then $\hat{\mu}^n$ is the characteristic function of the probability measure, namely the measure

$$\mu^{*n} := \overbrace{\mu * \cdots * \mu}^{n \text{ times}}. \tag{22.2}$$

Alternatively put, if $\{X_k\}_{k=1}^n$ are i.i.d. random variables with $\mu = P \circ X_k^{-1}$, then

$$f_{X_1 + \cdots + X_n}(\lambda) = f_{X_1}^n(\lambda).$$

Example 22.15. Suppose that $\{\mu_n\}_{n=0}^\infty$ are probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $\{p_n\}_{n=0}^\infty \subset [0,1]$ such that $\sum_{n=0}^\infty p_n = 1$. Then $\sum_{n=0}^\infty p_n \hat{\mu}_n$ is the characteristic function of the probability measure,

$$\mu := \sum_{n=0}^\infty p_n \mu_n.$$

Here is a more interesting interpretation of μ. Let $\{X_n\}_{n=0}^\infty \cup \{T\}$ be independent random variables with $P \circ X_n^{-1} = \mu_n$ and $P(T = n) = p_n$ for all $n \in \mathbb{N}_0$. Then $\mu(A) = P(X_T \in A)$, where $X_T(\omega) := X_{T(\omega)}(\omega)$. Indeed,

$$\begin{aligned}
\mu(A) = P(X_T \in A) &= \sum_{n=0}^\infty P(X_T \in A, T = n) = \sum_{n=0}^\infty P(X_n \in A, T = n) \\
&= \sum_{n=0}^\infty P(X_n \in A, T = n) = \sum_{n=0}^\infty p_n \mu_n(A).
\end{aligned}$$

Let us also observe that

$$\begin{aligned}
\hat{\mu}(\lambda) = \mathbb{E}\left[e^{i\lambda X_T}\right] &= \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda X_T} : T = n\right] = \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda X_n} : T = n\right] \\
&= \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda X_n}\right] P(T = n) = \sum_{n=0}^\infty p_n \hat{\mu}_n(\lambda).
\end{aligned}$$

Example 22.16. If μ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ then $\sum_{n=0}^\infty p_n \hat{\mu}^n$ is the characteristic function of a probability measure, ν, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. In this case, $\nu = \sum_{n=0}^\infty p_n \mu^{*n}$ where μ^{*n} is defined in Eq. (22.2). As an explicit example, if $a > 0$ and $p_n = \frac{a^n}{n!} e^{-a}$, then

$$\sum_{n=0}^\infty p_n \hat{\mu}^n = \sum_{n=0}^\infty \frac{a^n}{n!} e^{-a} \hat{\mu}^n = e^{-a} e^{a\hat{\mu}} = e^{a(\hat{\mu}-1)}$$

is the characteristic function of a probability measure. In other words,

$$f_{X_T}(\lambda) = \mathbb{E}\left[e^{i\lambda X_T}\right] = \exp\left(a\left(f_{X_1}(\lambda) - 1\right)\right).$$

22.3 Continuity Theorem

Lemma 22.17 (Tail Estimate). *Let $X : (\Omega, \mathcal{B}, P) \to \mathbb{R}$ be a random variable and $f_X(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ be its characteristic function. Then for $a > 0$,*

$$P(|X| \geq a) \leq \frac{a}{2} \int_{-2/a}^{2/a} (1 - f_X(\lambda)) d\lambda = \frac{a}{2} \int_{-2/a}^{2/a} (1 - \operatorname{Re} f_X(\lambda)) d\lambda \tag{22.3}$$

Proof. Recall that the Fourier transform of the uniform distribution on $[-c, c]$ is $\frac{\sin \lambda c}{\lambda c}$ and hence

$$\frac{1}{2c} \int_{-c}^{c} f_X(\lambda) d\lambda = \frac{1}{2c} \int_{-c}^{c} \mathbb{E}\left[e^{i\lambda X}\right] d\lambda = \mathbb{E}\left[\frac{\sin cX}{cX}\right].$$

Therefore,

$$\frac{1}{2c} \int_{-c}^{c} (1 - f_X(\lambda)) d\lambda = 1 - \mathbb{E}\left[\frac{\sin cX}{cX}\right] = \mathbb{E}[Y_c] \tag{22.4}$$

where

$$Y_c := 1 - \frac{\sin cX}{cX}.$$

Notice that $Y_c \geq 0$ (see Eq. (22.49)) and moreover, $Y_c \geq 1/2$ if $|cX| \geq 2$ ($|\sin cX| / |cX| \leq |\sin cX|/2 \leq 1/2$ if $|cX| \geq 2$). Hence we may conclude

$$\mathbb{E}[Y_c] \geq \mathbb{E}[Y_c : |cX| \geq 2] \geq \mathbb{E}\left[\frac{1}{2} : |cX| \geq 2\right] = \frac{1}{2} P(|X| \geq 2/c).$$

Combining this estimate with Eq. (22.4) shows,

$$\frac{1}{2c}\int_{-c}^{c} (1 - f_X(\lambda))\, d\lambda \geq \frac{1}{2} P(|X| \geq 2/c).$$

Taking $c = 2/a$ in this estimate proves Eq. (22.3). ∎

Exercise 22.1. Suppose now $X : (\Omega, \mathcal{B}, P) \to \mathbb{R}^d$ is a random vector and $f_X(\lambda) := \mathbb{E}\left[e^{i\lambda \cdot X}\right]$ is its characteristic function. Show for $a > 0$,

$$\begin{aligned} P(|X|_\infty \geq a) \leq& 2\left(\frac{a}{4}\right)^d \int_{[-2/a,2/a]^d} (1 - f_X(\lambda))\, d\lambda \\ =& 2\left(\frac{a}{4}\right)^d \int_{[-2/a,2/a]^d} (1 - \operatorname{Re} f_X(\lambda))\, d\lambda \end{aligned} \tag{22.5}$$

where $|X|_\infty = \max_i |X_i|$ and $d\lambda = d\lambda_1, \dots, d\lambda_d$.

Solution to Exercise (22.1). Working as above, we have

$$\left(\frac{1}{2c}\right)^d \int_{[-c,c]^d} \left(1 - e^{i\lambda \cdot X}\right) d\lambda = 1 - \prod_{j=1}^{d} \frac{\sin cX_j}{cX_j} =: Y_c, \tag{22.6}$$

where as before, $Y_c \geq 0$ and $Y_c \geq 1/2$ if $c|X_j| \geq 2$ for some j, i.e. if $c|X|_\infty \geq 2$. Therefore taking expectations of Eq. (22.6) implies,

$$\begin{aligned} \left(\frac{1}{2c}\right)^d \int_{[-c,c]^d} (1 - f_X(\lambda))\, d\lambda &= \mathbb{E}[Y_c] \geq \mathbb{E}[Y_c : |X|_\infty \geq 2/c] \\ &\geq \mathbb{E}\left[\frac{1}{2} : |X|_\infty \geq 2/c\right] = \frac{1}{2} P(|X|_\infty \geq 2/c). \end{aligned}$$

Taking $c = 2/a$ in this expression implies Eq. (22.5).

Theorem 22.18 (Continuity Theorem). *Suppose that $\{\mu_n\}_{n=1}^\infty$ is a sequence of probability measure on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$ and suppose that $f(\lambda) := \lim_{n\to\infty} \hat{\mu}_n(\lambda)$ exists for all $\lambda \in \mathbb{R}^d$. If f is continuous at $\lambda = 0$, then f is the characteristic function of a unique probability measure, μ, on $\mathcal{B}_{\mathbb{R}^d}$ and $\mu_n \implies \mu$ as $n \to \infty$.*

Proof. I will give the proof when $d = 1$ and leave the straight forward extension to the d – dimensional case to the reader.

By the continuity of f at $\lambda = 0$, for ever $\varepsilon > 0$ we may choose a_ε sufficiently large so that

$$\frac{1}{2} a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} f(\lambda))\, d\lambda \leq \varepsilon/2.$$

According to Lemma 22.17 and the DCT,

$$\begin{aligned} \mu_n(\{x : |x| \geq a_\varepsilon\}) &\leq \frac{1}{2} a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} \hat{\mu}_n(\lambda))\, d\lambda \\ &\to \frac{1}{2} a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} f(\lambda))\, d\lambda \leq \varepsilon/2 \text{ as } n \to \infty. \end{aligned}$$

Hence $\mu_n(\{x : |x| \geq a_\varepsilon\}) \leq \varepsilon$ for all sufficiently large n, say $n \geq N$. By increasing a_ε if necessary we can assure that $\mu_n(\{x : |x| \geq a_\varepsilon\}) \leq \varepsilon$ for all n and hence $\Gamma := \{\mu_n\}_{n=1}^\infty$ is tight.

By Theorem 21.56, we may find a subsequence, $\{\mu_{n_k}\}_{k=1}^\infty$ and a probability measure μ on $\mathcal{B}_{\mathbb{R}}$ such that $\mu_{n_k} \implies \mu$ as $k \to \infty$. Since $x \to e^{i\lambda x}$ is a bounded and continuous function, it follows that

$$\hat{\mu}(\lambda) = \lim_{k\to\infty} \hat{\mu}_{n_k}(\lambda) = f(\lambda) \text{ for all } \lambda \in \mathbb{R},$$

that is f is the characteristic function of a probability measure, μ.

We now claim that $\mu_n \implies \mu$ as $n \to \infty$. If not, we could find a bounded continuous function, g, such that $\lim_{n\to\infty} \mu_n(g) \neq \mu(g)$ or equivalently, there would exists $\varepsilon > 0$ and a subsequence $\{\mu'_k := \mu_{n_k}\}$ such that

$$|\mu(g) - \mu'_k(g)| \geq \varepsilon \text{ for all } k \in \mathbb{N}.$$

However by Theorem 21.56 again, there is a further subsequence, $\mu''_l = \mu'_{k_l}$ of μ'_k such that $\mu''_l \implies \nu$ for some probability measure ν. Since $\hat{\nu}(\lambda) = \lim_{l\to\infty} \hat{\mu}''_l(\lambda) = f(\lambda) = \hat{\mu}(\lambda)$, it follows that $\mu = \nu$. This leads to a contradiction since,

$$\varepsilon \leq \lim_{l\to\infty} |\mu(g) - \mu''_l(g)| = |\mu(g) - \nu(g)| = 0.$$

∎

Remark 22.19. One could also use Proposition 22.40 and Bochner's Theorem 22.43 below to conclude; if $f(\lambda) := \lim_{n\to\infty} \hat{\mu}_n(\lambda)$ exists and is continuous at 0, then f is the characteristic function of a probability measure. Indeed, the condition of a function being positive definite is preserved under taking pointwise limits.

Corollary 22.20. *Suppose that* $\{X_n\}_{n=1}^\infty \cup \{X\}$ *are random vectors in* $\mathbb{R}^d$*, then* $X_n \implies X$ *iff* $\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\cdot X_n}\right] = \mathbb{E}\left[e^{i\lambda\cdot X}\right]$ *for all* $\lambda \in \mathbb{R}^d$.

Proof. Since $f(x) := e^{i\lambda\cdot x}$ is in $BC(\mathbb{R}^d)$ for all $\lambda \in \mathbb{R}^d$, if $X_n \implies X$ then $\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\cdot X_n}\right] = \mathbb{E}\left[e^{i\lambda\cdot X}\right]$. Conversely if $\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\cdot X_n}\right] = \mathbb{E}\left[e^{i\lambda\cdot X}\right]$ for all $\lambda \in \mathbb{R}^d$ and $\mu_n := \mathrm{Law}(X_n)$ and which is equivalent to $X_n \implies X$. ■

The proof of the next corollary is a straightforward consequence of Corollary 22.20 used for dimension d and dimension 1.

Corollary 22.21. *Suppose that* $\{X_n\}_{n=1}^\infty \cup \{X\}$ *are random vectors in* $\mathbb{R}^d$*, then* $X_n \implies X$ *iff* $\lambda\cdot X_n \implies \lambda\cdot X$ *for all* $\lambda \in \mathbb{R}^d$.

Lemma 22.22. *If* $\{\mu_n\}_{n=1}^\infty$ *is a tight sequence of probability measures on* $\mathbb{R}^d$*, then the corresponding characteristic functions,* $\{\hat{\mu}_n\}_{n=1}^\infty$*, are equicontinuous on* $\mathbb{R}^d$.

Proof. By the tightness of the $\{\mu_n\}_{n=1}^\infty$, given $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that $\mu_n\left(\mathbb{R}^d \setminus [-M_\varepsilon, M_\varepsilon]^n\right) \le \varepsilon$ for all n. Let $\lambda, h \in \mathbb{R}^d$, then

$$\begin{aligned}|\hat{\mu}_n(\lambda+h) - \hat{\mu}_n(\lambda)| &\le \int_{\mathbb{R}^d} \left|e^{ix\cdot(\lambda+h)} - e^{ix\cdot\lambda}\right| d\mu_n(x) \\ &= \int_{\mathbb{R}^d} \left|e^{ix\cdot h} - 1\right| d\mu_n(x) \\ &\le 2\varepsilon + \sup_{x\in[-M_\varepsilon, M_\varepsilon]^n} \left|e^{ix\cdot h} - 1\right|.\end{aligned}$$

Therefore it follows that

$$\limsup_{h\to 0} \sup_{\lambda\in\mathbb{R}^d} |\hat{\mu}_n(\lambda+h) - \hat{\mu}_n(\lambda)| \le 2\varepsilon$$

and as $\varepsilon > 0$ was arbitrary the result follows. ■

Corollary 22.23 (Uniform Convergence). *If* $\mu_n \implies \mu$ *as* $n\to\infty$ *then* $\hat{\mu}_n(\lambda) \to \hat{\mu}(\lambda)$ *uniformly on compact subsets of* $\mathbb{R}$ $(\mathbb{R}^n)$.

Proof. This is a consequence of Theorem 22.18, Lemma 22.22, and the Arzela - Ascoli Theorem 45.36. For completeness here is a sketch of the proof.

Let K be a compact subset of $\mathbb{R}$ $(\mathbb{R}^n)$ and $\varepsilon > 0$ be given. Applying Lemma 22.22 to $\{\mu\} \cup \{\hat{\mu}_n\}$ we know that there exists $\delta > 0$ such that

$$\sup_\lambda |\hat{\mu}_n(\lambda+h) - \hat{\mu}_n(\lambda)| \le \varepsilon \text{ and } \sup_\lambda |\hat{\mu}(\lambda+h) - \hat{\mu}(\lambda)| \le \varepsilon \tag{22.7}$$

whenever $\|h\| \le \delta$. Let $F \subset K$ be a finite set such that $K \subset \cup_{\xi\in F} B(\xi,\delta)$. Since we already know that $\hat{\mu}_n \to \hat{\mu}$ pointwise we will have

$$\lim_{n\to\infty} \max_{\xi\in F} |\hat{\mu}_n(\xi) - \hat{\mu}(\xi)| = 0.$$

Since every point $\lambda \in K$ is within δ of a point in F we may use Eq. (22.7) to conclude that

$$\sup_{\lambda\in K} |\hat{\mu}_n(\lambda) - \hat{\mu}(\lambda)| \le 2\varepsilon + \max_{\xi\in F} |\hat{\mu}_n(\xi) - \hat{\mu}(\xi)|$$

and therefore, $\limsup_{n\to\infty} \sup_{\lambda\in K} |\hat{\mu}_n(\lambda) - \hat{\mu}(\lambda)| \le 2\varepsilon$. As $\varepsilon > 0$ was arbitrary the result follows. ■

The following lemma will be needed before giving our first applications of the continuity theorem.

Lemma 22.24. *Suppose that* $\{z_n\}_{n=1}^\infty \subset \mathbb{C}$ *satisfies,* $\lim_{n\to\infty} nz_n = \xi \in \mathbb{C}$*, then*

$$\lim_{n\to\infty} (1+z_n)^n = e^\xi.$$

Proof. Since $nz_n \to \xi$, it follows that $z_n \sim \frac{\xi}{n} \to 0$ as $n\to\infty$ and therefore by Lemma 22.45 below, $(1+z_n) = e^{\ln(1+z_n)}$ and

$$\ln(1+z_n) = z_n + O\left(z_n^2\right) = z_n + O\left(\frac{1}{n^2}\right).$$

Therefore,

$$(1+z_n)^n = \left[e^{\ln(1+z_n)}\right]^n = e^{n\ln(1+z_n)} = e^{n\left(z_n + O\left(\frac{1}{n^2}\right)\right)} \to e^\xi \text{ as } n\to\infty.$$

■

Proposition 22.25 (Weak Law of Large Numbers revisited). *Suppose that* $\{X_n\}_{n=1}^\infty$ *are i.i.d. integrable random variables. Then* $\frac{S_n}{n} \xrightarrow{P} \mathbb{E}X_1 =: c$.

Proof. Let $f(\lambda) := f_{X_1}(\lambda) = \mathbb{E}\left[e^{i\lambda X_1}\right]$ in which case

$$f_{\frac{S_n}{n}}(\lambda) = \left[f\left(\frac{\lambda}{n}\right)\right]^n.$$

By Taylor's theorem (see Appendix 22.7), $f(\lambda) = 1 + k(\lambda)\lambda$ where

$$\lim_{\lambda\to 0} k(\lambda) = k(0) = f'(0) = i\mathbb{E}[X_1].$$

It now follows from Lemma 22.24 that

$$f_{\frac{S_n}{n}}(\lambda) = \left[1 + k\left(\frac{\lambda}{n}\right)\frac{\lambda}{n}\right]^n \to e^{ic\lambda} \text{ as } n\to\infty$$

which is the characteristic function of the constant random variable, c. By the continuity Theorem 22.18, it follows that $\frac{S_n}{n} \implies c$ and since c is constant we may apply Lemma 21.25 to conclude $\frac{S_n}{n} \xrightarrow{P} c = \mathbb{E}X_1$. ■

We are now ready to continue are investigation of central limit theorems that was begun with Theorem 10.37 above.

Theorem 22.26 (The Basic Central Limit Theorem). *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. square integrable random variables such that $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 = 1$. Then $\frac{S_n}{\sqrt{n}} \implies N(0,1)$.*

Proof. If $f(\lambda) := \mathbb{E}\left[e^{i\lambda X_1}\right]$, then by Taylor's theorem (see Appendix 22.7),

$$f(\lambda) = f(0) + f'(0)\lambda + \frac{1}{2}k(\lambda)\lambda^2 = 1 + \frac{1}{2}k(\lambda)\lambda^2 \tag{22.8}$$

where

$$\lim_{\lambda\to 0} k(\lambda) = k(0) = f''(0) = -\mathbb{E}\left[X_1^2\right] = -1.$$

Hence, using Lemma 22.24, we find

$$\begin{aligned}\mathbb{E}\left[e^{i\lambda\frac{S_n}{\sqrt{n}}}\right] &= \left[f\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n \\ &= \left[1 + \frac{1}{2}k\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\lambda^2}{n}\right]^n \to e^{-\lambda^2/2}.\end{aligned}$$

Since $e^{-\lambda^2/2}$ is the characteristic function of $N(0,1)$ (Example 22.13), the result now follows from the continuity Theorem 22.18.

Alternative proof. Again it suffices to show

$$\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\frac{S_n}{\sqrt{n}}}\right] = e^{-\lambda^2/2} \text{ for all } \lambda \in \mathbb{R}.$$

We do this using Lemma 23.7 below as follows;

$$\begin{aligned}\left|f_{\frac{S_n}{\sqrt{n}}}(\lambda) - e^{-\lambda^2/2}\right| &= \left|\left[f\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n - \left[e^{-\lambda^2/2n}\right]^n\right| \\ &\le n\left|f\left(\frac{\lambda}{\sqrt{n}}\right) - e^{-\lambda^2/2n}\right| \\ &= n\left|1 - \frac{1}{2}\left(1 + \varepsilon\left(\frac{\lambda}{\sqrt{n}}\right)\right)\frac{\lambda^2}{n} - \left(1 - \frac{\lambda^2}{2n} + O\left(\frac{1}{n^2}\right)\right)\right| \\ &\to 0 \text{ as } n \to \infty.\end{aligned}$$

■

Corollary 22.27. *If $\{X_n\}_{n=1}^\infty$ are i.i.d. square integrable random variables such that $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 = 1$, then*

$$\sup_{\lambda\in\mathbb{R}}\left|P\left(\frac{S_n}{\sqrt{n}} \le y\right) - P(N(0,1) \le y)\right| \to 0 \text{ as } n \to \infty. \tag{22.9}$$

Proof. This is a direct consequence of Theorem 22.26 and Exercise 21.7. ■

Berry (1941) and Esseî (1942) showed there exists a constant, $C < \infty$, such that; if $\rho^3 := \mathbb{E}|X_1|^3 < \infty$, then

$$\sup_{\lambda\in\mathbb{R}}\left|P\left(\frac{S_n}{\sqrt{n}} \le y\right) - P(N(0,1) \le y)\right| \le C\left(\frac{\rho}{\sigma}\right)^3/\sqrt{n}.$$

In particular the rate of convergence is $n^{-1/2}$. The exact value of the best constant C is still unknown but it is known to be less than 1. We will not prove this theorem here. However we have seen a hint that such a result should be true in Theorem 10.37 above.

Remark 22.28 (Why normal?). It is now a reasonable question to ask "why" is the limiting random variable normal in Theorem 22.26. One way to understand this is, if under the assumptions of Theorem 22.26, we know $\frac{S_n}{\sqrt{n}} \implies L$ where L is some random variable with $\mathbb{E}L = 0$ and $\mathbb{E}L^2 = 1$, then

$$\begin{aligned}\frac{S_{2n}}{\sqrt{2n}} &= \frac{1}{\sqrt{2}}\left(\frac{\sum_{k=1,\ k \text{ odd}}^{2n} X_j}{\sqrt{n}} + \frac{\sum_{k=1,\ k \text{ even}}^{2n} X_j}{\sqrt{n}}\right) \\ &\implies \frac{1}{\sqrt{2}}(L_1 + L_2)\end{aligned} \tag{22.10}$$

where $L_1 \overset{d}{=} L \overset{d}{=} L_2$ and L_1 and L_2 are independent – see Exercise 21.11. In particular this implies that

$$f(\lambda) = \left[f\left(\frac{\lambda}{\sqrt{2}}\right)\right]^2 \text{ for all } \lambda \in \mathbb{R}. \tag{22.11}$$

We could also arrive at Eq. (22.11) by passing to the limit in the identity,

$$f_{\frac{S_{2n}}{\sqrt{2n}}}(\lambda) = f_{\frac{S_n}{\sqrt{n}}}\left(\frac{\lambda}{\sqrt{2}}\right) f_{\frac{S_n}{\sqrt{n}}}\left(\frac{\lambda}{\sqrt{2}}\right).$$

Iterating Eq. (22.11) and then Eq. (22.8) and Lemma 22.24 above we again deduce that,

$$\begin{aligned}f(\lambda) &= \left[f\left(\frac{\lambda}{(\sqrt{2})^n}\right)\right]^{2^n} \\ &= \left[1 + \frac{1}{2}k\left(\frac{\lambda}{2^{n/2}}\right)\frac{\lambda^2}{2^n}\right]^{2^n} \to e^{-\frac{1}{2}\lambda^2} = f_{N(0,1)}(\lambda).\end{aligned}$$

That is we must have $L \overset{d}{=} N(0,1)$. What we have proved is that if L is any square integrable random variable with zero mean and variance equal to one such that $L \overset{d}{=} \frac{1}{\sqrt{2}}(L_1 + L_2)$ where L_1 and L_2 are two independent copies of L, then $L \overset{d}{=} N(0,1)$.

Theorem 22.29 (The multi-dimensional Central Limit Theorem). *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. square integrable random vectors in $\mathbb{R}^d$ and let $m := \mathbb{E}X_1$ and $Q = \mathbb{E}\left[(X_1 - m)(X_1 - m)^{\mathrm{tr}}\right]$, that is $m \in \mathbb{R}^d$ and Q is the $d \times d$ matrix defined by*

$$\begin{aligned} m_j &:= \mathbb{E}(X_1)_j \text{ and} \\ Q_{ij} &:= \mathbb{E}\left[(X_1 - m)_i (X_1 - m)_j\right] = \operatorname{Cov}\left((X_1)_i, (X)_j\right) \end{aligned}$$

for all $1 \le i, j \le d$. Then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m) \implies Z \tag{22.12}$$

where $Z \overset{d}{=} N(0, Q)$, i.e. Z is a random vector such that

$$\mathbb{E}\left[e^{i\lambda \cdot Z}\right] = \exp\left(-\frac{1}{2} Q\lambda \cdot \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d.$$

Proof. Let $\lambda \in \mathbb{R}^d$, then

$$\lambda \cdot Z \overset{d}{=} N(0, Q\lambda \cdot \lambda) \overset{d}{=} \sqrt{Q\lambda \cdot \lambda} \cdot N(0,1)$$

and $\{\lambda \cdot X_k\}_{k=1}^\infty$ are i.i.d random variables with $\mathbb{E}[\lambda \cdot X_k] = \lambda \cdot m$ and $\operatorname{Var}(\lambda \cdot X_k) = Q\lambda \cdot \lambda$. If $Q\lambda \cdot \lambda = 0$ then $\lambda \cdot X_k = \lambda \cdot m$ a.s. and $\lambda \cdot Z = 0$ a.s. and we will have

$$\lambda \cdot \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m)\right) = 0 \implies 0 = \lambda \cdot Z. \tag{22.13}$$

If $Q\lambda \cdot \lambda > 0$ then $\left\{\frac{\lambda \cdot X_k - \lambda \cdot m}{\sqrt{Q\lambda \cdot \lambda}}\right\}_{k=1}^\infty$ satisfy the hypothesis of Theorem 22.26 and therefore,

$$\frac{1}{\sqrt{Q\lambda \cdot \lambda}} \lambda \cdot \left[\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m)\right] = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\lambda \cdot X_k - \lambda \cdot m}{\sqrt{Q\lambda \cdot \lambda}} \implies N(0,1)$$

which combined with Eq. (22.13) implies, for all $\lambda \in \mathbb{R}^d$ we have

$$\lambda \cdot \left[\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m)\right] \implies \sqrt{Q\lambda \cdot \lambda} \cdot N(0,1) \overset{d}{=} \lambda \cdot Z.$$

We may now apply Corollary 22.21 to conclude that Eq. (22.12) holds. ■

22.4 A Fourier Transform Inversion Formula

Corollary 8.11 guarantees the injectivity of the Fourier transform on the space of probability measures. Our next goal is to find an inversion formula for the Fourier transform. To motivate the construction below, let us first recall a few facts about Fourier series. To keep our exposition as simple as possible, we now restrict ourselves to the one dimensional case.

For $L > 0$, let $e_n^L(x) := e^{-i\frac{n}{L}x}$ and let

$$(f, g)_L := \frac{1}{2\pi L} \int_{-\pi L}^{\pi L} f(x) \bar{g}(x) \, dx$$

for $f, g \in L^2([-\pi L, \pi L], dx)$. Then it is well known (and fairly elementary to prove) that $\left\{e_n^L : n \in \mathbb{Z}\right\}$ is an orthonormal basis for $L^2([-\pi L, \pi L], dx)$. In particular, if $f \in C_c(\mathbb{R})$ with $\operatorname{supp}(f) \subset [-\pi L, \pi L]$, then for $x \in [-\pi L, \pi L]$,

$$\begin{aligned} f(x) &= \sum_{n \in \mathbb{Z}} \left(f, e_n^L\right)_L e_n^L(x) = \frac{1}{2\pi L} \sum_{n \in \mathbb{Z}} \left(\int_{-\pi L}^{\pi L} f(y) e^{i\frac{n}{L}y} dy\right) e^{-i\frac{n}{L}x} \\ &= \frac{1}{2\pi L} \sum_{n \in \mathbb{Z}} \hat{f}\left(\frac{n}{L}\right) e^{-i\frac{n}{L}x} \end{aligned} \tag{22.14}$$

where

$$\hat{f}(\lambda) = \int_{-\infty}^\infty f(y) e^{i\lambda y} dy.$$

Letting $L \to \infty$ in Eq. (22.14) then suggests that

$$\frac{1}{2\pi L} \sum_{n \in \mathbb{Z}} \hat{f}\left(\frac{n}{L}\right) e^{-i\frac{n}{L}x} \to \frac{1}{2\pi} \int_{-\infty}^\infty \hat{f}(\lambda) e^{-i\lambda x} d\lambda$$

and we are lead to expect,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^\infty \hat{f}(\lambda) e^{-i\lambda x} d\lambda. \tag{22.15}$$

Now suppose that $f(x) = \rho(x)$ where $\rho(x)$ is a probability density for a measure μ (i.e. $d\mu(x) := \rho(x) \, dx$) so that $\hat{\rho}(\lambda) = \hat{\mu}(\lambda)$. From Eq. (22.15) we expect that

$$\begin{aligned}\mu((a,b]) &= \int_a^b \rho(x)\,dx = \int_a^b \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mu}(\lambda)\,e^{-i\lambda x}d\lambda\right)dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mu}(\lambda)\left(\int_a^b e^{-i\lambda x}dx\right)d\lambda \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)d\lambda \\ &= \lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c} \hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)d\lambda. \qquad (22.16)\end{aligned}$$

We will prove this formula is essentially correct in Theorem 22.31 below. The following lemma is the key to computing the limit appearing in Eq. (22.16) which will be the heart of the proof of the inversion formula.

Lemma 22.30. *For $c>0$, let*

$$S(c) := \int_{-c}^{c} \frac{\sin\lambda}{\lambda}d\lambda. \qquad (22.17)$$

Then $S(c)$ is a continuous function such that $S(c)\to\pi$ boundedly as $c\to\infty$, see Figure 22.1. Moreover for any $y\in\mathbb{R}$ we have

$$\int_{-c}^{c} \frac{\sin\lambda y}{\lambda}d\lambda = \operatorname{sgn}(y)S(c|y|) \qquad (22.18)$$

where

$$\operatorname{sgn}(y) = \begin{cases} 1 & \text{if } y>0 \\ -1 & \text{if } y<0 \\ 0 & \text{if } y=0 \end{cases}.$$

Proof. The first assertion has already been dealt with in Example 9.11. We will repeat the argument here for the reader's convenience. By symmetry and Fubini's theorem,

$$\begin{aligned}S(c) &= 2\int_0^c \frac{\sin\lambda}{\lambda}d\lambda = 2\int_0^c \sin\lambda\cdot\left(\int_0^\infty e^{-\lambda t}dt\right)d\lambda \\ &= 2\int_0^\infty\left(\int_0^c \sin\lambda e^{-\lambda t}d\lambda\right)dt \\ &= 2\int_0^\infty\left(\frac{1}{1+t^2}\left[1-e^{-tc}(\cos c+t\sin c)\right]\right)dt \\ &= \pi - 2\int_0^\infty \frac{1}{1+t^2}e^{-tc}\left[\cos c+t\sin c\right]dt. \qquad (22.19)\end{aligned}$$

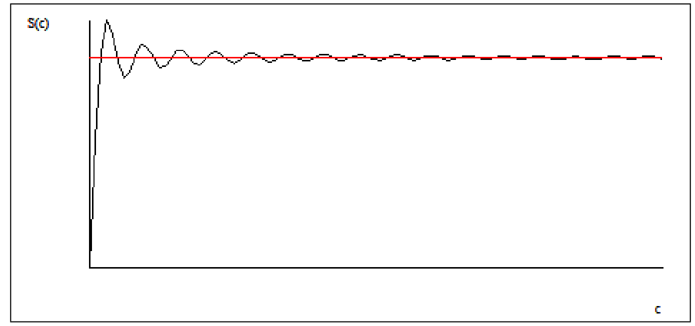


Fig. 22.1. The graph of $S(c)$ in black and π in red.

The the integral in Eq. (22.19) tends to 0 as $c\to\infty$ by the dominated convergence theorem. The second assertion in Eq. (22.18) is a consequence of the change of variables, $z=\lambda y$. ■

Theorem 22.31 (Fourier Inversion Formula). *If μ is a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ and $-\infty<a<b<\infty$, then*

$$\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)d\lambda = \mu((a,b)) + \frac{1}{2}\left(\mu(\{a\})+\mu(\{b\})\right). \qquad (22.20)$$

(At the end points, the limit picks up only half of the mass.)

Proof. Let $I(c)$ denote the integral appearing in Eq. (22.20). By Fubini's theorem and Lemma 22.30,

$$\begin{aligned}I(c) &:= \int_{-c}^{c}\hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)d\lambda \qquad (22.21) \\ &= \int_{-c}^{c}\left(\int_{\mathbb{R}} e^{i\lambda x}d\mu(x)\right)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)d\lambda \\ &= \int_{\mathbb{R}} d\mu(x)\int_{-c}^{c}d\lambda\ e^{i\lambda x}\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right) \\ &= \int_{\mathbb{R}} d\mu(x)\int_{-c}^{c}d\lambda\ \left(\frac{e^{-i\lambda(a-x)}-e^{-i\lambda(b-x)}}{i\lambda}\right).\end{aligned}$$

Since

$$\frac{e^{-i\lambda(\alpha-x)}}{i\lambda} = -\frac{i}{\lambda}\cos(\lambda(\alpha-x)) - \frac{1}{\lambda}\sin(\lambda(\alpha-x))$$

it follows that $\operatorname{Im}\left(\left[e^{-i\lambda(a-x)}-e^{-i\lambda(b-x)}\right]/i\lambda\right)$ is an odd function of λ and

$$\operatorname{Re}\left(\frac{e^{-i\lambda(a-x)}-e^{-i\lambda(b-x)}}{i\lambda}\right)=\frac{1}{\lambda}\left[\sin\left(\lambda\left(x-a\right)\right)-\sin\left(\lambda\left(x-b\right)\right)\right],$$

and therefore (using Lemma 22.30)

$$\begin{aligned}I\left(c\right)&=\int_{\mathbb{R}}d\mu\left(x\right)\int_{-c}^{c}d\lambda\operatorname{Re}\left(\frac{e^{-i\lambda(a-x)}-e^{-i\lambda(b-x)}}{i\lambda}\right)\\&=\int_{\mathbb{R}}d\mu\left(x\right)\int_{-c}^{c}d\lambda\left(\frac{\sin\lambda\left(x-a\right)-\sin\lambda\left(x-b\right)}{\lambda}\right)\\&=\int_{\mathbb{R}}d\mu\left(x\right)\left[\operatorname{sgn}(x-a)S\left(c\left|x-a\right|\right)-\operatorname{sgn}(x-b)S\left(c\left|x-b\right|\right)\right].\end{aligned}$$

Using Lemma 22.30 again along with the DCT we may pass to the limit as $c\uparrow\infty$ in the previous identity to get the result;

$$\begin{aligned}\lim_{c\to\infty}\frac{1}{2\pi}I\left(c\right)&=\frac{1}{2}\int_{\mathbb{R}}d\mu\left(x\right)\left[\operatorname{sgn}(x-a)-\operatorname{sgn}(x-b)\right]\\&=\frac{1}{2}\int_{\mathbb{R}}d\mu\left(x\right)\left[2\cdot 1_{(a,b)}\left(x\right)+1_{\{a\}}\left(x\right)+1_{\{b\}}\left(x\right)\right]\\&=\mu\left(\left(a,b\right)\right)+\frac{1}{2}\left[\mu\left(\{a\}\right)+\mu\left(\{b\}\right)\right].\end{aligned}$$

■

Corollary 22.32. *Suppose that μ is a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\hat{\mu}\in L^1\left(m\right)$, then $d\mu=\rho dm$ where ρ is the continuous probability density on $\mathbb{R}$ given by*

$$\rho\left(x\right):=\frac{1}{2\pi}\int_{\mathbb{R}}\hat{\mu}\left(\lambda\right)e^{-i\lambda x}d\lambda. \tag{22.22}$$

Proof. The function ρ defined in Eq. (22.22) is continuous by the dominated convergence theorem. Moreover for any $-\infty<a<b<\infty$ we have

$$\begin{aligned}\int_a^b\rho\left(x\right)dx&=\frac{1}{2\pi}\int_a^b dx\int_{\mathbb{R}}d\lambda\;\hat{\mu}\left(\lambda\right)e^{-i\lambda x}\\&=\frac{1}{2\pi}\int_{\mathbb{R}}d\lambda\;\hat{\mu}\left(\lambda\right)\int_a^b dxe^{-i\lambda x}\\&=\frac{1}{2\pi}\int_{\mathbb{R}}d\lambda\;\hat{\mu}\left(\lambda\right)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)\\&=\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\hat{\mu}\left(\lambda\right)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)d\lambda\\&=\mu\left(\left(a,b\right)\right)+\frac{1}{2}\left[\mu\left(\{a\}\right)+\mu\left(\{b\}\right)\right],\end{aligned}$$

wherein we have used Theorem 22.31 to evaluate the limit. Letting $a\uparrow b$ over $a\in\mathbb{R}$ such that $\mu\left(\{a\}\right)=0$ in this identity shows $\mu\left(\{b\}\right)=0$ for all $b\in\mathbb{R}$. Therefore we have shown

$$\mu\left(\left(a,b\right]\right)=\int_a^b\rho\left(x\right)dx\text{ for all }-\infty<a<b<\infty.$$

Using one of the multiplicative systems theorems, it is now easy to verify that $\mu\left(A\right)=\int_A\rho\left(x\right)dx$ for all $A\in\mathcal{B}_{\mathbb{R}}$ or $\int_{\mathbb{R}}hd\mu=\int_{\mathbb{R}}h\rho\,d\mu$ for all bounded measurable functions $h:\mathbb{R}\to\mathbb{R}$. This then implies that $\rho\geq 0$, m – a.e.[1] and the $d\mu=\rho dm$. ■

Example 22.33. Let $\rho\left(x\right)=\left(1-|x|\right)_+$ be the triangle density in Figure 22.2

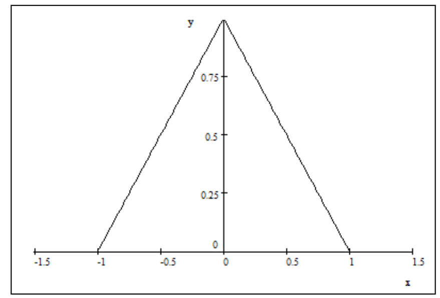


Fig. 22.2. The triangular density function.

Recall from Example 22.6 that

$$\int_{\mathbb{R}}e^{i\lambda x}\left(1-|x|\right)_+dx=2\frac{1-\cos\lambda}{\lambda^2}.$$

Alternatively by direct calculation,

$$\begin{aligned}\int_{\mathbb{R}}e^{i\lambda x}\left(1-|x|\right)_+dx&=2\operatorname{Re}\int_0^1e^{i\lambda x}\left(1-x\right)dx\\&=2\operatorname{Re}\left[\left(I-\frac{1}{i}\frac{d}{d\lambda}\right)\int_0^1e^{i\lambda x}dx\right]\\&=2\operatorname{Re}\left[\left(I-\frac{1}{i}\frac{d}{d\lambda}\right)\frac{e^{i\lambda}-1}{i\lambda}\right]\\&=2\frac{1-\cos\lambda}{\lambda^2}.\end{aligned}$$

[1] Since ρ is continuous we may further conclude that $\rho\left(x\right)\geq 0$ for every $x\in\mathbb{R}$.

Hence it follows[2] from Corollary 22.32 that

$$(1-|x|)_{+}=\frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\cos\lambda}{\lambda^{2}}e^{-i\lambda x}d\lambda. \tag{22.23}$$

Evaluating Eq. (22.23) at $x=0$ gives the identity

$$1=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos\lambda}{\lambda^{2}}d\lambda. \tag{22.24}$$

from which we deduce that

$$d\mu(x):=\frac{1}{\pi}\frac{1-\cos x}{x^{2}}dx \tag{22.25}$$

is a probability measure such that (from Eq. (22.23)) has characteristic function,

$$\hat{\mu}(\lambda)=(1-|\lambda|)_{+}. \tag{22.26}$$

Corollary 22.34. *For all random variables, X, we have*

$$\mathbb{E}|X|=\frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\operatorname{Re}f_{X}(\lambda)}{\lambda^{2}}d\lambda. \tag{22.27}$$

Proof. For $M\in\mathbb{R}\setminus\{0\}$, make the change of variables, $\lambda\to M\lambda$ in Eq. (22.24) for find

$$|M|=\frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\cos(\lambda M)}{\lambda^{2}}d\lambda. \tag{22.28}$$

Observe the identity holds for $M=0$ as well. Taking $M=X$ in Eq. (22.28) and then taking expectations implies,

$$\mathbb{E}|X|=\frac{1}{\pi}\int_{\mathbb{R}}\mathbb{E}\frac{1-\cos\lambda X}{\lambda^{2}}d\lambda=\frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\operatorname{Re}f_{X}(\lambda)}{\lambda^{2}}d\lambda.$$

■

Suppose that we did not know the value of $c:=\int_{-\infty}^{\infty}\frac{1-\cos\lambda}{\lambda^{2}}d\lambda$ is π, we could still proceed as above to learn

$$\mathbb{E}|X|=\frac{1}{c}\int_{\mathbb{R}}\frac{1-\operatorname{Re}f_{X}(\lambda)}{\lambda^{2}}d\lambda.$$

We could then evaluate c by making a judicious choice of X. For example if $X\overset{d}{=}N(0,1)$, we would have on one hand

$$\mathbb{E}|X|=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}|x|e^{-x^{2}/2}dx=\frac{2}{\sqrt{2\pi}}\int_{0}^{\infty}xe^{-x^{2}/2}dx=\sqrt{\frac{2}{\pi}}.$$

On the other hand, $f_{X}(\lambda)=e^{-\lambda^{2}/2}$ and so

$$\begin{aligned}\sqrt{\frac{2}{\pi}}&=-\frac{1}{c}\int_{\mathbb{R}}\left(1-e^{-\lambda^{2}/2}\right)d\left(\lambda^{-1}\right)=\frac{1}{c}\int_{\mathbb{R}}d\left(1-e^{-\lambda^{2}/2}\right)\left(\lambda^{-1}\right)\\&=\frac{1}{c}\int_{\mathbb{R}}e^{-\lambda^{2}/2}d\lambda=\frac{\sqrt{2\pi}}{c}\end{aligned}$$

from which it follows, again, that $c=\pi$.

Corollary 22.35. *Suppose X is a random variable and there exists $\varepsilon>0$ such that $u(\lambda):=\operatorname{Re}f_{X}(\lambda)=\mathbb{E}[\cos\lambda X]$ is continuously differentiable for $\lambda\in(-2\varepsilon,2\varepsilon)$. If we further assume that*

$$\int_{0}^{\varepsilon}\frac{|u'(\lambda)|}{\lambda}d\lambda<\infty, \tag{22.29}$$

then $\mathbb{E}|X|<\infty$ and $f_{X}\in C^{1}(\mathbb{R},\mathbb{C})$. (Since u is even, u' is odd and $u'(0)=0$. Hence if $u'(\lambda)$ were α – Hölder continuous for some $\alpha>0$, then Eq. (22.29) would hold.)

Proof. According to Eq. (22.27)

$$\pi\cdot\mathbb{E}|X|=\int_{\mathbb{R}}\frac{1-u(\lambda)}{\lambda^{2}}d\lambda=\int_{|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^{2}}d\lambda+\int_{|\lambda|>\varepsilon}\frac{1-u(\lambda)}{\lambda^{2}}d\lambda.$$

Since $0\le 1-u(\lambda)\le 2$ and $2/\lambda^{2}$ is integrable for $|\lambda|>\varepsilon$, to show $\mathbb{E}|X|<\infty$ we must show,

$$\infty>\int_{|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^{2}}d\lambda=\lim_{\delta\downarrow 0}\int_{\delta\le|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^{2}}d\lambda.$$

By an integration by parts we find

[2] This identity could also be verified directly using residue calculus techniques from complex variables.

$$\begin{aligned}
\int_{\delta\le|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda &= \int_{\delta\le|\lambda|\le\varepsilon}(1-u(\lambda))\,d\left(-\lambda^{-1}\right)\\
&= \frac{u(\lambda)-1}{\lambda}|_\delta^\varepsilon + \frac{u(\lambda)-1}{\lambda}|_{-\varepsilon}^{-\delta} - \int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda\\
&= -\int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda + \frac{u(\varepsilon)-1}{\varepsilon} - \frac{u(-\varepsilon)-1}{-\varepsilon}\\
&+ \frac{u(-\delta)-1}{-\delta} - \frac{u(\delta)-1}{\delta}.\\
&\to -\lim_{\delta\downarrow 0}\int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} + u'(0) - u'(0)\\
&\le \int_{|\lambda|\le\varepsilon}\frac{|u'(\lambda)|}{|\lambda|}d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon}\\
&= 2\int_0^\varepsilon\frac{|u'(\lambda)|}{\lambda}d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} < \infty.
\end{aligned}$$

Passing the limit as $\delta\downarrow 0$ using the fact that $u'(\lambda)$ is an odd function, we learn

$$\begin{aligned}
\int_{|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda &= \lim_{\delta\downarrow 0}\int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon}\\
&\le 2\int_0^\varepsilon\frac{|u'(\lambda)|}{\lambda}d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} < \infty.
\end{aligned}$$

■

22.5 Exercises

Exercise 22.2. For $x,\lambda\in\mathbb{R}$, let (also see Eq. (22.32))

$$\varphi(\lambda,x) := \begin{cases}\frac{e^{i\lambda x}-1-i\lambda x}{x^2} & \text{if } x\ne 0\\ -\frac{1}{2}\lambda^2 & \text{if } x=0.\end{cases} \tag{22.30}$$

Let $\{x_k\}_{k=1}^n\subset\mathbb{R}\setminus\{0\}$, $\{Z_k\}_{k=1}^n\cup\{N\}$ be independent random variables with $N\overset{d}{=}N(0,1)$ and Z_k being Poisson random variables with mean $a_k>0$, i.e. $P(Z_k=n)=e^{-a_k}\frac{a_k^n}{n!}$ for $n=0,1,2\dots$. With $Y:=\sum_{k=1}^n x_k(Z_k-a_k)+\alpha N$, show

$$f_Y(\lambda) := \mathbb{E}\left[e^{i\lambda Y}\right] = \exp\left(\int_{\mathbb{R}}\varphi(\lambda,x)\,d\nu(x)\right)$$

where ν is the discrete measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ given by

$$\nu = \alpha^2\delta_0 + \sum_{k=1}^n a_k x_k^2\delta_{x_k}. \tag{22.31}$$

It is easy to see that $\varphi(\lambda,0)=\lim_{x\to 0}\varphi(\lambda,x)$. In fact by Taylor's theorem with integral remainder we have

$$\varphi(\lambda,x) = -\frac{1}{2}\lambda^2\int_0^1 e^{it\lambda x}d\nu(t) \tag{22.32}$$

where $d\nu(t)=2(1-t)\,dt$ is a probability measure on $[0,1]$. From this formula it is clear that φ is a smooth function of (λ,x).

Exercise 22.3. To each finite and compactly supported measure, ν, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ show there exists a sequence $\{\nu_n\}_{n=1}^\infty$ of finitely supported finite measures on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\nu_n\implies\nu$. Here we say ν is compactly supported if there exists $M<\infty$ such that $\nu(\{x:|x|\ge M\})=0$ and we say ν is finitely supported if there exists a finite subset, $\Lambda\subset\mathbb{R}$ such that $\nu(\mathbb{R}\setminus\Lambda)=0$.

Exercise 22.4. Show that if ν is a finite measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$, then

$$f(\lambda) := \exp\left(\int_{\mathbb{R}}\varphi(\lambda,x)\,d\nu(x)\right) \tag{22.33}$$

is the characteristic function of a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$. Here is an outline to follow. (You may find the calculus estimates in Section 22.7 to be of help.)

1. Show $f(\lambda)$ is continuous.
2. Now suppose that ν is compactly supported. Show, using Exercises 22.2, 22.3, and the continuity Theorem 22.18 that $\exp\left(\int_{\mathbb{R}}\varphi(\lambda,x)\,d\nu(x)\right)$ is the characteristic function of a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$.
3. For the general case, approximate ν by a sequence of finite measures with compact support as in item 2.

Exercise 22.5 (Exercise 2.3 in [51]). Let μ be the probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$, such that $\mu(\{n\})=p(n)=c\frac{1}{n^2\ln|n|}1_{|n|\ge 2}$ with c chosen so that $\sum_{n\in\mathbb{Z}}p(n)=1$. Show that $\hat{\mu}\in C^1(\mathbb{R},\mathbb{C})$ even though $\int_{\mathbb{R}}|x|\,d\mu(x)=\infty$. To do this show,

$$g(t) := \sum_{n\ge 2}\frac{1-\cos nt}{n^2\ln n}$$

is continuously differentiable.

Exercise 22.6 (Polya's Criterion [4, Problem 26.3 on p. 305.] and [14, p. 104-107.]). Suppose $\varphi(\lambda)$ is a non-negative symmetric continuous function such that $\varphi(0)=1$, $\varphi(\lambda)$ is non-increasing and convex for $\lambda\ge 0$. Show $\varphi(\lambda)=\hat{\nu}(\lambda)$ for some probability measure, ν, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$.

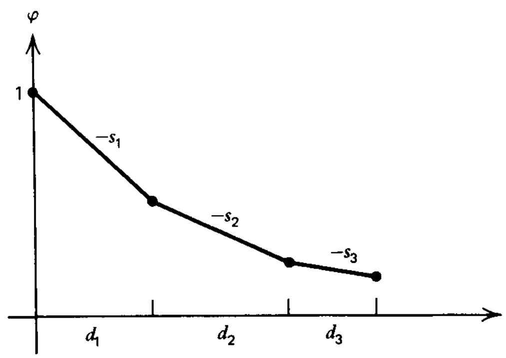


Fig. 22.3. Here is a piecewise linear convex function. We will assume that $d_n > 0$ for all n and that $\varphi(\lambda) = 0$ for λ sufficiently large. This last restriction may be removed later by a limiting argument.

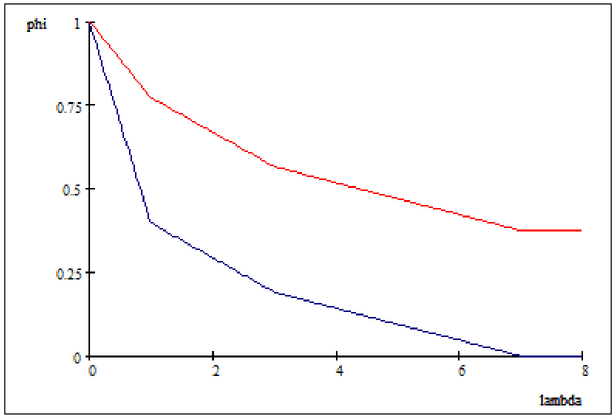


Fig. 22.4. The Fourier transform of $\nu = \frac{1}{2}\mu_1 + \frac{1}{6}\mu_3 + \frac{1}{3}\mu_7$ is in blue and the Fourier transform of $\nu = \frac{3}{8}\delta_0 + \frac{1}{8}\mu_1 + \frac{1}{6}\mu_3 + \frac{1}{3}\mu_7$ is given in red.

Solution to Exercise (22.6). Because of the continuity theorem and some simple limiting arguments, it suffices to prove the result for a function φ as pictured in Figure 22.3. From Example 22.33, we know that $(1-|\lambda|)_+ = \hat{\mu}(\lambda)$ where μ is the probability measure,

$$d\mu(x) := \frac{1}{\pi}\frac{1-\cos x}{x^2}dx.$$

For $a > 0$, let $\mu_a(A) = \mu(aA)$ in which case $\mu_a(f) = \mu\left(f\left(a^{-1}\cdot\right)\right)$ for all bounded measurable f and in particular,

$$\hat{\mu}_a(\lambda) = \hat{\mu}\left(a^{-1}\lambda\right) = \left(1 - \left|\frac{\lambda}{a}\right|\right)_+.$$

To finish the proof it suffices to show that $\varphi(\lambda)$ may be expressed as

$$\varphi(\lambda) = \sum_{n=1}^{\infty} p_n \hat{\mu}_{a_n}(\lambda) = \sum_{n=1}^{\infty} p_n \left(1 - \left|\frac{\lambda}{a_n}\right|\right)_+ \tag{22.34}$$

for some $a_n > 0$ and $p_n \geq 0$ such that $\sum_{n=1}^{\infty} p_n = 1$ Indeed, if this is the case we may take, $\nu := \sum_{n=1}^{\infty} p_n \mu_{a_n}$, see Figure 22.4.

It is pretty clear that we should take $a_n = d_1 + \cdots + d_n$ for all $n \in \mathbb{N}$. Since we are assuming $\varphi(\lambda) = 0$ for large λ, there is a first index, $N \in \mathbb{N}$, such that

$$0 = \varphi(a_N) = 1 - \sum_{n=1}^{N} d_n s_n. \tag{22.35}$$

Notice that $s_n = 0$ for all $n > N$.

Since

$$\varphi'(\lambda) = -\sum_{n=k}^{\infty} p_n \frac{1}{a_n} \text{ when } a_{k-1} < \lambda < a_k$$

we must require,

$$s_k = \sum_{n=k}^{\infty} p_n \frac{1}{a_n} \text{ for all } k$$

which then implies $p_k \frac{1}{a_k} = s_k - s_{k+1}$ or equivalently that

$$p_k = a_k (s_k - s_{k+1}). \tag{22.36}$$

Since φ is convex, we know that $-s_k \leq -s_{k+1}$ or $s_k \geq s_{k+1}$ for all k and therefore $p_k \geq 0$ and $p_k = 0$ for all $k > N$. Moreover,

$$\begin{aligned}\sum_{k=1}^{\infty} p_k &= \sum_{k=1}^{\infty} a_k (s_k - s_{k+1}) = \sum_{k=1}^{\infty} a_k s_k - \sum_{k=2}^{\infty} a_{k-1} s_k \\ &= a_1 s_1 + \sum_{k=2}^{\infty} s_k (a_k - a_{k-1}) = d_1 s_1 + \sum_{k=2}^{\infty} s_k d_k \\ &= \sum_{k=1}^{\infty} s_k d_k = 1\end{aligned}$$

where the last equality follows from Eq. (22.35). Working backwards with p_k defined as in Eq. (22.36) it is now easily shown that $\frac{d}{d\lambda}\sum_{n=1}^{\infty} p_n \left(1 - \left|\frac{\lambda}{a_n}\right|\right)_+ =$

$\varphi'(\lambda)$ for $\lambda \notin \{a_1, a_2, \dots\}$ and since both functions are equal to 1 at $\lambda = 0$ we may conclude that Eq. (22.34) is indeed valid.

22.6 Appendix: Bochner's Theorem

Definition 22.36. *A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have* ***rapid decay*** *or* ***rapid decrease*** *if*

$$\sup_{x\in\mathbb{R}^n} (1+|x|)^N |f(x)| < \infty \text{ for } N = 1, 2, \dots.$$

Equivalently, for each $N \in \mathbb{N}$ there exists constants $C_N < \infty$ such that $|f(x)| \le C_N(1+|x|)^{-N}$ for all $x \in \mathbb{R}^n$. A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have (at most) ***polynomial growth*** *if there exists $N < \infty$ such*

$$\sup (1+|x|)^{-N} |f(x)| < \infty,$$

i.e. there exists $N \in \mathbb{N}$ and $C < \infty$ such that $|f(x)| \le C(1+|x|)^N$ for all $x \in \mathbb{R}^n$.

Definition 22.37 (Schwartz Test Functions). *Let $\mathcal{S}$ denote the space of functions $f \in C^\infty(\mathbb{R}^n)$ such that f and all of its partial derivatives have rapid decay and let*

$$\|f\|_{N,\alpha} = \sup_{x\in\mathbb{R}^n} \left|(1+|x|)^N \partial^\alpha f(x)\right|$$

so that

$$\mathcal{S} = \left\{ f \in C^\infty(\mathbb{R}^n) : \|f\|_{N,\alpha} < \infty \text{ for all } N \text{ and } \alpha \right\}.$$

Also let $\mathcal{P}$ denote those functions $g \in C^\infty(\mathbb{R}^n)$ such that g and all of its derivatives have at most polynomial growth, i.e. $g \in C^\infty(\mathbb{R}^n)$ is in $\mathcal{P}$ iff for all multi-indices α, there exists $N_\alpha < \infty$ such

$$\sup (1+|x|)^{-N_\alpha} |\partial^\alpha g(x)| < \infty.$$

(Notice that any polynomial function on $\mathbb{R}^n$ is in $\mathcal{P}$.)

Definition 22.38. *A function $\chi : \mathbb{R}^n \to \mathbb{C}$ is said to be* ***positive (semi) definite*** *iff the matrices $A := \{\chi(\xi_k - \xi_j)\}_{k,j=1}^m$ are positive definite for all $m \in \mathbb{N}$ and $\{\xi_j\}_{j=1}^m \subset \mathbb{R}^n$.*

Lemma 22.39. *If μ is a finite positive measure on $\mathcal{B}_{\mathbb{R}^n}$, then $\chi := \hat{\mu} \in C(\mathbb{R}^n, \mathbb{C})$ is a positive definite function.*

Proof. The dominated convergence theorem implies $\hat{\mu} \in C(\mathbb{R}^n, \mathbb{C})$. Since μ is a positive measure (and hence real),

$$\hat{\mu}(-\xi) = \int_{\mathbb{R}^n} e^{i\xi\cdot x} d\mu(x) = \overline{\int_{\mathbb{R}^n} e^{-i\xi\cdot x} d\mu(x)} = \overline{\hat{\mu}(-\xi)}.$$

From this it follows that for any $m \in \mathbb{N}$ and $\{\xi_j\}_{j=1}^m \subset \mathbb{R}^n$, the matrix $A := \{\hat{\mu}(\xi_k - \xi_j)\}_{k,j=1}^m$ is self-adjoint. Moreover if $\lambda \in \mathbb{C}^m$,

$$\begin{aligned} \sum_{k,j=1}^m \hat{\mu}(\xi_k - \xi_j)\lambda_k \bar{\lambda}_j &= \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i(\xi_k-\xi_j)\cdot x} \lambda_k \bar{\lambda}_j d\mu(x) \\ &= \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i\xi_k\cdot x}\lambda_k \overline{e^{-i\xi_j\cdot x}\lambda_j} d\mu(x) \\ &= \int_{\mathbb{R}^n} \left| \sum_{k=1}^m e^{-i\xi_k\cdot x}\lambda_k \right|^2 d\mu(x) \ge 0 \end{aligned}$$

showing A is positive definite. ■

Proposition 22.40. *Suppose that $\chi : \mathbb{R}^n \to \mathbb{C}$ is said to be positive definite with $\chi(0) = 1$. If χ is continuous at 0 then in fact χ is uniformly continuous on all of $\mathbb{R}^n$.*

Proof. Taking $\xi_1 = x$, $\xi_2 = y$ and $\xi_3 = 0$ in Definition 22.38 we conclude that

$$A := \begin{bmatrix} 1 & \chi(x-y) & \chi(x) \\ \chi(y-x) & 1 & \chi(y) \\ \chi(-x) & \chi(-y) & 1 \end{bmatrix} = \begin{bmatrix} 1 & \chi(x-y) & \chi(x) \\ \bar{\chi}(x-y) & 1 & \chi(y) \\ \bar{\chi}(x) & \bar{\chi}(y) & 1 \end{bmatrix}$$

is positive definite. In particular,

$$\begin{aligned} 0 \le \det A = 1 &+ \chi(x-y)\chi(y)\bar{\chi}(x) + \chi(x)\bar{\chi}(x-y)\bar{\chi}(y) \\ &- |\chi(x)|^2 - |\chi(y)|^2 - |\chi(x-y)|^2. \end{aligned}$$

Combining this inequality with the identity,

$$|\chi(x) - \chi(y)|^2 = |\chi(x)|^2 + |\chi(y)|^2 - \chi(x)\bar{\chi}(y) - \chi(y)\bar{\chi}(x),$$

gives

$$\begin{aligned}
0 &\le 1-|\chi(x-y)|^2+\chi(x-y)\chi(y)\bar\chi(x)+\chi(x)\bar\chi(x-y)\bar\chi(y) \\
&\quad -\left\{|\chi(x)-\chi(y)|^2+\chi(x)\bar\chi(y)+\chi(y)\bar\chi(x)\right\} \\
&= 1-|\chi(x-y)|^2-|\chi(x)-\chi(y)|^2 \\
&\quad +\chi(x-y)\chi(y)\bar\chi(x)-\chi(y)\bar\chi(x)+\chi(x)\bar\chi(x-y)\bar\chi(y)-\chi(x)\bar\chi(y) \\
&= 1-|\chi(x-y)|^2-|\chi(x)-\chi(y)|^2+2\operatorname{Re}\left((\chi(x-y)-1)\chi(y)\bar\chi(x)\right) \\
&\le 1-|\chi(x-y)|^2-|\chi(x)-\chi(y)|^2+2|\chi(x-y)-1|.
\end{aligned}$$

Hence we have

$$\begin{aligned}
|\chi(x)-\chi(y)|^2 &\le 1-|\chi(x-y)|^2+2|\chi(x-y)-1| \\
&= (1-|\chi(x-y)|)(1+|\chi(x-y)|)+2|\chi(x-y)-1| \\
&\le 4|1-\chi(x-y)|
\end{aligned}$$

which completes the proof. ■

Remark 22.41. The function $f(\lambda)=1_{\{0\}}(\lambda)$ is positive definite since the matrix, $\{f(\lambda_i-\lambda_j)\}_{i,j=1}^n$ is the $n\times n$ identity matrix for all choices of distinct $\{\lambda_i\}_{i=1}^n$ in $\mathbb{R}$. Note however that f is not continuous at $\lambda=0$.

Lemma 22.42. *If $\chi\in C(\mathbb{R}^n,\mathbb{C})$ is a positive definite function, then*

1. $\chi(0)\ge 0$.
2. $\chi(-\xi)=\overline{\chi(\xi)}$ *for all* $\xi\in\mathbb{R}^n$.
3. $|\chi(\xi)|\le\chi(0)$ *for all* $\xi\in\mathbb{R}^n$.
4. *If we further assume that χ is continuous, then*

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}d\xi d\eta\ge 0 \tag{22.37}$$

for all $f\in\mathbb{S}(\mathbb{R}^d)$.

Proof. Taking $m=1$ and $\xi_1=0$ we learn $\chi(0)|\lambda|^2\ge 0$ for all $\lambda\in\mathbb{C}$ which proves item 1. Taking $m=2$, $\xi_1=\xi$ and $\xi_2=\eta$, the matrix

$$A:=\begin{bmatrix}\chi(0) & \chi(\xi-\eta)\\ \chi(\eta-\xi) & \chi(0)\end{bmatrix}$$

is positive definite from which we conclude $\chi(\xi-\eta)=\overline{\chi(\eta-\xi)}$ (since $A=A^*$ by definition) and

$$0\le\det\begin{bmatrix}\chi(0) & \chi(\xi-\eta)\\ \chi(\eta-\xi) & \chi(0)\end{bmatrix}=|\chi(0)|^2-|\chi(\xi-\eta)|^2.$$

and hence $|\chi(\xi)|\le\chi(0)$ for all ξ. This proves items 2. and 3. Item 4. follows by approximating the integral in Eq. (22.37) by Riemann sums,

$$\begin{aligned}
&\int_{\mathbb{R}^n\times\mathbb{R}^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}d\xi d\eta \\
&\qquad=\lim_{\varepsilon\downarrow 0}\varepsilon^{-2n}\sum_{\xi,\eta\in(\varepsilon\mathbb{Z}^n)\cap[-\varepsilon^{-1},\varepsilon^{-1}]^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}\ge 0.
\end{aligned}$$

The details are left to the reader keeping in mind this is where we must use the assumption that χ is continuous. ■

Theorem 22.43 (Bochner's Theorem). *Suppose $\chi\in C(\mathbb{R}^n,\mathbb{C})$ is positive definite function which is continuous at 0, then there exists a unique positive measure μ on $\mathcal{B}_{\mathbb{R}^n}$ such that $\chi=\hat\mu$.*

Proof. If $\chi(\xi)=\hat\mu(\xi)$, then for $f\in\mathcal{S}$ we would have

$$\int_{\mathbb{R}^n}fd\mu=\int_{\mathbb{R}^n}\left(f^\vee\right)\hat{}\,d\mu=\int_{\mathbb{R}^n}f^\vee(\xi)\hat\mu(\xi)\mathbf{d}\xi.$$

This suggests that we define

$$I(f):=\int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi)\mathbf{d}\xi\text{ for all }f\in\mathcal{S}.$$

We will now show I is positive in the sense if $f\in\mathcal{S}$ and $f\ge 0$ then $I(f)\ge 0$. For general $f\in\mathcal{S}$ we have

$$\begin{aligned}
I(|f|^2)&=\int_{\mathbb{R}^n}\chi(\xi)\left(|f|^2\right)^\vee(\xi)\mathbf{d}\xi=\int_{\mathbb{R}^n}\chi(\xi)\left(f^\vee\bigstar\bar f^\vee\right)(\xi)\mathbf{d}\xi \\
&=\int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi-\eta)\bar f^\vee(\eta)\mathbf{d}\eta\mathbf{d}\xi=\int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi-\eta)\overline{f^\vee(-\eta)}\mathbf{d}\eta\mathbf{d}\xi \\
&=\int_{\mathbb{R}^n}\chi(\xi-\eta)f^\vee(\xi)\overline{f^\vee(\eta)}\mathbf{d}\eta\mathbf{d}\xi\ge 0.
\end{aligned}\tag{22.38}$$

For $t>0$ let $p_t(x):=t^{-n/2}e^{-|x|^2/2t}\in\mathcal{S}$ and define

$$I_t(x):=I\bigstar p_t(x):=I(p_t(x-\cdot))=I(\left|\sqrt{p_t(x-\cdot)}\right|^2)$$

which is non-negative by Eq. (22.38) and the fact that $\sqrt{p_t(x-\cdot)}\in\mathcal{S}$. Using

$$\begin{aligned}
\left[p_t(x-\cdot)\right]^\vee(\xi)&=\int_{\mathbb{R}^n}p_t(x-y)e^{iy\cdot\xi}\mathbf{d}y=\int_{\mathbb{R}^n}p_t(y)e^{i(y+x)\cdot\xi}\mathbf{d}y \\
&=e^{ix\cdot\xi}p_t^\vee(\xi)=e^{ix\cdot\xi}e^{-t|\xi|^2/2},
\end{aligned}$$

$$\begin{aligned}\langle I_t, \psi\rangle &= \int_{\mathbb{R}^n} I(p_t(x-\cdot))\psi(x)\mathbf{d}x \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \chi(\xi)\left[p_t(x-\cdot)\right]^{\vee}(\xi)\psi(x)\mathbf{d}\xi\right)\mathbf{d}x \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \chi(\xi)e^{ix\cdot\xi}e^{-t|\xi|^2/2}\psi(x)\mathbf{d}\xi\right)\mathbf{d}x \\ &= \int_{\mathbb{R}^n} \chi(\xi)\psi^{\vee}(\xi)e^{-t|\xi|^2/2}\mathbf{d}\xi\end{aligned}$$

which coupled with the dominated convergence theorem shows

$$\langle I\bigstar p_t, \psi\rangle \to \int_{\mathbb{R}^n} \chi(\xi)\psi^{\vee}(\xi)\mathbf{d}\xi = I(\psi) \text{ as } t\downarrow 0.$$

Hence if $\psi \geq 0$, then $I(\psi) = \lim_{t\downarrow 0}\langle I_t, \psi\rangle \geq 0$.

Let $K \subset \mathbb{R}$ be a compact set and $\psi \in C_c(\mathbb{R}, [0,\infty))$ be a function such that $\psi = 1$ on K. If $f \in C_c^\infty(\mathbb{R}, \mathbb{R})$ is a smooth function with $\mathrm{supp}(f) \subset K$, then $0 \leq \|f\|_\infty \psi - f \in \mathcal{S}$ and hence

$$0 \leq \langle I, \|f\|_\infty \psi - f\rangle = \|f\|_\infty \langle I, \psi\rangle - \langle I, f\rangle$$

and therefore $\langle I, f\rangle \leq \|f\|_\infty \langle I, \psi\rangle$. Replacing f by $-f$ implies, $-\langle I, f\rangle \leq \|f\|_\infty \langle I, \psi\rangle$ and hence we have proved

$$|\langle I, f\rangle| \leq C(\mathrm{supp}(f))\,\|f\|_\infty \tag{22.39}$$

for all $f \in \mathcal{D}_{\mathbb{R}^n} := C_c^\infty(\mathbb{R}^n, \mathbb{R})$ where $C(K)$ is a finite constant for each compact subset of $\mathbb{R}^n$. Because of the estimate in Eq. (22.39), it follows that $I|_{\mathcal{D}_{\mathbb{R}^n}}$ has a unique extension I to $C_c(\mathbb{R}^n, \mathbb{R})$ still satisfying the estimates in Eq. (22.39) and moreover this extension is still positive. So by the Riesz – Markov Theorem **??**, there exists a unique Radon – measure μ on $\mathbb{R}^n$ such that such that $\langle I, f\rangle = \mu(f)$ for all $f \in C_c(\mathbb{R}^n, \mathbb{R})$.

To finish the proof we must show $\hat{\mu}(\eta) = \chi(\eta)$ for all $\eta \in \mathbb{R}^n$ given

$$\mu(f) = \int_{\mathbb{R}^n} \chi(\xi)f^{\vee}(\xi)\mathbf{d}\xi \text{ for all } f \in C_c^\infty(\mathbb{R}^n, \mathbb{R}). \tag{22.40}$$

Let $f \in C_c^\infty(\mathbb{R}^n, \mathbb{R}_+)$ be a radial function such $f(0) = 1$ and $f(x)$ is decreasing as $|x|$ increases. Let $f_\varepsilon(x) := f(\varepsilon x)$, then by Theorem **??**,

$$\mathcal{F}^{-1}\left[e^{-i\eta x}f_\varepsilon(x)\right](\xi) = \varepsilon^{-n}f^{\vee}(\frac{\xi-\eta}{\varepsilon})$$

and therefore, from Eq. (22.40),

$$\int_{\mathbb{R}^n} e^{-i\eta x}f_\varepsilon(x)d\mu(x) = \int_{\mathbb{R}^n} \chi(\xi)\varepsilon^{-n}f^{\vee}(\frac{\xi-\eta}{\varepsilon})\mathbf{d}\xi. \tag{22.41}$$

Because $\int_{\mathbb{R}^n} f^{\vee}(\xi)\mathbf{d}\xi = \mathcal{F}f^{\vee}(0) = f(0) = 1$, we may apply the approximate δ – function Theorem 22.44 below to Eq. (22.41) to find (using the continuity of χ here!)

$$\int_{\mathbb{R}^n} e^{-i\eta x}f_\varepsilon(x)d\mu(x) \to \chi(\eta) \text{ as } \varepsilon\downarrow 0. \tag{22.42}$$

On the the other hand, when $\eta = 0$, the monotone convergence theorem implies $\mu(f_\varepsilon) \uparrow \mu(1) = \mu(\mathbb{R}^n)$ and therefore $\mu(\mathbb{R}^n) = \mu(1) = \chi(0) < \infty$. Now knowing the μ is a finite measure we may use the dominated convergence theorem to concluded

$$\mu(e^{-i\eta x}f_\varepsilon(x)) \to \mu(e^{-i\eta x}) = \hat{\mu}(\eta) \text{ as } \varepsilon\downarrow 0$$

for all η. Combining this equation with Eq. (22.42) shows $\hat{\mu}(\eta) = \chi(\eta)$ for all $\eta \in \mathbb{R}^n$. ■

Theorem 22.44 (Approximate δ – functions). *Let $p \in [1,\infty]$, $\varphi \in L^1(\mathbb{R}^d)$, $a := \int_{\mathbb{R}^d}\varphi(x)dx$, and for $t > 0$ let $\varphi_t(x) = t^{-d}\varphi(x/t)$. Then*

1. *If $f \in L^p$ with $p < \infty$ then $\varphi_t * f \to af$ in L^p as $t\downarrow 0$.*
2. *If $f \in BC(\mathbb{R}^d)$ and f is uniformly continuous then $\|\varphi_t * f - af\|_\infty \to 0$ as $t\downarrow 0$.*
3. *If $f \in L^\infty$ and f is continuous on $U \subset_o \mathbb{R}^d$ then $\varphi_t * f \to af$ uniformly on compact subsets of U as $t\downarrow 0$.*

Proof. Making the change of variables $y = tz$ implies

$$\varphi_t * f(x) = \int_{\mathbb{R}^d} f(x-y)\varphi_t(y)dy = \int_{\mathbb{R}^d} f(x-tz)\varphi(z)dz$$

so that

$$\begin{aligned}\varphi_t * f(x) - af(x) &= \int_{\mathbb{R}^d}\left[f(x-tz) - f(x)\right]\varphi(z)dz \\ &= \int_{\mathbb{R}^d}\left[\tau_{tz}f(x) - f(x)\right]\varphi(z)dz.\end{aligned} \tag{22.43}$$

Hence by Minkowski's inequality for integrals (Theorem **??** of the analysis notes), Proposition **??** and the dominated convergence theorem,

$$\|\varphi_t * f - af\|_p \leq \int_{\mathbb{R}^d}\|\tau_{tz}f - f\|_p\,|\varphi(z)|\,dz \to 0 \text{ as } t\downarrow 0.$$

Item 2. is proved similarly. Indeed, form Eq. (22.43)

$$\|\varphi_t * f - af\|_\infty \le \int_{\mathbb{R}^d} \|\tau_{tz} f - f\|_\infty |\varphi(z)|\, dz$$

which again tends to zero by the dominated convergence theorem because $\lim_{t\downarrow 0} \|\tau_{tz} f - f\|_\infty = 0$ uniformly in z by the uniform continuity of f.

Item 3. Let $B_R = B(0,R)$ be a large ball in $\mathbb{R}^d$ and $K \sqsubset\sqsubset U$, then

$$\begin{aligned}
&\sup_{x\in K} |\varphi_t * f(x) - af(x)| \\
&\quad \le \left| \int_{B_R} [f(x-tz) - f(x)]\, \varphi(z) dz \right| + \left| \int_{B_R^c} [f(x-tz) - f(x)]\, \varphi(z) dz \right| \\
&\quad \le \int_{B_R} |\varphi(z)|\, dz \cdot \sup_{x\in K, z\in B_R} |f(x-tz) - f(x)| + 2\|f\|_\infty \int_{B_R^c} |\varphi(z)|\, dz \\
&\quad \le \|\varphi\|_1 \cdot \sup_{x\in K, z\in B_R} |f(x-tz) - f(x)| + 2\|f\|_\infty \int_{|z|>R} |\varphi(z)|\, dz
\end{aligned}$$

so that using the uniform continuity of f on compact subsets of U,

$$\limsup_{t\downarrow 0} \sup_{x\in K} |\varphi_t * f(x) - af(x)| \le 2\|f\|_\infty \int_{|z|>R} |\varphi(z)|\, dz \to 0 \text{ as } R \to \infty.$$

■

22.7 Appendix: Some Calculus Estimates

We end this section by gathering together a number of calculus estimates that we will need in the future.

1. Taylor's theorem with integral remainder states, if $f \in C^k(\mathbb{R})$ and $z, \Delta \in \mathbb{R}$ or f be holomorphic in a neighborhood of $z \in \mathbb{C}$ and $\Delta \in \mathbb{C}$ be sufficiently small so that $f(z + t\Delta)$ is defined for $t \in [0,1]$, then

$$\begin{aligned}
f(z+\Delta) &= \sum_{n=0}^{k-1} f^{(n)}(z) \frac{\Delta^n}{n!} + \Delta^k r_k(z,\Delta) &(22.44)\\
&= \sum_{n=0}^{k-1} f^{(n)}(z) \frac{\Delta^n}{n!} + \Delta^k \left[\frac{1}{k!} f^{(k)}(z) + \varepsilon(z,\Delta) \right] &(22.45)
\end{aligned}$$

where

$$\begin{aligned}
r_k(z,\Delta) &= \frac{1}{(k-1)!} \int_0^1 f^{(k)}(z+t\Delta)(1-t)^{k-1}\, dt &(22.46)\\
&= \frac{1}{k!} f^{(k)}(z) + \varepsilon(z,\Delta) &(22.47)
\end{aligned}$$

and

$$\varepsilon(z,\Delta) = \frac{1}{(k-1)!} \int_0^1 \left[f^{(k)}(z+t\Delta) - f^{(k)}(z) \right] (1-t)^{k-1}\, dt \to 0 \text{ as } \Delta \to 0. \tag{22.48}$$

To prove this, use integration by parts to show,

$$\begin{aligned}
r_k(z,\Delta) &= \frac{1}{k!} \int_0^1 f^{(k)}(z+t\Delta) \left(-\frac{d}{dt} \right) (1-t)^k\, dt \\
&= -\frac{1}{k!} \left[f^{(k)}(z+t\Delta)(1-t)^k \right]_{t=0}^{t=1} + \frac{\Delta}{k!} \int_0^1 f^{(k+1)}(z+t\Delta)(1-t)^k\, dt \\
&= \frac{1}{k!} f^{(k)}(z) + \Delta r_{k+1}(z,\Delta),
\end{aligned}$$

i.e.

$$\Delta^k r_k(z,\Delta) = \frac{1}{k!} f^{(k)}(z) \Delta^k + \Delta^{k+1} r_{k+1}(z,\Delta).$$

The result now follows by induction.

2. For $y \in \mathbb{R}$, $\sin y = y \int_0^1 \cos(ty)\, dt$ and hence

$$|\sin y| \le |y|. \tag{22.49}$$

3. For $y \in \mathbb{R}$ we have

$$\cos y = 1 + y^2 \int_0^1 -\cos(ty)(1-t)\, dt \ge 1 + y^2 \int_0^1 -(1-t)\, dt = 1 - \frac{y^2}{2}.$$

Equivalently put[3],

$$g(y) := \cos y - 1 + y^2/2 \ge 0 \text{ for all } y \in \mathbb{R}. \tag{22.50}$$

4. Since

[3] Alternatively,

$$|\sin y| = \left| \int_0^y \cos x dx \right| \le \left| \int_0^y |\cos x|\, dx \right| \le |y|$$

and for $y \ge 0$ we have,

$$\cos y - 1 = \int_0^y -\sin x dx \ge \int_0^y -x dx = -y^2/2.$$

This last inequality may also be proved as a simple calculus exercise following from; $g(\pm\infty) = \infty$ and $g'(y) = 0$ iff $\sin y = y$ which happens iff $y = 0$.

$$\begin{aligned}|e^z - 1 - z| &= \left|z^2 \int_0^1 e^{tz} (1-t)\, dt\right| \\ &\le |z|^2 \int_0^1 e^{t \operatorname{Re} z} (1-t)\, dt \\ &\le |z|^2 \int_0^1 e^{0 \vee \operatorname{Re} z} (1-t)\, dt\end{aligned}$$

we have shown

$$|e^z - 1 - z| \le e^{0 \vee \operatorname{Re} z} \cdot \frac{|z|^2}{2}. \tag{22.51}$$

In particular if $\operatorname{Re} z \le 0$, then

$$|e^z - 1 - z| \le |z|^2 /2. \tag{22.52}$$

5. Since $e^{iy} - 1 = iy \int_0^1 e^{ity} dt$, $\left|e^{iy} - 1\right| \le |y|$ and hence

$$\left|e^{iy} - 1\right| \le 2 \wedge |y| \text{ for all } y \in \mathbb{R}. \tag{22.53}$$

Lemma 22.45. *For $z = re^{i\theta}$ with $-\pi < \theta < \pi$ and $r > 0$, let $\ln z = \ln r + i\theta$. Then $\ln : \mathbb{C} \setminus (-\infty, 0] \to \mathbb{C}$ is a holomorphic function such that $e^{\ln z} = z$[4] and if $|z| < 1$ then*

$$|\ln (1+z) - z| \le |z|^2 \frac{1}{2(1-|z|)^2} \text{ for } |z| < 1. \tag{22.54}$$

Proof. Clearly $e^{\ln z} = z$ and $\ln z$ is continuous. Therefore by the inverse function theorem for holomorphic functions, $\ln z$ is holomorphic and

$$z \frac{d}{dz} \ln z = e^{\ln z} \frac{d}{dz} \ln z = 1.$$

Therefore, $\frac{d}{dz} \ln z = \frac{1}{z}$ and $\frac{d^2}{dz^2} \ln z = -\frac{1}{z^2}$. So by Taylor's theorem,

$$\ln(1+z) = z - z^2 \int_0^1 \frac{1}{(1+tz)^2} (1-t)\, dt. \tag{22.55}$$

If $t \ge 0$ and $|z| < 1$, then

$$\left|\frac{1}{(1+tz)}\right| \le \sum_{n=0}^\infty |tz|^n = \frac{1}{1 - t|z|} \le \frac{1}{1-|z|}.$$

and therefore,

$$\left|\int_0^1 \frac{1}{(1+tz)^2} (1-t)\, dt\right| \le \frac{1}{2(1-|z|)^2}. \tag{22.56}$$

Eq. (22.54) is now a consequence of Eq. (22.55) and Eq. (22.56). ■

Lemma 22.46. *For all $y \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$,*

$$\left|e^{iy} - \sum_{k=0}^n \frac{(iy)^k}{k!}\right| \le \frac{|y|^{n+1}}{(n+1)!} \tag{22.57}$$

and in particular,

$$\left|e^{iy} - 1\right| \le |y| \wedge 2 \tag{22.58}$$

and

$$\left|e^{iy} - \left(1 + iy - \frac{y^2}{2!}\right)\right| \le y^2 \wedge \frac{|y|^3}{3!}. \tag{22.59}$$

More generally for all $n \in \mathbb{N}$ we have

$$\left|e^{iy} - \sum_{k=0}^n \frac{(iy)^k}{k!}\right| \le \frac{|y|^{n+1}}{(n+1)!} \wedge \frac{2|y|^n}{n!}. \tag{22.60}$$

Proof. By Taylor's theorem (see Eq. (22.44) with $f(y) = e^{iy}$, $x = 0$ and $\Delta = y$) we have

$$\begin{aligned}\left|e^{iy} - \sum_{k=0}^n \frac{(iy)^k}{k!}\right| &= \left|\frac{y^{n+1}}{n!} \int_0^1 i^{n+1} e^{ity} (1-t)^n\, dt\right| \\ &\le \frac{|y|^{n+1}}{n!} \int_0^1 (1-t)^n\, dt = \frac{|y|^{n+1}}{(n+1)!}\end{aligned}$$

which is Eq. (22.57). Using Eq. (22.57) with $n = 0$ and the simple estimate; $\left|e^{iy} - 1\right| \le 2$ gives Eq. (22.58). Similarly, Eq. (22.57) follows from the estimates coming from Eq. (22.57) with $n = 1$ and $n = 2$ respectively;

[4] For the purposes of this lemma it suffices to define $\ln(1+z) = -\sum_{n=1}^\infty (-z)^n / n$ and to then observe: 1)

$$\frac{d}{dz} \ln(1+z) = \sum_{n=0}^\infty (-z)^n = \frac{1}{1+z},$$

and 2) the functions $1+z$ and $e^{\ln(1+z)}$ both solve

$$f'(z) = \frac{1}{1+z} f(z) \text{ with } f(0) = 1$$

and therefore $e^{\ln(1+z)} = 1 + z$.

$$\begin{aligned}\left|e^{iy}-\left(1+iy-\frac{y^2}{2!}\right)\right| &\leq \left|e^{iy}-(1+iy)\right|+\left|\frac{y^2}{2}\right| \\ &\leq \left|\frac{y^2}{2}\right|+\left|\frac{y^2}{2}\right|=y^2\end{aligned}$$

and

$$\left|e^{iy}-\left(1+iy-\frac{y^2}{2!}\right)\right| \leq \frac{|y|^3}{3!}.$$

Equation (22.60) is proved similarly and hence will be omitted. ∎

Lemma 22.47. *If X is a square integrable random variable, then*

$$\left|f(\lambda)-\left(1+i\lambda\mathbb{E}X-\frac{\lambda^2}{2!}\mathbb{E}\left[X^2\right]\right)\right| \leq \mathbb{E}\left|e^{i\lambda X}-\left(1+i\lambda X-\lambda^2\frac{X^2}{2!}\right)\right| \leq \lambda^2\varepsilon(\lambda) \tag{22.61}$$

where

$$\varepsilon(\lambda):=\mathbb{E}\left[X^2\wedge\frac{|\lambda|\,|X|^3}{3!}\right]\to 0 \text{ as } \lambda\to 0. \tag{22.62}$$

Proof. Using Eq. (22.59) with $y=\lambda X$ and taking expectations gives Eq. (22.61). The DCT, with $X^2\in L^1(P)$ being the dominating function, allows us to conclude that $\lim_{\varepsilon\to 0}\varepsilon(\lambda)=0$. ∎

23

Weak Convergence of Random Sums

Throughout this chapter, we will assume the following standing notation unless otherwise stated. For each $n \in \mathbb{N}$, let $\{X_{n,k}\}_{k=1}^{n}$ be independent random variables and let

$$S_n := \sum_{k=1}^{n} X_{n,k}. \tag{23.1}$$

Also let

$$f_{nk}(\lambda) := \mathbb{E}\left[e^{i\lambda X_{n,k}}\right] \tag{23.2}$$

denote the characteristic function of $X_{n,k}$. In Section 23.1 we are going to describe necessary and sufficient conditions on the array, $\{X_{n,k} : 1 \le k \le n < \infty\}$, so that $S_n \implies N(0,1)$. In the latter sections we are going to explore other possible limiting distributions for the $\{S_n\}_{n=1}^{\infty}$. This will lead us to the notions of infinitely divisible and stable random variables.

23.1 Lindeberg-Feller CLT

Assumption 2 *Until further notice we are going to assume* $\mathbb{E}[X_{n,k}] = 0$, $\sigma_{n,k}^2 = \mathbb{E}\left[X_{n,k}^2\right] < \infty$, *and* $\operatorname{Var}(S_n) = \sum_{k=1}^{n} \sigma_{n,k}^2 = 1$.

Example 23.1. Suppose $\{X_n\}_{n=1}^{\infty}$ are mean zero square integrable random variables with $\sigma_k^2 = \operatorname{Var}(X_k)$. If we let $s_n^2 := \sum_{k=1}^{n} \operatorname{Var}(X_k) = \sum_{k=1}^{n} \sigma_k^2$, $\sigma_{n,k}^2 := \sigma_k^2/s_n^2$, and $X_{n,k} := X_k/s_n$, then $\{X_{n,k}\}_{k=1}^{n}$ satisfy the above hypothesis and $S_n = \frac{1}{s_n}\sum_{k=1}^{n} X_k$.

Our main interest in this chapter is to consider the limiting behavior of S_n as $n \to \infty$. In order to do this, it will be useful to put conditions on the $\{X_{n,k}\}$ such that no one term dominates sum defining the sum defining S_n in Eq. (23.1) in the limit as $n \to \infty$.

Definition 23.2. *Let* $\{X_{n,k}\}$ *be as above.*

1. $\{X_{n,k}\}$ *satisfies the* ***Lindeberg Condition*** *(LC) iff*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] = 0 \text{ for all } t > 0. \tag{23.3}$$

(The (LC) condition is really a condition about small t.*)*

2. $\{X_{n,k}\}$ *satisfies* ***condition*** (M) *if*

$$D_n := \max\left\{\sigma_{n,k}^2 : k \le n\right\} \to 0 \text{ as } n \to \infty. \tag{23.4}$$

3. $\{X_{n,k}\}$ *is* ***uniformly asymptotic negligibility (UAN)*** *if for all* $\varepsilon > 0$,

$$\lim_{n\to\infty} \max_{k\le n} P(|X_{n,k}| > \varepsilon) = 0. \tag{23.5}$$

Clearly it suffices to test the Lindeberg condition for small t only. Each of these conditions imposes constraints on the size of the tails of the $\{X_{n,k}\}$, see Lemma 23.6 below where it is shown $(LC) \implies (M) \implies (UAN)$. Condition (M) asserts that all of the terms in the sum $\sum_{k=1}^{n} \sigma_{n,k}^2 = \operatorname{Var}(S_n) = 1$ are small so that no one term is contributing by itself.

Remark 23.3. The reader should observe that in order for condition (M) to hold in the setup in Example 23.1 it is necessary that $\lim_{n\to\infty} s_n^2 = \infty$.

Example 23.4. Suppose $\{X_n\}_{n=1}^{\infty}$ are i.i.d. with $\mathbb{E}X_n = 0$ and $\operatorname{Var}(X_n) = \sigma^2$. Then $\left\{X_{n,k} := \frac{1}{\sqrt{n}\sigma} X_k\right\}_{k=1}^{n}$ satisfy (LC). Indeed,

$$\begin{aligned}\sum_{k=1}^{n} \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] &= \frac{1}{n\sigma^2}\sum_{k=1}^{n} \mathbb{E}\left[X_k^2 : \left|\frac{X_k}{\sqrt{n}\sigma}\right| > t\right] \\ &= \frac{1}{\sigma^2}\mathbb{E}\left[X_1^2 : |X_1| > \sqrt{n}\sigma t\right]\end{aligned}$$

which tends to zero as $n \to \infty$ by DCT.

Lemma 23.5. *Let* $\{X_{n,k}\}_{k=1}^{n}$ *for* $n \in \mathbb{N}$ *be as in Assumption 2. If* $\{X_{n,k}\}_{k=1}^{n}$ *satisfy the* ***Liapunov condition****;*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}|X_{n,k}|^{\alpha} = 0 \text{ for some } \alpha > 2 \tag{23.6}$$

then (LC) *holds. More generally, if* $\{X_{n,k}\}$ *satisfies the Liapunov condition,*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[X_{n,k}^2 \varphi(|X_{n,k}|)\right] = 0$$

where $\varphi : [0,\infty) \to [0,\infty)$ *is a non-decreasing function such that* $\varphi(t) > 0$ *for all* $t > 0$*, then* $\{X_{n,k}\}$ *satisfies* (LC).

Proof. Assuming Eq. (23.6), then for any $t>0$,

$$\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^{2}:|X_{n,k}|>t\right]\leq\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^{2}\left|\frac{X_{n,k}}{t}\right|^{\alpha-2}:|X_{n,k}|>t\right]$$
$$\leq\frac{1}{t^{\alpha-2}}\sum_{k=1}^{n}\mathbb{E}\left[|X_{n,k}|^{\alpha}\right]=\frac{1}{t^{\alpha-2}}\sum_{k=1}^{n}\mathbb{E}\left|X_{n,k}\right|^{\alpha}\to0\text{ as }n\to\infty.$$

The generalization is proved similarly;

$$\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^{2}:|X_{n,k}|>t\right]\leq\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^{2}\frac{\varphi\left(|X_{n,k}|\right)}{\varphi\left(t\right)}:|X_{n,k}|>t\right]$$
$$\leq\frac{1}{\varphi\left(t\right)}\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^{2}\varphi\left(|X_{n,k}|\right)\right]\to0\text{ as }n\to\infty.$$

■

Lemma 23.6. *Let $\{X_{n,k}:1\leq k\leq n<\infty\}$ be as above, then $(LC)\implies(M)\implies(UAN)$. Moreover the Lindeberg Condition (LC) implies the following strong form of (UAN),*

$$\sum_{k=1}^{n}P\left(|X_{n,k}|>\varepsilon\right)\leq\frac{1}{\varepsilon^{2}}\sum_{k=1}^{n}\mathbb{E}\left[\left|X_{n,k}\right|^{2}:|X_{n,k}|>\varepsilon\right]\to0. \tag{23.7}$$

Proof. For $k\leq n$,

$$\sigma_{n,k}^{2}=\mathbb{E}\left[X_{n,k}^{2}\right]=\mathbb{E}\left[X_{n,k}^{2}1_{|X_{n,k}|\leq t}\right]+\mathbb{E}\left[X_{n,k}^{2}1_{|X_{n,k}|>t}\right]$$
$$\leq t^{2}+\mathbb{E}\left[X_{n,k}^{2}1_{|X_{n,k}|>t}\right]\leq t^{2}+\sum_{m=1}^{n}\mathbb{E}\left[X_{n,m}^{2}1_{|X_{n,m}|>t}\right]$$

and therefore using (LC) we find

$$\lim_{n\to\infty}\max_{k\leq n}\sigma_{n,k}^{2}\leq t^{2}\text{ for all }t>0.$$

This clearly implies (M) holds. For $\varepsilon>0$ we have by Chebyschev's inequality that

$$P\left(|X_{n,k}|>\varepsilon\right)\leq\frac{1}{\varepsilon^{2}}\mathbb{E}\left[\left|X_{n,k}\right|^{2}:|X_{n,k}|>\varepsilon\right]\leq\frac{1}{\varepsilon^{2}}\sigma_{n,k}^{2} \tag{23.8}$$

and therefore,

$$\max_{k\leq n}P\left(|X_{n,k}|>\varepsilon\right)\leq\frac{1}{\varepsilon^{2}}\max_{k\leq n}\sigma_{n,k}^{2}=\frac{1}{\varepsilon^{2}}D_{n}\to0\text{ as }n\to\infty$$

which shows $(M)\implies(UAN)$. Summing Eq. (23.8) on k gives Eq. (23.7) and the right member of this equation tends to zero as $n\to\infty$ by (LC). ■

We will need the following lemma for our subsequent applications of the continuity theorem.

Lemma 23.7. *Suppose that $a_i,b_i\in\mathbb{C}$ with $|a_i|,|b_i|\leq1$ for $i=1,2,\dots,n$. Then*

$$\left|\prod_{i=1}^{n}a_{i}-\prod_{i=1}^{n}b_{i}\right|\leq\sum_{i=1}^{n}|a_{i}-b_{i}|.$$

Proof. Let $a:=\prod_{i=1}^{n-1}a_i$ and $b:=\prod_{i=1}^{n-1}b_i$ and observe that $|a|,|b|\leq1$ and that

$$\begin{aligned}|a_{n}a-b_{n}b|&\leq|a_{n}a-a_{n}b|+|a_{n}b-b_{n}b|\\&=|a_{n}|\,|a-b|+|a_{n}-b_{n}|\,|b|\\&\leq|a-b|+|a_{n}-b_{n}|.\end{aligned}$$

The proof is now easily completed by induction on n. ■

Theorem 23.8 (Lindeberg-Feller CLT (I)). *Suppose $\{X_{n,k}\}$ satisfies (LC) and the hypothesis in Assumption 2, then*

$$S_{n}\implies N\left(0,1\right). \tag{23.9}$$

(See Theorem 23.13 for a converse to this theorem.)

To prove this theorem we must show

$$\mathbb{E}\left[e^{i\lambda S_{n}}\right]\to e^{-\lambda^{2}/2}\text{ as }n\to\infty. \tag{23.10}$$

Before starting the formal proof, let me give an informal explanation for Eq. (23.10). Using

$$f_{nk}\left(\lambda\right)\sim1-\frac{\lambda^{2}}{2}\sigma_{nk}^{2},$$

we might expect

$$\begin{aligned}\mathbb{E}\left[e^{i\lambda S_{n}}\right]&=\prod_{k=1}^{n}f_{nk}\left(\lambda\right)=e^{\sum_{k=1}^{n}\ln f_{nk}(\lambda)}\\&=e^{\sum_{k=1}^{n}\ln(1+f_{nk}(\lambda)-1)}\\&\overset{(A)}{\sim}e^{\sum_{k=1}^{n}(f_{nk}(\lambda)-1)}\left(=\prod_{k=1}^{n}e^{(f_{nk}(\lambda)-1)}\right)\\&\overset{(B)}{\sim}e^{\sum_{k=1}^{n}-\frac{\lambda^{2}}{2}\sigma_{nk}^{2}}=e^{-\frac{\lambda^{2}}{2}}.\end{aligned}$$

The question then becomes under what conditions are these approximations valid. It turns out that approximation (A), namely that

$$\lim_{n\to\infty}\left|\prod_{k=1}^{n} f_{nk}(\lambda) - \exp\left(\sum_{k=1}^{n}(f_{nk}(\lambda)-1)\right)\right| = 0, \tag{23.11}$$

is valid if condition (M) holds, see Lemma 23.11 below and the approximation (B) is valid, i.e.

$$\lim_{n\to\infty}\sum_{k=1}^{n}(f_{nk}(\lambda)-1) = -\frac{1}{2}\lambda^2,$$

if (LC) is satisfied, see Lemma 23.9. These observations would then constitute a proof of Theorem 23.8. The proof we give below of Theorem 23.8 will not quite follow this route and will not use Lemma 23.11 directly. However, this lemma will be used in the proofs of Theorems 23.13 and 23.22.

Proof. (Proof of Theorem 23.8) Since

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^{n} f_{nk}(\lambda) \text{ and } e^{-\lambda^2/2} = \prod_{k=1}^{n} e^{-\lambda^2\sigma_{n,k}^2/2},$$

we may use Lemma 23.7 to conclude,

$$\left|\mathbb{E}\left[e^{i\lambda S_n}\right] - e^{-\lambda^2/2}\right| \le \sum_{k=1}^{n}\left|f_{nk}(\lambda) - e^{-\lambda^2\sigma_{n,k}^2/2}\right| = \sum_{k=1}^{n}(A_{n,k}+B_{n,k})$$

where $A_{n,k}$ is defined in Eq. (23.12) and

$$B_{n,k} := \left|\left[1-\frac{\lambda^2\sigma_{n,k}^2}{2}\right] - e^{-\lambda^2\sigma_{n,k}^2/2}\right|.$$

Because of Lemma 23.9 below, in order to finish the proof it suffices to show $\lim_{n\to\infty}\sum_{k=1}^{n} B_{n,k} = 0$. To estimate $\sum_{k=1}^{n} B_{n,k}$, we use the estimate, $|e^{-u}-1+u| \le u^2/2$ valid for $u \ge 0$ (see Eq. (22.52) with $z=-u$). With this estimate we find,

$$\begin{aligned}\sum_{k=1}^{n} B_{n,k} &= \sum_{k=1}^{n}\left|\left[1-\frac{\lambda^2\sigma_{n,k}^2}{2}\right] - e^{-\lambda^2\sigma_{n,k}^2/2}\right| \\ &\le \sum_{k=1}^{n}\frac{1}{2}\left[\frac{\lambda^2\sigma_{n,k}^2}{2}\right]^2 = \frac{\lambda^4}{8}\sum_{k=1}^{n}\sigma_{n,k}^4 \\ &\le \frac{\lambda^4}{8}\max_{k\le n}\sigma_{n,k}^2\sum_{k=1}^{n}\sigma_{n,k}^2 = \frac{\lambda^4}{8}\max_{k\le n}\sigma_{n,k}^2 \to 0,\end{aligned}$$

wherein we have used (M) (which is implied by (LC)) in taking the limit as $n\to\infty$. ∎

Lemma 23.9. *Let*

$$A_{n,k} := \left|f_{nk}(\lambda) - 1 + \frac{\lambda^2\sigma_{n,k}^2}{2}\right| \tag{23.12}$$

and assume that $\{X_{n,k}\}_{k=1}^{n}$ *satisfies (LC), then*

$$\limsup_{n\to\infty}\sum_{k=1}^{n} A_{n,k} = 0 \text{ and}$$

$$\lim_{n\to\infty}\sum_{k=1}^{n}(f_{nk}(\lambda)-1) = -\lambda^2/2 \text{ for all } \lambda\in\mathbb{R}.$$

Proof. Rewriting $A_{n,k}$ and using Lemma 22.47 implies for every $\varepsilon > 0$ that,

$$\begin{aligned}A_{n,k} &= \left|\mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 + \frac{\lambda^2}{2}X_{n,k}^2\right]\right| \\ &\le \mathbb{E}\left|e^{i\lambda X_{n,k}} - 1 + \frac{\lambda^2}{2}X_{n,k}^2\right| \\ &\le \lambda^2\mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda|\,|X_{n,k}|^3}{3!}\right] \\ &\le \lambda^2\mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda|\,|X_{n,k}|^3}{3!} : |X_{n,k}| \le \varepsilon\right] + \lambda^2\mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda|\,|X_{n,k}|^3}{3!} : |X_{n,k}| > \varepsilon\right] \\ &\le \frac{|\lambda|^3}{3!}\varepsilon\cdot\mathbb{E}\left[|X_{n,k}|^2 : |X_{n,k}| \le \varepsilon\right] + \lambda^2\mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right] \\ &= \frac{|\lambda|^3\varepsilon}{6}\sigma_{n,k}^2 + \lambda^2\mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right].\end{aligned}$$

Summing this equation on k and making use of (LC) gives;

$$\limsup_{n\to\infty}\sum_{k=1}^{n} A_{n,k} \le \frac{\lambda^3\varepsilon}{6} \to 0 \text{ as } \varepsilon\downarrow 0. \tag{23.13}$$

The second limit follows from the first and the simple estimate;

$$\left|\sum_{k=1}^{n}(f_{nk}(\lambda)-1) + \lambda^2/2\right| = \left|\sum_{k=1}^{n}\left(f_{nk}(\lambda) - 1 + \frac{\lambda^2\sigma_{n,k}^2}{2}\right)\right| \le \sum_{k=1}^{n} A_{n,k}.$$

∎

As an application of Theorem 23.8 we can give half of the proof of Theorem 20.35.

Theorem 23.10 (Converse assertion in Theorem 20.35). *If $\{X_n\}_{n=1}^\infty$ are independent random variables and the random series, $\sum_{n=1}^\infty X_n$, is almost surely convergent, then for all $c>0$ the following three series converge;*

1. $\sum_{n=1}^\infty P(|X_n|>c)<\infty$,
2. $\sum_{n=1}^\infty \operatorname{Var}\left(X_n 1_{|X_n|\le c}\right)<\infty$, *and*
3. $\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n|\le c}\right)$ *converges.*

Proof. Since $\sum_{n=1}^\infty X_n$ is almost surely convergent, it follows that $\lim_{n\to\infty} X_n = 0$ a.s. and hence for every $c>0$, $P(\{|X_n|\ge c \text{ i.o.}\})=0$. According the Borel zero one law (Lemma 10.41) this implies for every $c>0$ that $\sum_{n=1}^\infty P(|X_n|>c)<\infty$. Since $X_n\to 0$ a.s., $\{X_n\}$ and $\left\{X_n^c := X_n 1_{|X_n|\le c}\right\}$ are tail equivalent for all $c>0$. In particular $\sum_{n=1}^\infty X_n^c$ is almost surely convergent for all $c>0$.

Fix $c>0$, let $Y_n := X_n^c - \mathbb{E}[X_n^c]$ and let

$$s_n^2 = \operatorname{Var}(Y_1+\dots+Y_n) = \sum_{k=1}^n \operatorname{Var}(Y_k) = \sum_{k=1}^n \operatorname{Var}(X_k^c) = \sum_{k=1}^n \operatorname{Var}\left(X_k 1_{|X_k|\le c}\right).$$

For the sake of contradictions, suppose $s_n^2\to\infty$ as $n\to\infty$. Since $|Y_k|\le 2c$, it follows that $\sum_{k=1}^n \mathbb{E}\left[Y_k^2 1_{|Y_k|>s_n t}\right]=0$ for all sufficiently large n and hence

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}\left[Y_k^2 1_{|Y_k|>s_n t}\right]=0,$$

i.e. $\{Y_{n,k} := Y_k/s_n\}_{n=1}^\infty$ satisfies (LC) – see Examples 23.1 and Remark 23.3. So by the central limit Theorem 23.8, it follows that

$$\frac{1}{s_n^2}\sum_{k=1}^n (X_n^c - \mathbb{E}[X_n^c]) = \frac{1}{s_n^2}\sum_{k=1}^n Y_k \implies N(0,1).$$

On the other hand we know

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^n X_n^c = \frac{\sum_{k=1}^\infty X_k^c}{\lim_{n\to\infty} s_n^2} = 0 \text{ a.s.}$$

and so by Slutsky's theorem,

$$\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}[X_n^c] = \frac{1}{s_n^2}\sum_{k=1}^n X_n^c - \frac{1}{s_n^2}\sum_{k=1}^n Y_k \implies N(0,1).$$

But it is not possible for constant (i.e. **non-random)** variables, $c_n := \frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}[X_n^c]$, to converge to a non-degenerate limit. (Think about this either in terms of characteristic functions or in terms of distribution functions.) Thus we must conclude that

$$\sum_{n=1}^\infty \operatorname{Var}\left(X_n 1_{|X_n|\le c}\right) = \sum_{n=1}^\infty \operatorname{Var}(X_n^c) = \lim_{n\to\infty} s_n^2 < \infty.$$

An application of Kolmogorov's convergence criteria (Theorem 20.11) implies that

$$\sum_{n=1}^\infty (X_n^c - \mathbb{E}[X_n^c]) \text{ is convergent } a.s.$$

Since we already know that $\sum_{n=1}^\infty X_n^c$ is convergent almost surely we may now conclude $\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n|\le c}\right)$ is convergent. ■

Let us now turn to the converse of Theorem 23.8, see Theorem 23.13 below.

Lemma 23.11. *Suppose that $\{X_{n,k}\}$ satisfies property (M), i.e. $D_n := \max_{k\le n}\sigma_{n,k}^2\to 0$. If we define,*

$$\varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1 = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1\right],$$

then;

1. $\lim_{n\to\infty}\max_{k\le n}|\varphi_{n,k}(\lambda)| = 0$ *and*
2. $f_{S_n}(\lambda) - \prod_{k=1}^n e^{\varphi_{n,k}(\lambda)} \to 0$ *as* $n\to\infty$, *where*

$$f_{S_n}(\lambda) = \mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^n f_{n,k}(\lambda).$$

Proof. For any $\varepsilon>0$ we have, making use of Eq. (22.58) and Chebyschev's inequality, that

$$\begin{aligned}|\varphi_{n,k}(\lambda)| &= |f_{n,k}(\lambda)-1| \le \mathbb{E}\left|e^{i\lambda X_{n,k}}-1\right| \le \mathbb{E}[2\wedge|\lambda X_{n,k}|]\\ &\le \mathbb{E}[2\wedge|\lambda X_{n,k}| : |X_{n,k}|\ge\varepsilon] + \mathbb{E}[2\wedge|\lambda X_{n,k}| : |X_{n,k}|<\varepsilon]\\ &\le 2P[|X_{n,k}|\ge\varepsilon] + |\lambda|\varepsilon \le \frac{2\sigma_{n,k}^2}{\varepsilon^2} + |\lambda|\varepsilon.\end{aligned}$$

Therefore,

$$\limsup_{n\to\infty}\max_{k\le n}|\varphi_{n,k}(\lambda)| \le \limsup_{n\to\infty}\left[\frac{2D_n}{\varepsilon^2} + |\lambda|\varepsilon\right] = |\lambda|\varepsilon \to 0 \text{ as } \varepsilon\downarrow 0.$$

For the second item, observe that $\operatorname{Re}\varphi_{n,k}(\lambda) = \operatorname{Re} f_{n,k}(\lambda) - 1 \le 0$ and hence $\left|e^{\varphi_{n,k}(\lambda)}\right| = e^{\operatorname{Re}\varphi_{n,k}(\lambda)} \le 1$. Therefore by Lemma 23.7 and the estimate (22.52) we find;

$$
\begin{aligned}
\left|\prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)} - \prod_{k=1}^{n} f_{n,k}(\lambda)\right| &\le \sum_{k=1}^{n} \left|e^{\varphi_{n,k}(\lambda)} - f_{n,k}(\lambda)\right| \\
&= \sum_{k=1}^{n} \left|e^{\varphi_{n,k}(\lambda)} - (1+\varphi_{n,k}(\lambda))\right| \\
&\le \frac{1}{2}\sum_{k=1}^{n} |\varphi_{n,k}(\lambda)|^2 \\
&\le \frac{1}{2}\max_{k\le n}|\varphi_{n,k}(\lambda)| \cdot \sum_{k=1}^{n} |\varphi_{n,k}(\lambda)| .
\end{aligned}
$$

Since $\mathbb{E}X_{n,k} = 0$ we may write express $\varphi_{n,k}$ as

$$\varphi_{n,k}(\lambda) = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}\right]$$

and then using estimate in Eq. (22.52) again shows

$$
\begin{aligned}
\sum_{k=1}^{n} |\varphi_{n,k}(\lambda)| &= \sum_{k=1}^{n} \left|\mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}\right]\right| \\
&\le \sum_{k=1}^{n} \left|\mathbb{E}\left[\frac{1}{2}|\lambda X_{n,k}|^2\right]\right| \le \frac{\lambda^2}{2}\sum_{k=1}^{n}\sigma_{n,k}^2 = \frac{\lambda^2}{2}.
\end{aligned}
$$

Thus we have shown,

$$\left|\prod_{k=1}^{n} f_{n,k}(\lambda) - \prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)}\right| \le \frac{\lambda^2}{4}\max_{k\le n}|\varphi_{n,k}(\lambda)|$$

and the latter expression tends to zero by item 1. ∎

Lemma 23.12. *Let X be a random variable such that $\mathbb{E}X^2 < \infty$ and $\mathbb{E}X = 0$. Further let $f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ and $u(\lambda) := \operatorname{Re}(f(\lambda) - 1)$. Then for all $c > 0$,*

$$u(\lambda) + \frac{\lambda^2}{2}\mathbb{E}\left[X^2\right] \ge \mathbb{E}\left[X^2\left[\frac{\lambda^2}{2} - \frac{2}{c^2}\right] : |X| > c\right] \tag{23.14}$$

or equivalently

$$\mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2\right] \ge \mathbb{E}\left[X^2\left[\frac{\lambda^2}{2} - \frac{2}{c^2}\right] : |X| > c\right]. \tag{23.15}$$

In particular if we choose $|\lambda| \ge \sqrt{6}/|c|$, then

$$\mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2\right] \ge \frac{1}{c^2}\mathbb{E}\left[X^2 : |X| > c\right]. \tag{23.16}$$

Proof. For all $\lambda \in \mathbb{R}$, we have (see Eq. (22.50)) $\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2 \ge 0$ and $\cos\lambda X - 1 \ge -2$. Therefore,

$$
\begin{aligned}
u(\lambda) + \frac{\lambda^2}{2}\mathbb{E}\left[X^2\right] &= \mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2\right] \\
&\ge \mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2 : |X| > c\right] \\
&\ge \mathbb{E}\left[-2 + \frac{\lambda^2}{2}X^2 : |X| > c\right] \\
&\ge \mathbb{E}\left[-2\frac{|X|^2}{c^2} + \frac{\lambda^2}{2}X^2 : |X| > c\right]
\end{aligned}
$$

which gives Eq. (23.14). ∎

Theorem 23.13 (Lindeberg-Feller CLT (II)). *Suppose $\{X_{n,k}\}$ satisfies (M) and also the central limit theorem in Eq. (23.9) holds, then $\{X_{n,k}\}$ satisfies (LC). So under condition (M), S_n converges to a normal random variable iff (LC) holds.*

Proof. By assumption we have

$$\lim_{n\to\infty}\max_{k\le n}\sigma_{n,k}^2 = 0 \text{ and } \lim_{n\to\infty}\prod_{k=1}^{n} f_{n,k}(\lambda) = e^{-\lambda^2/2}.$$

The second inequality combined with Lemma 23.11 implies,

$$\lim_{n\to\infty} e^{\sum_{k=1}^{n}\varphi_{n,k}(\lambda)} = \lim_{n\to\infty}\prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)} = e^{-\lambda^2/2}.$$

Taking the modulus of this equation then implies,

$$\lim_{n\to\infty} e^{\sum_{k=1}^{n}\operatorname{Re}\varphi_{n,k}(\lambda)} = \lim_{n\to\infty}\left|e^{\sum_{k=1}^{n}\varphi_{n,k}(\lambda)}\right| = e^{-\lambda^2/2}$$

from which we may conclude

$$\lim_{n\to\infty}\sum_{k=1}^{n}\operatorname{Re}\varphi_{n,k}(\lambda) = -\lambda^2/2.$$

We may write this last limit as

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}\left[\cos(\lambda X_{n,k}) - 1 + \frac{\lambda^2}{2}X_{n,k}^2\right] = 0$$

which by Lemma 23.12 implies

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^2 : |X_{n,k}|>c\right]=0$$

for all $c>0$ which is (LC). ■

As an application of Theorem 23.8 let us see what it has to say about Brownian motion. In what follows we say that $\{B_t\}_{t\geq 0}$ is a Gaussian process if for all finite subsets, $\Lambda\subset[0,\infty)$ the random variables $\{B_t\}_{t\in\Lambda}$ are **jointly** Gaussian. We will discuss Gaussian processes in more generality in Chapter 24.

Proposition 23.14. *Suppose that $\{B_t\}_{t\geq 0}$ is a stochastic process on some probability space, $(\Omega,\mathcal{B},P)$ such that;*

1. *$B_0=0$ a.s., $\mathbb{E}B_t=0$ for all $t\geq 0$,*
2. *$\mathbb{E}(B_t-B_s)^2=t-s$ for all $0\leq s\leq t<\infty$,*
3. *B has independent increments, i.e. if $0=t_0<t_1<\cdots<t_n<\infty$, then $\left\{B_{t_j}-B_{t_{j-1}}\right\}_{j=1}^n$ are independent random variables.*
4. *(Moment Condition) There exists $p>2$, $q>1$ and $c<\infty$ such that $\mathbb{E}|B_t-B_s|^p\leq c|t-s|^q$ for all $s,t\in\mathbb{R}_+$.*

Then $B_t-B_s\overset{d}{=}N(0,t-s)$ for all $0\leq s<t<\infty$. We call such a process satisfying these conditions a ***pre-Brownian motion.***

Proof. Let $0\leq s<t$ and for each $n\in\mathbb{N}$ and $1\leq k\leq n$ let $X_{n,k}:=B_{t_k}-B_{t_{k-1}}$ where $\{s=t_0<t_1<\cdots<t_n=t\}$ is the uniform partition of $[s,t]$. Under the moment condition hypothesis we find,

$$\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^p\right]\leq c\sum_{k=1}^{n}\left(\frac{t-s}{n}\right)^q=c\,(t-s)^q\,\frac{n}{n^q}\to 0\text{ as }n\to\infty.$$

Thus we have shown that $\{X_{n,k}\}$ satisfies a Liapunov condition which by Lemma 23.5 implies that $\{X_{n,k}\}$ satisfies (LC). Therefore, $B_t-B_s=\sum_{k=1}^n X_{n,k}\to N(0,t-s)$ as $n\to\infty$ by the Lindeberg-Feller central limit Theorem 23.8. ■

Remark 23.15 (Poisson Process). There certainly are other properties satisfying items 1.-3. other than a pre-Brownian motion. Indeed, if $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ (see Example 25.8), then $B_t:=\lambda^{-1}N_t-t$ satisfies the items 1.–3. above. Recall that $\mathrm{Var}(N_t-N_s)=(\lambda t-\lambda s)^2$ and $\mathbb{E}(N_t-N_s)=\lambda(t-s)$ so $\mathbb{E}B_t=0$ and

$$\begin{aligned}\mathbb{E}(B_t-B_s)^2&=\mathrm{Var}(B_t-B_s)=\mathrm{Var}\left(\lambda^{-1}(N_t-N_s)\right)\\&=\lambda^{-2}\,\mathrm{Var}(N_t-N_s)=\lambda^{-2}(\lambda t-\lambda s)^2=t-s.\end{aligned}$$

In this case one can show that $\mathbb{E}[(B_t-B_s)^p]\sim|t-s|$ for all $1\leq p<\infty$.

This last remark leads us to our next topic.

23.2 Infinitely Divisible Distributions

In the this section we are going to investigate the possible limiting distributions of the $\{S_n\}_{n=1}^\infty$ when we relax the Lindeberg condition. Let us begin with a simple example of the Poisson limit theorem.

Theorem 23.16 (A Poisson Limit Theorem). *For each $n\in\mathbb{N}$, let $\{Y_{n,k}\}_{k=1}^n$ be independent Bernoulli random variables with $P(Y_{n,k}=1)=p_{n,k}$ and $P(Y_{n,k}=0)=q_{n,k}:=1-p_{n,k}$. Suppose;*

1. *$\lim_{n\to\infty}\sum_{k=1}^n p_{n,k}=a\in(0,\infty)$ and*
2. *$\lim_{n\to\infty}\max_{1\leq k\leq n}p_{n,k}=0$. (So no one term is dominating the sums in item 1.)*

Then $S_n=\sum_{k=1}^n Y_{n,k}\implies Z$ where Z is a Poisson random variable with mean a. (See [14, Section 2.6] for more on this theorem.)

Proof. We will give two proofs of this theorem. The first proof relies on the law of rare events in Theorem 21.10 while the second uses Fourier transform methods.

First proof. Let $Z_n\overset{d}{=}Poi\left(\sum_{k=1}^n p_{n,k}\right)$, then by Theorem 21.10, we know that

$$d_{TV}(Z_n,S_n)\leq\sum_{k=1}^n p_{n,k}^2\leq\max_{1\leq k\leq n}p_{n,k}\cdot\sum_{k=1}^n p_{n,k}.$$

From the assumptions it follows that $\lim_{n\to\infty}d_{TV}(Z_n,S_n)=0$ and from part 3. of Exercise 21.6 we know that $\lim_{n\to\infty}d_{TV}(Z_n,Z)=0$. Therefore, $\lim_{n\to\infty}d_{TV}(Z,S_n)=0$.

Second proof. Recall from Example 22.11 that for any $a>0$,

$$\mathbb{E}\left[e^{i\lambda Z}\right]=\exp\left(a\left(e^{i\lambda}-1\right)\right).$$

Since

$$\mathbb{E}\left[e^{i\lambda Y_{n,k}}\right]=e^{i\lambda}p_{n,k}+(1-p_{n,k})=1+p_{n,k}\left(e^{i\lambda}-1\right),$$

it follows that

$$\mathbb{E}\left[e^{i\lambda S_n}\right]=\prod_{k=1}^n\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right].$$

Since $1+p_{n,k}\left(e^{i\lambda}-1\right)$ lies on the line segment joining 1 to $e^{i\lambda}$, it follows (see Figure 23.1) that

$$\left|1+p_{n,k}\left(e^{i\lambda}-1\right)\right|\leq 1.$$

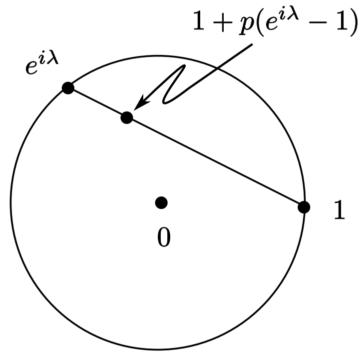


Fig. 23.1. Simple circle geometry reflecting the convexity of the disk.

Hence we may apply Lemma 23.7 to find

$$\begin{aligned}&\left|\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\\&\leq\sum_{k=1}^{n}\left|\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\\&=\sum_{k=1}^{n}\left|\exp\left(z_{n,k}\right)-\left[1+z_{n,k}\right]\right|\end{aligned}$$

where

$$z_{n,k}=p_{n,k}\left(e^{i\lambda}-1\right).$$

Since $\operatorname{Re} z_{n,k}=p_{n,k}\left(\cos\lambda-1\right)\leq 0$, we may use the calculus estimate in Eq. (22.52) to conclude,

$$\begin{aligned}&\left|\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\\&\leq\frac{1}{2}\sum_{k=1}^{n}\left|z_{n,k}\right|^{2}\leq\frac{1}{2}\max_{1\leq k\leq n}\left|z_{n,k}\right|\sum_{k=1}^{n}\left|z_{n,k}\right|\\&\leq 2\max_{1\leq k\leq n}p_{n,k}\sum_{k=1}^{n}p_{n,k}.\end{aligned}$$

Using the assumptions, we may conclude

$$\left|\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\to 0\text{ as }n\to\infty.$$

Since

$$\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)=\exp\left(\sum_{k=1}^{n}p_{n,k}\left(e^{i\lambda}-1\right)\right)\to\exp\left(a\left(e^{i\lambda}-1\right)\right),$$

we have shown

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[e^{i\lambda S_n}\right]&=\lim_{n\to\infty}\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\\&=\lim_{n\to\infty}\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)=\exp\left(a\left(e^{i\lambda}-1\right)\right).\end{aligned}$$

The result now follows by an application of the continuity Theorem 22.18. ■

Remark 23.17. Keeping the notation in Theorem 23.16, we have

$$\mathbb{E}\left[Y_{n,k}\right]=p_{n,k}\text{ and }\operatorname{Var}\left(Y_{n,k}\right)=p_{n,k}\left(1-p_{n,k}\right)$$

and

$$s_n^2:=\sum_{k=1}^{n}\operatorname{Var}\left(Y_{n,k}\right)=\sum_{k=1}^{n}p_{n,k}\left(1-p_{n,k}\right).$$

Under the assumptions of Theorem 23.16, we see that $s_n^2\to a$ as $n\to\infty$. Let us now center and normalize the $Y_{n,k}$ by setting;

$$X_{n,k}:=\frac{Y_{n,k}-p_{n,k}}{s_n}$$

so that

$$\sigma_{n,k}^2:=\operatorname{Var}\left(X_{n,k}\right)=\frac{1}{s_n^2}\operatorname{Var}\left(Y_{n,k}\right)=\frac{1}{s_n^2}p_{n,k}\left(1-p_{n,k}\right),$$

$\mathbb{E}\left[X_{n,k}\right]=0$, $\operatorname{Var}\left(\sum_{k=1}^{n}X_{n,k}\right)=1$, and the $\{X_{n,k}\}$ satisfy condition (M). On the other hand for small t and large n we have

$$\begin{aligned}\mathbb{E}\left[X_{n,k}^2:\left|X_{n,k}\right|>t\right]&=\mathbb{E}\left[X_{n,k}^2:\left|\frac{Y_{n,k}-p_{n,k}}{s_n}\right|>t\right]\\&=\mathbb{E}\left[X_{n,k}^2:\left|Y_{n,k}-p_{n,k}\right|>s_n t\right]\\&\geq\mathbb{E}\left[X_{n,k}^2:\left|Y_{n,k}-p_{n,k}\right|>2at\right]\\&=\mathbb{E}\left[X_{n,k}^2:Y_{n,k}=1\right]=p_{n,k}\left(\frac{1-p_{n,k}}{s_n}\right)^2\end{aligned}$$

from which it follows that

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}\left[X_{n,k}^2 : |X_{n,k}|>t\right]=\lim_{n\to\infty}\sum_{k=1}^{n}p_{n,k}\left(\frac{1-p_{n,k}}{s_n}\right)^2=a.$$

Therefore $\{X_{n,k}\}$ do **not** satisfy (LC). Nevertheless we have by Theorem 23.16 along with Slutzky's Theorem 21.39 that

$$\sum_{k=1}^{n}X_{n,k}=\frac{\sum_{k=1}^{n}Y_{n,k}-\sum_{k=1}^{n}p_{n,k}}{s_n}\implies\frac{Z-a}{a}$$

where Z is a Poisson random variable with mean a. Notice that the limit is **not** a normal random variable in agreement with Theorem 23.13.

Given this example it is natural to ask what are the possible limiting distribution of row sums of random arrays $\{X_{n,k}\}$. As it turns out the answer is often contained in the following definition.

Definition 23.18. *A probability distribution, μ, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ is **infinitely divisible** iff for all $n\in\mathbb{N}$ there exists i.i.d. nondegenerate random variables, $\{X_{n,k}\}_{k=1}^{n}$, such that $X_{n,1}+\cdots+X_{n,n}\overset{d}{=}\mu$. This can be formulated in the following two equivalent ways. For all $n\in\mathbb{N}$ there should exists a non-degenerate probability measure, μ_n, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\mu_n^{*n}=\mu$. For all $n\in\mathbb{N}$, $\hat{\mu}(\lambda)=[g(\lambda)]^n$ for some non-constant characteristic function, g.*

Theorem 23.19. *Suppose that μ is a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ and $\mathrm{Law}(X)=\mu$. Then μ is infinitely divisible iff there exits an array, $\{X_{n,k}:1\le k\le m_n\}$ with $\{X_{n,k}\}_{k=1}^{m_n}$ being i.i.d. such that $\sum_{k=1}^{m_n}X_{n,k}\implies X$ and $m_n\uparrow\infty$ as $n\to\infty$.*

Proof. The only non-trivial direction is $(\Longleftarrow)$. Let me do this in the case that $m_n=n$ for all n case. See Kallenberg [30, Lemma 15.13, p. 294] for the needed result involving the tail bounds needed to cover the full case. In this case given any k, we have $S_{nk}=\sum_{i=1}^{k}S_n^i$ where

$$S_n^i=\sum_{j=k(i-1)+1}^{ki}X_{n,j}.$$

Notice that $\{S_n^i\}_{i=1}^{k}$ are i.i.d. for each $n\in\mathbb{N}$ and since $S_{nk}\implies X$ as $n\to\infty$ we know that $\{S_{nk}\}_{n=1}^{\infty}$ is tight and there exists $\varepsilon(r)\downarrow 0$ as $r\uparrow\infty$ such that

$$P(|S_{nk}|>r)\le\varepsilon(r).$$

Since

$$\begin{aligned}P\left(S_n^1>r\right)^k&=P\left(S_n^i>r\text{ for }1\le i\le k\right)\\&\le P\left(S_{nk}>kr\right)\le P\left(|S_{nk}|>kr\right)\le\varepsilon(kr)\end{aligned}$$

and similarly,

$$\begin{aligned}P\left(-S_n^1>r\right)^k&=P\left(-S_n^i>r\text{ for }1\le i\le k\right)\\&\le P\left(-S_{nk}>kr\right)\le P\left(|S_{nk}|>kr\right)\le\varepsilon(kr)\end{aligned}$$

we see that $P\left(\left|S_n^1\right|>r\right)\le 2\varepsilon(kr)^{1.k}\to 0$ as $r\uparrow\infty$ which shows that $\left\{S_n^1\right\}_{n=1}^{\infty}$ has tight distributions as well. Thus there exists a subsequence $\{n_l\}$ such that $S_{n_l}^1\implies Y$ as $l\to\infty$. Let $\{Y_i\}_{i=1}^{k}$ be i.i.d. random variables with $Y_i\overset{d}{=}Y$. Then by Exercise 21.11 it follows that

$$S_{kn_l}=\sum_{i=1}^{k}S_{n_l}^i\implies Y_1+\dots Y_k$$

from which we conclude that $X\overset{d}{=}Y_1+\dots Y_k$. ■

It turns out that the characteristic function of an infinitely divisible distribution has to have a very special form. This will be the subject of the Lévy Kintchine formula in Theorem 23.21 below. First though, lets give a couple of examples.

Example 23.20 (Following Theorem 17.28). Suppose that $\{Z_n\}_{n=1}^{\infty}$ are i.i.d. random variables and $N_\alpha\overset{d}{=}Poi(\alpha)$ and $Y\overset{d}{=}\sigma N(0,1)+\mu$ are chosen so that $\{Z_n\}_{n=1}^{\infty}\cup\{N_\alpha,Y\}$ are all independent, then $S:=Y+\sum_{n\le N_\alpha}Z_n$ is infinitely divisible. Indeed we have

$$\begin{aligned}f_S(\lambda)&=\mathbb{E}\left[e^{i\lambda S}\right]=\mathbb{E}\left[e^{i\lambda Y}\right]\cdot\mathbb{E}\left[e^{i\lambda\sum_{k\le N_\alpha}Z_k}\right]\\&=\mathbb{E}\left[e^{i\lambda Y}\right]\cdot\sum_{k=0}^{\infty}\mathbb{E}\left[e^{i\lambda\sum_{k\le N_\alpha}Z_k}|N_\alpha=n\right]P(N_\alpha=n)\\&=\mathbb{E}\left[e^{i\lambda Y}\right]\cdot\sum_{n=0}^{\infty}\mathbb{E}\left[e^{i\lambda(Z_1+\cdots+Z_n)}\right]P(N_\alpha=n)\\&=\exp\left(-\frac{1}{2}\sigma^2\lambda^2+i\lambda\mu\right)\sum_{n=0}^{\infty}e^{-\alpha}\frac{\alpha^n}{n!}\hat{\mu}(\lambda)^n\\&=\exp\left(-\frac{1}{2}\sigma^2\lambda^2+i\lambda\mu+\alpha(\hat{\mu}(\lambda)-1)\right)=e^{\psi(\lambda)}\end{aligned}$$

where

$$\psi(\lambda)=-\frac{1}{2}\sigma^2\lambda^2+i\mu\lambda+\int_{\mathbb{R}}\left(e^{i\lambda x}-1\right)d\nu(x)$$

and $d\nu(x) := \alpha d\mu(x)$ is an arbitrary finite measure on $\mathbb{R}$. It is interesting to note that if $X_\alpha := \sum_{n \le N_\alpha} Z_n$ and $m \in \mathbb{N}$ then the law of the sum of m independent copies of $X_{\alpha/m}$ is the law of X_α. This explicitly shows that X_α is infinitely divisible.

In a related result, recall from Exercise 22.4 that for any finite measure ν on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, there exists a (necessarily unique) probability measure μ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that

$$\hat{\mu}(\lambda) = \exp\left(\int_\mathbb{R} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right).$$

Notice that any such probability measure μ is infinitely divisible since for each $n \in \mathbb{N}$ there exists a unique probability measure, μ_n, such that

$$\hat{\mu}_n(\lambda) = \exp\left(\frac{1}{n}\int_\mathbb{R} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right).$$

Keeping these two examples in mind should make the following important theorem plausible.

Theorem 23.21 (Lévy Kintchine formula). *A probability measure μ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is infinitely divisible iff $\hat{\mu}(\lambda) = e^{i\psi(\lambda)}$ where*

$$\psi(\lambda) = i\lambda b - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\le 1}\right) d\nu(x) \tag{23.17}$$

for some $b \in \mathbb{R}$, $a \ge 0$, and some measure ν on $\mathbb{R}\setminus\{0\}$ such that

$$\int_{\mathbb{R}\setminus\{0\}} \left(x^2 \wedge 1\right) d\nu(x) < \infty. \tag{23.18}$$

Proof. We will give the easy direction of this proof, namely the implication $(\Longleftarrow)$. Notice that if the measure ν appearing in Eq. (23.17) is a finite measure then

$$\psi(\lambda) = i\lambda b' - \frac{1}{2}a\lambda^2 + \int_\mathbb{R} \left(e^{i\lambda x} - 1\right) d\nu(x)$$

where

$$b' = b - \int_{|x|\le 1} x d\nu(x).$$

Thus we may use Example 23.20 in order to construct a random variable with distribution given by μ.

For general ν satisfying Eq. (23.18) let, for $\varepsilon > 0$, $d\nu_\varepsilon(x) := 1_{|x|\ge\varepsilon} d\nu(x)$ – a finite measure on $\mathbb{R}$. Thus by Example 23.20 there exits a probability measure, μ, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\hat{\mu}_\varepsilon = e^{\psi_\varepsilon}$ where

$$\psi_\varepsilon(\lambda) := i\lambda b - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\le 1}\right) 1_{|x|\ge\varepsilon} d\nu(x).$$

As $e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\le 1}$ is bounded and less than $C(\lambda)x^2$ for $|x| \le 1$, we may use DCT to show $\psi_\varepsilon(\lambda) \to \psi(\lambda)$. Furthermore the DCT also shows $\psi(\lambda)$ is continuous and therefore $e^{\psi(\lambda)}$ is continuous. Thus we have shown that $\mu_\varepsilon \to e^\psi$ where the limit is continuous and therefore by the continuity Theorem 22.18, there exists a probability measure μ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\hat{\mu} = e^\psi$.

The proof of the other implication ($\Longrightarrow$) will be discussed in Appendix 23.4 below. For more information about Poisson processes and Lévy processes; see Protter [42, Chapter I], [9, Chapter 9.5], and [18, Chapter XVII.2, p. 558-] for analytic proofs. Also see http://www.math.uconn.edu/˜bass/scdp.pdf, Kallenberg [30, Theorem 15.13, p. 294], and [1]. ∎

We are now going see that we may often drop the identically distributed assumption of the $\{X_{n,k}\}_{k=1}^n$ and yet still have that the weak limit of the sums of the form, $\sum_{k=1}^n X_{n,k}$, are still infinitely divisible distributions. In the next theorem we are going to see this is the case for weak limits under condition (M).

Theorem 23.22 (Limits under (M)). *Suppose $\{X_{n,k}\}_{k=1}^n$ satisfy property (M) and the normalizations in Assumption 2. If $S_n := \sum_{k=1}^n X_{n,k} \implies L$ for some random variable L, then*

$$f_L(\lambda) := \mathbb{E}\left[e^{i\lambda L}\right] = \exp\left(\int_\mathbb{R} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

for some finite positive measure, ν, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ with $\nu(\mathbb{R}) \le 1$.

Proof. As before, let $f_{n,k}(\lambda) = \mathbb{E}\left[e^{i\lambda X_{n,k}}\right]$ and $\varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1$. By the continuity theorem we are assuming

$$\lim_{n\to\infty} f_{S_n}(\lambda) = \lim_{n\to\infty} \prod_{k=1}^n f_{n,k}(\lambda) = f(\lambda)$$

where $f(\lambda)$ is continuous at $\lambda = 0$. We are also assuming property (M), i.e.

$$\lim_{n\to\infty} \max_{k\le n} \sigma_{n,k}^2 = 0.$$

Under condition (M), we expect $f_{n,k}(\lambda) \cong 1$ for n large. Therefore we expect

$$f_{n,k}(\lambda) = e^{\ln f_{n,k}(\lambda)} = e^{\ln[1 + (f_{n,k}(\lambda) - 1)]} \cong e^{(f_{n,k}(\lambda) - 1)}$$

and hence that

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^n f_{n,k}(\lambda) \cong \prod_{k=1}^n e^{(f_{n,k}(\lambda)-1)} = \exp\left(\sum_{k=1}^n (f_{n,k}(\lambda)-1)\right). \quad (23.19)$$

This is in fact correct, since Lemma 23.11 indeed implies

$$\lim_{n\to\infty}\left[\mathbb{E}\left[e^{i\lambda S_n}\right] - \exp\left(\sum_{k=1}^n (f_{n,k}(\lambda)-1)\right)\right] = 0. \quad (23.20)$$

Since $\mathbb{E}[X_{n,k}] = 0$,

$$\begin{aligned} f_{n,k}(\lambda) - 1 &= \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1\right] = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}\right] \\ &= \int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\mu_{n,k}(x) \end{aligned}$$

where $\mu_{n,k} := P \circ X_{n,k}^{-1}$ is the law of $X_{n,k}$. Therefore we have

$$\begin{aligned} \exp\left(\sum_{k=1}^n (f_{n,k}(\lambda)-1)\right) &= \exp\left(\sum_{k=1}^n \int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\mu_{n,k}(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) \sum_{k=1}^n d\mu_{n,k}(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\nu_n^*(x)\right) \quad (23.21) \end{aligned}$$

where $\nu_n^* := \sum_{k=1}^n \mu_{n,k}$. Let us further observe that

$$\int_{\mathbb{R}} x^2 d\nu_n^*(x) = \sum_{k=1}^n \int_{\mathbb{R}} x^2 d\mu_{n,k}(x) = \sum_{k=1}^n \sigma_{n,k}^2 = 1.$$

Hence if we define $d\nu_n(x) := x^2 d\nu_n^*(x)$, then ν_n is a probability measure and we have from Eqs. (23.20) and Eq. (23.21) that

$$\left| f_{S_n}(\lambda) - \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_n(x)\right)\right| \to 0. \quad (23.22)$$

Let

$$\varphi(\lambda, x) := \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} = -\frac{\lambda^2}{2}\int_0^1 e^{it\lambda x} 2(1-t)\,dt \quad (23.23)$$

(the second equality is from Taylor's theorem) and extend $\varphi(\lambda, \cdot)$ to $\bar{\mathbb{R}}$ by setting $\varphi(\lambda, \pm\infty) = 0$. Then $\{\varphi(\lambda, \cdot)\}_{\lambda\in\mathbb{R}} \subset C(\bar{\mathbb{R}})$ and therefore by Helly's selection Theorem 21.53 there is a probability measure $\bar{\nu}$ on $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ and a subsequence, $\{n_l\}$ of $\{n\}$ such that $\nu_{n_l}(\varphi(\lambda, \cdot)) \to \bar{\nu}(\varphi(\lambda, \cdot))$ for all $\lambda \in \mathbb{R}$ (in fact $\nu_{n_l}(h) \to \bar{\nu}(h)$ for all $h \in C(\bar{\mathbb{R}})$). Combining this with Eq. (23.22) allows us to conclude,

$$\begin{aligned} f_L(\lambda) &= \lim_{l\to\infty} \mathbb{E}\left[e^{i\lambda S_{n_l}}\right] = \lim_{l\to\infty} \exp\left(\int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\nu_{n_l}^*(x)\right) \\ &= \lim_{l\to\infty} \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu_{n_l}(x)\right) = \exp\left(\int_{\bar{\mathbb{R}}} \varphi(\lambda, x)\, d\bar{\nu}(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right) \end{aligned}$$

where $\nu := \bar{\nu}|_{\mathcal{B}_{\mathbb{R}}}$. The last equality follows from the fact that $\varphi(\lambda, \pm\infty) = 0$. The measure ν now satisfies, $\nu(\mathbb{R}) = \bar{\nu}(\mathbb{R}) \le \bar{\nu}(\bar{\mathbb{R}}) = 1$. ■

We are now going to drop the assumption that $\mathrm{Var}(S_n) = 1$ for all n and replace it with the following property.

Definition 23.23. *We say that $\{X_{n,k}\}_{k=1}^n$ has* ***bounded variation*** (BV) *iff*

$$\sup_n \mathrm{Var}(S_n) = \sup_n \sum_{k=1}^n \sigma_{n,k}^2 < \infty. \quad (23.24)$$

Corollary 23.24 (Limits under (BV)). *Suppose $\{X_{n,k}\}_{k=1}^n$ are independent mean zero random variables for each n which satisfy properties (M) and (BV). If $S_n := \sum_{k=1}^n X_{n,k} \implies L$ for some random variable L, then*

$$f_L(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right) \quad (23.25)$$

where ν – is a finite positive measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Proof. Let $s_n^2 := \mathrm{Var}(S_n)$. If $\lim_{n\to\infty} s_n = 0$, then $S_n \to 0$ in L^2 and hence weakly, therefore Eq. (23.25) holds with $\nu \equiv 0$. So let us now suppose $\lim_{n\to\infty} s_n \neq 0$. Since $\{s_n\}_{n=1}^\infty$ is bounded, we may by passing to a subsequence if necessary, assume $\lim_{n\to\infty} s_n = s > 0$. By replacing $X_{n,k}$ by $X_{n,k}/s_n$ and hence S_n by S_n/s_n, we then know by Slutzky's Theorem 21.39 that $S_n/s_n \implies L/s$. Hence by an application of Theorem 23.22, we may conclude

$$f_L(\lambda/s) = f_{L/s}(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

where ν – is a finite positive measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\nu(\mathbb{R}) \le 1$. Letting $\lambda \to s\lambda$ in this expression then implies

$$f_L(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda sx} - 1 - i\lambda sx}{x^2} d\nu(x)\right)$$
$$= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda sx} - 1 - i\lambda sx}{(sx)^2} s^2 d\nu(x)\right)$$
$$= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_s(x)\right)$$

where ν_s is the finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ defined by

$$\nu_s(A) := s^2 \nu\left(s^{-1}A\right) \text{ for all } A \in \mathcal{B}_{\mathbb{R}}.$$

■

From Eq. (23.23) we see that $\varphi(\lambda, x) := \left(e^{i\lambda x} - 1 - i\lambda x\right)/x^2$ is a smooth function of (λ, x). Moreover,

$$\frac{d}{d\lambda}\varphi(\lambda, x) = \frac{ixe^{i\lambda x} - ix}{x^2} = i\frac{e^{i\lambda x} - 1}{x}$$

and

$$\frac{d^2}{d\lambda^2}\varphi(\lambda, x) = i\frac{ixe^{i\lambda x}}{x} = -e^{i\lambda x}.$$

Using these remarks and the fact that $\nu(\mathbb{R}) < \infty$, it is easy to see that

$$f_L'(\lambda) = \left(\int_{\mathbb{R}} i\frac{e^{i\lambda x} - 1}{x} d\nu_s(x)\right) f_L(\lambda)$$

and

$$f_L''(\lambda) = \left(\int_{\mathbb{R}} -e^{i\lambda x} d\nu_s(x) + \left[\left(\int_{\mathbb{R}} i\frac{e^{i\lambda x} - 1}{x} d\nu_s(x)\right)^2\right]\right) f_L(\lambda)$$

and in particular, $f_L'(0) = 0$ and $f_L''(0) = -\nu_s(\mathbb{R})$. Therefore by Theorem 22.8 the probability measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\hat{\mu}(\lambda) = f_L(\lambda)$ has mean zero and variance, $\nu_s(\mathbb{R}) < \infty$. This later condition reflects the (BV) assumption that we made.

Theorem 23.25. *The following class of* ***symmetric*** *distributions on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *are equal;*

1. *C_1 – all possible limiting distributions under properties (M) and (BV).*
2. *C_2 – all distributions with characteristic functions of the form given in Corollary 23.24.*
3. *C_3 – all infinitely divisible distributions with mean zero and finite variance.*

Proof. The inclusion, $C_1 \subset C_2$, is the content of Corollary 23.24. For $C_2 \subset C_3$, observe that if

$$\hat{\mu}(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

then $\hat{\mu}(\lambda) = [\hat{\mu}_n(\lambda)]^n$ where μ_n is the unique probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that

$$\hat{\mu}_n(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} \frac{1}{n} d\nu(x)\right).$$

For $C_3 \subset C_1$, simply define $\{X_{n,k}\}_{k=1}^n$ to be i.i.d with $\mathbb{E}\left[e^{i\lambda X_{n,k}}\right] = \hat{\mu}_n(\lambda)$. In this case $S_n = \sum_{k=1}^n X_{n,k} \overset{d}{=} \mu$. ■

23.3 Stable Distributions

Definition 23.26. *A non-degenerate distribution $\mu = \mathrm{Law}(X)$ on $\mathbb{R}$ is* ***stable*** *if whenever X_1 and X_2 are independent copies of X, then for all $a, b \in \mathbb{R}$ there exists $c, d \in \mathbb{R}$ such that $aX_1 + bX_2 \overset{d}{=} cX + d$ with some constants c and d.*

Example 23.27. Any Gaussian random variable is stable. Indeed if $X \overset{d}{=} \sigma N + \mu$ where $\sigma > 0$ and $\mu \in \mathbb{R}$ and $N = N(0,1)$, then $X_i = \sigma N_i + \mu$ where N_1 and N_2 are independent with $N_i \overset{d}{=} N$ we will have $aX_1 + bX_2$ is Gaussian mean $(a+b)\mu$ and variance $\left(a^2 + b^2\right)\sigma^2$ so that

$$aX_1 + bX_2 \overset{d}{=} \sqrt{(a^2 + b^2)}\sigma N + (a+b)\mu$$
$$\overset{d}{=} \sqrt{(a^2 + b^2)}(X - \mu) + (a+b)\mu$$
$$= \sqrt{(a^2 + b^2)}X + \left(a + b - \sqrt{(a^2 + b^2)}\right)\mu.$$

Example 23.28. Poisson random variables are not stable. For suppose that $Z = Pois(\rho)$ and $Z_1 \overset{d}{=} Z_2 \overset{d}{=} Z$, then $Z_1 + Z_2 \overset{d}{=} Pois(2\rho)$. If we could find a, b such that

$$Pois(2\rho) \overset{d}{=} Z_1 + Z_2 \overset{d}{=} aZ + b$$

we would have

$$e^{-2\rho}\frac{(2\rho)^n}{n!} = P(aZ + b = n) = P\left(Z = \frac{n-b}{a}\right) \text{ for all } n.$$

In particular this implies that $\frac{n-b}{a} = k_n \in \mathbb{N}_0$ for all $n \in \mathbb{N}_0$ and the map $n \to k_n$ must be invertible so as probabilities are conserved. This can only be the case if $a = 1$ and $b = 0$ and we would conclude that $Z \overset{d}{=} Pois(2\rho)$ which is absurd.

Lemma 23.29. *Suppose that $\{X_i\}_{i=1}^n$ are i.i.d. random variables such that $X_1 + \cdots + X_n = c$ a.s., then $X_i = c/n$ a.s.*

Proof. Let $f(\lambda) := \mathbb{E}e^{i\lambda X_1}$, then

$$e^{i\lambda c} = \mathbb{E}\left[e^{i\lambda(X_1+\cdots+X_n)}|X_1\right] = e^{i\lambda X_1} f(\lambda) \text{ a.s.}$$

from which it follows that $f(\lambda) = e^{i\lambda(c-X_1)}$ a.s. and in particular for an ω where this equality holds we find, $f(\lambda) = e^{i\lambda(c-X_1(\omega))} = e^{i\lambda c'}$. By uniqueness of the Fourier transform it follows that $X_1 = c'$ a.s. and therefore $c = X_1 + \cdots + X_n = nc'$ a.s., i.e. $c' = c/n$. ■

Lemma 23.30. *If μ is a stable distribution then it is infinitely divisible.*

Proof. Let $\{X_n\}_{n=1}^N$ be i.i.d. random variables with $\operatorname{Law}(X_n) = \mu = \operatorname{Law}(X)$. As μ is stable we know that

$$X_1 + \cdots + X_N \stackrel{d}{=} aX + b. \tag{23.26}$$

As μ is non-degenerate, it follows from Lemma 23.29 that $a \neq 0$, Therefore form Eq. (23.26) we find,

$$X \stackrel{d}{=} \sum_{i=1}^N \frac{1}{a}(X_i - b/N)$$

and this shows that X is infinitely divisible. ■

The converse of this lemma is not true as is seen by considering Poisson random variables, see Example 23.28. The following characterization of the stable law may be found in [9, Chapter 9.9]. For a whole book about stable laws and their properties see Samorodnitsky and Taqqu [47].

Theorem 23.31. *A probability measure μ on $\mathbb{R}$ is a stable distribution iff μ is Gaussian or $\hat{\mu}(\lambda) = e^{\psi(\lambda)}$ where*

$$\psi(\lambda) = i\lambda b + \int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - \frac{i\lambda x}{1+x^2}\right) \frac{m_1 1_{x>0} - m_2 1_{x<0}}{|x|^{1+\alpha}} dx$$

for some constants, $0 < \alpha < 2$, $m_i \geq 0$ and $b \in \mathbb{R}$.

To get some feeling for this theorem. Let us consider the case of a stable random variable X which is **also** assumed to be symmetric. In this case if X_1, X_2 are independent copies of X and $a, b \in \mathbb{R}$ and $c = c(a_1, a_2)$ and $d = d(a_1, a_2)$ are then chosen so that $aX_1 + bX_2 \stackrel{d}{=} cX + d$, we must have that $d = 0$ and may take $c > 0$ by the symmetry assumption. Letting $f(\lambda) = \mathbb{E}\left[e^{i\lambda X}\right]$ we may now conclude that

$$f(a\lambda) \cdot f(b\lambda) = \mathbb{E}\left[e^{i\lambda(aX_1+bX_2)}\right] = \mathbb{E}\left[e^{i\lambda cX}\right] = f(c\lambda).$$

It turns out the solution to these functional equation are of the form $f(\lambda) = e^{-k|\lambda|^\alpha}$. If $f(\lambda)$ is of this form then

$$f(a\lambda) \cdot f(b\lambda) = \exp\left(-k\left(|a|^\alpha + |b|^\alpha\right)|\lambda|^\alpha\right) = f(c\lambda)$$

where $c = \left(|a|^\alpha + |b|^\alpha\right)^{1/\alpha}$. Moreover it turns out the f is a characteristic function when $0 < \alpha \leq 2$. The case $\alpha = 2$ is the Gaussian case, then case $\alpha = 1$ is the Cauchy distribution, for example if

$$d\mu(x) = \frac{1}{\pi(1+x^2)} dx \text{ then } \hat{\mu}(\lambda) = e^{-|\lambda|}.$$

For $\alpha \leq 1$ we find that we have

$$\begin{aligned} f'(\lambda) &= -k|\lambda|^{\alpha-1} f(\lambda) \leq 0 \text{ and} \\ f''(\lambda) &= \left[k^2|\lambda|^{2\alpha-2} - k(\alpha-1)|\lambda|^{\alpha-2}\right] f(\lambda) \geq 0 \end{aligned}$$

so that f is a decreasing convex symmetric function for $\lambda \geq 0$. Therefore by Polya's criteria of Exercise 22.6 it follows that $e^{-k|\lambda|^\alpha}$ is the characteristic function of a probability measure for $0 \leq \alpha \leq 1$. The full proof is not definitely not given here.

23.4 *Appendix: Lévy exponent and Lévy Process facts – Very Preliminary!!

We would like to characterize all processes with independent stationary increments with values in $\mathbb{R}$ or more generally $\mathbb{R}^d$. We begin with some more examples.

Proposition 23.32. *For every finite measure ν, the function*

$$f(\lambda) := \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

is the characteristic function of a probability measure, $\mu = \mu_\nu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The convention here is that

$$\frac{e^{i\lambda x} - 1 - i\lambda x}{x^2}|_{x=0} := \lim_{x\to 0} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} = -\frac{1}{2}\lambda^2.$$

Proof. This is the content of Exercise 22.4 ■

1. If $\{X_t\}_{t\geq 0}$ is a right continuous process with stationary and independent increments, then let $f_t(\lambda) := \mathbb{E}\left[e^{i\lambda(X_{t+\sigma}-X_\sigma)}\right]$ for any $\sigma \geq 0$. It then follows that

$$\begin{aligned} f_{t+s}(\lambda) &= \mathbb{E}\left[e^{i\lambda(X_{t+s}-X_0)}\right] = \mathbb{E}\left[e^{i\lambda(X_{t+s}-X_t+X_t-X_0)}\right] \\ &= \mathbb{E}\left[e^{i\lambda(X_{t+s}-X_t)}\right]\cdot\mathbb{E}\left[e^{i\lambda(X_t-X_0)}\right] \\ &= f_s(\lambda)\cdot f_t(\lambda). \end{aligned}$$

The right continuity of X_t now insures that f_t is also right continuous. The only solution to the above functional equation is therefore of the form, $f_t(\lambda) = e^{t\psi(\lambda)}$ for some function $\psi(\lambda)$. Since

$$e^{t\,\mathrm{Re}\,\psi(\lambda)} = |f_t(\lambda)| \leq 1$$

it follows that $\mathrm{Re}\,\psi(\lambda) \leq 0$. Let $\lambda \in \mathbb{R}$ be fixed and define $h(t) := f_t(\lambda)$, then h is right continuous, $h(0) = 1$, and $h(t+s) = h(t)h(s)$. Let ln be a branch of the logarithm defined near 1 such that $\ln 1 = 0$. Then there exists ε such that for all $t \leq \varepsilon$ we have $g(t) := \ln h(t)$ is well defined and $g(t)$ satisfies, $g(t+s) = g(t) + g(s)$ for all $0 \leq s, t \leq \varepsilon$. We now set $g_\varepsilon(t) := g(\varepsilon t)$ and then $g_\varepsilon(s+t) = g_\varepsilon(s) + g_\varepsilon(t)$ for all $0 \leq s, t \leq 1$ and is still right continuous. As usual it now follows that $g_\varepsilon(1) = g_\varepsilon(n\cdot 1/n) = n\cdot g_\varepsilon(1/n)$ for all n and therefore for all $0 \leq k \leq n$, we have $g_\varepsilon(k/n) = \frac{k}{n}g_\varepsilon(1)$. Using the right continuity of g_ε it now follows that $g_\varepsilon(t) = tg_\varepsilon(1)$ for all $0 \leq t < 1$. Thus we have shown $g(\varepsilon t) = tg(\varepsilon)$ for $0 \leq t < 1$ and therefore if we set $\theta := g(\varepsilon)/\varepsilon$ we have shown $g(t) = t\theta$ for $t \in [0, \varepsilon)$ that is ,

$$h(t) = e^{t\theta} \text{ for } 0 \leq t < \varepsilon.$$

This formula is now seen to be correct for all $t \geq 0$. Indeed if $t = k\varepsilon/2 + \tau$ with $0 \leq \tau < \varepsilon/2$, then

$$h(t) = h(\varepsilon/2)^k h(\tau) = \left[e^{\theta\varepsilon/2}\right]^k e^{\tau\theta} = e^{\theta[k\varepsilon/2+\tau]} = e^{t\theta}.$$

Thus we have shown that $f_t(\lambda) = e^{t\psi(\lambda)}$ for some function $\psi(\lambda)$. Let us further observe that

$$\psi(\lambda) = \lim_{t\downarrow 0}\frac{f_t(\lambda)-1}{t}$$

from which it follows that ψ must be measurable. Furthermore,

$$\psi(-\lambda) = \lim_{t\downarrow 0}\frac{f_t(-\lambda)-1}{t} = \lim_{t\downarrow 0}\frac{\overline{f_t(\lambda)}-1}{t} = \overline{\psi(\lambda)}.$$

We are going to show more.

2. Let $\{z_i\}_{i=1}^n \subset \mathbb{C}$ such that $\sum_{i=1}^n z_i = 1$ and $\{\lambda_i\}_{i=1}^n \subset \mathbb{R}$, then

$$\begin{aligned} \sum_{i,j=1}^n \psi(\lambda_i-\lambda_j) z_i\bar{z}_j &= \lim_{t\downarrow 0}\sum_{i,j=1}^n \frac{f_t(\lambda_i-\lambda_j)-1}{t} z_i\bar{z}_j \\ &= \lim_{t\downarrow 0}\frac{1}{t}\sum_{i,j=1}^n f_t(\lambda_i-\lambda_j) z_i\bar{z}_j \end{aligned}$$

while for any $\{z_i\}_{i=1}^n \subset \mathbb{C}$ we have

$$\begin{aligned} \sum_{i,j=1}^n f_t(\lambda_i-\lambda_j) z_i\bar{z}_j &= \sum_{i,j=1}^n \mathbb{E}\left[e^{i(\lambda_i-\lambda_j)X_t}\right] z_i\bar{z}_j \\ &= \sum_{i,j=1}^n \mathbb{E}\left[e^{i\lambda_i X_t} z_i\cdot e^{-i\lambda_j X_t}\bar{z}_j\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n e^{i\lambda_i X_t} z_i\cdot\sum_{j=1}^n e^{-i\lambda_j X_t}\bar{z}_j\right] \\ &= \mathbb{E}\left[\left|\sum_{i=1}^n e^{i\lambda_i X_t} z_i\right|^2\right] \geq 0. \end{aligned}$$

Therefore it follows that when $\sum_{i=1}^n z_i = 1$ then $\sum_{i,j=1}^n \psi(\lambda_i-\lambda_j) z_i\bar{z}_j \geq 0$. We say the ψ is conditionally positive definite in this case.
3. The Schoenberg correspondence says (see [1, Theorem 1.1.13]) that if ψ is continuous at zero, $\psi(-\lambda) = \overline{\psi(\lambda)}$ and ψ is conditionally positive definite, then $e^{t\psi(\lambda)}$ is a characteristic function. We will prove this below using Bochner's Theorem 22.43.
4. But first some examples;
   a) Let $\psi(\lambda) = i\lambda a - b\lambda^2$ with $a \in \mathbb{R}$ and $b \geq 0$.Then $\psi(-\lambda) = -i\lambda a - b\lambda^2 = \overline{\psi(\lambda)}$ and for $\sum_{i=1}^n z_i = 1$ we have

$$\sum_{i,j=1}^n \psi(\lambda_i-\lambda_j) z_i\bar{z}_j = \sum_{i,j=1}^n \left[-i(\lambda_i-\lambda_j)a - b(\lambda_i-\lambda_j)^2\right] z_i\bar{z}_j.$$

   Noting that

$$\sum_{i,j=1}^n \lambda_i z_i\bar{z}_j = \sum_{i=1}^n \lambda_i z_i\sum_{j=1}^n \bar{z}_j = \sum_{i=1}^n \lambda_i z_i\cdot 0 = 0$$

   and similarly that $\sum_{i,j=1}^n \left[\lambda_i^2\right] z_i\bar{z}_j = 0$, it follows that

$$\sum_{i,j=1}^{n}\psi(\lambda_i-\lambda_j)z_i\bar{z}_j=\sum_{i,j=1}^{n}\left[-b(-2\lambda_i\lambda_j)^2\right]z_i\bar{z}_j$$
$$=2b\left|\sum_{i,j=1}^{n}\lambda_i z_i\right|^2\geq 0.$$

b) Suppose that $\{Z_i\}_{i=1}^{\infty}$ are i.i.d. random variables and $\{N\}$ is an independent Poisson process with intensity λ. Let $X:=Z_1+\cdots+Z_N$, then

$$f_X(\lambda)=\mathbb{E}\left[e^{i\lambda X}\right]=\sum_{n=0}^{\infty}\mathbb{E}\left[e^{i\lambda X}:N=n\right]$$
$$=\sum_{n=0}^{\infty}\mathbb{E}\left[e^{i\lambda[Z_1+\cdots+Z_n]}:N=n\right]$$
$$=e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\left[f_{Z_1}(\lambda)\right]^n=\exp\left(\lambda\left(f_{Z_1}(\lambda)-1\right)\right).$$

So in this case $\psi(\lambda)=f_{Z_1}(\lambda)-1$ and we know by the theory above that $\psi(\lambda)$ is conditionally positive definite.

Lemma 23.33. *Suppose that $\{A_{ij}\}_{i,j=1}^{d}\subset\mathbb{C}$ is a matrix such that $A^*=A$ and $A\geq 0$. Then for all $n\in\mathbb{N}_0$, the matrix with entries $\left(A_{ij}^n\right)_{i,j=1}^n$ is positive semi-definite.*

Proof. Since $A_{ij}=(Ae_j,e_i)$ where $(v,w):=\sum_{j=1}^{d}v_j\bar{w}_j$ is the standard inner product on $\mathbb{C}^d$, it follows that

$$A_{ij}^n=\left(A^{\otimes n}e_j^{\otimes n},e_i^{\otimes n}\right)$$

and therefore,

$$\sum_{i,j=1}^{d}A_{ij}^n\bar{z}_i z_j=\sum_{i,j=1}^{d}\left(A^{\otimes n}e_j^{\otimes n},e_i^{\otimes n}\right)\bar{z}_i z_j=\left(A^{\otimes n}\psi,\psi\right)$$

where $\psi:=\sum_{j=1}^{d}z_j e_j^{\otimes n}\in\left(\mathbb{C}^d\right)^{\otimes n}$. So it suffices to show $A^{\otimes n}\geq 0$. To do this let $\{u_i\}_{i=1}^{d}$ be an O.N. basis for $\mathbb{C}^d$ such that $Au_i=\lambda_i u_i$ for all i. Since $A\geq 0$ we know that $\lambda_i\geq 0$ and therefore

$$A^{\otimes n}\left(u_{i_1}\otimes\cdots\otimes u_{i_d}\right)=\left(\lambda_{i_1}\dots\lambda_{i_d}\right)\left(u_{i_1}\otimes\cdots\otimes u_{i_d}\right)$$

where $\left(\lambda_{i_1}\dots\lambda_{i_d}\right)\geq 0$. This shows that $A^{\otimes n}$ is unitarily equivalent to a diagonal matrix with non-negative entries and hence is positive semi-definite. ∎

Proposition 23.34. *Suppose that $\{A_{ij}\}_{i,j=1}^{d}\subset\mathbb{C}$ is a matrix such that $A^*=A$ and A is conditionally positive definite, for example $A_{ij}:=\psi(\lambda_i-\lambda_j)$ as above. Then the matrix with entries, $\left(e^{A_{ij}}\right)_{i,j=1}^{d}$ is positive definite.*

Proof. Let $u:=(1,\dots,1)^{\mathrm{tr}}\in\mathbb{C}^d$. Let $\xi\in\mathbb{C}^d$ and write $\xi=z+\alpha u$ where $(z,u)=0$ and $\alpha:=(\xi,u)/d$. Letting $B:=\sqrt{A}$ on $u^{\perp}$ and 0 on $\mathbb{C}\cdot u$, we have

$$\begin{aligned}(A\xi,\xi)&=(A(z+\alpha u),z+\alpha u)\\&=(Az,z)+2\,\mathrm{Re}\left[\bar{\alpha}(Az,u)\right]+|\alpha|^2(Au,u)\\&=(Az,z)+2\,\mathrm{Re}\left[\bar{\alpha}\left(B^2z,u\right)\right]+|\alpha|^2(Au,u)\\&=(Az,z)+2\,\mathrm{Re}\left[\bar{\alpha}(Bz,B^*u)\right]+|\alpha|^2(Au,u)\\&\geq(Az,z)-2\|Bz\|\cdot|\alpha|\,\|B^*u\|+|\alpha|^2(Au,u)\\&\geq(Az,z)-\left[\|Bz\|^2+|\alpha|^2\|B^*u\|^2\right]+|\alpha|^2(Au,u)\\&=|\alpha|^2\left[(Au,u)-\|B^*u\|^2\right].\end{aligned}$$

Since

$$\left(u\,u^{\mathrm{tr}}\xi,\xi\right)=|\alpha|^2\left(u\,u^{\mathrm{tr}}u,u\right)=|\alpha|^2d^2,$$

it follows that

$$\left(\left(A+\lambda u\,u^{\mathrm{tr}}\right)\xi,\xi\right)\geq|\alpha|^2\left[(Au,u)-\|B^*u\|^2+\lambda d^2\right]\geq 0$$

provided $\lambda d^2\geq\|B^*u\|^2-(Au,u)$.

We now fix such a $\lambda\in\mathbb{R}$ so that $\left(A+\lambda u\,u^{\mathrm{tr}}\right)\geq 0$. It then follows from Lemma 23.33 that

$$e^{\lambda}e^{A_{ij}}=e^{A_{ij}+\lambda}=e^{\left(A+\lambda u\ u^{\mathrm{tr}}\right)_{ij}}=\sum_{n=0}^{\infty}\frac{\left(A+\lambda u\ u^{\mathrm{tr}}\right)_{ij}^n}{n!}$$

are the matrix entries of a positive definite matrix. Scaling this matrix by $e^{-\lambda}>0$ then gives the result that $\left(e^{A_{ij}}\right)_{i,j}\geq 0$. ∎

As a consequence it follows that $e^{t\psi(\lambda)}$ is a positive definite function whenever ψ is conditionally positive definite.

Proposition 23.35. *Suppose that $\{Z_i\}_{i=1}^{\infty}$ are i.i.d. random vectors in $\mathbb{R}^d$ with* Law$(Z_i)=\mu$ *and $\{N_t\}_{t\geq 0}$ be an independent Poisson process with intensity λ. Then $\{X_t:=S_{N_t}\}_{t\geq 0}$ is a Lévy process with $\mathbb{E}\left[e^{ik\cdot X_t}\right]=e^{t\psi(k)}$ where*

$$\psi(k)=\lambda\int_{\mathbb{R}^n}\left(e^{ik\cdot x}-1\right)d\mu(x)=\mathbb{E}\left[e^{ik\cdot Z_1}\right].$$

Proof. It has already been shown in Theorem 17.28 that $\{X_t\}_{t\geq 0}$ has stationary independent increments and being right continuous it is a Lévy process. It only remains to compute the Fourier transform,

$$\begin{aligned}\mathbb{E}\left[e^{ik\cdot X_t}\right] &= \left[Q_t\left(x \to e^{ik\cdot x}\right)\right](0) \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t}\mathbb{E}\left[e^{ik\cdot(Z_1+\cdots+Z_n)}\right] = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t}\hat{\mu}(k)^n \\ &= e^{t\lambda(\hat{\mu}(k)-1)} = \exp\left(t\lambda \int_{\mathbb{R}^n}\left(e^{ik\cdot x}-1\right)d\mu(x)\right).\end{aligned}$$

■

More generally, if we let B_t be Brownian motion in $\mathbb{R}^n$ with $\mathrm{Cov}\left(B_t^i, B_t^j\right) = A_{ij}t$ and $b \in \mathbb{R}^n$, then assuming B and X above are independent, then $X_t = bt + B_t + X_t$ is again a Levy process whose Fourier transform is given by,

$$\mathbb{E}\left[e^{ik\cdot X_t}\right] = \exp\left(ibt + Ak\cdot k + \lambda \int_{\mathbb{R}^n}\left(e^{ik\cdot x}-1\right)d\mu(x)\right).$$

Thus

$$\psi(\lambda) = ibt + Ak\cdot k + \lambda \int_{\mathbb{R}^n}\left(e^{ik\cdot x}-1\right)d\mu(x)$$

is a Lévy exponent for all choice of $b \in \mathbb{R}^n$, all $\lambda > 0$, probability measures μ on $\mathbb{R}^n$, and $A \geq 0$.

Lévy proved that in general $\psi(k)$ will be a Lévy exponent iff ψ has the form given in Eq. (23.27) below.

Theorem 23.36 (Lévy Kintchine formula). *If ψ is continuous at zero and conditionally positive definite, then*

$$\psi(\lambda) = i\lambda b - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}\setminus\{0\}}\left(e^{i\lambda x} - 1 - i\lambda x\cdot 1_{|x|\leq 1}\right)d\nu(x) \tag{23.27}$$

for some $b \in \mathbb{R}$, $a \geq 0$, and some measure ν such that

$$\int_{\mathbb{R}\setminus\{0\}}\left(x^2 \wedge 1\right)d\nu(x) < \infty.$$

Part V

Stochastic Processes II

We are now going to discuss continuous time stochastic processes in more detail. We will be using Poisson processes (Definition 11.7) and Brownian motion (Definition 17.21) as our model cases. Up to now we have not proved the existence of Brownian motion. This lapse will be remedied in the next couple of chapters. We are going to begin by constructing a process $\{B_t\}_{t\geq 0}$ satisfying all of the properties of a Brownian motion in Definition 17.21 except for the continuity of the sample paths. We will then use Kolmogorov's continuity criteria (Theorem 25.7) to show we can "modify" this process in such a way so as to produce an example of Brownian motion. We start with a class of random fields which are relatively easy to understand. (BRUCE – metion the free Euclidean field and its connections to SLE.)

Gaussian Random Fields

Recall from Section 9.8 (which the reader should review if necessary) that a random variable, $Y : \Omega \to \mathbb{R}$ is said to be **Gaussian** if

$$\mathbb{E}e^{i\lambda Y} = \exp\left(-\frac{1}{2}\lambda^2 \operatorname{Var}(Y) + i\lambda\mathbb{E}Y\right) \ \forall\ \lambda \in \mathbb{R}.$$

More generally a random vector, $X : \Omega \to \mathbb{R}^N$, is said to be **Gaussian** if $\lambda \cdot X$ is a Gaussian random variable for all $\lambda \in \mathbb{R}^N$. Equivalently put, $X : \Omega \to \mathbb{R}^N$ is **Gaussian** provided

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right] = \exp\left(-\frac{1}{2}\operatorname{Var}(\lambda \cdot X) + i\mathbb{E}(\lambda \cdot X)\right) \ \forall\ \lambda \in \mathbb{R}^N. \tag{24.1}$$

Remark 24.1. To conclude that a random vector, $X : \Omega \to \mathbb{R}^N$, is Gaussian it is **not** enough to check that each of its components are Gaussian random variables. The following simple counter example was provided by Nate Eldredge. Let $X \overset{d}{=} N(0,1)$ and Y be an independent Bernoulli random variable with $P(Y=1) = P(Y=-1) = 1/2$. Then the random vector, $(X, X\cdot Y)^{\mathrm{tr}}$ has Gaussian components but is not Gaussian.

Exercise 24.1 (Same as Exercise 9.9.). Prove the assertion made in Remark 24.1 by computing $\mathbb{E}\left[e^{i(\lambda_1 X + \lambda_2 XY)}\right]$. (Another proof that $(X, X\cdot Y)^{\mathrm{tr}}$ is not Gaussian follows from the fact that X and XY are uncorrelated but not independent[1] which would then contradict Lemma 10.24.)

24.1 Gaussian Integrals

The following theorem gives a useful way of computing Gaussian integrals of polynomials and exponential functions.

[1] To formally see that they are not independent, observe that $|X| \le \frac{1}{2}$ iff $|XY| \le \frac{1}{2}$ and therefore,

$$P\left(|X| \le \frac{1}{2} \text{ and } |XY| \le \frac{1}{2}\right) = P\left(|X| \le \frac{1}{2}\right) =: \alpha$$

while

$$P\left(|X| \le \frac{1}{2}\right) P\left(|XY| \le \frac{1}{2}\right) = \alpha^2 \neq \alpha.$$

Theorem 24.2. *Suppose $X \overset{d}{=} N(Q, 0)$ where Q is a $N \times N$ symmetric positive definite matrix. Let $L = L^Q := Q_{ij}\partial_i\partial_j$ (sum on repeated indices) where $\partial_i := \partial/\partial x_i$. Then for any polynomial function, $q : \mathbb{R}^N \to \mathbb{R}$,*

$$\mathbb{E}[q(X)] = \left(e^{\frac{1}{2}L}q\right)(0) := \sum_{n=0}^{\infty} \frac{1}{n!}\left(\left(\frac{L}{2}\right)^n q\right)(0) \ \textit{(a finite sum)}. \tag{24.2}$$

Proof. First Proof. The first proof is conceptually clear but technically a bit more difficult. In this proof we will begin by proving Eq. (24.2) when $q(x) = e^{i\lambda\cdot x}$ where $\lambda \in \mathbb{R}^N$. The function q is not a polynomial, but never mind. In this case,

$$\mathbb{E}[q(X)] = \mathbb{E}\left[e^{i\lambda\cdot X}\right] = e^{-\frac{1}{2}Q\lambda\cdot\lambda}.$$

On the other hand,

$$\left(\frac{1}{2}Lq\right)(x) = \frac{1}{2}Q_{ij}\partial_i\partial_j e^{i\lambda\cdot x} = \frac{1}{2}(Q\lambda\cdot\lambda)\, e^{i\lambda\cdot x} = -\frac{1}{2}(Q\lambda\cdot\lambda)\, q(x).$$

Therefore,

$$e^{\frac{1}{2}L}q = \sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n q = e^{-\frac{1}{2}Q\lambda\cdot\lambda}q$$

and hence

$$\left(e^{\frac{1}{2}L}q\right)(0) = e^{-\frac{1}{2}Q\lambda\cdot\lambda}.$$

Thus we have shown

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right] = e^{\frac{1}{2}L}e^{i\lambda\cdot x}|_{x=0}.$$

The result now formally follows by differentiating this equation in λ and then setting $\lambda = 0$. Indeed observe that

$$\begin{aligned}\mathbb{E}\left[(iX)^\alpha\right] &= \partial_\lambda^\alpha \mathbb{E}\left[e^{i\lambda\cdot X}\right]|_{\lambda=0} = \partial_\lambda^\alpha e^{\frac{1}{2}L}e^{i\lambda\cdot x}|_{x=0,\lambda=0} \\ &= e^{\frac{1}{2}L}\partial_\lambda^\alpha e^{i\lambda\cdot x}|_{x=0,\lambda=0} = e^{\frac{1}{2}L}(ix)^\alpha|_{x=0}.\end{aligned}$$

To justify this last equation we must show,

$$\partial_\lambda^\alpha e^{\frac{1}{2}L}e^{i\lambda\cdot x} = e^{\frac{1}{2}L}\partial_\lambda^\alpha e^{i\lambda\cdot x}$$

which is formally true since mixed partial derivatives commute. However there is also an infinite sum involved so we have to be a bit more careful. To see what is involved, on one hand

$$\partial_\lambda^\alpha e^{\frac{1}{2}L} e^{i\lambda\cdot x} = \partial_\lambda^\alpha \sum_{n=0}^\infty \frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}$$

while on the other,

$$\begin{aligned} e^{\frac{1}{2}L}\partial_\lambda^\alpha e^{i\lambda\cdot x} &= \sum_{n=0}^\infty \frac{1}{n!}\left(\left(\frac{L}{2}\right)^n \partial_\lambda^\alpha e^{i\lambda\cdot x}\right) = \sum_{n=0}^\infty \frac{1}{n!}\partial_\lambda^\alpha\left(\left(\frac{L}{2}\right)^n e^{i\lambda\cdot x}\right) \\ &= \sum_{n=0}^\infty \partial_\lambda^\alpha\left(\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}\right). \end{aligned}$$

Thus to complete the proof we must show,

$$\partial_\lambda^\alpha \sum_{n=0}^\infty \frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x} = \sum_{n=0}^\infty \partial_\lambda^\alpha\left(\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}\right).$$

Perhaps the easiest way to do this would be to use the Cauchy estimates[2] which allow one to show that if $\{f_n(\lambda)\}_{n=0}^\infty$ is a sequence of analytic functions such that $\sum_{n=0}^\infty f_n(\lambda)$ is uniformly convergent on compact subsets, then $\sum_{n=0}^\infty \partial_\lambda^\alpha f_n(\lambda)$ is also uniformly convergent on compact subsets and therefore,

$$\partial_\lambda^\alpha \sum_{n=0}^\infty f_n(\lambda) = \sum_{n=0}^\infty \partial_\lambda^\alpha f_n(\lambda).$$

Now apply this result with $f_n(\lambda) := \frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}$ to get the result. The details are left to the reader.

This proof actually shows more than what is claimed. Namely, 1. Q may be only non-negative definite and 2. Eq. (24.2) holds for $q(x) = p(x)e^{i\lambda\cdot x}$ where $\lambda\in\mathbb{R}^N$ and p is a polynomial.

Second Proof. Let

$$\begin{aligned} u(t,y) &:= \mathbb{E}\left[q\left(y+\sqrt{t}X\right)\right] = Z^{-1}\int_{\mathbb{R}^N} q\left(y+\sqrt{t}x\right)e^{-Q^{-1}x\cdot x/2}dx \qquad (24.3) \\ &= Z^{-1}\int_{\mathbb{R}^N} q(y+x)\frac{e^{-Q^{-1}x\cdot x/2t}}{t^{n/2}}dx. \qquad (24.4) \end{aligned}$$

[2] If you want to avoid the Cauchy estimates it would suffice to show by hand that

$$\sum_{n=0}^\infty \sup_{|\lambda|\le R}\left|\partial_\lambda^\alpha\left(\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}\right)\right| < \infty$$

for all multi- indices, α.

One now verifies that

$$\partial_t \frac{e^{-Q^{-1}x\cdot x/2t}}{t^{n/2}} = \frac{1}{2}L\frac{e^{-Q^{-1}x\cdot x/2t}}{t^{n/2}}.$$

Using this result and differentiating under the integral in Eq. (24.4) then shows,

$$\partial_t u(t,y) = \frac{1}{2}L_y u(t,y) \text{ with } u(0,y) = q(y).$$

Moreover, from Eq. (24.3), one easily sees that $u(t,y)$ is a polynomial in (t,y) and the degree in y is the same as the degree of q. On the other hand,

$$v(t,y) := \sum_{n=0}^\infty \frac{t^n}{n!}\left(\left(\frac{L}{2}\right)^n q\right)(y) = \left(e^{tL/2}q\right)(y)$$

satisfies the same equation as u in the same finite dimensional space of polynomials of degree less than or equal to $\deg(q)$. Therefore by uniqueness of solutions to ODE we must have $u(t,y) = v(t,y)$. The result now follows by taking $t=1$ and $y=0$ and observing that

$$\begin{aligned} u(1,0) &= \mathbb{E}\left[q\left(0+\sqrt{1}X\right)\right] = \mathbb{E}[q(X)] \text{ and} \\ v(1,0) &= \left(e^{L/2}q\right)(0). \end{aligned}$$

Third Proof. Let $u\in\mathbb{R}^N$. Since

$$\partial_u \exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right) = -\left(Q^{-1}x\cdot u\right)\exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right)$$

it follow by integration by parts that

$$\begin{aligned} \mathbb{E}\left[\left(Q^{-1}X\cdot u\right)p(X)\right] &= -\frac{1}{Z}\int_{\mathbb{R}^N} p(x)\,\partial_u \exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right)dx \\ &= \frac{1}{Z}\int_{\mathbb{R}^N}(\partial_u p)(x)\exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right)dx \\ &= \mathbb{E}\left[(\partial_u p)(X)\right]. \end{aligned}$$

Replacing u by Qu in this equation leads to important identity,

$$\mathbb{E}\left[(X\cdot u)\,p(X)\right] = \mathbb{E}\left[(\partial_{Qu}p)(X)\right]. \qquad (24.5)$$

It is clear that using this identity and induction it would be possible to compute $\mathbb{E}[p(X)]$ for any polynomial p. So to finish the proof it suffices to show

$$e^{L/2}\left((x\cdot u)\,p(x)\right)|_{x=0} = e^{L/2}\left((\partial_{Qu}p)(x)\right)|_x = 0.$$

This is correct notice that

$$e^{L/2}\left((x\cdot u)\,p(x)\right)=e^{L/2}\left((x\cdot u)\,e^{-L/2}e^{L/2}p(x)\right)=e^{L/2}M_{(x\cdot u)}e^{-L/2}e^{L/2}p(x)$$

Letting q be any polynomial and

$$F_t:=e^{tL/2}M_{(x\cdot u)}e^{-tL/2},$$

we have

$$\frac{d}{dt}F_t=e^{tL/2}\left[\frac{L}{2},M_{(x\cdot u)}\right]e^{-tL/2}=e^{tL/2}\partial_{Qu}e^{-tL/2}=\partial_{Qu}$$

and therefore,

$$e^{L/2}M_{(x\cdot u)}e^{-L/2}=F_1=M_{(x\cdot u)}+\partial_{Qu}.$$

Hence it follows that

$$\begin{aligned}e^{L/2}\left((x\cdot u)\,p(x)\right)|_{x=0}&=\left[\left(M_{(x\cdot u)}+\partial_{Qu}\right)e^{L/2}p(x)\right]_{x=0}=\left[\partial_{Qu}e^{L/2}p(x)\right]_{x=0}\\&=\left[e^{L/2}\partial_{Qu}p(x)\right]_{x=0}\end{aligned}\tag{24.6}$$

which is the same identity as in Eq. (24.5). ■

Example 24.3. Suppose $X\overset{d}{=}N(1,0)\in\mathbb{R}$, then

$$\mathbb{E}\left[X^{2n}\right]=\left[e^{\Delta/2}x^{2n}\right]_{x=0}=\frac{1}{n!\cdot 2^n}\Delta^n\|x\|^{2n}=\frac{(2n)!}{2^n\cdot n!}.$$

24.2 Existence of Gaussian Fields

Definition 24.4. *Let T be a set. A Gaussian random field indexed by T is a collection of random variables, $\{X_t\}_{t\in T}$ on some probability space $(\Omega,\mathcal{B},P)$ such that for any finite subset, $\Lambda\subset_f T$, $\{X_t:t\in\Lambda\}$ is a Gaussian random vector.*

Associated to a Gaussian random field, $\{X_t\}_{t\in T}$, are the two functions,

$$c:T\to\mathbb{R}\text{ and }Q:T\times T\to\mathbb{R}$$

defined by $c(t):=\mathbb{E}X_t$ and $Q(s,t):=\mathrm{Cov}(X_s,X_t)$. By the previous results, the functions (Q,c) uniquely determine the finite dimensional distributions $\{X_t:t\in T\}$, i.e. the joint distribution of the random variables, $\{X_t:t\in\Lambda\}$, for all $\Lambda\subset_f T$.

Definition 24.5. *Suppose T is a set and $\{X_t:t\in T\}$ is a random field. For any $\Lambda\subset T$, let $\mathcal{B}_\Lambda:=\sigma(X_t:t\in\Lambda)$.*

Proposition 24.6. *Suppose T is a set and $c:T\to\mathbb{R}$ and $Q:T\times T\to\mathbb{R}$ are given functions such that $Q(s,t)=Q(t,s)$ for all $s,t\in T$ and for each $\Lambda\subset_f T$*

$$\sum_{s,t\in\Lambda}Q(s,t)\,\lambda(s)\,\lambda(t)\geq 0\text{ for all }\lambda:\Lambda\to\mathbb{R}.$$

Then there exists a probability space, $(\Omega,\mathcal{B},P)$, and random variables, $X_t:\Omega\to\mathbb{R}$ for each $t\in T$ such that $\{X_t\}_{t\in T}$ is a Gaussian random process with

$$\mathbb{E}[X_s]=c(s)\text{ and }\mathrm{Cov}(X_s,X_t)=Q(s,t)\tag{24.7}$$

for all $s,t\in T$.

Proof. Since we will construct $(\Omega,\mathcal{B},P)$ by Kolmogorov's extension Theorem 17.54, let $\Omega:=\mathbb{R}^T$, $\mathcal{B}=\mathcal{B}_{\mathbb{R}^T}$, and $X_t(\omega)=\omega_t$ for all $t\in T$ and $\omega\in\Omega$. Given $\Lambda\subset_f T$, let μ_Λ be the unique Gaussian measure on $\left(\mathbb{R}^\Lambda,\mathcal{B}_\Lambda:=\mathcal{B}_{\mathbb{R}^\Lambda}\right)$ such that

$$\begin{aligned}&\int_{\mathbb{R}^\Lambda}e^{i\sum_{t\in\Lambda}\lambda(t)x(t)}d\mu_\Lambda(x)\\&=\exp\left(-\frac{1}{2}\sum_{s,t\in\Lambda}Q(s,t)\,\lambda(s)\,\lambda(t)+i\sum_{s\in\Lambda}c(s)\,\lambda(s)\right).\end{aligned}$$

The main point now is to show $\left\{\left(\mathbb{R}^\Lambda,\mathcal{B}_\Lambda,\mu_\Lambda\right)\right\}_{\Lambda\subset_f T}$ is a consistent family of measures. For this, suppose $\Lambda\subset\Gamma\subset_f T$ and $\pi:\mathbb{R}^\Gamma\to\mathbb{R}^\Lambda$ is the projection map, $\pi(x)=x|_\Lambda$. For any $\lambda\in\mathbb{R}^\Lambda$, let $\tilde{\lambda}\in\mathbb{R}^\Gamma$ be defined so that $\tilde{\lambda}=\lambda$ on Λ and $\tilde{\lambda}=0$ on $\Gamma\setminus\Lambda$. We then have,

$$\begin{aligned}&\int_{\mathbb{R}^\Lambda}e^{i\sum_{t\in\Lambda}\lambda(t)x(t)}d\left(\mu_\Gamma\circ\pi^{-1}\right)(x)\\&=\int_{\mathbb{R}^\Gamma}e^{i\sum_{t\in\Lambda}\lambda(t)\pi(x)(t)}d\mu_\Gamma(x)\\&=\int_{\mathbb{R}^\Gamma}e^{i\sum_{t\in\Gamma}\tilde{\lambda}(t)x(t)}d\mu_\Gamma(x)\\&=\exp\left(-\frac{1}{2}\sum_{s,t\in\Gamma}Q(s,t)\,\tilde{\lambda}(s)\,\tilde{\lambda}(t)+i\sum_{s\in\Gamma}c(s)\,\tilde{\lambda}(s)\right)\\&=\exp\left(-\frac{1}{2}\sum_{s,t\in\Lambda}Q(s,t)\,\lambda(s)\,\lambda(t)+i\sum_{s\in\Lambda}c(s)\,\lambda(s)\right)\\&=\int_{\mathbb{R}^\Lambda}e^{i\sum_{t\in\Lambda}\lambda(t)x(t)}d\mu_\Lambda(x).\end{aligned}$$

Since this is valid for all $\lambda \in \mathbb{R}^{\Lambda}$, it follows that $\mu_{\Gamma} \circ \pi^{-1} = \mu_{\Lambda}$ as desired. Hence by Kolmogorov's theorem, there exists a unique probability measure, P on $(\Omega, \mathcal{B})$ such that

$$\int_{\Omega} f(\omega|_{\Lambda})\, dP(\omega) = \int_{\mathbb{R}^{\Lambda}} f(x)\, d\mu_{\Lambda}(x)$$

for all $\Lambda \subset_f T$ and all bounded measurable functions, $f : \mathbb{R}^{\Lambda} \to \mathbb{R}$. In particular, it follows that

$$\begin{aligned}\mathbb{E}\left[e^{i\sum_{t\in\Lambda}\lambda(t)X_t}\right] &= \int_{\Omega} e^{i\sum_{t\in\Lambda}\lambda(t)\omega(t)} dP(\omega)\\ &= \exp\left(-\frac{1}{2}\sum_{s,t\in\Lambda} Q(s,t)\,\lambda(s)\,\lambda(t) + i\sum_{s\in\Lambda} c(s)\,\lambda(s)\right)\end{aligned}$$

for all $\lambda \in \mathbb{R}^{\Lambda}$. From this it follows that $\{X_t\}_{t\in T}$ is a Gaussian random field satisfying Eq. (24.7). ∎

Exercise 24.2. Suppose $T = [0, \infty)$ and $\{X_t : t \in T\}$ is a mean zero Gaussian random field (process). Show that $\mathcal{B}_{[0,\sigma]} \underset{X_\sigma}{\perp\!\!\!\perp} \mathcal{B}_{[\sigma,\infty)}$ for all $0 \le \sigma < \infty$ iff

$$Q(s,\sigma)\, Q(\sigma,t) = Q(\sigma,\sigma)\, Q(s,t) \ \ \forall\ 0 \le s \le \sigma \le t < \infty. \tag{24.8}$$

Hint: see use Exercises 10.6 and 10.4.

24.3 Gaussian Field Interpretation of Pre-Brownian Motion

Lemma 24.7. *Suppose that $\{B_t\}_{t\ge 0}$ is a pre-Brownian motion as described in Proposition 23.14, also see Corollary 17.22. Then $\{B_t\}_{t\ge 0}$ is a mean zero Gaussian random process with $\mathbb{E}[B_t B_s] = s \wedge t$ for all $s, t \ge 0$.*

Proof. Suppose we are given $0 = t_0 < t_1 < \cdots < t_n < \infty$ and recall from Proposition 23.14 that $B_0 = 0$ a.s. and $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^n$ are independent mean zero Gaussian random variables. Hence it follows from Corollary 10.25 that $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^n$ is a Gaussian random vector. Since the random vector $\left\{B_{t_j}\right\}_{j=0}^n$ is a linear transformation $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^n$ it follows from Lemma 9.36 that $\left\{B_{t_j}\right\}_{j=0}^n$ is a Gaussian random vector. Since $0 = t_0 < t_1 < \cdots < t_n < \infty$ was arbitrary, it follows that $\{B_t\}_{t\ge 0}$ is a Gaussian process. Since $B_t = B_t - B_0 \overset{d}{=} N(0,t)$ we see that $\mathbb{E}B_t = 0$ for all t. Moreover we have for $0 \le s < t < \infty$ that

$$\begin{aligned}\mathbb{E}[B_t B_s] &= \mathbb{E}[(B_t - B_s + B_s - B_0)(B_s - B_0)]\\ &= \mathbb{E}[(B_t - B_s)(B_s - B_0)] + \mathbb{E}\left[(B_s - B_0)^2\right]\\ &= \mathbb{E}[B_t - B_s]\cdot\mathbb{E}[B_s - B_0] + s = 0\cdot 0 + s\end{aligned}$$

which completes the proof. ∎

Theorem 24.8. *The function $Q(s,t) := s \wedge t$ defined on $s, t \ge 0$ is positive definite.*

Proof. We are going to give a six proofs of this theorem.

1. Choose any independent square integrable random variables, $\{X_j\}_{j=1}^n$, such that $\mathbb{E}X_j = 0$ and $\operatorname{Var}(X_j) = t_j - t_{j-1}$. Let $Y_j := X_1 + \cdots + X_j$ for $j = 1, 2, \ldots, n$. We then have, for $j \le k$ that
$$\begin{aligned}\operatorname{Cov}(Y_j, Y_k) &= \sum_{m\le j,\ n\le k} \operatorname{Cov}(X_m, X_n) = \sum_{m\le j,\ n\le k} \delta_{m,n}(t_m - t_{m-1})\\ &= \sum_{m\le j}(t_m - t_{m-1}) = t_j,\end{aligned}$$
i.e. $t_j \wedge t_k = \operatorname{Cov}(Y_j, Y_k)$. But such covariance matrices are always positive definite. Indeed,
$$\begin{aligned}\sum_{j,k\le n} t_j \wedge t_k \lambda_j \lambda_k &= \sum_{j,k\le n} \lambda_j \lambda_k \operatorname{Cov}(Y_j, Y_k)\\ &= \operatorname{Var}(\lambda_1 Y_1 + \cdots + \lambda_n Y_n) \ge 0\end{aligned}$$
with equality holding iff $\lambda_1 Y_1 + \cdots + \lambda_n Y_n = 0$ from which it follows that
$$0 = \mathbb{E}[Y_j(\lambda_1 Y_1 + \cdots + \lambda_n Y_n)] = \lambda_j (t_j - t_{j-1}),$$
i.e. $\lambda_j = 0$.
2. According to Exercise 21.12 we can find stochastic processes $\left\{B_n(t) = \sqrt{n} S_{[nt]}\right\}_{n=1}^{\infty}$ such that $\mathbb{E}[B_n(t)\, B_n(s)] \to s \wedge t$ as $n \to \infty$ and therefore
$$\begin{aligned}\sum_{s,t\in\Lambda}(s\wedge t)\,\lambda_s\lambda_t &= \lim_{n\to\infty}\sum_{s,t\in\Lambda}\mathbb{E}[B_n(t)\, B_n(s)]\,\lambda_s\lambda_t\\ &= \lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{t\in\Lambda} B_n(t)\,\lambda_t\right)^2\right] \ge 0.\end{aligned}$$

3. Appealing to Corollary 17.22, there exists a time homogeneous Markov processes $\{B_t\}_{t\geq 0}$ with Markov transition kernels given by
$$Q_t(x,dy) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y-x|^2} dy. \tag{24.9}$$
It is now easy to see that $s \wedge t = \operatorname{Cov}(B_s, B_t)$ which is automatically non-negative as we saw in the proof of item 2.
4. Let $\Lambda = \{0 < t_1 < \dots < t_n < \infty\}$ and $\{\lambda_i\}_{i=1}^n \subset \mathbb{R}$ be given. Further let $\alpha_j := \lambda_i + \lambda_{i+1} + \dots + \lambda_n$ with the convention that $\alpha_{n+1} = 0$. We then have,
$$\sum_{i=1}^n t_i \wedge t_j \lambda_i = \sum_{i=1}^n t_i \wedge t_j (\alpha_i - \alpha_{i+1})$$
$$= \sum_{i=1}^n [t_i \wedge t_j - t_{i-1} \wedge t_j] \alpha_i = \sum_{1\leq i\leq j} [t_i - t_{i-1}] \alpha_i$$
where $t_0 := 0$. Hence it follows that
$$\sum_{i,j=1}^n t_i \wedge t_j \lambda_i \lambda_j = \sum_{j=1}^n \sum_{1\leq i\leq j} [t_i - t_{i-1}] \alpha_i \lambda_j = \sum_{1\leq i\leq j\leq n} [t_i - t_{i-1}] \alpha_i \lambda_j$$
$$= \sum_{1\leq i\leq n} [t_i - t_{i-1}] \alpha_i^2 \geq 0$$
with equality iff $\alpha_i = 0$ for all i which is equivalent to $\lambda_i = 0$ for all i.
5. Let $h_t(\tau) := t \wedge \tau$ be as after Theorem 32.9 below and using the results and notation proved there we find,
$$\sum_{s,t\in\Lambda} (s\wedge t) \lambda_s \lambda_t = \sum_{s,t\in\Lambda} \langle h_t, h_s \rangle_T \lambda_s \lambda_t = \left\| \sum_{t\in\Lambda} \lambda_t h_t \right\|_T^2 \geq 0.$$
This shows Q is positive semi-definite and equality holds iff $\sum_{t\in\Lambda} \lambda_t h_t = 0$. After taking the derivative of this identity, it is not hard to see that $\lambda_t = 0$ for all t so that Q is positive definite.
6. The function $Q(s,t) = s \wedge t$ restricted to $s,t \in [0,T]$ for some $T < \infty$ is the Green's function for the positive definite second order differential operator $-\frac{d^2}{dt^2}$ which is equipped with Dirichlet boundary condition at $t=0$ and Neumann boundary conditions at $t=T$.

■

We have already given a Markov process proof of the existence of Pre-Brownian motion in Corollary 17.22. Given Theorem 24.8 we can also give a Gaussian process proof of the existence of pre-Brownian motion which we summarize in the next proposition.

Proposition 24.9 (Pre-Brownian motion). *Let $\{B_t\}_{t\geq 0}$ be a mean zero Gaussian process such that* $\operatorname{Cov}(B_s, B_t) = s \wedge t$ *for all $s,t \geq 0$ and let $\mathcal{B}_t := \sigma(B_s : s \leq t)$ and $\mathcal{B}_{t+} := \cap_{\sigma > t} \mathcal{B}_\sigma$ Then;*

1. *$B_0 = 0$ a.s.*
2. *$\{B_t\}_{t\geq 0}$ has independent increments with $B_t - B_s \overset{d}{=} N(0,(t-s))$ for all $0 \leq s < t < \infty$.*
3. *For all $t \geq s \geq 0$, $B_t - B_s$ is independent of $\mathcal{B}_{s+}$.*
4. *$\{B_t\}_{t\geq 0}$ is a time homogeneous Markov process with transition kernels $\{Q_t(x,dy)\}_{t\geq 0}$ given as in Eq. (24.9).*

Proof. See Exercise 24.3 – 24.5. ■

Exercise 24.3 (Independent increments). Let
$$\mathcal{P} := \{0 = t_0 < t_1 < \dots < t_n = T\}$$
be a partition of $[0,T]$, $\Delta_i B := B_{t_i} - B_{t_{i-1}}$ and $\Delta_i t := t_i - t_{i-1}$. Show $\{\Delta_i B\}_{i=1}^n$ are independent mean zero normal random variables with $\operatorname{Var}(\Delta_i B) = \Delta_i t$.

Exercise 24.4 (Increments independent of the past). Let $\mathcal{B}_t := \sigma(B_s : s \leq t)$. For each $s \in (0,\infty)$ and $t > s$, show;

1. $B_t - B_s$ is independent of $\mathcal{B}_s$ and
2. more generally show, $B_t - B_s$ is independent of $\mathcal{B}_{s+} := \cap_{\sigma > s} \mathcal{B}_\sigma$.

Exercise 24.5 (The simple Markov property). Show $B_t - B_s$ is independent of $\mathcal{B}_s$ for all $t \geq s$. Use this to show, for any bounded measurable function, $f : \mathbb{R} \to \mathbb{R}$ that
$$\mathbb{E}[f(B_t)|\mathcal{B}_{s+}] = \mathbb{E}[f(B_t)|\mathcal{B}_s] = \mathbb{E}[f(B_t)|B_s]$$
$$= (p_{t-s} * f)(B_s) =: \left(e^{(t-s)\Delta/2} f\right)(B_s) \text{ a.s.,}$$
where
$$p_t(x) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}$$
so that $p_t * f = Q_t(\cdot, f)$. This problem verifies that $\{B_t\}_{t\geq 0}$ is a "**Markov process**" with transition kernels $\{Q_t\}_{t\geq 0}$ which have $\frac{1}{2}\Delta = \frac{1}{2}\frac{d^2}{dx^2}$ as there "**infinitesimal generator.**"

Exercise 24.6. Let
$$\mathcal{P} := \{0 = t_0 < t_1 < \dots < t_n = T\}$$
and $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded measurable function. Show

$$\mathbb{E}\left[f\left(B_{t_1},\dots,B_{t_n}\right)\right]=\int_{\mathbb{R}^n} f\left(x_1,\dots,x_n\right) q_{\mathcal{P}}\left(x\right) dx$$

where

$$q_{\mathcal{P}}\left(x\right):=p_{t_1}\left(x_1\right) p_{t_2-t_1}\left(x_2-x_1\right)\dots p_{t_n-t_{n-1}}\left(x_n-x_{n-1}\right).$$

Hint: Either use Exercise 24.3 by writing

$$f\left(x_1,\dots,x_n\right)=g\left(x_1,x_2-x_1,x_3-x_2,\dots,x_n-x_{n-1}\right)$$

for some function, g or use Exercise 24.5 first for functions, f of the form,

$$f\left(x_1,\dots,x_n\right)=\prod_{j=1}^{n}\varphi_j\left(x_j\right).$$

Better yet, do it by both methods!

25

Versions and Modifications

We need to introduce a bit of terminology which we will use throughout this part of the book. As before we will let T be an index space which will typically be $\mathbb{R}_+$ or $[0,1]$ in this part of the book. We further sill suppose that $(\Omega, \mathcal{B}, P)$ is a given probability space, (S, ρ) is a separable (for simplicity) metric state space, and $X_t : \Omega \to S$ is a measurable stochastic processes.

Definition 25.1 (Versions). *Suppose, $X_t : \Omega \to S$ and $\tilde{X}_t : \Omega \to S$ are two processes defined on T. We say that $\tilde{X}$ is a* version *or a* modification *of X provided, for each $t \in T$, $X_t = \tilde{X}_t$ a.s.. (Notice that the null set may depend on the parameter t in the uncountable set, T.)*

Definition 25.2. *We say two processes are **indistinguishable** iff $P^*(Y_\cdot \neq X_\cdot) = 0$, i.e. iff there is a measurable set, $E \subset \Omega$, such that $P(E) = 0$ and $\{Y_\cdot \neq X_\cdot\} \subset E$ where*

$$\begin{aligned} \{Y_\cdot \neq X_\cdot\} &= \{\omega \in \Omega : Y_t(\omega) \neq X_t(\omega) \text{ for some } t \in [0,\infty)\} \\ &= \cup_{t\in[0,\infty)} \{\omega \in \Omega : Y_t(\omega) \neq X_t(\omega)\}. \end{aligned} \tag{25.1}$$

So Y is a modification of X iff

$$0 = \sup_{t\in T} P\left(X_t \neq \tilde{X}_t\right) = \sup_{t\in T} P\left(\left\{\rho\left(X_t, \tilde{X}_t\right) > 0\right\}\right)$$

while Y is indistinguishable for X iff

$$0 = P^*\left(X_t \neq \tilde{X}_t \ \forall\, t\right) = P^*\left(\left\{\sup_{t\in T} \rho\left(X_t, \tilde{X}_t\right) > 0\right\}\right).$$

Thus the formal difference between the two notions is simply whether the supremum is taken outside or inside the probabilities. See Exercise 28.1 for an example of two processes which are modifications of each other but are not indistinguishable.

Exercise 25.1. Suppose $\{Y_t\}_{t\geq 0}$ is a version of a process, $\{X_t\}_{t\geq 0}$. Further suppose that $t \to Y_t(\omega)$ and $t \to X_t(\omega)$ are both right continuous everywhere. Show $E := \{Y_\cdot \neq X_\cdot\}$ is a measurable set such that $P(E) = 0$ and hence X and Y are indistinguishable. **Hint:** replace the union in Eq. (25.1) by an appropriate countable union.

Exercise 25.2. Suppose that $\{X_t : \Omega \to S\}_{t\geq 0}$ is a process such that for each $N \in \mathbb{N}$ there is a right continuous modification, $\left\{\tilde{X}_t^{(N)}\right\}_{0\leq t<N}$ of $\{X_t\}_{0\leq t<N}$. Show that X admits a right continuous modifications, $\tilde{X}$, defined for all $t \geq 0$.

25.1 Kolmolgorov's Continuity Criteria

Let $\mathbb{D}_n := \left\{\frac{i}{2^n} : i \in \mathbb{Z}\right\}$ and $\mathbb{D} := \cup_{n=0}^\infty \mathbb{D}_n$ be the dyadic rational numbers.

Lemma 25.3. *Let $\mathbb{D}_+ = \mathbb{D} \cap [0,\infty)$ and $s \in \mathbb{D}_+ = \mathbb{D} \cap [0,\infty)$ and $n \in \mathbb{N}_0$ be given, then;*

1. *there exists a unique $i = i(n,s) \in \mathbb{N}_0$ such that $i2^{-n} \leq s < (i+1)2^{-n}$ and*
2. *s may be uniquely written as*

$$s = \frac{i}{2^n} + \sum_{k=1}^\infty \frac{a_k}{2^{n+k}},$$

where $a_k = a_k(n,s) \in \{0,1\}$ with $a_k = 0$ for all sufficiently large k.

Example 25.4. Suppose that $s = 85/32 = 85/2^5 \in \mathbb{D}$ and $2 \in \mathbb{N}_0$ are given, then $2^2 s = 85/8 = 10 + 5/8$, i.e.

$$s = \frac{10}{2^2} + \frac{5}{2^5}.$$

Similarly, $2^3 \cdot 5/2^5 = 5/4 = 1 + 1/4$ so that $5/2^5 = 1/2^3 + 1/2^5$ and we have expressed s as

$$s = \frac{10}{2^2} + \frac{1}{2^3} + \frac{0}{2^4} + \frac{1}{2^5}.$$

Proof. The first assertion follows from the fact that $\mathbb{D}_+$ is partitioned by $\left[\{i2^{-n}, (i+1)2^{-n}\right) \cap \mathbb{D}\}_{i\in\mathbb{N}_0}$. For the second assertion define the $a_1 = 1$ if $\frac{i}{2^n} + \frac{a_1}{2^n} \leq s$ and 0 otherwise, then choose $a_2 = 1$ if $\frac{i}{2^n} + \frac{a_1}{2^{n+1}} + \frac{a_2}{2^{n+2}} \leq s$ and 0 otherwise, etc. It is easy to check that $s_m := \frac{i}{2^n} + \sum_{k=1}^m \frac{a_k}{2^{n+k}}$ so constructed satisfies, $s - \frac{1}{2^{n+m}} < s_m \leq s$ for all $m \in \mathbb{N}$. As $s_m \in \mathbb{D}_{m+n}$ and $s \in \mathbb{D}_N$ for some N, if $m + n \geq N$ then we must have $s_m = s$ because $s - s_m < \frac{1}{2^{n+m}}$ and $s, s_m \in \mathbb{D}_{m+n}$.

■

Suppose now that (S,ρ) is a metric space and $x : Q := \mathbb{D} \cap [0,1] \to S$. For $n \in \mathbb{N}_0$ let

$$\begin{aligned} \Delta_n(x) &= \max\left\{\rho\left(x\left(i2^{-n}\right), x\left((i-1)2^{-n}\right)\right) : 1 \le i \le 2^n\right\} \\ &= \max\left\{\rho\left(x(t), x(s)\right) : s,t \in Q \cap \mathbb{D}_n \text{ with } |s-t| \le \frac{1}{2^n}\right\}. \end{aligned}$$

If $\gamma \in (0,1)$ and $x : \mathbb{D}_1 := \mathbb{D} \cap [0,1] \to S$ is a $\gamma-$ **Hölder continuous** function, i.e.

$$\rho\left(x(t), x(s)\right) \le K |t-s|^{\gamma}$$

for some and $K < \infty$, then $\Delta_n(x) \le K 2^{-n\gamma}$ for all n and in particular for all $\alpha \in (0,\gamma)$ we have

$$\sum_{n=0}^{\infty} 2^{n\alpha} \Delta_n(x) \le K \sum_{n=0}^{\infty} 2^{n\alpha} 2^{-n\gamma} = K\left(1 - \frac{1}{2^{\gamma-\alpha}}\right)^{-1} < \infty.$$

Our next goal is to produce the following "converse" to this statement.

Lemma 25.5. *Suppose $\alpha > 0$ and $x : Q := \mathbb{D} \cap [0,1] \to S$ is a map such that $\sum_{n=0}^{\infty} 2^{n\alpha} \Delta_n(x) < \infty$, then*

$$\rho\left(x(t), x(s)\right) \le 2^{1+\alpha} \cdot \sum_{k=0}^{\infty} 2^{\alpha k} \Delta_k(x) \cdot |t-s|^{\alpha} \text{ for all } s,t \in Q. \tag{25.2}$$

Moreover there exists a unique continuous function $\tilde{x} : [0,1] \to S$ extending x and this extesion is still α – Hölder continuous. (When $\alpha > 1$ it follows by Exercise 25.3 that $x(t)$ is constant.)

Proof. Let $s,t \in Q$ with $s < t$ and choose n so that

$$\frac{1}{2^{n+1}} < t - s \le \frac{1}{2^n} \tag{25.3}$$

and observe that if $s \in [i2^{-n}, (i+1)2^{-n})$, then $t \in [i2^{-n}, (i+2)2^{-n})$ and therefore

$$s = \frac{i}{2^n} + \sum_{k=1}^{\infty} \frac{a_k}{2^{n+k}} \text{ and } t = \frac{j}{2^n} + \sum_{k=1}^{\infty} \frac{b_k}{2^{n+k}}$$

where $j \in \{i, i+1\}$ and $a_k, b_k \in \{0,1\}$ with $a_k = b_k = 0$ for a.a. k. Letting

$$s_m := \frac{i}{2^n} + \sum_{k=1}^{m} \frac{a_k}{2^{n+k}}$$

as above we have $s_N = s$ for large N. Since $s_m, s_{m+1} \in Q \cap \mathbb{D}_{n+m+1}$ with $|s_m - s_{m+1}| \le 2^{-(n+m+1)}$, it follows from the definition of Δ_{n+m+1} that $\rho\left(x(s_{m+1}), x(s_m)\right) \le \Delta_{n+m+1}(x)$ which combined with the triangle inequality shows

$$\rho\left(x(s), x(s_0)\right) \le \sum_{m=1}^{N} \rho\left(x(s_m), x(s_{m-1})\right) \le \sum_{m=1}^{\infty} \Delta_{n+m}(x).$$

Similarly $\rho\left(x(t), x(t_0)\right) \le \sum_{m=1}^{\infty} \Delta_{n+m}(x)$ while $\rho\left(x(s_0), x(t_0)\right) \le \Delta_n(x)$. One more application of the triangle inequality now shows,

$$\begin{aligned} \rho\left(x(t), x(s)\right) &\le \Delta_n(x) + 2 \cdot \sum_{m=1}^{\infty} \Delta_{n+m}(x) \\ &\le 2 \cdot \sum_{k=n}^{\infty} \Delta_k(x) = 2 \cdot \sum_{k=n}^{\infty} 2^{-\alpha k} \cdot 2^{\alpha k} \Delta_k(x) \\ &\le 2 \cdot \left(2^{-n}\right)^{\alpha} \sum_{k=n}^{\infty} 2^{\alpha k} \Delta_k(x). \end{aligned}$$

Combining this with the lower bound in Eq. (25.3) gives the estimate in Eq. (25.2).

For the last assertion we define $\tilde{x}(t) := \lim_{Q \ni s \to t} x(s)$. This limit exists since for any sequence $\{s_n\}_{n=1}^{\infty} \subset Q$ with $s_n \to t \in [0,1]$, the sequence $\{x(s_n)\}_{n=1}^{\infty}$ is Cauchy in S because of Eq. (25.2) and hence convergent in S. It is easy to check that $\lim_{n\to\infty} x(s_n)$ is independent of the choice of the sequence $\{s_n\}_{n=1}^{\infty}$. A simple limiting argument now shows that

$$\rho\left(\tilde{x}(t), \tilde{x}(s)\right) \le 2^{1+\alpha} \cdot \sum_{k=0}^{\infty} 2^{\alpha k} \Delta_k(x) \cdot |t-s|^{\alpha} \text{ for all } s,t \in [0,1]$$

which shows that $\tilde{x}$ is Hölder continuous. As we had no choice but to define $\tilde{x}$ the way we did if $\tilde{x}$ is to be continuous, the extension is unique. ■

Exercise 25.3. Show; if $x : Q \to S$ is α – Hölder continuous for some $\alpha > 1$, then x is constant.

Solution to Exercise (25.3). Let $s,t \in Q$ with $s < t$ and let $\varepsilon > 0$ be given. Choose a partition $\{s = t_0 < t_1 < \dots \dots < t_n = t\} \subset Q$ such that $\Delta_i := t_i - t_{i-1} \le \varepsilon$. Then by the triangle inequality and the assumed Hölder continuity

$$\begin{aligned}
\rho\left(x\left(t\right),x\left(s\right)\right) &\leq \sum_{i=1}^{n} \rho\left(x\left(t_{i-1}\right),x\left(t_{i}\right)\right) \\
&\leq \sum_{i=1}^{n} C\Delta_i^{\alpha} \leq C\varepsilon^{\alpha-1}\sum_{i=1}^{n}\Delta_i \\
&\leq C\varepsilon^{\alpha-1}\left(t-s\right).
\end{aligned}$$

As $\alpha > 1$ and $\varepsilon > 0$ is arbitrary, we may let $\varepsilon \downarrow 0$ in this equation in order to learn $\rho\left(x\left(t\right),x\left(s\right)\right) = 0$. As $s,t \in Q$ is arbitrary it follows that $x\left(t\right)$ is constant in t.

Theorem 25.6 (Kolmogorov's Continuity Criteria). *Let $\{X_t\}_{t\in Q}$ be an S – valued stochastic process and suppose there exists $\gamma, \varepsilon > 0$ such that*

$$\mathbb{E}\left[\rho\left(X_t,X_s\right)^{\gamma}\right] \leq C\left|t-s\right|^{1+\varepsilon} \text{ for all } s,t \in Q. \tag{25.4}$$

Then for all $\alpha \in (0,\varepsilon/\gamma)$ there exists $0 \leq K_{\alpha}\left(X\right) \in L^{\gamma}\left(P\right)$ such that

$$\rho\left(X_t,X_s\right) \leq K_{\alpha}\left(X\right)\left|t-s\right|^{\alpha} \text{ for all } s,t \in Q. \tag{25.5}$$

and there is a function, $M\left(C,\gamma,\varepsilon,\alpha\right) < \infty$, such that

$$\mathbb{E}\left[K_{\alpha}\left(X\right)^{\gamma}\right] \leq M\left(C,\gamma,\varepsilon,\alpha\right). \tag{25.6}$$

Proof. According to Exercise 25.4 below when $\gamma < 1$ or more generally when $\gamma < 1+\varepsilon$ it actually follows that $X_t = X_0$ a.s. and therefore Eq. (25.5) holds for some K_α which is equal to zero almost surely. So we may now suppose that $\gamma \geq 1+\varepsilon \geq 1$ and let $\alpha \in (0,\gamma/\varepsilon)$ and

$$\kappa_{\alpha} := \sum_{k=0}^{\infty} 2^{\alpha k}\Delta_k\left(X\right).$$

The estimate in Eq. (25.5) with $K_{\alpha}\left(X\right) = 2^{-(1+\alpha)\gamma}\cdot\kappa_{\alpha}$ is now a consequence of Lemma 25.5. So it only remains to show $\mathbb{E}\left[\kappa_{\alpha}^{\gamma}\right] < \infty$ when $\alpha \in (0,\varepsilon/\gamma)$.

From the following simple estimate,

$$\Delta_k\left(X\right)^{\gamma} = \max_{1\leq i\leq 2^k} \rho\left(X_{i2^{-k}},X_{(i-1)2^{-k}}\right)^{\gamma} \leq \sum_{i=1}^{2^k}\rho\left(X_{i2^{-k}},X_{(i-1)2^{-k}}\right)^{\gamma},$$

we find

$$\mathbb{E}\left[\Delta_k\left(X\right)^{\gamma}\right] \leq \sum_{i=1}^{2^k}\mathbb{E}\left[\rho\left(X_{i2^{-k}},X_{(i-1)2^{-k}}\right)^{\gamma}\right] \leq 2^k\cdot C\left(2^{-k}\right)^{1+\varepsilon} = C2^{-k\varepsilon} \tag{25.7}$$

and therefore $\left\|\Delta_k\left(X\right)\right\|_{\gamma} \leq C^{1/\gamma}2^{-k\varepsilon/\gamma}$. Combining this inequality with Minikowski's inequality shows

$$\begin{aligned}
\left\|\kappa_{\alpha}\right\|_{\gamma} &= \left\|\sum_{k=0}^{\infty}\left[2^{\alpha k}\Delta_k\left(X\right)\right]\right\|_{\gamma} \leq \sum_{k=0}^{\infty}2^{\alpha k}\left\|\Delta_k\left(X\right)\right\|_{\gamma} \\
&\leq C^{1/\gamma}\sum_{k=0}^{\infty}2^{\alpha k}2^{-k\varepsilon/\gamma} = C^{1/\gamma}\sum_{k=0}^{\infty}\left(2^{\alpha-\varepsilon/\gamma}\right)^{k}.
\end{aligned}$$

This is again finite provided that $\alpha < \varepsilon/\gamma$. ■

Theorem 25.7 (Kolmogorov's Continuity Criteria). *Let $T \in \mathbb{N}$, $D = [0,T] \subset \mathbb{R}$, (S,ρ) be a complete separable metric space and suppose that $X_t : \Omega \to S$ is a process for $t \in D$. Assume there exists $\gamma, C, \varepsilon > 0$ such that such that*

$$\mathbb{E}\left[\rho\left(X_t,X_s\right)^{\gamma}\right] \leq C\left|t-s\right|^{1+\varepsilon} \text{ for all } s,t \in D. \tag{25.8}$$

Then there is a modification, $\tilde{X}$, of X which is α–Hölder continuous for all $\alpha \in (0,\varepsilon/\gamma)$ and for each such α there is a random variable $K_{\alpha}\left(X\right) \in L^{\gamma}\left(P\right)$ such that

$$\rho(\tilde{X}_t,\tilde{X}_s) \leq K_{\alpha}\left(X\right)\left|t-s\right|^{\alpha} \text{ for all } s,t \in D. \tag{25.9}$$

(Again according to Exercise 25.4, we will have $X_t = X_0$ a.s. for all $t \in D$ unless $\gamma \geq 1+\varepsilon$.)

Proof. From Theorem 25.6 we know for all $\alpha \in (0,\varepsilon/\gamma)$ there is a random variable $K_{\alpha}\left(X\right) \in L^{\gamma}\left(P\right)$ such that

$$\rho(X_t,X_s) \leq K_{\alpha}\left(X\right)\left|t-s\right|^{\alpha} \text{ for all } s,t \in D\cap\mathbb{D}.$$

On the set $\{K_{\alpha}\left(X\right) < \infty\}$, $\{X_t\}_{t\in D\cap\mathbb{D}}$ has a unique continuous extension to D which we denote by $\left\{\tilde{X}_t\right\}_{t\in D}$. Moroever this extenson is easily seen to satisfy Eq. (25.9). Lastly we have for $s \in D\cap\mathbb{D}$ and $t \in D$ that

$$\begin{aligned}
\rho\left(X_t,\tilde{X}_t\right)^{\gamma} &\leq \liminf_{D\cap\mathbb{D}\ni s\to t}\left[\rho\left(X_t,X_s\right)+\rho\left(X_s,\tilde{X}_t\right)\right]^{\gamma} \\
&= \liminf_{D\cap\mathbb{D}\ni s\to t}\rho\left(X_t,X_s\right)^{\gamma}
\end{aligned}$$

and so by Fatou's lemma,

$$\mathbb{E}\left[\rho\left(X_t,\tilde{X}_t\right)^{\gamma}\right] \leq \liminf_{D\cap\mathbb{D}\ni s\to t}\mathbb{E}\left[\rho\left(X_t,X_s\right)^{\gamma}\right] \leq \liminf_{D\cap\mathbb{D}\ni s\to t} C\left|t-s\right|^{1+\varepsilon} = 0.$$

This certainly implies that $\rho\left(X_t,\tilde{X}_t\right) = 0$ a.s. for every $t \in D$ and therefore that $\tilde{X}$ is a modification of X. ■

Our construction of Brownian motion in Theorem 26.3 below will give us an opportunity to apply Theorem 25.7. At this time let us observe that it is important that ε is greater than 0 in the previous two theorems.

Example 25.8. Recall that a Poisson process, $\{N_t\}_{t\geq 0}$, with parameter λ satisfies (by definition): (i) N has *independent increments*, and (ii) if $0 \leq u < v$ then $N_v - N_u$ has the Poisson distribution with parameter $\lambda(v-u)$. Using the generating function (or the Laplace or Fourier transform, see Example 22.11), one can show that for any $k \in \mathbb{N}$, that

$$\mathbb{E}\left|N_t - N_s\right|^k \sim \lambda\left|t-s\right| \text{ for } \left|t-s\right| \text{ small.} \tag{25.10}$$

Notice that we can not use Eq. (25.10) for any $k \in \mathbb{N}$ to satisfy the hypothesis of Theorem 25.7 which is good since $\{N_t\}_{t\geq 0}$ is integer value and does not have a continuous modification. However, see Example 28.27 below where it is shown that $\{N_t\}_{t\geq 0}$ has a right continuous modification.

Exercise 25.4. Let $T \in \mathbb{N}$, $D = [0,T] \subset \mathbb{R}$, (S,ρ) be a complete separable metric space and suppose that $X_t : \Omega \to S$ is a process for $t \in D$. Assume there exists $\gamma > 0$, $C > 0$, and $\varepsilon > 0$ such that $(1+\varepsilon)/\gamma > 1$ and

$$\mathbb{E}\left[\rho\left(X_t, X_s\right)^\gamma\right] \leq C\left|t-s\right|^{1+\varepsilon} \text{ for all } s,t \in D.$$

Show $X_t = X_0$ a.s. for each $t \in D$. **Hint:** for $\gamma \in (0,1)$ use the inequality[1] $(a+b)^\gamma \leq a^\gamma + b^\gamma$ for all $a,b \geq 0$ while for $\gamma \geq 1$ use Minikowski's inequality. (This Exercise was inspired by questions posed by Dennis Leung.)

Solution to Exercise (25.3). Case 1 where $\gamma \in (0,1)$. Let $t \in D$ and suppose that $\{0 = t_0 < t_1 < \dots\dots < t_n = t\}$ is a partition of $[0,t]$ and let $\delta := \max_i |t_i - t_{i-1}|$. To simplify notation let $\Delta_i := t_i - t_{i-1} \leq \delta$. Then by the triangle inequality and the estimate $(a+b)^\gamma \leq a^\gamma + b^\gamma$ for all $a,b \geq 0$ we find

$$\begin{aligned}\mathbb{E}\left[\rho\left(X_t,X_0\right)^\gamma\right] &\leq \mathbb{E}\left[\sum_{i=1}^n \rho\left(X_{t_{i-1}}, X_{t_i}\right)\right]^\gamma \\ &\leq \sum_{i=1}^n \mathbb{E}\left[\rho\left(X_{t_{i-1}}, X_{t_i}\right)^\gamma\right] \\ &\leq \sum_{i=1}^n C\Delta_i^{1+\varepsilon} \leq C\delta^\varepsilon \sum_{i=1}^n \Delta_i = C\delta^\varepsilon t.\end{aligned}$$

Since we are free to choose the partition as we wish we may let $\delta \downarrow 0$ in the previous estimate in order to discover that $\mathbb{E}\left[\rho\left(X_t,X_0\right)^\gamma\right] = 0$ for all $t \in D$. Therefore $X_t = X_0$ a.s. for all $t \in D$.

Case 2. Suppose that $\gamma \geq 1$, then again by the triangle inequality for ρ and Minikowski's inequality;

$$\begin{aligned}\left\|\rho\left(X_t,X_0\right)\right\|_\gamma &\leq \left\|\sum_{i=1}^n \rho\left(X_{t_{i-1}}, X_{t_i}\right)\right\|_\gamma \leq \sum_{i=1}^n \left\|\rho\left(X_{t_{i-1}}, X_{t_i}\right)\right\|_\gamma \\ &\leq C\sum_{i=1}^n \Delta_i^{\frac{1+\varepsilon}{\gamma}} \leq C\delta^{\frac{1+\varepsilon}{\gamma}-1} \sum_{i=1}^n \Delta_i = C\delta^{\frac{1+\varepsilon}{\gamma}-1} t.\end{aligned}$$

As before, since $(1+\varepsilon)/\gamma - 1 > 0$, the same argument as before show that $\left\|\rho\left(X_t,X_0\right)\right\|_\gamma = 0$ and therefore $X_t = X_0$ a.s. again.

25.2 Kolmolgorov's Tightness Criteria

Before leaving this chapter let us record a compactness result which follows easily from what we have done so far.

Theorem 25.9 (Tightness Criteria). *Let S be a complete metric space satisfying the Heine–Borel property[2] (for example $S = \mathbb{R}^d$ for some $d < \infty$). Suppose that $\{B_n(t) : 0 \leq t \leq 1\}_{n=1}^\infty$ is a sequence of S – valued continuous stochastic processes and suppose there exists $\gamma, \varepsilon > 0$ and $C < \infty$ such that*

$$\sup_n \mathbb{E}\left[\rho\left(B_n(t), B_n(s)\right)^\gamma\right] \leq C\left|t-s\right|^{1+\varepsilon} \text{ for all } 0 \leq s,t \leq 1 \tag{25.11}$$

and for some point $s_0 \in S$ we have

$$\lim_{N\uparrow\infty} \sup_n P\left[\rho\left(B_n(0), s_0\right) > N\right] = 0. \tag{25.12}$$

Then the collection of measures, $\{\mu_n := \mathrm{Law}(B_n)\}_{n=1}^\infty$ on $C([0,1],S)$ are tight.

[1] If $f(x) = x^\gamma$ for $\gamma \in (0,1)$, then f is an increasing function which is concave down, i.e. f' is decreasing. For $a > 0$ let $g(x) := f(x+a) - f(x)$. By looking at a picture or just noting that $g'(x) \leq 0$ since f' is decreasing, it follows that g is a decreasing function of x. In particular it follows that

$$f(b+a) - f(b) = g(b) \leq g(0) = f(a) - f(0) = f(a).$$

[2] The Heine–Borel property means that closed and bounded sets are compact.

Proof. Let $\alpha \in (0, \varepsilon/\gamma)$ and for $\omega \in C([0,1], S)$ let

$$K_\alpha(\omega) = 2^{-(1+\alpha)\gamma} \sum_{k=0}^{\infty} 2^{\alpha k} \Delta_k(\omega)$$

where

$$\Delta_k(\omega) := \max\left\{\rho\left(\omega\left(\frac{j-1}{2^k}\right), \omega\left(\frac{j}{2^k}\right)\right) : 1 \le j \le 2^k\right\}.$$

The assumptions of this theorem allows us to apply Theorem 25.6 in order to learn;

$$\sup_n \mathbb{E}\left[K_\alpha(B_n)^\gamma\right] \le M(C, \gamma, \varepsilon, \alpha) < \infty.$$

Now let Ω_N denote those $\omega \in C([0,1], S)$ such that $\rho(\omega(0), s_0) \le N$ and $K_\alpha(\omega) \le N$. We then have that

$$\begin{aligned}\mu_n(\Omega_N^c) &= P(B_n \notin \Omega_N) \\ &= P(\rho(B_n(0), s_0) > N \text{ or } K_\alpha(B_n) > N) \\ &\le P(\rho(B_n(0), s_0) > N) + P(K_\alpha(B_n) > N) \\ &\le P[\rho(B_n(0), s_0) > N] + \frac{1}{N^\gamma}\mathbb{E}K_\alpha(B_n)^\gamma \\ &\le P[\rho(B_n(0), s_0) > N] + \frac{1}{N^\gamma} M(C, \gamma, \varepsilon, \alpha).\end{aligned}$$

From this inequality and the hypothesis of the theorem it follows that $\lim_{N\to\infty} \sup_n \mu_n(\Omega_N^c) = 0$. To complete the proof it suffices to observe that for $\omega \in \Omega_N$ we have $\rho(\omega(0), s_0) \le N$ and

$$\rho(\omega(t), \omega(s)) \le K_\alpha(\omega) |t-s|^\alpha \le N |t-s|^\alpha \ \forall\, 0 \le s, t \le 1.$$

Therefore by the Arzela - Ascoli Theorem 45.36 and Remark 45.37, it follows that Ω_N is precompact inside of the complete separable metric space, $C([0,1], S)$. ■

25.2.1 Appendix: Alternate Proofs (please ignore)

(BRUCE: Add Garsia Rumsey stuff here as well – see Chapter **??**.) Let $T \in \mathbb{N}$ and $D = [0, T] \subset \mathbb{R}$,

$$\mathbb{D} = \left\{\frac{i}{2^n} : i, n \in \mathbb{N}_0\right\} \text{ – the Dyadic Rationals in } [0, \infty),$$

and (S, ρ) be a complete metric space. Suppose $\gamma \in (0,1)$ and $x : D \to S$ is a γ – **Hölder continuous** function, i.e. there exists $K < \infty$ such that

$$\rho(x(t), x(s)) \le K |t-s|^\gamma \text{ for all } s, t \in D.$$

Then for any $\alpha \in (0, \gamma)$ and $n \ge \nu := n \ge (\log_2 K) / (\gamma - \alpha)$ we have,

$$\rho\left(x\left(\frac{i+1}{2^n}\right), x\left(\frac{i}{2^n}\right)\right) \le K 2^{-n\gamma} = K 2^{-n(\gamma-\alpha)} 2^{-n\alpha} \le 2^{-n\alpha}$$

provided $\frac{i}{2^n} < T$.

Lemma 25.10. *Suppose that $x : D \cap \mathbb{D} \to S$ is a given function. Assume there exists $\nu \in \mathbb{N}$ such that*

$$\rho\left(x\left(\frac{i+1}{2^n}\right), x\left(\frac{i}{2^n}\right)\right) \le 2^{-n\alpha} \text{ for all } n \ge \nu \tag{25.13}$$

provided $\frac{i}{2^n} < T$. Then,

$$\rho(x(s), x(t)) \le C |t-s|^\alpha \ \forall\, s, t \in D \cap \mathbb{D} \text{ with } |t-s| \le 2^{-\nu}, \tag{25.14}$$

where

$$C = C(\alpha) = \frac{1 + 2^\alpha}{1 - 2^{-\alpha}}.$$

In particular x uniquely extends to a continuous function on D which is α– Hölder continuous. Moreover, the extension, x, satisfies the Hölder estimate,

$$\rho(x(s), x(t)) \le C_1(\alpha, T)\, 2^{\nu(1-\alpha)} |t-s|^\alpha \text{ for all } s, t \in D \tag{25.15}$$

where

$$C_1(\alpha, T) := 2C(\alpha)\, T^{1-\alpha}. \tag{25.16}$$

Proof. Let $n \ge \nu$ and $s \in \mathbb{D} \cap D$ and express s as,

$$s = \frac{i}{2^n} + \sum_{k=1}^{\infty} \frac{a_k}{2^{n+k}}$$

where $i = i_n(s) \in \mathbb{N}_0$ is chosen so that $i 2^{-n} \le s < (i+1) 2^{-n}$, and $a_k = 0$ or 1 with $a_k = 0$ for almost all k. Set

$$s_m = \frac{i}{2^n} + \sum_{k=1}^{m} \frac{a_k}{2^{n+k}}$$

and notice that $s_0 = \frac{i}{2^n}$ and $s_m = s$ for all m sufficiently large. Therefore with m_0 sufficiently large, we have

$$\rho\left(x(s), x\left(\frac{i}{2^n}\right)\right) = \rho(x(s_{m_0}), x(s_0)) \le \sum_{k=0}^{m_0-1} \rho(x(s_{k+1}), x(s_k)).$$

Since $s_{k+1} - s_k = \frac{a_{k+1}}{2^{n+k+1}} \le 2^{-\nu}$, it follows from Eq. (25.13) that

$$\rho\left(x(s), x\left(\frac{i}{2^n}\right)\right) \le \sum_{k=0}^{\infty} \left(\frac{1}{2^{n+1+k}}\right)^{\alpha} = 2^{-\alpha(n+1)} \frac{1}{1-2^{-\alpha}} = \frac{2^{-\alpha}}{1-2^{-\alpha}} 2^{-\alpha n}.$$

If $0 < t - s \le 2^{-\nu}$, let $n \ge \nu$ be chosen so that $2^{-(n+1)} < t - s \le 2^{-n}$. For this choice of n, if $i = i_n(s)$ then $j = i_n(t) \in \{i, i+1\}$, see Figure 25.1. Therefore

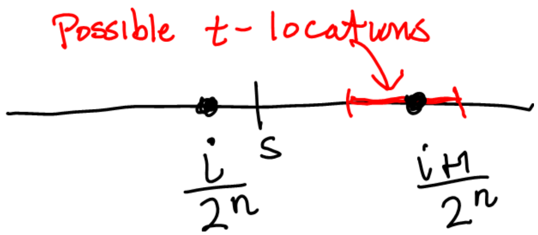


Fig. 25.1. The geometry of s and t with $2^{-(n+1)} < t - s \le 2^{-n}$.

$$\begin{aligned}\rho(x(s), x(t)) &\le \rho(x(s), x(i2^{-n})) + \rho(x(i2^{-n}), x(j2^{-n})) + \rho(x(j2^{-n}), x(t)) \\ &\le \frac{2^{-\alpha}}{1-2^{-\alpha}} 2^{-\alpha n} + 2^{-n\alpha} + \frac{2^{-\alpha}}{1-2^{-\alpha}} 2^{-\alpha n} = 2^{-n\alpha} \frac{1+2^{-\alpha}}{1-2^{-\alpha}}.\end{aligned}$$

Since

$$2^{-n\alpha} = 2^{\alpha} \cdot 2^{-(n+1)\alpha} < 2^{\alpha} (t-s)^{\alpha},$$

we may conclude that

$$\rho(x(s), x(t)) \le (t-s)^{\alpha} \frac{1+2^{-\alpha}}{1-2^{-\alpha}} 2^{\alpha} = C(\alpha) |t-s|^{\alpha}.$$

From this estimate it follows that x has an α–Hölder continuous extension to D. We will continue to denote this extension by x.

If $s, t \in D \cap \mathbb{D}$ with $t - s > 2^{-\nu}$, choose $k \in \mathbb{N}$ such that $t - s = k2^{-\nu} + \delta$ with $0 \le \delta < 2^{-\nu}$. It then follows that

$$\begin{aligned}\rho(x(s), x(t)) &= \rho\left(x(s), x\left(s + k2^{-\nu} + \delta\right)\right) \\ &\le \sum_{j=1}^{k} \rho\left(x\left(s + (j-1)2^{-\nu}\right), x\left(s + j2^{-\nu}\right)\right) \\ &\quad + \rho\left(x\left(s + k2^{-\nu}\right), x\left(s + k2^{-\nu} + \delta\right)\right) \\ &\le C(\alpha)\left(k2^{-\nu\alpha} + \delta^{\alpha}\right) \le 2C(\alpha) k 2^{-\nu\alpha}.\end{aligned}$$

Since

$$k \le 2^{\nu}(t-s) \le 2^{\nu}(t-s)^{\alpha} T^{1-\alpha},$$

we may conclude

$$\rho(x(s), x(t)) \le 2C(\alpha) 2^{-\nu\alpha} 2^{\nu} T^{1-\alpha} |t-s|^{\alpha} = C_1(\alpha, T) 2^{\nu(1-\alpha)} |t-s|^{\alpha}$$

where $C_1(\alpha, T)$ is given as in Eq. (25.16). As x is continuous and $D \cap \mathbb{D}$ is dense in D, the above estimate extends to all $s, t \in D$. ∎

Theorem 25.11 (Kolmogorov's Continuity Criteria). *Suppose that $X_t : \Omega \to S$ is a process for $t \in D$. Assume there exists positive constants, ε, β, and C, such that*

$$\mathbb{E}[\rho(X_t, X_s)^{\varepsilon}] \le C |t-s|^{1+\beta} \tag{25.17}$$

for all $s, t \in D$. Then for any $\alpha \in (0, \beta/\varepsilon)$ there is a modification, $\tilde{X}$, of X which is α–Hölder continuous. Moreover, there is a random variable K_α such that,

$$\rho(\tilde{X}_t, \tilde{X}_s) \le K_\alpha |t-s|^{\alpha} \text{ for all } s, t \in D \tag{25.18}$$

and $\mathbb{E}K_\alpha^p < \infty$ for all $p < \frac{\beta - \alpha\varepsilon}{1-\alpha}$.

Proof. Using Chebyshev's inequality,

$$\begin{aligned}P&\left(\rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right) \ge 2^{-n\alpha}\right) \\ &= P\left(\rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right)^{\varepsilon} \ge 2^{-n\alpha\varepsilon}\right) \\ &\le 2^{n\alpha\varepsilon} \mathbb{E}\left[\rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right)^{\varepsilon}\right] \\ &\le C 2^{-n(1+\beta-\alpha\varepsilon)}.\end{aligned} \tag{25.19}$$

Letting

$$\begin{aligned}A_n &= \left\{\max_{0 \le i \le T2^n - 1} \rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right) \ge 2^{-n\alpha}\right\} \\ &= \cup_{0 \le i \le T2^n - 1} \left\{\rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right) \ge 2^{-n\alpha}\right\},\end{aligned}$$

it follows from Eq. (25.19), that

$$\begin{aligned}P(A_n) &\le \sum_{0 \le i \le T2^n - 1} P\left(\rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right) \ge 2^{-n\alpha}\right) \\ &\le T2^n C 2^{-n(1+\beta-\alpha\varepsilon)} = CT2^{-n(\beta-\alpha\varepsilon)}.\end{aligned} \tag{25.20}$$

Since,

$$\sum_{n=0}^{\infty} P(A_n) \le CT \frac{1}{1-2^{-(\beta-\alpha\varepsilon)}} < \infty,$$

it follows by the first Borel Cantelli lemma that $P(A_n \text{ i.o.}) = 0$ or equivalently put, if

$$\Omega_0 := \{A_n^c \text{ a.a.}\} = \left\{ \max_{0\le i\le T2^n-1} \rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right) < 2^{-n\alpha} \text{ a.a.}\right\},$$

then $P(\Omega_0) = 1$.

For $\omega \in \Omega_0$, let

$$\nu(\omega) = \min\{n : \omega \in A_m^c \text{ for all } m \ge n\} < \infty.$$

On Ω_0, we know that

$$\max_{0\le i\le T2^n-1} \rho\left(X\left(\frac{i+1}{2^n}\right), X\left(\frac{i}{2^n}\right)\right) < 2^{-n\alpha} \text{ for all } n \ge \nu$$

and hence by Lemma 25.10,

$$\rho(X_t, X_s) \le |t-s|^\alpha \text{ when } |t-s| \le 2^{-\nu}. \tag{25.21}$$

Hence on Ω_0, if we define $\tilde{X}_t := \lim_{\tau\in\mathbb{D}\cap D;\ \tau\to t} X_t$, the resulting process, $\tilde{X}_t$, will be α – Hölder continuous on D. To complete the definition of $\tilde{X}_t$, fix a point $y \in S$ and set $\tilde{X}_t(\omega) = y$ for all $t \in D$ and $\omega \notin \Omega_0$.

For $t \in D$ and $s \in D \cap \mathbb{D}$, we have that

$$\rho(\tilde{X}_t, X_t) \le \rho(\tilde{X}_t, \tilde{X}_s) + \rho(\tilde{X}_s, X_s) + \rho(X_s, X_t) = \rho(\tilde{X}_t, \tilde{X}_s) + \rho(X_s, X_t) \text{ a.e.}$$

By continuity, $\lim_{s\to t} \tilde{X}_s = \tilde{X}_t$ and by Eq. (25.17) it follows that $\lim_{s\to t} X_s = X_t$ in measure and hence we may conclude that $\rho(\tilde{X}_t, X_t) = 0$ a.s., i.e. $X_t = \tilde{X}_t$ a.e. and $\tilde{X}$ is a version of X.

It is only left to prove the quantitative estimate in Eq. (25.18). Because of Eq. (25.15) we have the following estimate:

$$\rho(\tilde{X}_t, \tilde{X}_s) \le K_\alpha |t-s|^\alpha \text{ for all } s, t \in D, \tag{25.22}$$

where

$$K_\alpha := 1_{\Omega_0} \cdot C_1(\alpha, T)\, 2^{\nu(1-\alpha)}.$$

Since $\{\nu > N\} = \cup_{m=N}^{\infty} A_n$, it follows from Eq. (25.20), that

$$P(\nu > N) \le \sum_{m\ge N} P(A_m) \le \sum_{m\ge N} C 2^{-m(\beta-\alpha\varepsilon)} \le C \frac{2^{-N(\beta-\alpha\varepsilon)}}{1-2^{-(\beta-\alpha\varepsilon)}}. \tag{25.23}$$

Using this estimate we find

$$\begin{aligned} \mathbb{E}K_\alpha^p &= C_1(\alpha,T)^p\, \mathbb{E}\left[2^{\nu(1-\alpha)p}\right] \\ &= C_1(\alpha,T)^p \cdot \sum_{n=0}^{\infty} 2^{n(1-\alpha)p} P(\nu = n) \\ &\le C_1(\alpha,T)^p \cdot \left(1 + \sum_{n=1}^{\infty} 2^{n(1-\alpha)p} P(\nu > n-1)\right) \\ &\le C_1(\alpha,T)^p \cdot \left(1 + \frac{C}{1-2^{-(\beta-\alpha\varepsilon)}} \sum_{n=1}^{\infty} 2^{n(1-\alpha)p} 2^{-(n-1)(\beta-\alpha\varepsilon)}\right) \end{aligned}$$

which is finite provided that $(1-\alpha)p - (\beta - \alpha\varepsilon) < 0$. ■

26

Brownian Motion I

Our next goal is to prove existence of Brownian motion and then describe some of its basic path properties.

Definition 26.1 (Brownian Motion). *A* ***Brownian motion*** $\{B_t\}_{t\geq 0}$ *is an adapted mean zero Gaussian random process on some filtered probability space,* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, *satisfying; 1) for each* $\omega \in \Omega$, $t \to B_t(\omega)$ *is continuous, and 2)*

$$\mathbb{E}\left[B_t B_s\right] = t \wedge s \text{ for all } s, t \geq 0. \tag{26.1}$$

So a Brownian motion is a pre-Brownian motion with continuous sample paths.

Remark 26.2. If no filtration is given, we can use the process to construct one. Namely, let $\mathcal{B}_t^0 := \sigma(B_s : s \leq t)$ and replace $\mathcal{B}$ by $\vee \mathcal{B}_t^0$ if necessary. We call $\left\{\mathcal{B}_t^0\right\}$ the raw filtration associated to $\{B_t\}$.

Theorem 26.3 (Wiener 1923). *Brownian motions exists. Moreover for any* $\alpha \in (0, 1/2)$, $t \to B_t$ *is locally* α *– Hölder continuous almost surely.*

Proof. For $0 \leq s < t < \infty$, $\tilde{B}_t - \tilde{B}_s$ is a mean zero Gaussian random variable with

$$\mathbb{E}\left[\left(\tilde{B}_t - \tilde{B}_s\right)^2\right] = \mathbb{E}\left[\tilde{B}_t^2 + \tilde{B}_s^2 - 2B_s\tilde{B}_t\right] = t + s - 2s = t - s.$$

Hence if N is a standard normal random variable, then $\tilde{B}_t - \tilde{B}_s \overset{d}{=} \sqrt{t-s}N$ and therefore, for any $p \in [1, \infty)$,

$$\mathbb{E}\left|\tilde{B}_t - \tilde{B}_s\right|^p = (t-s)^{p/2}\,\mathbb{E}\,|N|^p. \tag{26.2}$$

Hence an application of Theorem 25.7 shows, with $\varepsilon = p > 2$, $\beta = p/2 - 1$, $\alpha \in \left(0, \frac{p/2-1}{p}\right) = \left(0, \frac{1}{2} - 1/p\right)$, there exists a modification, B of $\tilde{B}$ such that

$$|B_t - B_s| \leq C_{\alpha,T}\,|t-s|^{\alpha} \text{ for } s, t \in [0, T).$$

By applying this result with $T = N \in \mathbb{N}$, we find there exists a continuous version, B, of $\tilde{B}$ for all $t \in [0, \infty)$ and this version is locally Hölder continuous with Hölder constant $\alpha < 1/2$. ■

For the rest of this chapter we will assume that $\{B_t\}_{t\geq 0}$ is a Brownian motion on some probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ and $\mathcal{B}_t := \sigma(B_s : s \leq t)$.

26.1 Donsker's Invariance Principle

In this section we will see that Brownian motion may be thought of as a limit of random walks – this is the content of Donsker's invariance principle or the so called functional central limit theorem. The setup is to start with a random walk, $S_n := X_1 + \cdots + X_n$ where $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables with zero mean and variance one. We then define for each $n \in \mathbb{N}$ the following continuous process,

$$B_n(t) := \frac{1}{\sqrt{n}}\left(S_{[nt]} + (nt - [nt])\,X_{[nt]+1}\right) \tag{26.3}$$

where for $\tau \in \mathbb{R}_+$, $[\tau]$ is the integer part of τ, i.e. the nearest integer to τ which is no greater than τ. The first step in this program is to prove convergence in the sense of finite dimensional distributions.

Proposition 26.4. *Let* B *be a standard Brownian motion, then* $B_n \overset{f.d.}{\Longrightarrow} B$.

Proof. In Exercise 21.12 you showed that $\left\{\frac{1}{\sqrt{n}}S_{[nt]}\right\}_{t\geq 0} \overset{\text{f.d.}}{\Longrightarrow} B$ as $n \to \infty$. Suppose that $0 < t_1 < t_2 < \cdots < t_k < \infty$ are given and we let

$$X_n := \frac{1}{\sqrt{n}}\left(S_{[nt_1]}, \ldots, S_{[nt_k]}\right),$$
$$Y_n := \left(B_n(t_1), \ldots, B_n(t_k)\right)$$

and $\varepsilon_n := Y_n - X_n$. From Eq. (26.3) and Chebyshev's inequality it follows that

$$P\left(|(\varepsilon_n)_i| > \varepsilon\right) \leq \frac{1}{\varepsilon}\mathbb{E}\left[\frac{1}{\sqrt{n}}\left(nt_i - [nt_i]\right)\left|X_{[nt]+1}\right|\right] \leq \sqrt{n}\frac{\mathbb{E}\,|X_1|}{\varepsilon} \to 0$$

as $n \to \infty$. This then easily implies that $\varepsilon_n \overset{P}{\to} 0$ as $n \to \infty$ and therefore by Slutzky's Theorem 21.39 it follows that $Y_n = X_n + \varepsilon_n \implies (B(t_1), \ldots, B(t_k))$. ■

Let $\Omega := C([0, \infty), \mathbb{R})$ which becomes a complete metric space in the metric defined by,

$$\rho(\omega_1, \omega_2) := \sum_{n=1}^{\infty} \frac{1}{2^n} \max_{0\leq t\leq n} |\omega_1(t) - \omega_2(t)| \wedge 1.$$

Theorem 26.5 (Donsker's invariance principle). *Let Ω, B_n, and B be as above. Then for $B_n \implies B$, i.e.*

$$\lim_{n\to\infty} \mathbb{E}\left[F\left(B_n\right)\right] = \mathbb{E}\left[F\left(B\right)\right]$$

for all bounded continuous functions $F : \Omega \to \mathbb{R}$.

One method of proof, see [31] and [49], goes by the following two steps. 1) Show that the finite dimensional distributions of B_n converge to a B as was done in Proposition 26.4. 2) Show the distributions of B_n are tight. A proof of the tightness may be based on Ottaviani's maximal inequality in Corollary 20.48, [30, Corollary 16.7]. Another possible proof or this theorem is based on "Skorokhod's representation," see (for example) [30, Theorem 14.9 on p. 275] or [41]. Rather than give the full proof here I will give the proof of a slightly weaker version of the theorem under the more stringent restriction that the X_n posses fourth moments.

Proposition 26.6 (Random walk approximation bounds). *Suppose that $\{X_n\}_{n=1}^\infty \subset L^4(P)$ are i.i.d. random variables with $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = 1$ and $\gamma := \mathbb{E}X_n^4 < \infty$. Then there exists $C < \infty$ such that*

$$\mathbb{E}\left|B_n(t) - B_n(s)\right|^4 \le C\left|t-s\right|^2 \text{ for all } s,t \in \mathbb{R}_+. \tag{26.4}$$

Exercise 26.1. Provide a proof of Proposition 26.6. **Hints:** Use the results of Exercise 10.9 to verify that Eq. (26.4) holds for $s,t \in D_n := \frac{1}{n}\mathbb{N}_0$. Take care of the case where $s,t \ge 0$ with $|t-s| < 1/n$ by hand and finish up using these results along with Minikowski's inequality.

Theorem 26.7 (Baby Donsker Theorem). *Continuing the notation used in Proposition 26.6 and let $T < \infty$ and B be a Brownian motion. Then*

$$B_n|_{[0,T]} \implies B|_{[0,T]}, \tag{26.5}$$

i.e. $\mathrm{Law}\left(B_n|_{[0,T]}\right) \implies \mathrm{Law}\left(B|_{[0,T]}\right)$ *as distributions on $\Omega_T = C([0,T],\mathbb{R})$ – a complete separable metric space. (See the simulation file* ***Random Walks to BM.xls.)***

Proof. If Eq. (26.5) fails to hold there would exists $g \in BC(\Omega_T)$ and a subsequence $B'_k = B_{n_k}$ such that

$$\varepsilon := \inf_k \left|\mathbb{E}\left[g\left(B'_k|_{[0,T]}\right)\right] - \mathbb{E}g\left(B|_{[0,T]}\right)\right| > 0. \tag{26.6}$$

Since $B_n(0) = 0$ for all n and the estimate in Eq. (26.4) of Proposition 26.6 holds, it follows from Theorem 25.9 that $\left\{B_n|_{[0,T]}\right\}_{n=1}^\infty$ is tight. So by Prokhorov's Theorem 21.61, there is a further subsequence $B''_l = B'_{k_l}$ which is weakly convergent to some Ω_T – valued process X. Replacing B'_k bit B''_l in Eq. (26.6) and then letting $l \to \infty$ in the resulting equation shows

$$\left|\mathbb{E}\left[g\left(X|_{[0,T]}\right)\right] - \mathbb{E}g\left(B|_{[0,T]}\right)\right| \ge \varepsilon > 0. \tag{26.7}$$

On the other hand by Proposition 26.4 we know that $B_n \overset{\text{f.d.}}{\implies} B$ as $n \to \infty$ and therefore X and B are continuous processes on $[0,T]$ with the same finite dimensional distributions and hence are indistinguishable by Exercise 25.1. However this is in contradiction to Eq. (26.7). ■

26.2 Path Regularity Properties of BM

Definition 26.8. *Let $(V, \|\cdot\|)$ be a normed space and $Z \in C([0,T],V)$. For $1 \le p < \infty$, the* ***p - variation*** *of Z is;*

$$v_p(Z) := \sup_{\Pi}\left(\sum_{j=1}^n \left\|Z_{t_j} - Z_{t_{j-1}}\right\|^p\right)^{1/p}$$

where the supremum is taken over all partitions, $\Pi := \{0 = t_0 < t_1 < \dots < t_n = T\}$, of $[0,T]$.

Lemma 26.9. *The function $v_p(Z)$ is a decreasing function of p.*

Proof. Let $a := \{a_j\}_{j=1}^n$ be a sequence of non-negative numbers and set

$$\|a\|_p := \left(\sum_{j=1}^n a_j^p\right)^{1/p}.$$

It will suffice to show $\|a\|_p$ is a decreasing function of p. To see this is true, $q = p + r$. Then

$$\|a\|_q^q = \sum_{j=1}^n a_j^{p+r} \le \left(\max_j a_j\right)^r \cdot \sum_{j=1}^n a_j^p \le \|a\|_p^r \cdot \|a\|_p^p = \|a\|_p^q,$$

wherein we have used,

$$\max_j a_j = \left(\max_j a_j^p\right)^{1/p} \le \left(\sum_{j=1}^n a_j^p\right)^{1/p} = \|a\|_p.$$

■

Notation 26.10 (Partitions) *Given $\mathcal{P} := \{0 = t_0 < t_1 < \cdots < t_n = T\}$, a partition of $[0,T]$, let*

$$\Delta_i B := B_{t_i} - B_{t_{i-1}}, \text{ and } \Delta_i t := t_i - t_{i-1}$$

for all $i = 1, 2, \ldots, n$. Further let $\operatorname{mesh}(\mathcal{P}) := \max_i |\Delta_i t|$ denote the mesh of the partition, $\mathcal{P}$.

Corollary 26.11. *For all $p > 2$ and $T < \infty$, $v_p\left(B|_{[0,T]}\right) < \infty$ a.s. (We will see later that $v_p\left(B|_{[0,T]}\right) = \infty$ a.s. for all $p < 2$.)*

Proof. By Theorem 26.3, there exists $K_p < \infty$ a.s. such that

$$|B_t - B_s| \le K_p |t-s|^{1/p} \text{ for all } 0 \le s, t \le T. \tag{26.8}$$

Thus we have

$$\sum_i |\Delta_i B|^p \le \sum_i \left(K_p |t_i - t_{i-1}|^{1/p}\right)^p \le \sum_i K_p^p |t_i - t_{i-1}| = K_p^p T$$

and therefore, $v_p\left(B|_{[0,T]}\right) \le K_p^p T < \infty$ a.s. ∎

Exercise 26.2 (Quadratic Variation). Let

$$\mathcal{P}_m := \left\{0 = t_0^m < t_1^m < \cdots < t_{n_m}^m = T\right\}$$

be a sequence of partitions such that $\operatorname{mesh}(\mathcal{P}_m) \to 0$ as $m \to \infty$. Further let

$$Q_m := \sum_{i=1}^{n_m} (\Delta_i^m B)^2 := \sum_{i=1}^{n_m} \left(B_{t_i^m} - B_{t_{i-1}^m}\right)^2. \tag{26.9}$$

Show

$$\lim_{m\to\infty} \mathbb{E}\left[(Q_m - T)^2\right] = 0$$

and $\lim_{m\to\infty} Q_m = T$ a.s. if $\sum_{m=1}^\infty \operatorname{mesh}(\mathcal{P}_m) < \infty$. This result is often abbreviated by the writing, $dB_t^2 = dt$. **Hint:** it is useful to observe; 1)

$$Q_m - T = \sum_{i=1}^{n_m} \left[(\Delta_i^m B)^2 - \Delta_i t\right]$$

and 2) using Eq. (26.2) there is a constant, $c < \infty$ such that

$$\mathbb{E}\left[(\Delta_i^m B)^2 - \Delta_i t\right]^2 = c\,(\Delta_i t)^2.$$

Proposition 26.12. *Suppose that $\{\mathcal{P}_m\}_{m=1}^\infty$ is a sequence of partitions of $[0,T]$ such that $\mathcal{P}_m \subset \mathcal{P}_{m+1}$ for all m and $\operatorname{mesh}(\mathcal{P}_m) \to 0$ as $m \to \infty$. Then $Q_m \to T$ a.s. where Q_m is defined as in Eq. (26.9).*

Proof. It is always possible to find another sequence of partitions, $\{\mathcal{P}'_n\}_{n=1}^\infty$, of $[0,T]$ such that $\mathcal{P}'_n \subset \mathcal{P}'_{n+1}$, $\operatorname{mesh}(\mathcal{P}'_n) \to 0$ as $n \to \infty$, $\#\left(\mathcal{P}'_{n+1}\right) = \#\left(\mathcal{P}'_n\right) + 1$, and $\mathcal{P}_m = \mathcal{P}'_{n_m}$ where $\{n_m\}_{m=1}^\infty$ is a subsequence of $\mathbb{N}$. If we let Q'_n denote the quadratic variations associated to $\mathcal{P}'_n$ and we can shown $Q'_n \to T$ a.s. then we will also have $Q_m = Q'_{n_m} \to T$ a.s. as well. So with these comments we may now assume that $\#\left(\mathcal{P}_{n+1}\right) = \#\left(\mathcal{P}_n\right) + 1$.

We already know form Exercise 26.2 that $Q_m \to T$ in $L^2(P)$. So it suffices to show Q_m is almost surely convergent. We will do this by showing $\{Q_m\}_{m=1}^\infty$ is a backwards martingale relative to the filtration,

$$\mathcal{F}_m := \sigma\left(Q_m, Q_{m+1}, \ldots\right).$$

To do this, suppose that $\mathcal{P}_{m+1} = \mathcal{P}_m \cup \{v\}$ and $u = t_{i-1}, v = t_{i+1} \in \mathcal{P}_m$ such that $u < v < w$. Let $X := B_v - B_w$ and $Y := B_w - B_u$. Then

$$\begin{aligned} Q_m &= Q_{m+1} - (B_v - B_w)^2 - (B_w - B_u)^2 + (B_v - B_u)^2 \\ &= Q_{m+1} - X^2 - Y^2 + (X+Y)^2 \\ &= Q_{m+1} + 2XY \end{aligned}$$

therefore,

$$\mathbb{E}\left[Q_m | \mathcal{F}_{m+1}\right] = Q_{m+1} + 2\mathbb{E}\left[XY | \mathcal{F}_{m+1}\right].$$

So to finish the proof it suffices to show $\mathbb{E}\left[XY | \mathcal{F}_{m+1}\right] = 0$ a.s.

To do this let

$$b_t := \begin{cases} B_t & \text{if } t \le v \\ B_v - (B_t - B_v) & \text{if } t \ge v, \end{cases}$$

that is after $t = v$, the increments of b are the **reflections** of the increments of B. Clearly b_t is still a continuous process and it is easily verified that $\mathbb{E}\left[b_t b_s\right] = s \wedge t$. Thus $\{b_t\}_{t\ge 0}$ is still a Brownian motion. Moreover, if $Q_{m+n}(b)$ is the quadratic variation of b relative to $\mathcal{P}_{m+n}$, then

$$Q_{m+n}(b) = Q_{m+n} = Q_{m+n}(B) \text{ for all } n \in \mathbb{N}.$$

On the other hand, under this transformation, $X \to X$ and $Y \to -Y$. Since $(X, Y, Q_{m+1}, Q_{m+2}, \ldots)$ and $(-X, Y, Q_{m+1}, Q_{m+2}, \ldots)$ have the same distribution, if we write

$$\mathbb{E}\left[XY | \mathcal{F}_{m+1}\right] = f\left(Q_{m+1}, Q_{m+2}, \ldots\right) \text{ a.s.}, \tag{26.10}$$

then it follows from Exercise 14.6, that

$$\mathbb{E}\left[-XY|\mathcal{F}_{m+1}\right]=f\left(Q_{m+1},Q_{m+2},\dots\right) \text{ a.s.} \tag{26.11}$$

Hence we may conclude,

$$\mathbb{E}\left[XY|\mathcal{F}_{m+1}\right]=\mathbb{E}\left[-XY|\mathcal{F}_{m+1}\right]=-\mathbb{E}\left[XY|\mathcal{F}_{m+1}\right],$$

and thus $\mathbb{E}\left[XY|\mathcal{F}_{m+1}\right]=0$ a.s. ■

Corollary 26.13. *If $p<2$, then $v_p\left(B|_{[0,T]}\right)=\infty$ a.s.*

Proof. Choose partitions, $\{\mathcal{P}_m\}$, of $[0,T]$ such that $\lim_{m\to\infty}Q_m=T$ a.s. where Q_m is as in Eq. (26.9) and let $\Omega_0:=\{\lim_{m\to\infty}Q_m=T\}$ so that $P(\Omega_0)=1$. If $v_p\left(B|_{[0,T]}(\omega)\right)<\infty$ for some $\omega\in\Omega_0$, then, with $b=B(\omega)$, we would have

$$\begin{aligned}Q_m(\omega)&=\sum_{i=1}^{n_m}\left(\Delta_i^m b\right)^2=\sum_{i=1}^{n_m}\left|\Delta_i^m b\right|^p\cdot\left|\Delta_i^m b\right|^{2-p}\\&\le\max_i\left|\Delta_i^m b\right|^{2-p}\cdot\sum_{i=1}^{n_m}\left|\Delta_i^m b\right|^p\le\left[v_p(b)\right]^p\cdot\max_i\left|\Delta_i^m b\right|^{2-p}\end{aligned}$$

which tends to zero as $m\to\infty$ by the uniform continuity of b. But this contradicts the fact that $\lim_{m\to\infty}Q_m(\omega)=T$. Thus we must $v_p\left(B|_{[0,T]}\right)=\infty$ on Ω_0. ■

Remark 26.14. The reader may find a proof that Corollary 26.13 also holds for $p=2$ in [19, Theorem 13.69 on p. 382]. You should consider why this result in not in contradiction with Exercise 26.2 and Theorem 26.12. **Hint:** unlike the case of $p=1$, when $p>1$ the quantity;

$$v_p^{\Pi}(Z):=\left(\sum_{j=1}^{n}\left\|Z_{t_j}-Z_{t_{j-1}}\right\|^p\right)^{1/p}$$

does not increase under refinement of partitions so that

$$v_p(Z)=\sup_{\Pi}v_p^{\Pi}(Z)\neq\lim_{|\Pi|\to 0}v_p^{\Pi}(Z)$$

when $p>1$.

Corollary 26.15 (Roughness of Brownian Paths). *A Brownian motion, $\{B_t\}_{t\ge 0}$, is **not** almost surely α – Hölder continuous for any $\alpha>1/2$.*

Proof. According to Exercise 26.2, we may choose partition, $\mathcal{P}_m$, such that $\operatorname{mesh}(\mathcal{P}_m)\to 0$ and $Q_m\to T$ a.s. If B were α – Hölder continuous for some $\alpha>1/2$, then

$$\begin{aligned}Q_m&=\sum_{i=1}^{n_m}\left(\Delta_i^m B\right)^2\le C\sum_{i=1}^{n_m}\left(\Delta_i^m t\right)^{2\alpha}\le C\max\left(\left[\Delta_i t\right]^{2\alpha-1}\right)\sum_{i=1}^{n_m}\Delta_i^m t\\&\le C\left[\operatorname{mesh}(\mathcal{P}_m)\right]^{2\alpha-1}T\to 0 \text{ as } m\to\infty\end{aligned}$$

which contradicts the fact that $Q_m\to T$ as $m\to\infty$. ■

Lemma 26.16. *For any $\alpha>1/2$, $\limsup_{t\downarrow 0}|B_t|/t^\alpha=\infty$ a.s. (See Exercise 31.3 below to see that $\alpha=1/2$ would work as well.)*

Proof. If $\limsup_{t\downarrow 0}|\omega_t|/t^\alpha<\infty$ then there would exists $C<\infty$ such that $|\omega_t|\le Ct^\alpha$ for all $t\le 1$ and in particular, $\left|\omega_{1/n}\right|\le Cn^{-\alpha}$ for all $n\in\mathbb{N}$. Hence we have shown

$$\left\{\limsup_{t\downarrow 0}|B_t|/t^\alpha<\infty\right\}\subset\cup_{C\in\mathbb{N}}\cap_{n\in\mathbb{N}}\left\{\left|B_{1/n}\right|\le Cn^{-\alpha}\right\}.$$

This completes the proof because,

$$\begin{aligned}P\left(\cap_{n\in\mathbb{N}}\left\{\left|B_{1/n}\right|\le Cn^{-\alpha}\right\}\right)&\le\liminf_{n\to\infty}P\left(\left|B_{1/n}\right|\le Cn^{-\alpha}\right)\\&=\liminf_{n\to\infty}P\left(\frac{1}{\sqrt{n}}|B_1|\le Cn^{-\alpha}\right)\\&=\liminf_{n\to\infty}P\left(|B_1|\le Cn^{1/2-\alpha}\right)=P(|B_1|=0)=0\end{aligned}$$

if $\alpha>1/2$. ■

Theorem 26.17 (Nowhere $1/2+\varepsilon$ – Hölder Continuous). *Let*

$$W:=\{\omega\in C([0,\infty)\to\mathbb{R}):\omega(0)=0\},$$

*$\mathcal{B}$ denote the σ – field on W generated by the projection maps, $b_t(\omega)=\omega(t)$ for all $t\in[0,\infty)$, and μ be **Wiener measure** on $(W,\mathcal{B})$, i.e. μ is the Law of a Brownian motion. For $\alpha>1/2$ and E_α denote the set of $\omega\in W$ such that ω is α–Hölder continuous at some point $t=t_\omega\in[0,1]$. when $\mu^*(E_\alpha)=0$, i.e. there exists a set $\tilde{E}_\alpha\in\mathcal{B}$ such that*

$$E_\alpha=\left\{\inf_{0\le t\le 1}\limsup_{h\to 0}\frac{|\omega(t+h)-\omega(t)|}{|h|^\alpha}<\infty\right\}\subset\tilde{E}_\alpha$$

and $\mu\left(\tilde{E}_\alpha\right)=0$. In particular, μ is concentrated on $\tilde{E}_\alpha^c$ which is a subset of the collection paths which are nowhere differentiable on $[0,1]$.

Proof. Let $\alpha \in (0,1)$ and $\nu \in \mathbb{N}$ – to be chosen more specifically later. If $\omega \in E_\alpha$, then there exists, $t \in [0,1]$, $C < \infty$, such that

$$|\omega(t) - \omega(s)| \le C\,|t-s|^\alpha \text{ for all } |s| \le \nu+1.$$

For all $n \in \mathbb{N}$ we may choose $i \ge 0$ so that $\left|t - \frac{i}{n}\right| < \frac{1}{n}$. By the triangle inequality, for all $j = 1,2,\dots,\nu$, we have

$$\begin{aligned}\left|\omega\left(\frac{i+j}{n}\right) - \omega\left(\frac{i+j-1}{n}\right)\right| &\le \left|\omega\left(\frac{i+j}{n}\right) - \omega(t)\right| + \left|\omega(t) - \omega\left(\frac{i+j-1}{n}\right)\right| \\ &\le C\left[\left|\frac{i+j}{n} - t\right|^\alpha + \left|\frac{i+j-1}{n} - t\right|^\alpha\right] \\ &\le Cn^{-\alpha}\left[|\nu+1|^\alpha + |\nu|^\alpha\right] =: Dn^{-\alpha}.\end{aligned}$$

Therefore, $\omega \in E_\alpha$ implies there exists $D \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ there exists $i \le n$ such that

$$\left|\omega\left(\frac{i+j}{n}\right) - \omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\ \forall\ j = 1,2,\dots,\nu.$$

Letting

$$A_D := \cap_{n=1}^\infty \cup_{i\le n} \cap_{j=1}^\nu \left\{\omega : \left|\omega\left(\frac{i+j}{n}\right) - \omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\right\},$$

we have shown that $E_\alpha \subset \cup_{D\in\mathbb{N}} A_D$. We now complete the proof by showing $P(A_D) = 0$. To do this, we compute,

$$\begin{aligned}P(A_D) &\le \liminf_{n\to\infty} P\left(\cup_{i\le n}\cap_{j=1}^\nu\left\{\omega : \left|\omega\left(\frac{i+j}{n}\right) - \omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\right\}\right) \\ &\le \liminf_{n\to\infty} \sum_{i\le n}\prod_{j=1}^\nu P\left(\omega : \left|\omega\left(\frac{i+j}{n}\right) - \omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\right) \\ &= \liminf_{n\to\infty} n\left[P\left(\frac{1}{\sqrt{n}}|N| \le Dn^{-\alpha}\right)\right]^\nu \\ &= \liminf_{n\to\infty} n\left[P\left(|N| \le Dn^{\frac{1}{2}-\alpha}\right)\right]^\nu \\ &\le \liminf_{n\to\infty} n\left[Cn^{\frac{1}{2}-\alpha}\right]^\nu = C^\nu \liminf_{n\to\infty} n^{1+\left(\frac{1}{2}-\alpha\right)\nu}. \qquad (26.12)\end{aligned}$$

wherein we have used

$$\mu(|N| \le \delta) = \frac{1}{\sqrt{2\pi}}\int_{|x|\le\delta} e^{-\frac{1}{2}x^2}dx \le \frac{1}{\sqrt{2\pi}}2\delta.$$

The last limit in Eq. (26.12) is zero provided we choose $\alpha > \frac{1}{2}$ and $\nu\left(\alpha - \frac{1}{2}\right) > 1$. ■

26.3 Scaling Properties of B. M.

Theorem 26.18 (Transformations preserving B. M.). *Let $\{B_t\}_{t\ge0}$ be a Brownian motion and $\mathcal{B}_t := \sigma(B_s : s \le t)$. Then;*

1. *$b_t = -B_t$ is again a Brownian motion.*
2. *if $c > 0$ and $b_t := c^{-1/2}B_{ct}$ is again a Brownian motion.*
3. *$b_t := tB_{1/t}$ for $t > 0$ and $b_0 = 0$ is a Brownian motion. In particular, $\lim_{t\downarrow0} tB_{1/t} = 0$ a.s.*
4. *for all $T \in (0,\infty)$, $b_t := B_{t+T} - B_T$ for $t \ge 0$ is again a Brownian motion which is independent of $\mathcal{B}_T$.*
5. *for all $T \in (0,\infty)$, $b_t := B_{T-t} - B_T$ for $0 \le t \le T$ is again a Brownian motion on $[0,T]$.*

Proof. It is clear that in each of the four cases above $\{b_t\}_{t\ge0}$ is still a Gaussian process. Hence to finish the proof it suffices to verify, $\mathbb{E}[b_t b_s] = s \wedge t$ which is routine in all cases. Let us work out item 3. in detail to illustrate the method. For $0 < s < t$,

$$\mathbb{E}[b_s b_t] = st\mathbb{E}[B_{s^{-1}}B_{t^{-1}}] = st\left(s^{-1}\wedge t^{-1}\right) = st\cdot t^{-1} = s.$$

Notice that $t \to b_t$ is continuous for $t > 0$, so to finish the proof we must show that $\lim_{t\downarrow0} b_t = 0$ a.s. However, this follows from Kolmogorov's continuity criteria. Since $\{b_t\}_{t\ge0}$ is a pre-Brownian motion, we know there is a version, $\tilde{b}$ which is a.s. continuous for $t \in [0,\infty)$. By exercise 25.1, we know that

$$E := \left\{\omega \in \Omega : b_t(\omega) \ne \tilde{b}_t(\omega) \text{ for some } t > 0\right\}$$

is a null set. Hence $\omega \notin E$ it follows that

$$\lim_{t\downarrow0} b_t(\omega) = \lim_{t\downarrow0}\tilde{b}_t(\omega) = 0.$$

■

Corollary 26.19 (B. M. Law of Large Numbers). *Suppose $\{B_t\}_{t\ge0}$ is a Brownian motion, then almost surely, for each $\beta > 1/2$,*

$$\limsup_{t\to\infty}\frac{|B_t|}{t^\beta} = \begin{cases} 0 & \text{if } \beta > 1/2 \\ \infty & \text{if } \beta \in (0,1/2). \end{cases} \qquad (26.13)$$

Proof. Since $b_t := tB_{1/t}$ for $t > 0$ and $b_0 = 0$ is a Brownian motion, we know that for all $\alpha < 1/2$ there exists, $C_\alpha(\omega) < \infty$ such that, almost surely,

$$t\left|B_{1/t}\right| = \left|tB_{1/t}\right| = |b_t| \le C_\alpha|t|^\alpha \text{ for all } t \le 1.$$

Replacing t by $1/t$ in this inequality implies, almost surely, that

$$\frac{1}{t}|B_t| \le \frac{C_\alpha}{|t|^\alpha} \text{ for all } t \ge 1.$$

or equivalently that

$$|B_t| \le C_\alpha t^{1-\alpha} \text{ for all } t \ge 1. \tag{26.14}$$

Hence if $\beta > 1/2$, let $\alpha < 1/2$ such that $\beta < 1-\alpha$. Then Eq. (26.13) follows from Eq. (26.14).

On the other hand, taking $\alpha > 1/2$, we know by Lemma 26.16 (or Theorem 26.17) that

$$\limsup_{t\downarrow 0} \frac{t\left|B_{1/t}\right|}{t^\alpha} = \limsup_{t\downarrow 0} \frac{|b_t|}{t^\alpha} = \infty \text{ a.s.}$$

This may be expressed as saying

$$\infty = \limsup_{t\to\infty} \frac{t^{-1}|B_t|}{t^{-\alpha}} = \limsup_{t\to\infty} \frac{|B_t|}{t^{1-\alpha}} \text{ a.s.}$$

Since $\beta := 1-\alpha$ is any number less that $1/2$, the proof is complete. ∎

27

Filtrations and Stopping Times

For our later development we need to go over some measure theoretic preliminaries about processes indexed by $\mathbb{R}_+ := [0,\infty)$. We will continue this discussion in more depth later. For this chapter we will always suppose that $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}\right)$ is a **filtered measurable space**, i.e. Ω is as set, $\mathcal{B} \subset 2^{\Omega}$ is a σ algebra, and $\{\mathcal{B}_t\}_{t\in\mathbb{R}_+}$ is a **filtration** which to say each $\mathcal{B}_t$ is a sub – σ – algebra of $\mathcal{B}$ and $\mathcal{B}_s \subset \mathcal{B}_t$ for all $s \le t$.

27.1 Measurability Structures

Notation 27.1 ($\mathcal{B}_t^{\pm}$) *Let*

$$\mathcal{B}_\infty = \mathcal{B}_{\infty+} = \vee_{t\in\mathbb{R}_+}\mathcal{B}_t = \sigma\left(\cup_{t\in\mathbb{R}_+}\mathcal{B}_t\right) \subset \mathcal{B},$$

and for $t \in \mathbb{R}_+$, let

$$\mathcal{B}_t^+ = \mathcal{B}_{t+} := \cap_{s>t}\mathcal{B}_s.$$

Also let $\mathcal{B}_{0-} := \mathcal{B}_0$ and for $t \in (0,\infty]$ let

$$\mathcal{B}_{t-} := \vee_{s<t}\mathcal{B}_s = \sigma\left(\cup_{s<t}\mathcal{B}_s\right).$$

(Observe that $\mathcal{B}_{\infty-} = \mathcal{B}_\infty$.)

The filtration, $\left\{\mathcal{B}_t^+\right\}_{t\in\mathbb{R}_+}$, "peaks" infinitesimally into the future while $\mathcal{B}_t^-$ limits itself to knowing about the state of the system up to the times infinitesimally before time t.

Definition 27.2 (Right continuous filtrations). *The filtration $\{\mathcal{B}_t\}_{t\ge0}$ is **right continuous** if $\mathcal{B}_t^+ := \mathcal{B}_{t+} = \mathcal{B}_t$ for all $t \ge 0$.*

The next result is trivial but we record it as a lemma nevertheless.

Lemma 27.3 (Right continuous extension). *Suppose $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\ge0}\right)$ is a filtered space and $\mathcal{B}_t^+ := \mathcal{B}_{t+} := \cap_{s>t}\mathcal{B}_s$. Then $\left\{\mathcal{B}_t^+\right\}_{t\ge0}$ is right continuous. (We refer to $\left\{\mathcal{B}_t^+\right\}_{t\in\mathbb{R}_+}$ as the **right continuous filtration** associated to $\{\mathcal{B}_t\}_{t\in\mathbb{R}_+}$.)*

Exercise 27.1. Suppose $(\Omega, \mathcal{F})$ is a measurable space, (S,ρ) is a separable metric space[1], and $\mathcal{S}$ is the Borel σ – algebra on S – i.e. the σ – algebra generated by all open subset of S.

1. Let $D \subset S$ be a countable dense set and $\mathbb{Q}_+ := \mathbb{Q} \cap \mathbb{R}_+$. Show $\mathcal{S}$ may be described as the σ – algebra generated by all open (or closed) balls of the form
$$B(a,\varepsilon) := \{s \in S : \rho(s,a) < \varepsilon\} \tag{27.1}$$
$$(\text{or } C(a,\varepsilon) := \{s \in S : \rho(s,a) \le \varepsilon\}) \tag{27.2}$$
with $a \in D$ and $\varepsilon \in \mathbb{Q}_+$.
2. Show a function, $Y : \Omega \to S$, is $\mathcal{F}/\mathcal{S}$ – measurable iff the functions, $\Omega \ni \omega \to \rho(x, Y(\omega)) \in \mathbb{R}_+$ are measurable for all $x \in D$. **Hint:** show, for each $x \in S$, that $\rho(x,\cdot) : S \to \mathbb{R}_+$ is a measurable map.
3. If $X_n : \Omega \to S$ is a sequence of $\mathcal{F}/\mathcal{S}$ – measurable maps such that $X(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists in S for all $\omega \in \Omega$, then the limiting function, X, is $\mathcal{F}/\mathcal{S}$ – measurable as well. (**Hint:** use item 2.)

Definition 27.4. *Suppose S is a metric space, $\mathcal{S}$ is the Borel σ – algebra on S, and $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}\right)$ is a filtered measurable space. A process, $X_t : \Omega \to S$ for $t \in \mathbb{R}_+$ is;*

1. ***adapted** if X_t is $\mathcal{B}_t/\mathcal{S}$– measurable for all $t \in \mathbb{R}_+$,*
2. ***right continuous** if $t \to X_t(\omega)$ is right continuous for all $\omega \in \Omega$,*
3. ***left continuous** if $t \to X_t(\omega)$ is left continuous for all $\omega \in \Omega$, and*
4. ***progressively measurable**, if for all $T \in \mathbb{R}_+$, the map $\varphi^T : [0,T] \times \Omega \to S$ defined by $\varphi^T(t,\omega) := X_t(\omega)$ is $\mathcal{B}_{[0,T]} \otimes \mathcal{B}_T/\mathcal{S}$ – measurable.*

Lemma 27.5. *Let $\varphi(t,\omega) := X_t(\omega)$ where we are continuing the notation in Definition 27.4. If $X_t : \Omega \to S$ is a progressively measurable process then $X.$ is adapted and $\varphi : \mathbb{R}_+ \times \Omega \to S$ is $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}/\mathcal{S}$ – measurable and the X is adapted.*

Proof. For $T \in \mathbb{R}_+$, let $\eta_T : \Omega \to [0,T] \times \Omega$, be defined by $\eta_T(\omega) := (T,\omega)$. If $a \in [0,T]$ and $A \in \mathcal{B}_T$, then $\eta_T^{-1}([0,a] \times A) = \emptyset \in \mathcal{B}_T$ if $a \ne T$

[1] If you are unconfortable with this much generality, you may assume S is a subset of $\mathbb{R}^d$ and $\rho(x,y) := \|x-y\|$ for all $x, y \in S$.

and $\eta_T^{-1}([0,a]\times A) = A \in \mathcal{B}_T$ if $a = T$. This shows η_T is $\mathcal{B}_T/\mathcal{B}_{[0,T]}\otimes\mathcal{B}$ – measurable. Therefore, the composition, $\varphi^T\circ\eta_T = X_T$ is $\mathcal{B}_T/\mathcal{S}$ – measurable for all $T\in\mathbb{R}_+$ which is the statement that X is adapted.

For $V\in\mathcal{S}$ and $T<\infty$, we have

$$\varphi^{-1}(V)\cap([0,T]\times\Omega) = \left(\varphi^T\right)^{-1}(V)\in\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T\subset\mathcal{B}_{[0,T]}\otimes\mathcal{B}. \tag{27.3}$$

Since $\mathcal{B}_{[0,T]}\otimes\mathcal{B}$ and $\left(\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right)_{[0,T]\times\Omega}$ are σ – algebras which are generated by sets of the form $[0,a]\times A$ with $a\in[0,T]$ and $A\in\mathcal{B}$, they are equal – $\mathcal{B}_{[0,T]}\otimes\mathcal{B} = \left(\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right)_{[0,T]\times\Omega}$. This observation along with Eq. (27.3) then implies,

$$\varphi^{-1}(V)\cap([0,T]\times\Omega)\in\left(\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right)_{[0,T]\times\Omega}\subset\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}$$

and therefore,

$$\varphi^{-1}(V) = \cup_{T\in\mathbb{N}}\left[\varphi^{-1}(V)\cap([0,T]\times\Omega)\right]\in\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}.$$

This shows φ is $\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}/\mathcal{S}$ – measurable as claimed. ■

Lemma 27.6. *Suppose S is a separable metric space, $\mathcal{S}$ is the Borel σ – algebra on S, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\in\mathbb{R}_+}\right)$ is a filtered measurable space, and $X_t:\Omega\to S$ for $t\in\mathbb{R}_+$ is an adapted right continuous process. Then $X_\cdot$ is progressively measurable and the map, $\varphi:\mathbb{R}_+\times\Omega\to S$ defined by $\varphi(t,\omega) = X_t(\omega)$ is $\left[\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right]/\mathcal{S}$ – measurable.*

Proof. Let $T\in\mathbb{R}_+$. To each $n\in\mathbb{N}$ let $\varphi_n(0,\omega) = X_0(\omega)$ and

$$\varphi_n(t,\omega) := X_{\frac{kT}{2^n}}(\omega) \text{ if } \frac{(k-1)T}{2^n} < t \le \frac{kT}{2^n} \text{ for } k\in\{1,2,\dots,2^n\}.$$

Then

$$\varphi_n^{-1}(A) = \left[\{0\}\times X_0^{-1}(A)\right]\cup_{k=1}^\infty\left[\left(\frac{(k-1)T}{2^n},\frac{kT}{2^n}\right]\times X^{-1}_{Tk2^{-n}}(A)\right]\in\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T,$$

showing that φ_n is $\left[\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T\right]/\mathcal{S}$ – measurable. Therefore, by Exercise 27.1, $\varphi^T = \lim_{n\to\infty}\varphi_n$ is also $\left[\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T\right]/\mathcal{S}$ – measurable. The fact that φ is $\left[\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right]/\mathcal{S}$ – measurable now follows from Lemma 27.5. ■

Lemma 27.7. *Suppose that $T\in(0,\infty)$, $\Omega_T := C([0,T],\mathbb{R})$, and $\mathcal{F}_T = \sigma\left(B_t^T : t\le T\right)$, where $B_t^T(\omega) = \omega(t)$ for all $t\in[0,T]$ and $\omega\in\Omega_T$. Then;*

1. *The map, $\pi:\Omega\to\Omega_T$ defined by $\pi(\omega) := \omega|_{[0,T]}$ is $\mathcal{B}_T/\mathcal{F}_T$ – measurable.*
2. *A function, $F:\Omega\to\mathbb{R}$ is $\mathcal{B}_T$ – measurable iff there exists a function, $f:\Omega_T\to\mathbb{R}$ which is $\mathcal{F}_T$ – measurable such that $F = f\circ\pi$.*
3. *Let $\|\omega\|_T := \max_{t\in[0,T]}|\omega(t)|$ so that $(\Omega_T,\|\cdot\|_T)$ is a Banach space. The Borel σ – algebra, $\mathcal{B}_{\Omega_T}$ on Ω_T is the same as $\mathcal{F}_T$.*
4. *If $F = f\circ\pi$ where $f:\Omega_T\to\mathbb{R}$ is a $\|\cdot\|_T$ – continuous function, then F is $\mathcal{B}_T$ – measurable.*

Proof. 1. Since $B_t^T\circ\pi = B_t$ is $\mathcal{B}_T$ – measurable for all $t\in[0,T]$, it follows that π is measurable.

2. Clearly if $f:\Omega_T\to\mathbb{R}$ is $\mathcal{F}_T$ – measurable, then $F = f\circ\pi:\Omega\to\mathbb{R}$ is $\mathcal{B}_T$ – measurable. For the converse assertion, let $\mathbb{H}$ denote the bounded $\mathcal{B}_T$ – measurable functions of the form $F = f\circ\pi$ with $f:\Omega_T\to\mathbb{R}$ being $\mathcal{F}_T$ – measurable. It is a simple matter to check that $\mathbb{H}$ is a vector space which is closed under bounded convergence and contains all cylinder functions of the form, $G(B_{t_1},\dots B_{t_n}) = G\left(B_{t_1}^T,\dots B_{t_n}^T\right)\circ\pi$ with $\{t_i\}_{i=1}^n\subset[0,T]$. The latter set of functions generates the σ – algebra, $\mathcal{B}_T$, and so by the multiplicative systems theorem, $\mathbb{H}$ contains all bounded $\mathcal{B}_T$ – measurable functions. For a general $\mathcal{B}_T$ – measurable function, $F:\Omega\to\mathbb{R}$, the truncation by $N\in\mathbb{N}$, $F_N = -N\vee(F\wedge N)$, is of the form $F_N = f_N\circ\pi$ for some $\mathcal{F}_T$ – measurable function, $f_N:\Omega_T\to\mathbb{R}$. Since every $\omega\in\Omega_T$ extends to an element of $\tilde\omega\in\Omega$, it follows that $\lim_{N\to\infty}f_N(\omega) = \lim_{N\to\infty}F_N(\tilde\omega) = F(\tilde\omega)$ exists. Hence if we let $f := \lim_{N\to\infty}f_N$, we will have $F = f\circ\pi$ with f being a $\mathcal{F}_T$ – measurable function.

3. Recall that $\mathcal{B}_{\Omega_T} = \sigma(\text{open sets})$. Since $B_s:\Omega\to\mathbb{R}$ is continuous for all s, it follows that $\sigma\left(B_s^T\right)\subset\mathcal{B}_{\Omega_T}$ for all s and hence $\mathcal{F}_T\subset\mathcal{B}_{\Omega_T}$. Conversely, since

$$\|\omega\| := \sup_{t\in\mathbb{Q}\cap[0,T]}|\omega(t)| = \sup_{t\in\mathbb{Q}\cap[0,T]}\left|B_t^T(\omega)\right|,$$

it follows that $\|\cdot-\omega_0\| = \sup_{t\in\mathbb{Q}\cap[0,T]}\left|B_t^T(\cdot)-\omega_0(t)\right|$ is $\mathcal{F}_T$ – measurable for every $\omega_0\in\Omega$. From this we conclude that each open ball, $B(\omega_0,r) := \{\omega\in\Omega : \|\omega-\omega_0\| < r\}$, is in $\mathcal{F}_T$. By the classical Weierstrass approximation theorem we know that Ω is separable and hence we may now conclude that $\mathcal{F}_T$ contains all open subsets of Ω. This shows that $\mathcal{B}_{\Omega_T} = \sigma(\text{open sets})\subset\mathcal{F}_T$.

4. Any continuous function, $f:\Omega\to\mathbb{R}$ is $\mathcal{B}_{\Omega_T} = \mathcal{F}_T$ – measurable and therefore, $F = f\circ\pi$ is $\mathcal{B}_T$ – measurable since it is the composition of two measurable functions. ■

27.2 Stopping and optional times

Definition 27.8. *A random time $T:\Omega\to[0,\infty]$ is a **stopping time** iff $\{T\le t\}\in\mathcal{B}_t$ for all $t\ge 0$ and is an **optional time** iff $\{T<t\}\in\mathcal{B}_t$ for all $t\ge 0$.*

If T is an optional time, the condition $\{T < 0\} = \emptyset \in \mathcal{B}_{0+}$ is vacuous. Moreover, since $\{T < t\} \downarrow \{T = 0\}$ as $t \downarrow 0$, it follows that $\{T = 0\} \in \mathcal{B}_{0+}$ when T is an optional time.

Proposition 27.9. *Suppose $T : \Omega \to [0, \infty]$ is a random time. Then;*

1. *If $T(\omega) = s$ with $s \geq 0$ for all ω, then T is a stopping time.*
2. *Every stopping time is optional.*
3. *T is a $\{\mathcal{B}_t\}$ – optional time iff T is a $\{\mathcal{B}_t^+\}$ – stopping time. In particular, if $\mathcal{B}_t$ is right continuous, i.e. $\mathcal{B}_t^+ = \mathcal{B}_t$ for all t, then the notion of optional time and stopping time are the same.*

Proof. 1.
$$\{T \leq t\} = \begin{cases} \emptyset \text{ if } t < s \\ \Omega \text{ if } t \geq s \end{cases}$$
which show $\{T \leq t\}$ is in any σ – algebra on Ω.

2. If T is a stopping time, $t > 0$ and $t_n \in (0, t)$ with $t_n \uparrow t$, then
$$\{T < t\} = \cup_n \{T \leq t_n\} \in \mathcal{B}_{t-} \subset \mathcal{B}_t.$$
This shows T is an optional time.

3. If T is $\{\mathcal{B}_t\}$ – optional and $t \geq 0$, choose $t_n > t$ such that $t_n \downarrow t$. Then $\{T < t_n\} \downarrow \{T \leq t\}$ which implies $\{T \leq t\} \in \mathcal{B}_{t+} = \mathcal{B}_t^+$. Conversely if T is an $\{\mathcal{B}_{t+}\}$ – stopping time, $t > 0$, and $t_n \in (0, t)$ with $t_n \uparrow t$, then $\{T \leq t_n\} \in \mathcal{B}_{t_n+} \subset \mathcal{B}_t$ for all n and therefore,
$$\{T < t\} = \cup_{n=1}^{\infty} \{T \leq t_n\} \in \mathcal{B}_t.$$
■

Exercise 27.2. Suppose, for all $t \in \mathbb{R}_+$, that $X_t : \Omega \to \mathbb{R}$ is a function. Let $\mathcal{B}_t := \mathcal{B}_t^X := \sigma(X_s : s \leq t)$ and $\mathcal{B} = \mathcal{B}_\infty := \vee_{0 \leq t < \infty} \mathcal{B}_t$. (Recall that the general element, $A \in \mathcal{B}_t$ is of the form, $A = X_\Lambda^{-1}\left(\tilde{A}\right)$ where Λ is a countable subset of $[0, \infty)$, $\tilde{A} \subset \mathbb{R}^\Lambda$ is a measurable set relative to the product σ – algebra on $\mathbb{R}^\Lambda$, and $X_\Lambda : \Omega \to \mathbb{R}^\Lambda$ is defined by, $X_\Lambda(\omega)(t) = X_t(\omega)$ for all $t \in \Lambda$.) If T is a stopping time and $\omega, \omega' \in \Omega$ satisfy $X_t(\omega) = X_t(\omega')$ for all $t \in [0, T(\omega)] \cap \mathbb{R}$, then show $T(\omega) = T(\omega')$.

Definition 27.10. *Given a process, $X_t : \Omega \to S$ and $A \subset S$, let*
$$T_A(\omega) := \inf\{t > 0 : X_t(\omega) \in A\} \text{ and}$$
$$D_A(\omega) := \inf\{t \geq 0 : X_t(\omega) \in A\}$$
be the ***first hitting time*** *and* ***Debut (first entrance time)*** *of A. As usual the infimum of the empty set is taken to be infinity.*

Clearly, $D_A \leq T_A$ and if $D_A(\omega) > 0$ or more generally if $X_0(\omega) \notin A$, then $T_A(\omega) = D_A(\omega)$. Hence we will have $D_A = T_A$ iff $T_A(\omega) = 0$ whenever $X_0(\omega) \in A$.

In the sequel will typically assume that (S, ρ) is a metric space and $\mathcal{S}$ is the Borel σ – algebra on S. We will also typically assume (or arrange) for our processes to have right continuous sample paths. If A is an open subset of S and $t \to X_t(\omega)$ is right continuous, then $T_A = D_A$. Indeed, if $X_0(\omega) \in A$, then by the right continuity of $X_\cdot(\omega)$, we know that $\lim_{t \downarrow 0} X_t(\omega) = X_0(\omega) \in A$ and hence $X_t(\omega) \in A$ for all $t > 0$ sufficiently close to 0 and therefore, $T_A(\omega) = 0$. On the other hand, if A is a closed set and $X_0(\omega) \in \mathrm{bd}(A)$, there is no need for $T_A(\omega) = 0$ and hence in this case, typically $D_A \lneqq T_A$.

Proposition 27.11. *Suppose $\left(\Omega, \{\mathcal{B}_t\}_{t \geq 0}, \mathcal{B}\right)$ is a filtered measurable space, (S, ρ) is a metric space, and $X_t : \Omega \to S$ is a right continuous $\{\mathcal{B}_t\}_{t \geq 0}$ – adapted process. Then;*

1. *If $A \subset S$ is a open set, $T_A = D_A$ is an optional time.*
2. *If $A \subset S$ is closed, on $\{T_A < \infty\}$ $(\{D_A < \infty\})$, $X_{T_A} \in A$ $(X_{D_A} \in A)$.*
3. *If $A \subset S$ is closed and X is a continuous process, then D_A is a stopping time.*
4. *If $A \subset S$ is closed and X is a continuous process, then T_A is an optional time. In fact, $\{T_A \leq t\} \in \mathcal{B}_t$ for all $t > 0$ while $\{T_A = 0\} \in \mathcal{B}_{0+}$, see Figure 27.1.*

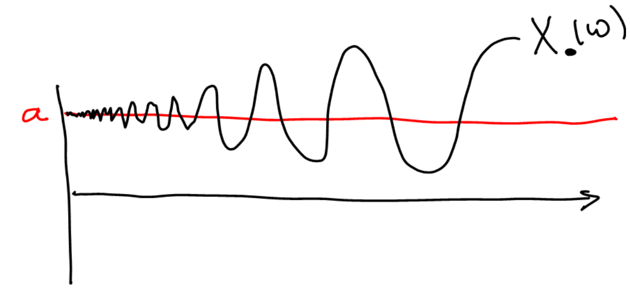


Fig. 27.1. A sample point, $\omega \in \Omega$, where $T_A(\omega) = 0$ with $A = \{a\} \subset \mathbb{R}$.

Proof. 1. By definition, $D_A(\omega) < t$ iff $X_s(\omega) \in A$ for some $s < t$, which by right continuity of X happens iff $X_s(\omega) \in A$ for some $s < t$ with $s \in \mathbb{Q}$. Therefore,
$$\{D_A < t\} = \bigcup\nolimits_{\mathbb{Q} \ni s < t} X_s^{-1}(A) \in \mathcal{B}_{t-} \subset \mathcal{B}_t.$$

2. If $A \subset S$ is closed and $T_A(\omega) < \infty$ (or $D_A(\omega) < \infty$), there exists $t_n > 0$ ($t_n \geq 0$) such that $X_{t_n} \in A$ and $t_n \downarrow T_A(\omega)$. Since X is right continuous and A is closed, $X_{t_n} \to X_{T_A(\omega)} \in A$ ($X_{t_n} \to X_{D_A}(\omega)$).

For the rest of the argument we will now assume that X is a continuous process and A is a closed subset of S.

3. Observe that $D_A(\omega) > t$ iff $X_{[0,t]}(\omega) \cap A = \emptyset$. Since X is continuous, $X_{[0,t]}(\omega)$ is a compact subset of S and therefore

$$\varepsilon := \rho\left(X_{[0,t]}(\omega), A\right) > 0$$

where

$$\rho(A, B) := \inf\{\rho(a, b) : a \in A \text{ and } b \in B\}.$$

Hence we have shown,

$$\begin{aligned}\{D_A > t\} &= \cup_{n=1}^{\infty}\left\{\omega : \rho\left(X_{[0,t]}(\omega), A\right) \geq 1/n\right\} \\ &= \cup_{n=1}^{\infty} \cap_{s\in\mathbb{Q}\cap[0,t]}\{\rho(X_s, A) \geq 1/n\} \in \mathcal{B}_t\end{aligned}$$

wherein we have used $\rho(\cdot, A) : S \to \mathbb{R}_+$ is continuous and hence measurable. As $\{D_A \leq t\} = \{D_A > t\}^c \in \mathcal{B}_t$ for all t, we have shown D_A is an stopping time.

4. Suppose $t > 0$. Then $T_A(\omega) > t$ iff $X_{(0,t]}(\omega) \cap A = \emptyset$ which happens iff for all $\delta \in (0, t)$ we have $X_{[\delta,t]}(\omega) \cap A = \emptyset$ or equivalently iff for all $\delta \in (0, t)$, $\varepsilon := \rho\left(X_{[\delta,t]}(\omega), A\right) > 0$. Using these observations we find,

$$\begin{aligned}\{T_A > t\} &= \cap_{n>1/t} \cup_{m=1}^{\infty}\left\{\omega : \rho\left(X_{[1/n,t]}(\omega), A\right) \geq 1/m\right\} \\ &= \cap_{n>1/t} \cup_{m=1}^{\infty} \cap_{s\in\mathbb{Q}\cap[1/n,t]}\{\rho(X_s, A) \geq 1/m\} \in \mathcal{B}_t.\end{aligned}$$

This shows $\{T_A \leq t\} = \{T_A > t\}^c \in \mathcal{B}_t$ for all $t > 0$. Since, for $t > 0$, $\{T_A < t\} = \cup_{s\in\mathbb{Q}(0,t)}\{T_A \leq s\} \in \mathcal{B}_t$ we see that T_A is an optional time. ■

The only thing keeping T_A from being a stopping time in item 4 above is the fact that $\{T_A = 0\} \in \mathcal{B}_{0+}$ rather than $\{T_A = 0\} \in \mathcal{B}_0$. It should be clear that, in general, $\{T_A = 0\} \notin \mathcal{B}_0$ for $\{T_A = 0\} \in \mathcal{B}_0$ iff $1_{T_A=0} = f(X_0)$ for some measurable function, $f : S \to \{0, 1\} \subset \mathbb{R}$. But it is clearly impossible to determine whether $T_A = 0$ by only observing X_0.

Notation 27.12 *If $\tau : \Omega \to [0, \infty]$ is a random $\mathcal{B}_\infty$ – measurable time, let*

$$\mathcal{B}_\tau := \{A \in \mathcal{B}_\infty : \{\tau \leq t\} \cap A \in \mathcal{B}_t \textit{ for all } t \in [0, \infty]\}. \tag{27.4}$$

and

$$\mathcal{B}_{\tau+} := \{A \in \mathcal{B}_\infty : A \cap \{\tau < t\} \in \mathcal{B}_t \textit{ for all } t \leq \infty\}.$$

Exercise 27.3. If τ is a stopping time then $\mathcal{B}_\tau$ is a sub-σ – algebra of $\mathcal{B}_\infty$ and if τ is an optional time then $\mathcal{B}_{\tau+}$ is a sub-σ – algebra of $\mathcal{B}_\infty$.

Exercise 27.4. Suppose $\tau : \Omega \to [0, \infty]$ is the constant function, $\tau = s$, show $\mathcal{B}_\tau = \mathcal{B}_s$ and $\mathcal{B}_{\tau+} = \cap_{t>s}\mathcal{B}_t =: \mathcal{B}_{s+}$ so that the notation introduced in Notation 27.12 is consistent with the previous meanings of $\mathcal{B}_s$ and $\mathcal{B}_{s+}$.

Exercise 27.5. Suppose that τ is an optional time and let

$$\mathcal{B}_\tau^+ := \left\{A \in \mathcal{B}_\infty : A \cap \{\tau \leq t\} \in \mathcal{B}_{t+} = \mathcal{B}_t^+ \text{ for all } t \leq \infty\right\}.$$

Show $\mathcal{B}_{\tau+} = \mathcal{B}_\tau^+$. Hence $\mathcal{B}_{\tau+}$ is precisely the stopped σ – algebra of the stopping time, τ, relative to the filtration $\left\{\mathcal{B}_t^+\right\}$.

Lemma 27.13. *Suppose $T : \Omega \to [0, \infty]$ is a random time.*

1. *If T is a $\{\mathcal{B}_t\}$ – stopping time, then T is $\mathcal{B}_T$ – measurable.*
2. *If T is a $\{\mathcal{B}_t\}$ – optional time, then T is $\mathcal{B}_{T+} = \mathcal{B}_T^+$ – measurable.*

Proof. Because of Exercise 27.5, it suffices to prove the first assertion. For all $s, t \in \mathbb{R}_+$, we have

$$\{T \leq t\} \cap \{T \leq s\} = \{T \leq s \wedge t\} \in \mathcal{B}_{s\wedge t} \subset \mathcal{B}_s.$$

This shows $\{T \leq t\} \in \mathcal{B}_T$ for all $t \in \mathbb{R}_+$ and therefore that T is $\mathcal{B}_T$ – measurable. ■

Lemma 27.14. *If τ is a $\{\mathcal{B}_t\}$ – stopping time and $X_t : \Omega \to S$ is a $\{\mathcal{B}_t\}$ – progressively measurable process, then X_τ defined on $\{\tau < \infty\}$ is $(\mathcal{B}_\tau)_{\{\tau<\infty\}}/\mathcal{S}$ – measurable. Similarly, if τ is a $\{\mathcal{B}_t\}$ – optional time and $X_t : \Omega \to S$ is a $\left\{\mathcal{B}_t^+\right\}$ – progressively measurable process, then X_τ defined on $\{\tau < \infty\}$ is $(\mathcal{B}_{\tau+})_{\{\tau<\infty\}}/\mathcal{S}$ – measurable.*

Proof. In view of Proposition 27.9 and Exercise 27.5, it suffices to prove the first assertion. For $T \in \mathbb{R}_+$, let $\psi_T : \{\tau \leq T\} \to [0, T] \times \Omega$ be defined by $\psi_T(\omega) = (\tau(\omega), \omega)$ and $\varphi^T : [0, T] \times \Omega \to S$ be defined by $\varphi^T(t, \omega) = X_t(\omega)$. By definition φ^T is $\mathcal{B}_{[0,T]} \otimes \mathcal{B}_T/\mathcal{S}$ – measurable. Since, for all $A \in \mathcal{B}_T$ and $a \in [0, T]$,

$$\psi_T^{-1}([0, a] \times A) = \{\tau \leq a\} \cap A \in (\mathcal{B}_T)_{\{\tau\leq T\}},$$

it follows that ψ_T is $(\mathcal{B}_T)_{\{\tau\leq T\}}/\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T$ – measurable and therefore, $\varphi^T \circ \psi_T : \{\tau \leq T\} \to S$ is $(\mathcal{B}_T)_{\{\tau\leq T\}}/\mathcal{S}$ – measurable.

For $V \in \mathcal{S}$ and $T \in \mathbb{R}_+$,

$$\begin{aligned}X_\tau^{-1}(A) \cap \{\tau \leq T\} &= \left\{\omega \in \Omega : \tau(\omega) \leq T \text{ and } X_{\tau(\omega)}(\omega) \in A\right\} \\ &= \left\{\omega \in \Omega : \tau(\omega) \leq T \text{ and } \varphi^T \circ \psi_T(\omega) \in A\right\} \\ &= \{\tau \leq T\} \cap \left\{\varphi^T \circ \psi_T \in A\right\} \in (\mathcal{B}_T)_{\{\tau\leq T\}} \subset \mathcal{B}_T.\end{aligned}$$

This is true for arbitrary $T \in \mathbb{R}_+$ we conclude that $X_\tau^{-1}(A) \in \mathcal{B}_\tau$ and since, by definition, $X_\tau^{-1}(A) \subset \{\tau < \infty\}$, it follows that $X_\tau^{-1}(A) \in (\mathcal{B}_\tau)_{\{\tau<\infty\}}$. This completes the proof since $A \in \mathcal{S}$ was arbitrary. ■

Lemma 27.15 (Properties of Optional/Stopping times). *Let T and S be optional times and $\theta > 0$. Then;*

1. *$T + \theta$ is a stopping time.*
2. *$T + S$ is an optional time.*
3. *If $T > 0$ and T is a stopping time then $T + S$ is again a stopping time.*
4. *If $T > 0$ and $S > 0$, then $T + S$ is a stopping time.*
5. *If we further assume that S and T are stopping times, then $T \wedge S$, $T \vee S$, and $T + S$ are stopping times.*
6. *If $\{T_n\}_{n=1}^{\infty}$ are optional times, then*
$$\sup_{n\geq 1} T_n, \quad \inf_{n\geq 1} T_n, \quad \liminf_{n\to\infty} T_n, \quad \text{and} \quad \limsup_{n\to\infty} T_n$$
are all optional times. If $\{T_n\}_{n=1}^{\infty}$ are stopping times, then $\sup_{n\geq 1} T_n$ is a stopping time.

Proof. 1. This follows from the observation that
$$\{T + \theta \leq t\} = \{T \leq t - \theta\} \in \mathcal{B}_{(t-\theta)+} \subset \mathcal{B}_t.$$
Notice that if $t < \theta$, then $\{T + \theta \leq t\} = \emptyset \in \mathcal{B}_0$.

2. – 4. For item 2., if $\tau > 0$, then
$$\{T + S < \tau\} = \cup \{T < t, S < s : s, t \in \mathbb{Q} \cap (0, \tau] \text{ with } s + t < \tau\} \in \mathcal{B}_{\tau-} \subset \mathcal{B}_\tau$$
and if $\tau = 0$, then $\{T + S < 0\} = \emptyset \in \mathcal{B}_0$. If $T > 0$ and T is a stopping time and $\tau > 0$, then
$$\{T + S \leq \tau\} = \{S = 0, T \leq \tau\} \cup \{0 < S, S + T \leq \tau\}$$
and $\{S = 0, T \leq \tau\} \in \mathcal{B}_\tau$. Hence it suffices to show $\{0 < S, S + T \leq \tau\} \in \mathcal{B}_\tau$. To this end, observe that $0 < S$ and $S + T \leq \tau$ happens iff there exists $m \in \mathbb{N}$ such that for all $n \geq m$, there exists a $r = r_n \in \mathbb{Q}$ such that
$$0 < r_n < S < r_n + 1/n < \tau \text{ and } T \leq \tau - r_n.$$
Indeed, if the latter condition holds, then $S + T \leq \tau - r_n + (r_n + 1/n) = \tau + 1/n$ for all n and therefore $S + T \leq \tau$. Thus we have shown
$$\{S + T \leq \tau\} = \cup_{m\in\mathbb{N}} \cap_{n\geq m} \cup \{r < S < r + 1/n, T \leq \tau - r : 0 < r < r + 1/n < \tau\}$$
which is in $\mathcal{B}_{\tau-} \subset \mathcal{B}_\tau$. In showing $\{0 < S, S + T \leq \tau\} \in \mathcal{B}_\tau$ we only need for S and T to be optional times and so if $S > 0$ and $T > 0$, then
$$\{T + S \leq \tau\} = \{0 < S, S + T \leq \tau\} \in \mathcal{B}_{\tau-} \subset \mathcal{B}_\tau.$$

5. If T and S are stopping times and $\tau \geq 0$, then
$$\{T \wedge S \leq \tau\} = \{T \leq \tau\} \cup \{S \leq \tau\} \in \mathcal{B}_\tau,$$
$$\{T \vee S \leq \tau\} = \{T \leq \tau\} \cap \{S \leq \tau\} \in \mathcal{B}_\tau,$$
and
$$\begin{aligned}\{S + T > \tau\} &= \{T = 0, S > \tau\} \cup \{0 < T, T + S > \tau\} \\ &= \{T = 0, S > \tau\} \cup \{0 < T < \tau, T + S > \tau\} \cup \{T \geq \tau, T + S > \tau\} \\ &= \{T = 0, S > \tau\} \cup \{0 < T < \tau, T + S > \tau\} \cup \\ &\qquad \cup \{S = 0, T > \tau\} \cup \{S > 0, T \geq \tau\}.\end{aligned}$$
The first, third, and fourth events are easily seen to be in $\mathcal{B}_\tau$. As for the second event,
$$\{0 < T < \tau, T + S > \tau\} = \cup \{r < T < \tau, S > \tau - r : r \in \mathbb{Q} \text{ with } 0 < r < \tau\}.$$

6. We have
$$\left\{\sup_n T_n \leq t\right\} = \cap_{n=1}^{\infty} \{T_n \leq t\},$$
$$\left\{\inf_n T_n < t\right\} = \cup_{n=1}^{\infty} \{T_n < t\}$$
which shows that $\sup_n T_n$ is a stopping time if each T_n is a stopping time and that $\inf_n T_n$ is optional if each T_n is optional. Moreover, if each T_n is optional, then T_n is a $\mathcal{B}_{t+}$ stopping time and hence $\sup_n T_n$ is an $\mathcal{B}_{t+}$ stopping time and hence $\sup_n T_n$ is an $\mathcal{B}_t$ optional time, wherein we have used Proposition 27.9 twice. ■

Lemma 27.16 (Stopped σ – algebras). *Suppose σ and τ are stopping times.*

1. *$\mathcal{B}_\tau = \mathcal{B}_t$ on $\{\tau = t\}$.*
2. *If $t \in [0, \infty]$, then $\tau \wedge t$ is $\mathcal{B}_t$ – measurable.*
3. *If $\sigma \leq \tau$, then $\mathcal{B}_\sigma \subset \mathcal{B}_\tau$.*
4. *$(\mathcal{B}_\sigma)_{\{\sigma\leq\tau\}} \subset \mathcal{B}_{\sigma\wedge\tau}$ and in particular $\{\sigma \leq \tau\}$, $\{\sigma < \tau\}$, $\{\tau \leq \sigma\}$, and $\{\tau < \sigma\}$ are all in $\mathcal{B}_{\sigma\wedge\tau}$.*
5. *$(\mathcal{B}_\sigma)_{\{\sigma<\tau\}} \subset \mathcal{B}_{\sigma\wedge\tau}$.*
6. *$\mathcal{B}_\sigma \cap \mathcal{B}_\tau = \mathcal{B}_{\sigma\wedge\tau}$.*
7. *If $\mathbb{D}$ is a countable set and $\tau : \Omega \to \mathbb{D} \subset [0, \infty]$ is a function, then τ is a stopping time iff $\{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$.*
8. *If the range of τ is a countable subset, $\mathbb{D} \subset [0, \infty]$, then $A \subset \Omega$ is in $\mathcal{B}_\tau$ iff $A \cap \{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$.*
9. *If the range of τ is a countable subset, $\mathbb{D} \subset [0, \infty]$, then a function $f : \Omega \to \mathbb{R}$ is $\mathcal{B}_\tau$ – measurable iff $1_{\{\tau=t\}} f$ is $\mathcal{B}_t$ – measurable for all $t \in \mathbb{D}$.*

Moreover, all of the above results hold if σ and τ are optional times provided every occurrence of the letter $\mathcal{B}$ is replaced by $\mathcal{B}^+$.

Proof. Recall from Definition 14.21 that if $\mathcal{G}$ is a σ – algebra on Ω and $A \subset \Omega$, then $\mathcal{G}_A := \{B \cap A : B \in \mathcal{G}\}$ – a sub – σ – algebra of 2^A. Moreover if $\mathcal{G}$ and $\mathcal{F}$ are two σ – algebras on Ω and $A \in \mathcal{G} \cap \mathcal{F}$, then (by definition) $\mathcal{G} = \mathcal{F}$ on A iff $\mathcal{G}_A = \mathcal{F}_A$.

1. If $A \in \mathcal{B}_\tau$, then
$$A \cap \{\tau = t\} = A \cap \{\tau \le t\} \cap \{\tau < t\}^c \in \mathcal{B}_t.$$
Conversely if $A \in \mathcal{B}_t$ and $s \in \mathbb{R}_+$,
$$A \cap \{\tau = t\} \cap \{\tau \le s\} = \begin{cases} \emptyset & \text{if } s < t \\ A \cap \{\tau = t\} & \text{if } s \ge t \end{cases}$$
from which it follows that $A \cap \{\tau = t\} \in \mathcal{B}_\tau$.
2. To see $\tau \wedge t$ is $\mathcal{B}_t$ – measurable simply observe that
$$\{\tau \wedge t \le s\} = \begin{cases} \Omega \in \mathcal{B}_t & \text{if } t \le s \\ \{\tau \le s\} \in \mathcal{B}_s \subset \mathcal{B}_t & \text{if } t > s \end{cases}$$
and hence $\{\tau \wedge t \le s\} \in \mathcal{B}_t$ for all $s \in [0, \infty]$.
3. If $A \in \mathcal{B}_\sigma$ and $\sigma \le \tau$, then
$$A \cap \{\tau \le t\} = [A \cap \{\sigma \le t\}] \cap \{\tau \le t\} \in \mathcal{B}_t$$
for all $t \le \infty$ and therefore $A \in \mathcal{B}_\tau$.
4. If $A \in \mathcal{B}_\sigma$ then $A \cap \{\sigma \le \tau\}$ is the generic element of $(\mathcal{B}_\sigma)_{\{\sigma \le \tau\}}$. We now have
$$\begin{aligned}(A \cap \{\sigma \le \tau\}) \cap \{\tau \wedge \sigma \le t\} &= (A \cap \{\sigma \le \tau\}) \cap \{\sigma \le t\} \\ &= (A \cap \{\sigma \wedge t \le \tau \wedge t\}) \cap \{\sigma \le t\} \\ &= (A \cap \{\sigma \le t\}) \cap \{\sigma \wedge t \le \tau \wedge t\} \in \mathcal{B}_t\end{aligned}$$
since $(A \cap \{\sigma \le t\}) \in \mathcal{B}_t$ and $\sigma \wedge t$ and $\tau \wedge t$ are $\mathcal{B}_t$ – measurable and hence $\{\sigma \wedge t \le \tau \wedge t\} \in \mathcal{B}_t$. Since $\Omega \in \mathcal{B}_\sigma$, it follows from what we have just proved that $\{\sigma \le \tau\} = \Omega \cap \{\sigma \le \tau\} \in \mathcal{B}_{\sigma \wedge \tau}$ and hence also $\{\tau < \sigma\} = \{\sigma \le \tau\}^c \in \mathcal{B}_{\sigma \wedge \tau}$. By symmetry we may also conclude that $\{\tau \le \sigma\}$ and $\{\sigma < \tau\}$ are in $\mathcal{B}_{\sigma \wedge \tau}$.
5. By item 4., if $A \in \mathcal{B}_\sigma$, then
$$A \cap \{\sigma < \tau\} = A \cap \{\sigma \le \tau\} \cap \{\sigma < \tau\} \in \mathcal{B}_{\sigma \wedge \tau}.$$
6. Since $\sigma \wedge \tau$ is a stopping time which is no larger than either σ or τ, it follows that from item 2. that $\mathcal{B}_{\sigma \wedge \tau} \subset \mathcal{B}_\sigma \cap \mathcal{B}_\tau$. Conversely, if $A \in \mathcal{B}_\sigma \cap \mathcal{B}_\tau$ then
$$\begin{aligned}A \cap \{\sigma \wedge \tau \le t\} &= A \cap [\{\sigma \le t\} \cup \{\tau \le t\}] \\ &= [A \cap \{\sigma \le t\}] \cup [A \cap \{\tau \le t\}] \in \mathcal{B}_t\end{aligned}$$
for all $t \le \infty$. From this it follows that $A \in \mathcal{B}_{\sigma \wedge \tau}$.
7. If τ is a stopping time and $t \in \mathbb{D}$, then $\{\tau = t\} = \{\tau \le t\} \setminus [\cup_{\mathbb{D} \ni s < t} \{\tau \le s\}] \in \mathcal{B}_t$. Conversely if $\{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$ and $s \in \mathbb{R}_+$, then
$$\{\tau \le s\} = \cup_{\mathbb{D} \ni t \le s} \{\tau = t\} \in \mathcal{B}_s$$
showing τ is a stopping time.
8. If $A \cap \{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$, then for any $s \le \infty$,
$$A \cap \{\tau \le s\} = \cup_{\mathbb{D} \ni t \le s} [A \cap \{\tau = t\}] \in \mathcal{B}_s$$
which shows $A \in \mathcal{B}_\tau$. Conversely if $A \in \mathcal{B}_\tau$ and $t \in \mathbb{D}$, then
$$A \cap \{\tau = t\} = [A \cap \{\tau \le t\}] \setminus [\cup_{\mathbb{D} \ni s < t} (A \cap \{\tau \le s\})] \in \mathcal{B}_t.$$
9. If $f : \Omega \to \mathbb{R}$ is $\mathcal{B}_\tau$ measurable, then f is a limit of $\mathcal{B}_\tau$ – simple functions, say $f_n \to f$. By item 7. it easily follows that $1_{\{\tau = t\}} f_n$ is $\mathcal{B}_t$ – measurable for each $t \in \mathbb{D}$ and therefore $1_{\{\tau = t\}} f = \lim_{n \to \infty} 1_{\{\tau = t\}} f_n$ is $\mathcal{B}_t$ – measurable for each $t \in \mathbb{D}$.
Conversely if $f : \Omega \to \mathbb{R}$ is a function such that $1_{\{\tau = t\}} f$ is $\mathcal{B}_t$ – measurable for each $t \in \mathbb{D}$, then for every $A \in \mathcal{B}_\mathbb{R}$ with $0 \notin A$ we have
$$\{\tau = t\} \cap \{f \in A\} = \left\{1_{\{\tau = t\}} f \in A\right\} \in \mathcal{B}_t \text{ for all } t \in \mathbb{D}.$$
Hence it follows by item 7. that $\{f \in A\} \in \mathcal{B}_\tau$. Similalry,
$$\{\tau = t\} \cap \{f = 0\} = \left\{1_{\{\tau = t\}} f = 0\right\} \cap \{\tau = t\} \in \mathcal{B}_t \text{ for all } t \in \mathbb{D}$$
and so again $\{f = 0\} \in \mathcal{B}_\tau$ by item 7. This suffices to show that f is $\mathcal{B}_\tau$ – measurable.

■

Corollary 27.17. *If σ and τ are stopping times and F is a $\mathcal{B}_\sigma$ – measurable function then $1_{\{\sigma \le \tau\}} F$ and $1_{\{\sigma < \tau\}} F$ are $\mathcal{B}_{\sigma \wedge \tau}$ – measurable.*

Proof. If $F = 1_A$ with $A \in \mathcal{B}_\sigma$, then the assertion follows from items 4. and 5. from Lemma 27.16. By linearity, the assertion holds if F is a $\mathcal{B}_\sigma$ – measurable simple function and then, by taking limits, for all $\mathcal{B}_\sigma$ – measurable functions. ■

Lemma 27.18 (Optional time approximation lemma). *Let τ be a $\{\mathcal{B}_t\}_{t\geq 0}$ – optional time and for $n\in\mathbb{N}$, let $\tau_n:\Omega\to[0,\infty]$ be defined by*

$$\tau_n := \frac{1}{2^n}\left[2^n\tau\right] = \infty 1_{\tau=\infty} + \sum_{k=1}^{\infty}\frac{k}{2^n}1_{\frac{k-1}{2^n}\leq\tau<\frac{k}{2^n}}. \tag{27.5}$$

Then $\{\tau_n\}_{n=1}^{\infty}$ are stopping times such that;

1. *$\tau_n\downarrow\tau$ as $n\to\infty$,*
2. *$\mathcal{B}_\tau^+\subset\mathcal{B}_{\tau_n}$ for all n, and*
3. *$\{\tau_n=\infty\}=\{\tau=\infty\}$ for all n.*

Proof. If $A\in\mathcal{B}_\tau^+$ then

$$\begin{aligned}A\cap\left\{\tau_n=k2^{-n}\right\} &= A\cap\left\{(k-1)\,2^{-n}\leq\tau<k2^{-n}\right\}\\ &= \left[A\cap\left\{\tau<k2^{-n}\right\}\right]\setminus\left\{\tau<(k-1)\,2^{-n}\right\}\in\mathcal{B}_{k2^{-n}}.\end{aligned}$$

Taking $A=\Omega$ in this equation shows $\{\tau_n=k2^{-n}\}\in\mathcal{B}_{k2^{-n}}$ for all $k\in\mathbb{N}$ and so is a stopping time by Lemma 27.16. Moreover this same lemma shows that $A\in\mathcal{B}_{\tau_n}$. The fact that $\tau_n\downarrow\tau$ as $n\to\infty$ and $\tau_n=\infty$ iff $\tau=\infty$ should be clear. ■

27.3 Filtration considerations

For this sction suppose that $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ is a given filtered probability space.

Notation 27.19 (Null sets) *Let $\mathcal{N}^P:=\{N\in\mathcal{B}:P(N)=0\}$ – be the collection of **null sets** of P.*

Definition 27.20. *If $\mathcal{A}\subset\mathcal{B}$ is a sub-sigma-algebra of $\mathcal{B}$, then the **augmentation of** $\mathcal{A}$ is the σ – algebra, $\bar{\mathcal{A}}:=\mathcal{A}\vee\mathcal{N}:=\sigma(\mathcal{A}\cup\mathcal{N})$.*

Definition 27.21 (Usual hypothesis). *A filtered probability space, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$, is said to satisfy the **weak usual hypothesis** if:*

1. *For each $t\in\mathbb{R}_+$, $\mathcal{N}^P\subset\mathcal{B}_t$, i.e. $\mathcal{B}_t$ contains all of the P – null sets.*
2. *The filtration, $\{\mathcal{B}_t\}_{t\in\mathbb{R}_+}$ is **right continuous**, i.e. $\mathcal{B}_{t+}=\mathcal{B}_t$.*

*If in addition, $(\Omega,\mathcal{B},P)$ is complete (i.e. if $N\in\mathcal{N}^P$, $A\subset N$, then $A\in\mathcal{N}^P$), then we say $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ satisfies the **usual hypothesis.***

It is always possible to make an arbitrary filtered probability space, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$, into one satisfying the (weak) usual hypothesis by "augmenting" the filtration by the null sets and taking the "right continuous extension." We are going to develop these two concepts now. (For even more information on the usual hypothesis, see [43, pages 34-36.].)

Lemma 27.22 (Augmentation lemma). *Continuing the notation in Definition 27.20, we have*

$$\bar{\mathcal{A}}:=\{B\in\mathcal{B}:\ \exists\ A\in\mathcal{A}\ \ni\ A\bigtriangleup B\in\mathcal{N}\}. \tag{27.6}$$

Proof. Let $\mathcal{G}$ denote the right side of Eq. (27.6). If $B\in\mathcal{G}$ and $A\in\mathcal{A}$ such that $N:=A\bigtriangleup B\in\mathcal{N}$, then

$$B=[A\cap B]\cup[A\setminus B]=[A\setminus(A\setminus B)]\cup[B\setminus A]. \tag{27.7}$$

Since $A\setminus B\subset N$ and $B\setminus A\subset N$ implies $A\setminus B$ and $B\setminus A$ are in $\mathcal{N}$, it follows that $B\in\mathcal{A}\vee\mathcal{N}=\bar{\mathcal{A}}$. Thus we have shown, $\mathcal{G}\subset\bar{\mathcal{A}}$. Since it is clear that $\mathcal{A}\subset\mathcal{G}$ and $\mathcal{N}\subset\mathcal{G}$, to finish the proof it suffices to show $\mathcal{G}$ is a σ – algebra. For if we do this, then $\bar{\mathcal{A}}=\mathcal{A}\vee\mathcal{N}\subset\mathcal{G}$.

Since $A^c\bigtriangleup B^c=A\bigtriangleup B$, we see that $\mathcal{G}$ is closed under complementation. Moreover, if $B_j\in\mathcal{G}$, there exists $A_j\in\mathcal{A}$ such that $A_j\bigtriangleup B_j\in\mathcal{N}$ for all j. So letting $A=\cup_jA_j\in\mathcal{A}$ and $B=\cup_jB_j\in\mathcal{B}$, we have

$$\mathcal{B}\ni A\bigtriangleup B\subset\cup_j\left[A_j\bigtriangleup B_j\right]\in\mathcal{N}$$

from which we conclude that $A\bigtriangleup B\in\mathcal{N}$ and hence $B\in\mathcal{G}$. This shows that $\mathcal{G}$ is closed under countable unions, complementation, and contains $\mathcal{A}$ and hence the empty set and Ω, thus $\mathcal{G}$ is a σ – algebra. ■

Lemma 27.23 (Commutation lemma). *If $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ is a filtered probability space, then $\bar{\mathcal{B}}_{t+}=\overline{\mathcal{B}_{t+}}$. In words the augmentation procedure and the right continuity extension procedure commute.*

Proof. Since for any $s>t$, $\mathcal{B}_{t+}\subset\mathcal{B}_s$ it follows that $\overline{\mathcal{B}_{t+}}\subset\bar{\mathcal{B}}_s$ and therefore that

$$\overline{\mathcal{B}_{t+}}\subset\cap_{s>t}\bar{\mathcal{B}}_s=\bar{\mathcal{B}}_{t+}.$$

Conversely if $B\in\bar{\mathcal{B}}_{t+}=\cap_{s>t}\bar{\mathcal{B}}_s$ and $t_n>t$ such that $t_n\downarrow 0$, then for each $n\in\mathbb{N}$ there exists $A_n\in\mathcal{B}_{t_n}$ such that $A_n\bigtriangleup B\in\mathcal{N}$. We will now show that $B\in\overline{\mathcal{B}_{t+}}$, by showing $B\bigtriangleup A\in\mathcal{N}$ where

$$A:=\{A_n\text{ i.o.}\}=\cap_{m\in\mathbb{N}}\cup_{n\geq m}A_n\in\mathcal{B}_{t+}.$$

To prove this let $A_m':=\cup_{n\geq m}A_n$ so that $A_n'\downarrow A$ as $n\uparrow\infty$. Then

$$B \triangle A = B \triangle [\cap_m A'_m] = (B \setminus [\cap_m A'_m]) \cup ([\cap_m A'_m] \setminus B)$$
$$\subset [\cup_m (B \setminus A'_m)] \cup (A'_1 \setminus B) \in \mathcal{N}$$

because measurable subsets of elements in $\mathcal{N}$ are still in $\mathcal{N}$, $\mathcal{N}$ is closed under countable unions,

$$B \setminus A'_m \subset B \setminus A_m \subset B \triangle A_m \in \mathcal{N}, \text{ and}$$
$$A'_1 \setminus B = \cup_{n=1}^{\infty} [A_n \setminus B] \subset \cup_{n=1}^{\infty} [A_n \triangle B] \in \mathcal{N}.$$

■

27.3.1 ***More Augmentation Results (This subsection neeed serious editing.)

In this subsection we generalize the augmentation results above to the setting where we adjoin our favorite collection of "null like" sets.

Definition 27.24. *Suppose $(\Omega, \mathcal{B})$ is a measurable space. A collection of subsets, $\mathcal{N} \subset \mathcal{B}$ is a* ***null like collection in*** *$\mathcal{B}$ if; 1) $\mathcal{N}$ is closed under countable union, and 2) if $A \in \mathcal{B}$ and there exists $N \in \mathcal{N}$ such that $A \subset N$, then $A \in \mathcal{N}$.*

Example 27.25. Let $\{P_i\}_{i \in I}$ be any collection of probability measures on a measurable space, $(\Omega, \mathcal{B})$. Then

$$\mathcal{N} := \{N \in \mathcal{B} : P_i(N) = 0 \text{ for all } i \in I\}$$

is a null like collections of subsets in $\mathcal{B}$.

Example 27.26. If $\mathcal{N}$ is a null like collection in $\mathcal{B}$, then

$$\bar{\mathcal{N}} := \left\{A \subset 2^{\Omega} : A \subset N \text{ for some } N \in \mathcal{N}\right\}$$

is a null like collection in 2^{Ω}.

Example 27.27. Let $\{P_i\}_{i \in I}$ be any collection of probability measures on a measurable space, $(\Omega, \mathcal{B})$. Then

$$\mathcal{N} := \left\{N \subset 2^{\Omega} : \exists\, B \in \mathcal{B} \ni P_i(B) = 0 \text{ and } N \subset B \text{ for all } i \in I\right\}$$

is a null like collections of subsets of 2^{Ω}. Similarly,

$$\mathcal{N} := \left\{N \subset 2^{\Omega} : \exists\, B_i \in \mathcal{B} \ni P_i(B_i) = 0 \text{ and } N \subset B_i \text{ for all } i \in I\right\}$$

is a null like collections of subsets of 2^{Ω}. These two collections are easily seen to be the same if I is countable, otherwise they may be different.

Example 27.28. If $\mathcal{N}_i \subset \mathcal{B}$ are null like collections in $\mathcal{B}$ for all $i \in I$, then $\mathcal{N} := \cap_{i \in I} \mathcal{N}_i$ is another null like collection in $\mathcal{B}$. Indeed, if $B \in \mathcal{B}$ and $B \subset N \in \mathcal{N}$, then $B \subset N \in \mathcal{N}_i$ for all i and therefore, $B \in \mathcal{N}_i$ for all i and hence $B \in \mathcal{N}$. Moreover, it is clear that $\mathcal{N}$ is still closed under countable unions.

Definition 27.29. *If $(\Omega, \mathcal{B})$ is a measurable space, $\mathcal{A} \subset \mathcal{B}$ is a sub-sigma-algebra of $\mathcal{B}$ and $\mathcal{N} \subset \mathcal{B}$ is a null like collection in $\mathcal{B}$, we say $\mathcal{A}^{\mathcal{N}} := \mathcal{A} \vee \mathcal{N}$ is the* ***augmentation of*** *$\mathcal{A}$ by $\mathcal{N}$.*

Lemma 27.30 (Augmentation lemma). *If $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$ and $\mathcal{N}$ is a null like collection in $\mathcal{B}$, then the augmentation of $\mathcal{A}$ by $\mathcal{N}$ is*

$$\mathcal{A}^{\mathcal{N}} := \{B \in \mathcal{B} : \exists\, A \in \mathcal{A} \ni A \triangle B \in \mathcal{N}\}. \tag{27.8}$$

Proof. Let $\mathcal{G}$ denote the right side of Eq. (27.8). If $B \in \mathcal{G}$ and $A \in \mathcal{A}$ such that $N := A \triangle B \in \mathcal{N}$, then

$$B = [A \cap B] \cup [A \setminus B] = [A \setminus (A \setminus B)] \cup [B \setminus A]. \tag{27.9}$$

Since $A \setminus B \subset N$ and $B \setminus A \subset N$ implies $A \setminus B$ and $B \setminus A$ are in $\mathcal{N}$, it follows that $B \in \mathcal{A} \vee \mathcal{N} = \mathcal{A}^{\mathcal{N}}$. Thus we have shown, $\mathcal{G} \subset \mathcal{A}^{\mathcal{N}}$. Since it is clear that $\mathcal{A} \subset \mathcal{G}$ and $\mathcal{N} \subset \mathcal{G}$, to finish the proof it suffices to show $\mathcal{G}$ is a $\sigma -$ algebra. For if we do this, then $\mathcal{A}^{\mathcal{N}} = \mathcal{A} \vee \mathcal{N} \subset \mathcal{G}$.

Since $A^c \triangle B^c = A \triangle B$, we see that $\mathcal{G}$ is closed under complementation. Moreover, if $B_j \in \mathcal{G}$, there exists $A_j \in \mathcal{A}$ such that $A_j \triangle B_j \in \mathcal{N}$ for all j. So letting $A = \cup_j A_j \in \mathcal{A}$ and $B = \cup_j B_j \in \mathcal{B}$, we have

$$\mathcal{B} \ni A \triangle B \subset \cup_j [A_j \triangle B_j] \in \mathcal{N}$$

from which we conclude that $A \triangle B \in \mathcal{N}$ and hence $B \in \mathcal{G}$. This shows that $\mathcal{G}$ is closed under countable unions, complementation, and contains $\mathcal{A}$ and hence the empty set and Ω, thus $\mathcal{G}$ is a $\sigma -$ algebra. ■

Lemma 27.31 (Commutation lemma). *Let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}\right)$ be a filtered space and $\mathcal{N} \subset \mathcal{B}$ be null like collection and for $\mathcal{G} \subset \mathcal{B}$, let $\bar{\mathcal{G}} := \mathcal{G} \vee \mathcal{N}$. Then $\bar{\mathcal{B}}_{t+} = \overline{\mathcal{B}_{t+}}$.*

Proof. Since for any $s > t$, $\mathcal{B}_{t+} \subset \mathcal{B}_s$ it follows that $\overline{\mathcal{B}_{t+}} \subset \bar{\mathcal{B}}_s$ and therefore that

$$\overline{\mathcal{B}_{t+}} \subset \cap_{s>t} \bar{\mathcal{B}}_s = \bar{\mathcal{B}}_{t+}.$$

Conversely if $B \in \bar{\mathcal{B}}_{t+} = \cap_{s>t} \bar{\mathcal{B}}_s$ and $t_n > t$ such that $t_n \downarrow 0$, then for each $n \in \mathbb{N}$ there exists $A_n \in \mathcal{B}_{t_n}$ such that $A_n \triangle B \in \mathcal{N}$. We will now show that $B \in \overline{\mathcal{B}_{t+}}$, by shown $B \triangle A \in \mathcal{N}$ where

$$A := \{A_n \text{ i.o.}\} = \cap_{m \in \mathbb{N}} \cup_{n \geq m} A_n \in \mathcal{B}_{t+}.$$

To prove this let $A'_m := \cup_{n\geq m} A_n$ so that $A'_n \downarrow A$ as $n \uparrow \infty$. Then

$$B \triangle A = B \triangle [\cap_m A'_m] = (B \setminus [\cap_m A'_m]) \cup ([\cap_m A'_m] \setminus B) \\ \subset [\cup_m (B \setminus A'_m)] \cup (A'_1 \setminus B) \in \mathcal{N}$$

because measurable subsets of elements in $\mathcal{N}$ are still in $\mathcal{N}$, $\mathcal{N}$ is closed under countable unions,

$$B \setminus A'_m \subset B \setminus A_m \subset B \triangle A_m \in \mathcal{N}, \text{ and} \\ A'_1 \setminus B = \cup_{n=1}^\infty [A_n \setminus B] \subset \cup_{n=1}^\infty [A_n \triangle B] \in \mathcal{N}.$$

■

Corollary 27.32. *Suppose $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$ and $\mathcal{N}$ is a null like collection in $\mathcal{B}$ with the additional property, that for all $N \in \mathcal{N}$ there exists $N' \in \mathcal{A} \cap \mathcal{N}$ such that $N \subset N'$. Then*

$$\mathcal{A}^{\mathcal{N}} := \{A \cup N : A \in \mathcal{A} \text{ and } N \in \mathcal{N}\}. \tag{27.10}$$

Proof. Let $\mathcal{G}$ denote the right side of Eq. (27.10). It is clear that $\mathcal{G} \subset \mathcal{A}^{\mathcal{N}}$. Conversely if $B \in \mathcal{A}^{\mathcal{N}}$, we know by Lemma 27.22 that there exists, $A \in \mathcal{A}$ such that $N := A \triangle B \in \mathcal{N}$. Since $C := A \setminus B \subset N$, $C \in \mathcal{N}$ and so by assumption there exists $N' \in \mathcal{A} \cap \mathcal{N}$ such that $C \subset N'$. Therefore, according to Eq. (27.7), we have

$$B = [A \setminus C] \cup [B \setminus A] = [A \setminus N'] \cup [(A \cap N') \setminus C] \cup [B \setminus A].$$

Since $A \setminus N' \in \mathcal{A}$ and $[(A \cap N') \setminus C] \cup [B \setminus A] \in \mathcal{N}$, it follows that $B \in \mathcal{G}$. ■

Example 27.33. Let ν be a probability measure on $\mathbb{R}$. As in Notation **??**, let $P_\nu := \int_{\mathbb{R}} d\nu(x)\ P_x$ be the Wiener measure on $\Omega := C([0,\infty), \mathbb{R})$, $B_t : \Omega \to \mathbb{R}$ be the projection map, $B_t(\omega) = \omega(t)$, $\mathcal{B}_t = \sigma(B_s : s \leq t)$, and $\mathcal{N}_{t+}(\nu) := \{N \in \mathcal{B}_{t+} : P_\nu(N) = 0\}$. Then by Corollary 29.13, $\mathcal{B}_{t+} = \mathcal{B}_t \vee \mathcal{N}_{t+}(\nu)$. Hence if we let

$$\mathcal{N}(\nu) := \{N \in \mathcal{B} : P_\nu(N) = 0\}$$

and

$$\bar{\mathcal{N}}(\nu) := \left\{B \subset 2^\Omega : B \subset N \text{ for some } N \in \mathcal{N}(\nu)\right\}$$

then $\mathcal{B}_{t+} \vee \mathcal{N}(\nu) = \mathcal{B}_t \vee \mathcal{N}(\nu) = \bar{\mathcal{B}}_t$ and $\mathcal{B}_{t+} \vee \bar{\mathcal{N}}(\nu) = \mathcal{B}_t \vee \bar{\mathcal{N}}(\nu)$ for all $t \in \mathbb{R}_+$. This shows that the augmented Brownian filtration, $\left\{\bar{\mathcal{B}}_t\right\}_{t\geq 0}$, is already right continuous.

Definition 27.34. *Recall from Proposition 5.50, if $(\Omega, \mathcal{B}, P)$ is a probability space and $\bar{\mathcal{N}}^P := \{A \subset \Omega : P^*(A) = 0\}$, then the **completion,** $\bar{P} := P^*|_{\mathcal{B} \vee \bar{\mathcal{N}}^P}$, is a probability measure on $\mathcal{B} \vee \bar{\mathcal{N}}^P$ which extends P so that $\bar{P}(A) = 0$ for all $A \in \bar{\mathcal{N}}^P$.*

Suppose that $(\Omega, \mathcal{B})$ is a measurable space and $\mathcal{N} \subset \mathcal{B}$ is a collection of sets closed under countable union and also satisfying, $A \in \mathcal{N}$ if $A \in \mathcal{B}$ with $A \subset N$ for some $N \in \mathcal{N}$. The main example that we will use below is to let $\{P_i\}_{i\in I}$ to be a collection of probability measures on $(\Omega, \mathcal{B})$ and then let

$$\mathcal{N} := \{N \in \mathcal{B} : P_i(N) = 0 \text{ for all } i \in I\}.$$

Let us also observe that if $\mathcal{N}_i$ is a collection of null sets above for each $i \in I$, then $\mathcal{N} = \cap_i \mathcal{N}_i$ is also a collection of null sets. Indeed, if $B \in \mathcal{B}$ and $B \subset N \in \mathcal{N}$, then $B \subset N \in \mathcal{N}_i$ for all i and therefore, $B \in N_i$ for all i and hence $B \in \mathcal{N}$. Moreover, it is clear that $\mathcal{N}$ is still closed under countable unions.

Lemma 27.35 (Augmentation). *Let us not suppose that $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$. Then the **augmentation** of $\mathcal{A}$ by $\mathcal{N}$,*

$$\mathcal{A}^{\mathcal{N}} := \{B \in \mathcal{B} : \exists\ A \in \mathcal{A}\ \ni\ A \triangle B \in \mathcal{N}\},$$

is a sub-sigma-algebra of $\mathcal{B}$. Moreover if $\mathcal{N} = \cap \mathcal{N}_i$ and $\mathcal{A}_i = \mathcal{A}^{\mathcal{N}_i}$ is the augmentation of $\mathcal{A}$ by $\mathcal{N}_i$, then

$$\mathcal{A}^{\mathcal{N}} = \cap_i \mathcal{A}_i.$$

Proof. To prove this, first observe that

$$A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cap B^c) \cup (B \cap A^c) \\ = (B^c \setminus A^c) \cup (A^c \setminus B^c) = A^c \triangle B^c$$

from which it follows that $\mathcal{A}^{\mathcal{N}}$ is closed under complementation. Moreover, if $B_j \in \mathcal{A}^{\mathcal{N}}$, then there exists $A_j \in \mathcal{A}$ such that $A_j \triangle B_j \in \mathcal{N}$ for all j. So letting $A = \cup_j A_j \in \mathcal{A}$ and $B = \cup_j B_j \in \mathcal{B}$, we have

$$\mathcal{B} \ni A \triangle B \subset \cup_j [A_j \Delta B_j] \in \mathcal{N}$$

from which we conclude that $A \triangle B \in \mathcal{N}$ and hence $B \in \mathcal{A}^{\mathcal{N}}$. This shows that $\mathcal{A}^{\mathcal{N}}$ is closed under unions and hence we have show $\mathcal{A} \subset \mathcal{A}^{\mathcal{N}} \subset \mathcal{B}$ and $\mathcal{A}^{\mathcal{N}}$ is sigma algebra.

???Now to prove the second assertion of this lemma. It is clear that if $\mathcal{N} \subset \mathcal{N}'$, then $\mathcal{A}^{\mathcal{N}} \subset \mathcal{A}^{\mathcal{N}'}$ and hence it follows that

$$\mathcal{A}^{\mathcal{N}} \subset \cap_i \mathcal{A}_i = \cap_i \mathcal{A}^{\mathcal{N}_i}.$$

For the converse inclusion, suppose that $B \in \cap_i \mathcal{A}^{\mathcal{N}_i}$ in which case there exists $A_i \in \mathcal{A}$ such that $B \triangle A_i \in \mathcal{N}_i$ for all $i \in I$. ■

Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space and $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$. The **augmentation,** $\mathcal{A}^P$, of $\mathcal{A}$ by the $P-$null sets of $\mathcal{B}$ is the collection of sets:

$$\mathcal{A}^P := \{B \in \mathcal{B} : \exists\ A \in \mathcal{A}\ \ni\ P(B \Delta A) = 0\}.$$

Notation 27.36 *Let $\bar{\mathcal{B}}^P$ denote the completion of $\mathcal{B}$. Let $\mathcal{B}_t^P$ denote the augmentation of $\mathcal{B}_t$ by the P – null subsets of $\mathcal{B}$. We also let $\bar{\mathcal{B}}_t^P$ denote the augmentation of $\mathcal{B}_t$ by the P – null subsets of $\bar{\mathcal{B}}_t^P$.*

Continuous time (sub)martingales

For this chapter, let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}, P\right)$ be a filtered probability space as described in Chapter 27.

Definition 28.1. *Given a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, an adapted process, $X_t : \Omega \to \mathbb{R}$, is said to be a $(\{\mathcal{B}_t\})$ **martingale** provided, $\mathbb{E}|X_t| < \infty$ for all t and $\mathbb{E}[X_t - X_s|\mathcal{B}_s] = 0$ for all $0 \leq s \leq t < \infty$. If $\mathbb{E}[X_t - X_s|\mathcal{B}_s] \geq 0$ or $\mathbb{E}[X_t - X_s|\mathcal{B}_s] \leq 0$ for all $0 \leq s \leq t < \infty$, then X is said to be **submartingale** or **supermartingale** respectively.*

Remark 28.2. If σ and τ are two $\{\mathcal{B}_t\}$ – optional times, then $\sigma \wedge \tau$ is as well. Indeed, it $t \in \mathbb{R}_+ \cup \{\infty\}$, then

$$\{\sigma \wedge \tau < t\} = \{\sigma < t\} \cup \{\tau < t\} \in \mathcal{B}_t.$$

The following results are of fundamental importance for a number or results in this chapter. The first result is a simple consequence of the optional sampling Theorem 18.39.

Proposition 28.3 (Discrete optional sampling). *Suppose $\{X_t\}_{t\in\mathbb{R}_+}$ is a submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, and τ and σ are two $\{\mathcal{B}_t\}_{t\geq 0}$ – stopping times with values in $\mathbb{D}_n := \left\{\frac{k}{2^n} : k \in \bar{\mathbb{N}}\right\}$ for some $n \in \mathbb{N}$. If $M := \sup_\omega \tau(\omega) < \infty$, then $X_\tau \in L^1(\Omega, \mathcal{B}_\tau, P)$, $X_{\sigma\wedge\tau} \in L^1(\Omega, \mathcal{B}_{\sigma\wedge\tau}, P)$, and*

$$X_{\sigma\wedge\tau} \leq \mathbb{E}[X_\tau|\mathcal{B}_\sigma].$$

Proof. For $k \in \bar{\mathbb{N}}$, let $\mathcal{F}_k := \mathcal{B}_{k2^{-n}}$ and $Y_k := X_{k2^{-n}}$. Then $\{Y_k\}_{k=0}^\infty$ is a $\{\mathcal{F}_k\}$ – submartingale and $2^n\sigma$, $2^n\tau$ are two $\bar{\mathbb{N}}$ – valued stopping times with $2^n\tau \leq 2^n M < \infty$. Therefore we may apply the optional sampling Theorem 18.39 to find

$$X_{\sigma\wedge\tau} = Y_{(2^n\sigma)\wedge(2^n\tau)} \leq \mathbb{E}[Y_{2^n\tau}|\mathcal{F}_{2^n\sigma}] = \mathbb{E}[X_\tau|\mathcal{B}_\sigma].$$

We have used $\mathcal{F}_{2^n\sigma} = \mathcal{B}_\sigma$ (you prove) in the last equality. ■

Lemma 28.4 (L^1 – convergence I). *Suppose $\{X_t\}_{t\in\mathbb{R}_+}$ is a submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$. If $t \in \mathbb{R}_+$ and $\{t_n\}_{n=1}^\infty \subset (t, \infty)$ such that $t_n \downarrow t$, then $\lim_{n\to\infty} X_{t_n}$ exists almost surely and in $L^1(P)$.*

Proof. Let $Y_n := X_{t_{-n}}$ and $\mathcal{F}_n := \mathcal{B}_{t_{-n}}$ for $n \in -\mathbb{N}$. Then $\{(Y_n, \mathcal{F}_n)\}_{n\in-\mathbb{N}}$ is a backwards submartingale such that $\inf \mathbb{E}Y_n \geq \mathbb{E}X_t$ and hence the result follows by by Theorem 18.75. ■

Lemma 28.5 (L^1 – convergence II). *Suppose $\{X_t\}_{t\in\mathbb{R}_+}$ is a submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, τ is a bounded $\{\mathcal{B}_t\}$ – optional time, and $\{\tau_n\}_{n=1}^\infty$ is the sequence of approximate stopping times defined in Lemma 27.18. Then $X_{\tau+} := \lim_{n\to\infty} X_{\tau_n}$ exists a.s. and in $L^1(P)$.*

Proof. Let $M := \sup_{\omega\in\Omega} \tau(\omega)$. If $m < n$, then τ_m and τ_n take values in $\mathbb{D}_n$, $0 \leq \tau_m \leq \tau_n$, and $\tau_n \leq M + 1$. Therefore by Proposition 28.3, $X_{\tau_m} \leq \mathbb{E}[X_{\tau_n}|\mathcal{B}_{\tau_m}]$ and $X_0 \leq \mathbb{E}[X_{\tau_n}|\mathcal{B}_0]$. Hence if we let $Y_n := X_{\tau_{-n}}$ and $\mathcal{F}_n := \mathcal{B}_{\tau_{-n}}$ for $n \in -\mathbb{N}$. Then $\{(Y_n, \mathcal{F}_n)\}_{n\in-\mathbb{N}}$ is a backwards submartingale such that

$$\inf_{n\in-\mathbb{N}} \mathbb{E}Y_n = \inf_{n\in\mathbb{N}} \mathbb{E}X_{\tau_n} \geq \mathbb{E}X_0 > -\infty.$$

The result now follows by an application of Theorem 18.75. ■

Lemma 28.6 (L^1 – convergence III). *Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\{\mathcal{B}_n\}_{n=1}^\infty$ is a decreasing sequence of sub– σ– algebras of $\mathcal{B}$. Then for all $Z \in L^1(P)$,*

$$\lim_{n\to\infty} \mathbb{E}[Z|\mathcal{B}_n] = \mathbb{E}[Z| \cap_{n=1}^\infty \mathcal{B}_n] \tag{28.1}$$

where the above convergence is almost surely and in $L^1(P)$.

Proof. This is a special case of Corollary 18.77 applied to the reverse martingale, $M_m = \mathbb{E}[Z|\mathcal{F}_m]$ where, for $m \in -\mathbb{N}$, $\mathcal{F}_m := \mathcal{B}_{-m}$. This may also be proved by Hilbert space projection methods when $Z \in L^2(P)$ and then by a limiting argument for all $Z \in L^1(P)$. ■

Proposition 28.7. *Suppose that $Z \in L^1(\Omega, \mathcal{B}, P)$ and σ and τ are two stopping times. Then*

1. *$\mathbb{E}[Z|\mathcal{B}_\sigma] = \mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]$ on $\{\sigma \leq \tau\}$ and hence on $\{\sigma < \tau\}$.*
2. *$\mathbb{E}[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau] = \mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]$.*

Moreover, both results hold if σ and τ are optional times provided every occurrence of the letter $\mathcal{B}$ is replaced by $\mathcal{B}^+$.

Proof. 1. From Corollary 27.17, $1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma]$ is $\mathcal{B}_{\sigma\wedge\tau}$ – measurable and therefore,

$$\begin{aligned}1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma] &= \mathbb{E}[1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_{\sigma\wedge\tau}]\\ &= 1_{\sigma\le\tau}\mathbb{E}[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_{\sigma\wedge\tau}] = 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]\end{aligned}$$

as desired.

2. Writing

$$Z = 1_{\sigma\le\tau}Z + 1_{\tau<\sigma}Z$$

we find, using item 1. that,

$$\begin{aligned}\mathbb{E}[Z|\mathcal{B}_\sigma] &= 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma] + 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_\sigma]\\ &= 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}] + 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_\sigma].\end{aligned} \tag{28.2}$$

Another application of item 1. shows,

$$\begin{aligned}\mathbb{E}[1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau] &= 1_{\tau<\sigma}\mathbb{E}[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau]\\ &= 1_{\tau<\sigma}\mathbb{E}[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_{\tau\wedge\sigma}] = 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_{\tau\wedge\sigma}].\end{aligned}$$

Using this equation and the fact that $1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]$ is $\mathcal{B}_{\sigma\wedge\tau}$ – measurable, we may condition Eq. (28.2) on $\mathcal{B}_{\tau\wedge\sigma}$ to find

$$\mathbb{E}[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau] = 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}] + 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_{\tau\wedge\sigma}] = \mathbb{E}[Z|\mathcal{B}_{\tau\wedge\sigma}].$$

■

Lemma 28.8. *Suppose σ is an optional time and $\{\sigma_m\}_{m=1}^\infty$ are stopping times such that $\sigma_m \downarrow \sigma$ as $m\uparrow\infty$ and $\sigma<\sigma_m$ on $\{\sigma<\infty\}$ for all $m\in\mathbb{N}$. Then $\mathcal{B}_{\sigma_m}\downarrow\mathcal{B}_\sigma^+$ as $m\to\infty$, i.e. $\mathcal{B}_{\sigma_m}$ is decreasing in m and*

$$\mathcal{B}_\sigma^+ = \cap_{m=1}^\infty \mathcal{B}_{\sigma_m}. \tag{28.3}$$

Proof. If $A\in\mathcal{B}_\sigma^+$, then $A\in\mathcal{B}_\infty$ and for all $t\in\mathbb{R}_+$ and $m\in\mathbb{N}$ we have

$$A\cap\{\sigma_m\le t\} = A\cap\{\sigma<t\}\cap\{\sigma_m\le t\}\in\mathcal{B}_t.$$

This shows $A\in\cap_{m=1}^\infty\mathcal{B}_{\sigma_m}$. For the converse, observe that

$$\{\sigma<t\} = \cup_{m=1}^\infty\{\sigma_m\le t\}\ \forall\, t\in\mathbb{R}_+.$$

Therefore if $A\in\cap_{m=1}^\infty\mathcal{B}_{\sigma_m}$ then $A\in\mathcal{B}_\infty$ and

$$A\cap\{\sigma<t\} = \cup_{m=1}^\infty[A\cap\{\sigma_m\le t\}]\in\mathcal{B}_t\ \forall\, t\in\mathbb{R}_+.$$

■

Theorem 28.9 (Continuous time optional sampling theorem). *Let $\{X_t\}_{t\geqslant 0}$ be a right continuous $\{\mathcal{B}_t\}$ (or $\{\mathcal{B}_t^+\}$) – submartingale and σ and τ be two $\{\mathcal{B}_t\}$ – optional (or stopping) times such that $M := \sup_{\omega\in\Omega}\tau(\omega)<\infty$.*[1] *Then $X_\tau\in L^1(\Omega,\mathcal{B}_\tau^+,P)$, $X_{\sigma\wedge\tau}\in L^1(\Omega,\mathcal{B}_{\sigma\wedge\tau}^+,P)$ and*

$$X_{\sigma\wedge\tau}\le\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma^+\right]. \tag{28.4}$$

Proof. Let $\{\sigma_m\}_{m=1}^\infty$ and $\{\tau_n\}_{n=1}^\infty$ be the sequences of approximate times for σ and τ respectively defined Lemma 27.18, i.e.

$$\tau_n := \infty 1_{\tau=\infty} + \sum_{k=1}^\infty \frac{k}{2^n}1_{\frac{k-1}{2^n}\le\tau<\frac{k}{2^n}}.$$

By the discrete optional sampling Proposition 28.3, we know that

$$X_{\sigma_m\wedge\tau_n}\le\mathbb{E}[X_{\tau_n}|\mathcal{B}_{\sigma_m}]\text{ a.s.} \tag{28.5}$$

Since X_t is right continuous, $X_{\tau_n}(\omega)\to X_\tau(\omega)$ for all $\omega\in\Omega$ which combined with Lemma 28.5 implies $X_{\tau_n}\to X_\tau$ in $L^1(P)$ and in particular $X_\tau\in L^1(P)$. Similarly, $X_{\sigma_n\wedge\tau_n}\to X_{\sigma\wedge\tau}$ in $L^1(P)$ and therefore $X_{\sigma\wedge\tau}\in L^1(P)$. Using the $L^1(P)$ – contractivity of conditional expectation along with the fact that $X_{\sigma_m\wedge\tau_n}\to X_{\sigma_m\wedge\tau}$ on Ω, we may pass to the limit $(n\to\infty)$ in Eq. (28.5) to find

$$X_{\sigma_m\wedge\tau}\le\mathbb{E}[X_\tau|\mathcal{B}_{\sigma_m}]\text{ a.s.} \tag{28.6}$$

From the right continuity of $\{X_t\}$ and making use of Lemma 28.6 (or Corollary 18.77) and Lemma 28.8, we may let $m\to\infty$ in Eq. (28.6) to find

$$X_{\sigma\wedge\tau}\le\lim_{m\to\infty}\mathbb{E}[X_\tau|\mathcal{B}_{\sigma_m}] = \mathbb{E}[X_\tau|\cap_{m=1}^\infty\mathcal{B}_{\sigma_m}] = \mathbb{E}\left[X_\tau|\mathcal{B}_\sigma^+\right]$$

which is Eq. (28.4). ■

Corollary 28.10 (Optional stopping). *Let $\{X_t\}_{t\geqslant 0}$ be a right continuous $\{\mathcal{B}_t\}$ (or $\{\mathcal{B}_t^+\}$) – submartingale and τ be any $\{\mathcal{B}_t\}$ – optional (or stopping) time. Then the stopped process, $X_t^\sigma := X_{\sigma\wedge t}$ is a right continuous $\{\mathcal{B}_t^+\}$ – submartingale.*

Proof. Let $0\le s\le t<\infty$ and apply Theorem 28.9 with to the two stopping times, $\sigma\wedge s$ and $\sigma\wedge t$ to find

$$X_s^\sigma = X_{\sigma\wedge s}\le\mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_{\sigma\wedge s}^+\right] = \mathbb{E}\left[X_t^\sigma|\mathcal{B}_{\sigma\wedge s}^+\right].$$

From Proposition 28.7,

$$\mathbb{E}\left[X_t^\sigma|\mathcal{B}_{\sigma\wedge s}^+\right] = \mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_{\sigma\wedge t\wedge s}^+\right] = \mathbb{E}\left[\mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_{\sigma\wedge t}^+\right]|\mathcal{B}_s^+\right] = \mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_s^+\right]$$

and therefore, we have shown $X_s^\sigma\le\mathbb{E}[X_t^\sigma|\mathcal{B}_s^+]$. Since $X_s^\sigma = X_{\sigma\wedge s}$ is $\mathcal{B}_{\sigma\wedge s}^+$ measurable and $\mathcal{B}_{\sigma\wedge s}^+\subset\mathcal{B}_s^+$, it follows that X_s^σ is $\mathcal{B}_s^+$ – measurable. ■

[1] We will see below in Theorem 28.28, that the boundedness restriction on τ may be replaced by the assumption that $\{X_t^+\}_{t\ge 0}$ is uniformly integrable.

28.1 Submartingale Inequalities

Let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}, P\right)$ be a filtered probability space, $\mathbb{D}$ be any dense subset of $\mathbb{R}_+$ containing 0, and let $\mathbb{T}$ denote either $\mathbb{D}$ or $\mathbb{R}_+$. Throughout this section, $\{X_t\}_{t\in\mathbb{T}}$ will be a submartingale which is *assumed* to be *right continuous* if $\mathbb{T}=\mathbb{R}_+$. To keep the notation unified, for $T\in\mathbb{R}_+$, we will simply denote $\sup_{\mathbb{D}\ni t\le T} X_t = \inf_{\mathbb{D}\ni t\le T} X_t$, and $\sup_{s\in\mathbb{D}\cap[0,T]}|X_s|$ by $\sup_{t\le T} X_t$, $\inf_{t\le T} X_t$, and X_T^* respectively. It is worth observing that if $\mathbb{T}=\mathbb{R}_+$ and $T\in\mathbb{D}$, we have (by the assumed right continuity of X_t) that

$$\sup_{t\le T} X_t = \sup_{\mathbb{D}\ni t\le T} X_t,\ \ \inf_{t\le T} X_t = \inf_{\mathbb{D}\ni t\le T} X_t \text{ and },\ \sup_{s\in[0,T]}|X_s| = \sup_{s\in\mathbb{D}\cap[0,T]}|X_s|. \tag{28.7}$$

Our immediate goal is to generalize the submartingale inequalities of Section 18.5 to this context.

Proposition 28.11 (Maximal Inequalities of Bernstein and Lévy). *With $\mathbb{T}=\mathbb{D}$ for $\mathbb{T}=\mathbb{R}_+$, for any $a\ge 0$ and $T\in\mathbb{T}$, we have,*

$$aP\left(\sup_{t\le T} X_t \ge a\right) \le \mathbb{E}\left[X_T : \sup_{t\le T} X_t \ge a\right] \le \mathbb{E}\left[X_T^+\right], \tag{28.8}$$

$$aP\left(\inf_{t\le T} X_t \le -a\right) \le \mathbb{E}\left[X_T : \inf_{t\le T} X_t > -a\right] - \mathbb{E}\left[X_0\right] \tag{28.9}$$

$$\le \mathbb{E}\left[X_T^+\right] - \mathbb{E}\left[X_0\right], \tag{28.10}$$

and

$$aP\left(X_T^*\ge a\right) \le 2\mathbb{E}\left[X_T^+\right] - \mathbb{E}\left[X_0\right]. \tag{28.11}$$

In particular if $\{M_t\}_{t\in\mathbb{T}}$ is a martingale and $a>0$, then

$$P\left(M_T^*\ge a\right) \le \frac{1}{a}\mathbb{E}\left[|M|_T : M_T^*\ge a\right] \le \frac{1}{a}\mathbb{E}\left[|M_T|\right] \tag{28.12}$$

Proof. First assume $\mathbb{T}=\mathbb{D}$. For each $k\in\mathbb{N}$ let

$$\Lambda_k = \{0=t_0<t_1<\cdots<t_m=T\}\subset\mathbb{D}\cap[0,T]$$

be a finite subset of $\mathbb{D}\cap[0,T]$ containing $\{0,T\}$ such that $\Lambda_k\uparrow\mathbb{D}\cap[0,T]$. Noting that $\{X_{t_n}\}_{n=0}^m$ is a discrete $\left(\Omega,\mathcal{B},\{\mathcal{B}_{t_n}\}_{n=0}^m, P\right)$ submartingale, Proposition 18.42 implies all of the inequalities in Eqs. (28.8) – (28.11) hold provided we replace $\sup_{t\le T}X_t$ by $\max_{t\in\Lambda_k}X_t$, $\inf_{t\le T}X_t$ by $\min_{t\in\Lambda_k}X_t$, and X_T^* by $\max_{t\in\Lambda_k}|X_t|$. Since $\max_{t\in\Lambda_k}X_t\uparrow\sup_{t\le T}X_t$, $\max_{t\in\Lambda_k}|X_t|\uparrow X_T^*$, and $\min_{t\in\Lambda_k}X_t\downarrow\inf_{t\le T}X_t$, we may use the MCT and the DCT to pass to the limit $(k\to\infty)$ in order to conclude Eqs. (28.8) – (28.11) are valid as stated. Equation (28.12) follows from Eq. (28.8) applied to $X_t := |M_t|$.

Now suppose that $\{X_t\}_{t\in\mathbb{R}_+}$ and $\{M_t\}_{t\in\mathbb{R}_+}$ are right continuous. Making use of the observations in Eq. (28.7), we see that Eqs. (28.8) – (28.12) remain valid for $\mathbb{T}=\mathbb{R}_+$ by what we have just proved in the case $\mathbb{T}=\mathbb{D}\cup\{T\}$. ∎

Proposition 28.12 (Doob's Inequality). *Suppose that X_t is a non-negative submartingale (for example $X_t=|M_t|$ where M_t is a martingale) and $1<p<\infty$, then for any $T\in\mathbb{T}$,*

$$\mathbb{E}X_T^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X_T^p. \tag{28.13}$$

Proof. Using the notation in the proof of Proposition 28.11, it follows from Corollary 18.46 that

$$\mathbb{E}\left[\max_{t\in\Lambda_k}|X_t|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X_T^p.$$

Using the MCT, we may let $k\uparrow\infty$ in this equation to arrive at Eq. (28.13) when $\mathbb{T}=\mathbb{D}$. The case when $\mathbb{T}=\mathbb{R}_+$ follows immediately using the comments at the end of the proof of Proposition 28.11. ∎

Lemma 28.13. *Suppose that F_n is a sequence of bounded functions on $[a,b)$ which are uniformly convergent to a function F. If $\xi_n := \lim_{t\downarrow a}F_n(t)$ exists for all n, then $\xi := \lim_{t\downarrow a}F(t)$ exists and $\xi_n\to\xi$ as $n\to\infty$. An analogous statement holds for left limits. In particular right (left) continuous functions are preserved under uniform limits.*

Proof. Let $\varepsilon_n := \sup_{t\in[a,b)}|F(t)-F_n(t)|$ which by assumption tends to zero as $n\to\infty$. Thus for $s,t>a$, we have

$$\begin{aligned}|F(t)-F(s)| &\le |F(t)-F_n(t)| + |F_n(t)-F_n(s)| + |F_n(s)-F(s)| \\ &\le 2\varepsilon_n + |F_n(t)-F_n(s)|.\end{aligned}$$

Therefore we have

$$\limsup_{s,t\downarrow a}|F(t)-F(s)| \le 2\varepsilon_n \to 0 \text{ as } n\to\infty$$

which shows that $\xi := \lim_{t\downarrow a}F(t)$ exists. Similarly, for any $t>a$,

$$\begin{aligned}|\xi-\xi_n| &\le |\xi-F(t)| + |F(t)-F_n(t)| + |F_n(t)-\xi_n| \\ &\le |\xi-F(t)| + |F_n(t)-\xi_n| + \varepsilon_n\end{aligned}$$

and hence by passing to the limit as $t\downarrow a$ in the previous inequality we have $|\xi-\xi_n|\le\varepsilon_n\to 0$ as $n\to\infty$. ∎

Corollary 28.14. *Suppose that* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ *is a filtered probability space such that* $\mathcal{B}_t$ *contains all* P *– null subsets*[2] *of* $\mathcal{B}$ *for all* $t \in \mathbb{R}_+$. *For any* $T \in \mathbb{R}_+$, *let* $\mathbb{M}_T$ *denote the collection of (right) continuous* L^2 *– martingales,* $M := \{M_t\}_{t\leq T}$ *equipped with the inner product,*

$$(M, N)_T := \mathbb{E}\left[M_T N_T\right].$$

(More precisely, two (right) continuous L^2 *– martingales,* M *and* N, *are taken to be equal if* $P\left(M_t = N_t \ \forall\ t \leq T\right) = 1$.*) Then the space,* $\left(\mathbb{M}_T, (\cdot,\cdot)_T\right)$, *is a Hilbert space and the map,* $U : \mathbb{M}_T \to L^2\left(\Omega, \mathcal{B}_T, P\right)$ *defined by* $UM := M_T$, *is an isometry.*

Proof. Since $M_t = \mathbb{E}\left[M_T|\mathcal{B}_t\right]$ a.s., if $(M_T, M_T) = \mathbb{E}\left|M_T\right|^2 = 0$ then $M = 0$ in $\mathbb{M}_T$. This shows that U is injective and by definition U is an isometry and $(\cdot,\cdot)_T$ is an inner product on $\mathbb{M}_T$. To finish the proof, we need only show $H := \mathrm{Ran}\,(U)$ is a closed subspace of $L^2\left(\Omega, \mathcal{B}_T, P\right)$ or equivalently that $\mathbb{M}_T$ is complete.

Suppose that $\{M^n\}_{n=1}^\infty$ is a Cauchy sequence in $\mathbb{M}_T$, then by Doob's inequality (Proposition 28.12) and Hölder's inequality, we have

$$\begin{aligned}\mathbb{E}\left[(M^n - M^m)_T^*\right] &\leq \sqrt{\mathbb{E}\left[(M^n - M^m)_T^{*2}\right]} \\ &\leq \sqrt{4\mathbb{E}\left|M_T^n - M_T^m\right|^2} = 2\left\|M^n - M^m\right\|_T \to 0 \text{ as } m, n \to \infty.\end{aligned}$$

By passing to a subsequence if necessary, we may assume

$$\sum_{n=1}^\infty \mathbb{E}\left[\left(M^{n+1} - M^n\right)_T^*\right] \leq 2\sum_{n=1}^\infty \left\|M^{n+1} - M^n\right\|_T < \infty$$

from which it follows that

$$\mathbb{E}\left[\sum_{n=1}^\infty \left(M^{n+1} - M^n\right)_T^*\right] = \sum_{n=1}^\infty \mathbb{E}\left[\left(M_\cdot^{n+1} - M_\cdot^n\right)_T^*\right] < \infty.$$

So if we let

$$\Omega_0 := \left\{\sum_{n=1}^\infty \left(M_\cdot^{n+1} - M_\cdot^n\right)_T^* < \infty\right\},$$

then $P\left(\Omega_0\right) = 1$. Hence if $m < l$, the triangle inequality implies

$$\left(M_\cdot^l - M_\cdot^m\right)_T^* \leq \sum_{n=m}^{l-1} \left(M_\cdot^{n+1} - M_\cdot^n\right)_T^* \to 0 \text{ on } \Omega_0 \text{ as } m, l \to \infty,$$

which shows that $\{M_\cdot^n(\omega)\}_{n=1}^\infty$ is a uniformly Cauchy sequence and hence uniformly convergent for all $\omega \in \Omega_0$. Therefore by Lemma 28.13, $t \to M_t(\omega)$ is (right) continuous for all $\omega \in \Omega_0$. We complete the definition of M by setting $M_\cdot(\omega) \equiv 0$ for $\omega \notin \Omega_0$. Since $\mathcal{B}_t$ contains all of the null subset in $\mathcal{B}$, it is easy to see that $M_\cdot$ is a $\mathcal{B}_t$ – adapted process. Moreover, by Fatou's lemma, we have

$$\begin{aligned}\mathbb{E}\left[(M_\cdot - M_\cdot^m)_T^{*2}\right] &= \mathbb{E}\left[\liminf_{n\to\infty} (M_\cdot^n - M_\cdot^m)_T^{*2}\right] \\ &\leq \liminf_{n\to\infty} \mathbb{E}\left[(M_\cdot^n - M_\cdot^m)_T^{*2}\right] \to 0 \text{ as } m \to \infty.\end{aligned}$$

In particular $M_t^m \to M_t$ in $L^2(P)$ for all $t \leq T$ from which follows that M is still an L^2 – martingale. As M is (right) continuous, $M \in \mathbb{M}_T$ and

$$\|M - M^n\|_T = \|M_T - M_T^n\|_{L^2(P)} \to 0 \text{ as } n \to \infty.$$

■

28.2 Regularizing a submartingale

Lemma 28.15. *Suppose that* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ *is a filtered probability space and* $\{X_t\}_{t\geq 0}$ *be a* $\{\mathcal{B}_t\}_{t\geq 0}$ *– submartingale. Then* $\{X_t\}_{t\geq 0}$ *is also a* $\left\{\bar{\mathcal{B}}_t\right\}_{t\geq 0}$ *– submartingale. Moreover, we may first replace* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ *by its completion,* $\left(\Omega, \bar{\mathcal{B}}, \bar{P}\right)$ *(see Proposition 5.50), then* $\{X_t\}_{t\geq 0}$ *is still a submartingale relative to the filtration* $\bar{\mathcal{B}}_t := \mathcal{B}_t \vee \bar{\mathcal{N}}$ *where*

$$\bar{\mathcal{N}} := \left\{B \in \bar{\mathcal{B}} : \bar{P}(B) = 0\right\}.$$

Proof. It suffices to prove the second assertion. By the augmentation Lemma 27.22 we know that $B \in \bar{\mathcal{B}}_s := \mathcal{B}_s \vee \bar{\mathcal{N}}$ iff there exists $A \in \mathcal{B}_s$ such that $B \triangle A \in \bar{\mathcal{N}}$. Then for any $t > s$ we have

$$\mathbb{E}_{\bar{P}}\left[X_t - X_s : B\right] = \mathbb{E}_{\bar{P}}\left[X_t - X_s : A\right] = \mathbb{E}_P\left[X_t - X_s : A\right] \geq 0.$$

■

Proposition 28.16. *Suppose that* $\{X_t\}_{t\in\mathbb{R}_+}$ *is an* $\{\mathcal{B}_t\}$ *– submartingale such that* $t \to X_t$ *is right continuous in probability, i.e.* $X_t \xrightarrow{P} X_s$ *as* $t \downarrow s$ *for all* $s \in \mathbb{R}_+$. *(For example, this hypothesis will hold if there exists* $\varepsilon > 0$ *such that* $\lim_{t\downarrow s} \mathbb{E}\left|X_t - X_s\right|^\varepsilon = 0$ *for all* $s \in \mathbb{R}_+$.*) Then* $\{X_t\}_{t\in\mathbb{R}_+}$ *is also an* $\left\{\mathcal{B}_t^+\right\}$ *– submartingale.*

[2] Lemma 28.15 below shows that this hypothesis can always be fulfilled if one is willing to "augment" the filtration by the P – null sets.

Proof. Let $0 \le s < t < \infty$, $A \in \mathcal{B}_s^+$, and $s_n \in (s,t)$ such that $s_n \downarrow s$. By Lemma 28.4 we know that $\hat{X}_s := \lim_{n\to\infty} X_{s_n}$ exists a.s. and in $L^1(P)$ and using the assumption that $X_{s_n} \xrightarrow{P} X_s$ we may conclude that $X_{s_n} \to X_s$ in $L^1(P)$. Since $A \in \mathcal{B}_s^+ \subset \mathcal{B}_{s_n}$ for all n, we have

$$\mathbb{E}\left[X_t - X_s : A\right] = \lim_{n\to\infty} \mathbb{E}\left[X_t - X_{s_n} : A\right] \ge 0.$$

■

Corollary 28.17. *Suppose* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\ge 0}, P\right)$ *is a filtered probability space and* $\{X_t\}_{t\in\mathbb{R}_+}$ *is an* $\{\mathcal{B}_t\}$ *– submartingale such that* $X_t \xrightarrow{P} X_s$ *as* $t \downarrow s$ *for all* $s \in \mathbb{R}_+$. *Let* $\left(\Omega, \bar{\mathcal{B}}, \bar{P}\right)$ *denote the completion of* $(\Omega, \mathcal{B}, P)$, $\mathcal{N}$ *and* $\bar{\mathcal{N}}$ *be the* P *and* $\bar{P}$ *null sets respectively, then* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_{t+} \vee \mathcal{N}\}_{t\ge 0}, P\right)$ *satisfies the weak usual hypothesis (see Definition 27.21),* $\left(\Omega, \bar{\mathcal{B}}, \left\{\mathcal{B}_{t+} \vee \bar{\mathcal{N}}\right\}_{t\ge 0}, \bar{P}\right)$ *satisfies the usual hypothesis, and* $\{X_t\}_{t\ge 0}$ *is a submartingale relative to each of these filtrations.*

Proof. This follows directly from Proposition 28.16, Lemma 28.15, and Lemma 27.23. We use Lemma 27.23 to guarantee that $\{\mathcal{B}_{t+} \vee \mathcal{N}\}_{t\ge 0}$ and $\left\{\mathcal{B}_{t+} \vee \bar{\mathcal{N}}\right\}_{t\ge 0}$ are right continuous. ■

In all of the examples of submartingales appearing in this book, the hypothesis and hence the conclusions of Proposition 28.16 will apply. For this reason there is typically no harm in assuming that our filtration is right continuous. By Corollary 28.17 we may also assume that $\mathcal{B}_t$ contains all P – null sets. The results in the following exercise are useful to keep in mind as you are reading the rest of this section.

Exercise 28.1 (Continuous version of Example 18.7). Suppose that $\Omega = (0,1]$, $\mathcal{B} = \mathcal{B}_{(0,1]}$, and $P = m$ – Lebesgue measure. Further suppose that $\varepsilon : [0,\infty) \to \{0,1\}$ is **any** function of your choosing. Then define, for $t \ge 0$ and $x \in \Omega$,

$$M_t^\varepsilon(x) := e^t \left(\varepsilon(t) 1_{0<x\le e^{-t}} + (1-\varepsilon(t)) 1_{0<x<e^{-t}}\right) = e^t \left(1_{0<x<e^{-t}} + \varepsilon(t) 1_{x=e^{-t}}\right).$$

Further let $\mathcal{B}_t^\varepsilon := \sigma(M_s^\varepsilon : s \le t)$ for all $t \ge 0$ and for $a \in (0,1]$ let

$$\mathcal{F}_{(0,a]} := \left\{[0,a] \cup A : A \in \mathcal{B}_{(a,1]}\right\} \cup \mathcal{B}_{(a,1]}$$

and

$$\mathcal{F}_{(0,a)} := \left\{(0,a) \cup A : A \in \mathcal{B}_{[a,1]}\right\} \cup \mathcal{B}_{[a,1]}.$$

Show:

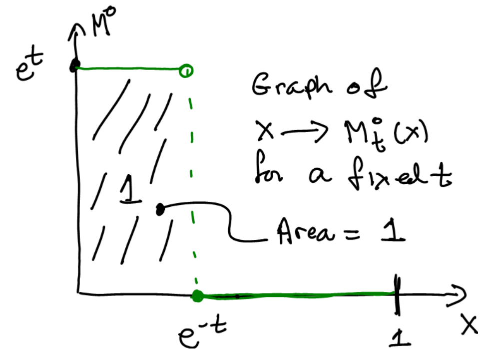


Fig. 28.1. The graph of $x \to M_t^0(x)$ for some fixed t.

1. $\mathcal{F}_{(0,a]}$ and $\mathcal{F}_{(0,a)}$ are sub – sigma – algebras of $\mathcal{B}$ such that $\mathcal{F}_{(0,a]} \subsetneq \mathcal{F}_{(0,a)}$ and
$$\mathcal{B} = \vee_{a\in(0,1]} \mathcal{F}_{(0,a]} = \vee_{a\in(0,1]} \mathcal{F}_{(0,a)}.$$
2. For all $b \in (0,1]$,
$$\mathcal{F}_{(0,b)} = \cap_{a<b} \mathcal{F}_{(0,a]} = \cap_{a<b} \mathcal{F}_{(0,a)}. \tag{28.14}$$
3. $M_{t+}^\varepsilon = M_t^0$ for all $t \ge 0$ and $M_{t-}^\varepsilon = M_t^1$ for all $t > 0$. In particular, the sample paths, $t \to M_{t+}^\varepsilon(x)$, are right continuous and possess left limits for all $x \in \Omega$.

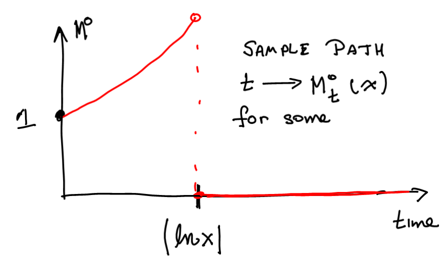


Fig. 28.2. A typical sample path of $M_\cdot^0(x)$.

4. $\mathcal{B}_t^\varepsilon = \mathcal{F}_{(0,e^{-t}]}$ if $\varepsilon(t) = 1$ and $\mathcal{B}_t^\varepsilon = \mathcal{F}_{(0,e^{-t})}$ if $\varepsilon(t) = 0$.
5. No matter how ε is chosen, $\mathcal{B}_{t+}^\varepsilon = \mathcal{B}_t^0 := \mathcal{F}_{(0,e^{-t})}$ for all $t \ge 0$.
6. M_t^ε is a $\{\mathcal{B}_t^\varepsilon\}_{t\ge 0}$ – martingale and in fact it is a $\left\{\mathcal{B}_{t+}^\varepsilon = \mathcal{B}_t^0\right\}_{t\ge 0}$ – martingale.
7. The map, $[0,\infty) \times (0,1] \ni (t,x) \to M_t^\varepsilon(x) \in \mathbb{R}_+$ is measurable iff $\{t \in [0,\infty) : \varepsilon(t) = 1\} \in \mathcal{B}_{\mathbb{R}_+}$.

8. Let
$$\mathcal{N} := \left\{x : M_t^\varepsilon(x) \neq M_{t+}^\varepsilon(x) \text{ for some } t \geq 0\right\}.$$
Show $\mathcal{N} = \{x : \varepsilon(|\ln x|) = 1\}$ and observe that $\mathcal{N}$ is measurable iff ε is measurable. Also observe that if $\varepsilon \equiv 1$, then $P(\mathcal{N}) = 1$ and hence M_{t+}^ε and M_t^ε are certainly not indistinguishable, see Definition 25.2.
9. Show $\{M_t^\varepsilon\}_{t\geq 0}$ is **not** uniformly integrable.
10. Let $Z \in L^1(\Omega, \mathcal{B}, P)$ find a version, N_t^ε, of $\mathbb{E}[Z|\mathcal{B}_t^\varepsilon]$. Verify that for any sequence, $\{t_n\}_{n=1}^\infty \subset [1,\infty)$, that $N_{t_n}^\varepsilon \to Z$ almost surely and in $L^1(P)$ as $n \to \infty$.

Solution to Exercise (28.1).

1. It is routine to check that $\mathcal{F}_{(0,a]} \subsetneqq \mathcal{F}_{(0,a)} \subset \mathcal{B}_{(0,1]}$ are sigma algebras. Since $\vee_{a\in(0,1]}\mathcal{F}_{(0,a]}$ contains all sets of the form, $\{(a,1] : 0 \leq a \leq 1\}$, it contains $\mathcal{B}_{(0,1]}$. In fact this is most easily done by observing that $\mathcal{F}_{(0,a)}$ consists those $A \in \mathcal{B}_{(0,1]}$ such that $(0,a) \subset A$ or $(0,a) \subset A^c$ with a similar characterization of $\mathcal{F}_{(0,a]}$.
2. Since $\mathcal{F}_{(0,a]} \subset \mathcal{F}_{(0,a)}$, $\mathcal{F}_{(0,a]}$ and $\mathcal{F}_{(0,a)}$ are increasing as a decreases, and $\mathcal{F}_{(0,b)} \subset \mathcal{F}_{(0,a]}$ if $a < b$,
$$\mathcal{F}_{(0,b)} \subset \cap_{a<b}\mathcal{F}_{(0,a]} \subset \cap_{a<b}\mathcal{F}_{(0,a)}.$$
Now suppose $\tilde{A} \in \cap_{a<b}\mathcal{F}_{(0,a)}$ and $a_n < b$ with $a_n \uparrow b$. Then for each n, there exists an $A_n \in \mathcal{B}_{[a_n,1]}$ such that $\tilde{A} = (0,a_n) \cup A_n$ or $\tilde{A} = A_n$. There are now only two alternatives, either 1) $\tilde{A} = (0,a_n) \cup A_n$ for all n or 2) $\tilde{A} = A_n$ for all n. In the first case, we must have
$$(0,b) = \cup_{n=1}^\infty (0,a_n) \subset \tilde{A}$$
and therefore $\tilde{A} = (0,b) \cup A_\infty$ for some $A_\infty \in B_{[b,1]}$ and thus $\tilde{A} \in \mathcal{F}_{(0,b)}$. In the second case, we know that $\tilde{A} \cap (0,a_n) = \emptyset$ for all n and therefore, $\tilde{A} \cap (0,b) = \emptyset$ from which it again follows that $\tilde{A} \in \mathcal{F}_{(0,b)}$. Therefore we have shown
$$\mathcal{F}_{(0,b)} \subset \cap_{a<b}\mathcal{F}_{(0,a]} \subset \cap_{a<b}\mathcal{F}_{(0,a)} \subset \mathcal{F}_{(0,b)}$$
which implies Eq. (28.14).
3. Let $x \in \Omega = (0,1]$ be fixed and observe that
$$\begin{aligned} M_t^\varepsilon(x) &= e^t\left(1_{t<-\ln(x)} + \varepsilon(t)\,1_{t=-\ln(x)}\right) \\ &= e^t\left(1_{t<|\ln(x)|} + \varepsilon(|\ln(x)|)\,1_{t=|\ln(x)|}\right). \end{aligned}$$
Therefore, no matter the value of $\varepsilon(|\ln(x)|)$, $M_t^\varepsilon(x) = 0$ if $t > |\ln(x)|$ and $M_t^\varepsilon(x) = e^t$ if $t < |\ln(x)|$. Hence it follows that
$$M_{t+}^\varepsilon(x) = e^t 1_{t<-\ln(x)} = M_t^0(x) \tag{28.15}$$
and
$$M_{t-}^\varepsilon(x) = e^t 1_{t\leq-\ln(x)} = M_t^1(x). \tag{28.16}$$
4. Since $e^{-t}M_t^0(x) = 1_{0<x<e^{-t}}$ and $e^{-t}M_t^1(x) = 1_{0<x\leq e^{-t}}$, the reader may easily show, $\mathcal{B}_t^0 = \mathcal{F}_{(0,e^{-t})}$ and $\mathcal{B}_t^1 = \mathcal{F}_{(0,e^{-t}]}$ for all $t \geq 0$. Also, no matter how ε is chosen, $M_s^1 = M_{s-}^\varepsilon$ is $\mathcal{B}_t^\varepsilon$ measurable for all $0 < s \leq t$ and since $M_0^1 \equiv 1$ we may conclude that M_0^1 is $\mathcal{B}_0^\varepsilon \subset \mathcal{B}_t^\varepsilon$ measurable as well. It is also simple to verify M_t^ε is $\mathcal{F}_{(0,e^{-t})}$ – measurable for all $t \geq 0$. From these observations we may conclude, no matter how ε is chosen, that
$$\mathcal{F}_{(0,e^{-t}]} = \mathcal{B}_t^1 \subset \mathcal{B}_t^\varepsilon \subset \mathcal{F}_{(0,e^{-t})}.$$
As M_s^ε is $\mathcal{F}_{(0,e^{-t}]}$ – measurable for all $s < t$ and M_t^ε is $\mathcal{F}_{(0,e^{-t}]}$ – measurable if $\varepsilon(t) = 1$ we may conclude, $\mathcal{B}_t^\varepsilon = \mathcal{F}_{(0,e^{-t}]}$ if $\varepsilon(t) = 1$. On the other hand, if $\varepsilon(t) = 0$, then $\mathcal{B}_t^\varepsilon$ contains $\mathcal{F}_{(0,e^{-t}]}$ and all M_t^0 – measurable functions. This then implies that
$$\mathcal{F}_{(0,e^{-t})} = \left\{e^{-t}\right\} \vee \mathcal{F}_{(0,e^{-t}]} \subset \mathcal{B}_t^\varepsilon \subset \mathcal{F}_{(0,e^{-t})}$$
which forces $\mathcal{B}_t^\varepsilon = \mathcal{F}_{(0,e^{-t})}$ if $\varepsilon(t) = 0$.
5. By items 2. and 4.,
$$\mathcal{B}_t^0 = \mathcal{B}_{t+}^1 \subset \mathcal{B}_{t+}^\varepsilon \subset \mathcal{B}_{t+}^0 = \mathcal{B}_t^0$$
from which it follows that $\mathcal{B}_{t+}^\varepsilon = \mathcal{B}_t^0 := \mathcal{F}_{(0,e^{-t})}$ for all $t \geq 0$.
6. If $0 \leq s \leq t$ and $A \in \mathcal{B}_s^0 = \mathcal{F}_{(0,e^{-s})}$, then
$$\mathbb{E}[M_t^\varepsilon : A] = \begin{cases} 0 \text{ if } A \cap (0,e^{-s}) = \emptyset \\ 1 \text{ if } \quad (0,e^{-s}) \subset A. \end{cases}$$
Since this is constant in t for $t \geq s$ and $A \in \mathcal{B}_s^0$, we wee that M_t^ε is a $\left\{\mathcal{B}_t^0\right\}_{t\geq 0}$ – martingale.
7. Since $(t,x) \to M_t^\varepsilon(x)$ is measurable iff $(t,x) \to e^{-t}M_t^\varepsilon(x)$ is measurable and as the latter function is an indicator function this is equivalent to the set, $B := \{(t,x) : e^{-t}M_t^\varepsilon(x) = 1\}$ being measurable. Now,
$$\begin{aligned} B &= \{(t,x) : (1_{0<x<e^{-t}} + \varepsilon(t)\,1_{x=e^{-t}}) = 1\} \\ &= \{(t,x) : 1_{0<x<e^{-t}} = 1\} \cup \{(t,x) : \varepsilon(t)\,1_{x=e^{-t}} = 1\} \\ &= \{(t,x) : 1_{0<x<e^{-t}} = 1\} \cup A \end{aligned}$$
where $A := \{(t,e^{-t}) : \varepsilon(t) = 1\} \subset [0,\infty) \times (0,1]$. As A is disjoint from the measurable set, $\{(t,x) : 1_{0<x<e^{-t}} = 1\}$, it follows that B is measurable iff

A is measurable. If A is measurable, let $\psi : \mathbb{R}_+ \to \mathbb{R}_+ \times (0,1]$ be the continuous map defined by $\psi(t) = (t, e^{-t})$. Then $\{t : \varepsilon(t) = 1\} = \psi^{-1}(A)$ is measurable and because ε is an indicator function this shows ε is measurable. Conversely if ε is measurable, then

$$A = \left\{\left(t, e^{-t}\right) : t \geq 0\right\} \cap (\{t : \varepsilon(t) = 1\} \times (0,1])$$

which is measurable.

8. Since $M^\varepsilon_{t+}(x) = M^0_t(x)$, we see that $M^\varepsilon_t(x) \neq M^0_t(x)$ can only happen if $x = e^{-t}$ and $\varepsilon(t) = 1$, i.e. only if
$$1 = \varepsilon(-\ln x) = \varepsilon(|\ln x|).$$
9. Since $M^\varepsilon_t \to 0$ as $t \to \infty$ while $\mathbb{E}M^\varepsilon_t = 1$ for all t, see Figure 28.1, $\{M^\varepsilon_t\}_{t\geq 0}$ can not be uniformly integrable.
10. Let $Z \in L^1(\Omega, \mathcal{B}, P)$ one easily shows that
$$\mathbb{E}\left[Z|\mathcal{B}^0_t\right](x) = 1_{0<x<e^{-t}} \cdot e^t \int_0^{e^{-t}} Z(y)\,dy + 1_{e^{-t}\leq x<\infty} \cdot Z(x)$$
and
$$\mathbb{E}\left[Z|\mathcal{B}^1_t\right](x) = 1_{0<x\leq e^{-t}} \cdot e^t \int_0^{e^{-t}} Z(y)\,dy + 1_{e^{-t}<x<\infty} \cdot Z(x)$$
will do the trick.

Definition 28.18 (Upcrossings). *Let $\{x_t\}_{t\in\mathbb{T}}$ be a real valued function which is right continuous if $\mathbb{T} = \mathbb{R}_+$. Given $-\infty < a < b < \infty$, $T \in \mathbb{T}$, and a finite subset, F, of $[0,T] \cap \mathbb{T}$, let $U^x_F(a,b)$ denote the* ***number of upcrossings of*** *$\{x_t\}_{t\in F}$* ***across*** *$[a,b]$, see Section 18.6. Also let*

$$U^x_T(a,b) := \sup\{U^x_F(a,b) : F \subset_f \mathbb{T} \cap [0,T]\} \tag{28.17}$$

be the ***number of upcrossings of*** *$\{x_t\}_{t\in\mathbb{T}\cap[0,T]}$* ***across*** *$[a,b]$.*

Lemma 28.19. *If $\mathbb{T} = \mathbb{D}$ and $\{F_n\}_{n=1}^\infty$ is a sequence of finite subsets of $\mathbb{D}\cap[0,T]$ such that $F_n \uparrow \mathbb{D}\cap[0,T]$, then*

$$U^x_T(a,b) := \lim_{n\to\infty} U^x_{F_n}(a,b). \tag{28.18}$$

In particular, $U^X_T(a,b)$ is a $\mathcal{B}_T$ – measurable random variable when $\mathbb{T} = \mathbb{D}$.

Proof. It is clear that $U^x_{F_n}(a,b) \leq U^x_T(a,b)$ for all n and $U^x_{F_n}(a,b)$ is increasing with n and therefore the limit in Eq. (28.18) exists and satisfies, $\lim_{n\to\infty} U^x_{F_n}(a,b) \leq U^x_T(a,b)$. Moreover, for any $F \subset_f \mathbb{D}\cap[0,T]$ we may find an $n \in \mathbb{N}$ sufficiently large so that $F \subset F_n$. For this n we will have

$$U^x_F(a,b) \leq U^x_{F_n}(a,b) \leq \lim_{n\to\infty} U^x_{F_n}(a,b).$$

Taking supremum over all $F \subset_f \mathbb{D}\cap[0,T]$ in this estimate then shows $U^x_T(a,b) \leq \lim_{n\to\infty} U^x_{F_n}(a,b)$. ■

Remark 28.20. It is easy to see that if $\mathbb{T} = \mathbb{R}_+$, x_t is right continuous, and $a < \alpha < \beta < b$, then

$$U^x_T(a,b) \leq \sup\{U^x_F(\alpha,\beta) : F \subset_f \mathbb{D}\cap[0,T]\}.$$

Lemma 28.21. *Let $T \in \mathbb{R}_+$ and $\{x_t\}_{t\in\mathbb{D}}$ be a real valued function such that $U^x_T(a,b) < \infty$ for all $-\infty < a < b < \infty$ with $a, b \in \mathbb{Q}$. Then*

$$x_{t-} := \lim_{\mathbb{D}\ni s\uparrow t} x_s \text{ exists in } \bar{\mathbb{R}} \text{ for } t \in (0,T] \text{ and} \tag{28.19}$$

$$x_{t+} := \lim_{\mathbb{D}\ni s\downarrow t} x_s \text{ exists in } \bar{\mathbb{R}} \text{ for } t \in [0,T). \tag{28.20}$$

Moreover, if we let $U^x_\infty(a,b) = \lim_{T\uparrow\infty} U^x_T(a,b)$ and further assume that $U^x_\infty(a,b) < \infty$ for all $-\infty < a < b < \infty$ with $a, b \in \mathbb{Q}$, then $x_\infty := \lim_{t\uparrow\infty} x_t$ exists in $\bar{\mathbb{R}}$ as well.

Proof. I will only prove the statement in Eq. (28.19) since all of the others are similar. If x_{t-} does not exists in $\bar{\mathbb{R}}$ then we can find $a, b \in \mathbb{Q}$ such that

$$\liminf_{\mathbb{D}\ni s\uparrow t} x_s < a < b < \limsup_{\mathbb{D}\ni s\uparrow t} x_s.$$

From the definition of the lim inf and the lim sup, it follows that for every $\varepsilon \in (0,t)$ there are infinitely many $s \in (t-\varepsilon, t)$ such that $x_s < a$ and infinitely may $s \in (t-\varepsilon, t)$ such that $x_s > b$. From this observation it is easy to see that $\infty = U^x_t(a,b) \leq U^x_T(a,b)$. ■

Lemma 28.22. *Suppose that $\mathbb{T} = \mathbb{D}$, S is a metric space, and $\{x_t \in S\}_{t\in\mathbb{D}}$.*

1. *If for all $t \in \mathbb{R}_+$,*
$$x^+_t := x_{t+} = \lim_{\mathbb{D}\ni s\downarrow t} x_s \text{ exists in } S,$$
then $\mathbb{R}_+ \ni t \to x^+_t \in S$ is right continuous.
2. *If we further assume that*
$$x_{t-} := \lim_{\mathbb{D}\ni s\uparrow t} x_s \text{ exists in } S$$
for all $t > 0$, then $\lim_{\tau\uparrow t} x_{\tau+} = x_{t-}$ for all $t > 0$.

3. Moreover, if $\lim_{\mathbb{D}\ni t\uparrow\infty} x_t$ exists in S then again $\lim_{t\uparrow\infty} x_{t+} = \lim_{\mathbb{D}\ni t\uparrow\infty} x_t$.

Proof. 1. Suppose $t \in \mathbb{R}_+$ and $\varepsilon > 0$ is given. By assumption, there exists $\delta > 0$ such that for $s \in (t, t+\delta) \cap \mathbb{D}$, we have $\rho(x_{t+}, x_s) \le \varepsilon$. Therefore if $\tau \in (t, t+\delta)$, then

$$\rho(x_{t+}, x_{\tau+}) = \lim_{\mathbb{D}\ni s\downarrow\tau} \rho(x_{t+}, x_s) \le \varepsilon$$

from which it follows that $x_{\tau+} \to x_{t+}$ as $\tau \downarrow t$.

2. Now suppose $t > 0$ such that x_{t-} exists in S. Then for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $\rho(x_{t-}, x_s) \le \varepsilon$ if $s \in (t-\delta, t) \cap \mathbb{D}$. Hence, if $\tau \in (t-\delta, t)$ we may conclude,

$$\rho(x_{t-}, x_{\tau+}) = \lim_{s\downarrow\tau} \rho(x_{t+}, x_s) \le \varepsilon$$

from which it follows that $x_{\tau+} \uparrow x_{t-}$ as $\tau \uparrow t$.

3. Now suppose $x_\infty := \lim_{\mathbb{D}\ni s\uparrow\infty} x_s$ exists in S. Then for every $\varepsilon > 0$, there exists $M = M(\varepsilon) < \infty$ such that $\rho(x_\infty, x_s) \le \varepsilon$ if $s \in \mathbb{D} \cap (M, \infty)$. Hence if $t \in (M, \infty)$ we have

$$\rho(x_\infty, x_{t+}) = \lim_{\mathbb{D}\ni s\downarrow t} \rho(x_\infty, x_s) \le \varepsilon$$

from which we conclude that, $\lim_{t\uparrow\infty} x_{t+}$ exists in S and is equal to x_∞. ■

Theorem 28.23 (Doob's upcrossing inequality). *Let $\{X_t\}_{t\in\mathbb{D}}$ be a submartingale and $-\infty < a < b < \infty$. Then for all $T \in \mathbb{D}$,*

$$\mathbb{E}\left[U_T^X(a,b)\right] \le \frac{1}{b-a}\left[\mathbb{E}(X_T - a)_+ - \mathbb{E}(X_0 - a)_+\right]. \tag{28.21}$$

Proof. Let $\{F_n\}_{n=1}^\infty$ be a sequence as in Lemma 28.19 and assume without loss of generality that $0, T \in F_n$ for all n. It then follows From Theorem 18.51 that

$$\mathbb{E}\left[U_{F_n}^X(a,b)\right] \le \frac{1}{b-a}\left[\mathbb{E}(X_T - a)_+ - \mathbb{E}(X_0 - a)_+\right] \ \forall\, n \in \mathbb{N}.$$

By letting $n \uparrow \infty$, Eq. (28.21) follows from this inequality, Lemma 28.19, and the MCT. ■

Theorem 28.24. *Let $\{X_t\}_{t\in\mathbb{D}}$ be a submartingale,*

$$\Omega_0 := \bigcap_{T\in\mathbb{N}} \left(\left\{ \sup_{\mathbb{D}\ni t\le T} |X_t| < \infty \right\} \cap \left[\bigcap \{U_T^X(a,b) < \infty : \ a < b \text{ with } a, b \in \mathbb{Q}\} \right] \right), \tag{28.22}$$

for all $t \in \mathbb{R}_+$,

$$Y_t := \limsup_{\mathbb{D}\ni s\downarrow t} X_s \text{ and } \bar{X}_t := Y_t \cdot 1_{|Y_t|<\infty}. \tag{28.23}$$

Then;

1. *$P(\Omega_0) = 1$.*
2. *on Ω_0, $\sup_{t\le T} |X_t| < \infty$ and X_{t+} and X_{t-} exist for all $t \in \mathbb{R}_+$ where by convention $X_{0-} := X_0$.*
3. *$\{X_{t+}(\omega)\}_{t\in\mathbb{R}_+}$ is right continuous with left hand limits for all $\omega \in \Omega_0$.*
4. *For any $t \in \mathbb{R}_+$ and any sequence $\{s_n\}_{n=1}^\infty \subset \mathbb{D} \cap (t, \infty)$ such that $s_n \downarrow t$, then $X_{s_n} \to X_{t+}$ in $L^1(P)$ as $n \to \infty$.*
5. *The process $\{\bar{X}_t\}_{t\in\mathbb{R}_+}$ is a $\{\mathcal{B}_t^+\}_{t\ge 0}$ submartingale such that $t \to \bar{X}_t$ is right continuous and has left limits on Ω_0.*
6. *$X_t \le \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s. for all $t \in \mathbb{D}$ with equality at some $t \in \mathbb{D}$ iff $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}X_s = \mathbb{E}X_t$.*
7. *If $X_s \xrightarrow{P} X_t$ as $\mathbb{D} \ni s \downarrow t$ at some $t \in \mathbb{D}$,[3] then $\bar{X}_t = X_t$ a.s.*
8. *If $C := \sup_{t\in\mathbb{D}} \mathbb{E}|X_t| < \infty$ (or equivalently $\sup_{t\in\mathbb{D}} \mathbb{E}X_t^+ < \infty$), then $X_\infty := \lim_{\mathbb{D}\ni t\uparrow\infty} \bar{X}_t = \lim_{\mathbb{D}\ni t\uparrow\infty} X_t$ exists in $\mathbb{R}$ a.s. and $\mathbb{E}|X_\infty| < C < \infty$.*
   ***Note:** if $\{X_t^+\}_{t\in\mathbb{D}}$ is uniformly integrable then $\sup_{t\in\mathbb{D}} \mathbb{E}\left|X_t^+\right| < \infty$.*
9. *If $\{X_t^+\}_{t\in\mathbb{D}}$ is uniformly integrable iff there exists $X_\infty \in L^1(\Omega, \mathcal{B}, P)$ such that $\{X_t\}_{t\in\mathbb{D}\cup\{\infty\}}$ is a submartingale. In other words, $\{X_t^+\}_{t\in\mathbb{D}}$ is uniformly integrable iff here exists $X_\infty \in L^1(\Omega, \mathcal{B}, P)$ such that $X_t \le \mathbb{E}[X_\infty|\mathcal{B}_t]$ a.s. for all $t \in \mathbb{D}$.*

Proof. 1. – 3. The fact that $P(\Omega_0) = 1$ follows from Doob's upcrossing inequality and the maximal inequality in Eq. (28.11). The assertions in items 2. and 3. are now a consequence of the definition of Ω_0 and Lemmas 28.21 and 28.22.

4. Let $Y_n := X_{s_{-n}}$ and $\mathcal{F}_n := \mathcal{B}_{s_{-n}}$ for $-n \in \mathbb{N}$. Then $\{(Y_n, \mathcal{F}_n)\}_{n\in-\mathbb{N}}$ is a backwards submartingale such that $\inf \mathbb{E}Y_n \ge \mathbb{E}X_t$ and hence by Theorem 18.75, $Y_n = X_{s_{-n}} \to X_{t+}$ in $L^1(P)$ as $n \to -\infty$.

5. Since $\bar{X}_t = X_{t+}$ on Ω_0 and X_{t+} is right continuous with left hand limits, $\bar{X}$ has these properties on Ω_0 as well. Now let $0 \le s < t < \infty$, $\{s_n\}, \{t_n\} \subset \mathbb{D}$, such that $s_n \downarrow s$, $t_n \downarrow t$ with $s_n < t$ for all n. Then by item 4. and the submartingale property of X,

$$\mathbb{E}\left[\bar{X}_t - \bar{X}_s : A\right] = \mathbb{E}\left[X_{t+} - X_{s+} : A\right] = \lim_{n\to\infty} \mathbb{E}\left[X_{t_n} - X_{s_n} : A\right] \ge 0$$

for all and $A \in \mathcal{B}_{s+}$.

6. Let $A \in \mathcal{B}_t$ and $\{t_n\} \subset \mathbb{D}$ with $t_n \downarrow t \in \mathbb{D}$, then

$$\mathbb{E}\left[\bar{X}_t : A\right] = \lim_{n\to\infty} \mathbb{E}\left[X_{t_n} : A\right] \ge \lim_{n\to\infty} \mathbb{E}\left[X_t : A\right].$$

Since $A \in \mathcal{B}_t$ is arbitrary it follows that $X_t \le \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s. If equality holds, then, taking $A = \Omega$ above, we find

[3] For example, this will hold if $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}|X_t - X_s| = 0$.

$$\mathbb{E}X_t = \mathbb{E}\bar{X}_t = \lim_{n\to\infty} \mathbb{E}\left[X_{t_n}\right].$$

Since $\{t_n\} \subset \mathbb{D}$ with $t_n \downarrow t$ was arbitrary, we may conclude that $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}X_s = \mathbb{E}X_t$. Conversely if $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}X_s = \mathbb{E}X_t$, then along any sequence, $\{s_n\} \subset \mathbb{D}$ with $s_n \downarrow s$, we have

$$\mathbb{E}X_t = \lim_{n\to\infty} \mathbb{E}X_{s_n} = \mathbb{E} \lim_{n\to\infty} X_{s_n} = \mathbb{E}\bar{X}_t = \mathbb{E}\mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right].$$

As $X_t \le \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s. this identity implies $X_t = \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s.

7. Let $t_n \in \mathbb{D}$ such that $t_n \downarrow t$, then as we have already seen $X_{t_n} \to \bar{X}_t$ in $L^1(P)$. However by assumption, $X_{t_n} \xrightarrow{P} X_t$, and therefore we must have $\bar{X}_t = X_t$ a.s. since limits in probability are unique up to null sets.

The proof or items 8. and 9. will closely mimic their discrete versions given in Corollary 18.54.

8. The proof here mimics closely the discrete version given in Corollary 18.54. For any $-\infty < a < b < \infty$, Doob's upcrossing inequality (Theorem 28.23) and the MCT implies,

$$\begin{aligned}\mathbb{E}\left[U^X_\infty(a,b)\right] &= \lim_{\mathbb{D}\ni T\to\infty} \mathbb{E}\left[U^X_T(a,b)\right]\\ &\le \frac{1}{b-a}\left[\sup_{T\in\mathbb{D}} \mathbb{E}\left(X_T - a\right)_+ - \mathbb{E}\left(X_0 - a\right)_+\right] < \infty\end{aligned}$$

where

$$U^X_\infty(a,b) = \lim_{\mathbb{D}\ni T\to\infty} U^X_T(a,b)$$

is the total number of upcrossings of X across $[a,b]$. In particular it follows that

$$\tilde{\Omega}_0 := \cap\left\{U^X_\infty(a,b) < \infty : a,b \in \mathbb{Q} \text{ with } a < b\right\}$$

has probability one. Hence by Lemma 28.21, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{\mathbb{D}\ni t\to\infty} X_t(\omega)$ exists in $\bar{\mathbb{R}}$. By Fatou's lemma with $\mathbb{D} \ni t_n \uparrow \infty$, it follows that

$$\mathbb{E}\left[|X_\infty|\right] = \mathbb{E}\left[\liminf_{n\to\infty} |X_n|\right] \le \liminf_{n\to\infty} \mathbb{E}\left[|X_n|\right] \le C < \infty$$

and therefore that $X_\infty \in \mathbb{R}$ a.s.

9. If $\left\{X_t^+\right\}_{t\ge0}$ is uniformly integrable, then, by Vitalli's convergence Theorem 12.44 and the fact that $X_t^+ \to X_\infty^+$ a.s. (as we have already shown), $X_t^+ \to X_\infty^+$ in $L^1(P)$. Therefore for $A \in \mathcal{B}_t$ we have, by Fatou's lemma, that

$$\begin{aligned}\mathbb{E}\left[X_t 1_A\right] &\le \limsup_{\mathbb{D}\ni s\to\infty} \mathbb{E}\left[X_s 1_A\right] = \limsup_{\mathbb{D}\ni s\to\infty} \left(\mathbb{E}\left[X_s^+ 1_A\right] - \mathbb{E}\left[X_s^- 1_A\right]\right)\\ &= \mathbb{E}\left[X_\infty^+ 1_A\right] - \liminf_{\mathbb{D}\ni s\to\infty} \mathbb{E}\left[X_s^- 1_A\right] \le \mathbb{E}\left[X_\infty^+ 1_A\right] - \mathbb{E}\left[\liminf_{\mathbb{D}\ni s\to\infty} X_s^- 1_A\right]\\ &= \mathbb{E}\left[X_\infty^+ 1_A\right] - \mathbb{E}\left[X_\infty^- 1_A\right] = \mathbb{E}\left[X_\infty 1_A\right].\end{aligned}$$

Since $A \in \mathcal{B}_t$ was arbitrary we may conclude that $X_t \le \mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$ a.s. for all $t \in \mathbb{R}_+$.

Conversely if we suppose that $X_t \le \mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$ a.s. for all $t \in \mathbb{R}_+$, then by Jensen's inequality, $X_t^+ \le \mathbb{E}\left[X_\infty^+|\mathcal{B}_t\right]$ and therefore $\left\{X_t^+\right\}_{t\ge0}$ is uniformly integrable by Proposition 18.8 and Exercise 12.5. ■

Example 28.25. In this example we show that there exists a right continuous submartingale, $\{X_t\}_{t\ge0}$, such that $\{X_{s_n}\}_{n=1}^\infty$ is **not** uniformly integrable for some bounded increasing sequence $\{s_n\}_{n=1}^\infty$. Indeed, let

$$X_t := -M^0_{\tan\left(\frac{\pi}{2}t\wedge1\right)},$$

where $\{M_t\}_{t\ge0}$ is the martingale constructed in Exercise 28.1. Then it is easily checked that $\{X_t\}_{t\ge0}$ is a $\left\{\mathcal{B}_{\tan\left(\frac{\pi}{2}t\wedge1\right)}\right\}_{t\ge0}$ – submartingale. Moreover if $s_n \in [0,1)$ with $s_n \uparrow 1$, the collection, $\{X_{s_n}\}_{n=1}^\infty$ is not uniformly integrable for if it were we would have

$$-1 = \lim_{n\to\infty} \mathbb{E}X_{s_n} = \mathbb{E}\left[\lim_{n\to\infty} X_{s_n}\right] = \mathbb{E}[0] = 0.$$

In particular this shows that in item 4. of Theorem 28.24, we can **not** suppose $\{s_n\}_{n=1}^\infty \subset \mathbb{D}\cap[0,t)$ with $s_n \uparrow t$.

Exercise 28.2. If $\{X_t\}_{t\ge0}$ is a right continuous submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\ge0}, P\right)$, then $s \to \mathbb{E}X_s$ is right continuous at t and X is a $\left\{\mathcal{B}_t^+\right\}$ – submartingale.

Solution to Exercise (28.2). Let $\{t_n\}_{n=1}^\infty$ be any decreasing sequence in (t,∞) such that $t_n \downarrow t$ as $n\to\infty$. By Lemma 28.4, we know $Y := \lim_{n\to\infty} X_{t_n}$ exists a.s and in $L^1(P)$. As $X_{t_n} \to X_t$ as $n\to\infty$, it follows that $X_{t_n} \to X_t$ in $L^1(P)$ and therefore, $\lim_{n\to\infty} \mathbb{E}X_{t_n} = \mathbb{E}X_t$. Since $\{t_n\}_{n=1}^\infty \subset (t,\infty)$ with $t_n \downarrow t$ was an arbitrary sequence, we may conclude that $t \to \mathbb{E}X_t$ is right continuous. Similarly, if $0 \le s < t < \infty$, $s_n \in (s,t)$ with $s_n \downarrow s$, then for any $B \in \mathcal{B}_{s+}$ we have

$$\mathbb{E}\left[X_t - X_s : B\right] = \lim_{n\to\infty} \mathbb{E}\left[X_t - X_{s_n} : B\right] \ge 0.$$

Exercise 28.3. Let $\{X_t\}_{t\ge0}$ be a submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\ge0}, P\right)$, $t \in \mathbb{R}_+$, and $X_s \xrightarrow{P} X_t$ as $s \downarrow t$, then $s \to \mathbb{E}X_s$ is right continuous at t.

Solution to Exercise (28.3). Let $\{t_n\}_{n=1}^\infty$ be a decreasing sequence in (t,∞) such that $t_n \downarrow t$ as $n\to\infty$. By Lemma 28.4, we know $Y := \lim_{n\to\infty} X_{t_n}$ exists a.s and in $L^1(P)$. As $X_{t_n} \xrightarrow{P} X_t$ as $n\to\infty$, it follows that $X_{t_n} \to X_t$ in $L^1(P)$

and a.s. Therefore, $\lim_{n\to\infty}\mathbb{E}X_{t_n}=\mathbb{E}X_t$. Since $\{t_n\}_{n=1}^{\infty}\subset(t,\infty)$ with $t_n\downarrow t$ was an arbitrary sequence, we may conclude that $s\to\mathbb{E}X_s$ is right continuous at t.

Theorem 28.26 (Regularizing Submartingales). *Let $\{X_t\}_{t\geq 0}$ be a submartingale on a filtered probability space, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$, and let Ω_0 and $\bar{X}_t$ be as in Theorem 28.24 applied to $\{X_t\}_{t\in\mathbb{D}}$. Further let*

$$\hat{X}_t(\omega):=\begin{cases}\bar{X}_t(\omega) & \text{if } \omega\in\Omega_0\\ 0 & \text{if } \omega\notin\Omega_0,\end{cases}$$

and $\left\{\bar{\mathcal{B}}_{t+}:=\mathcal{B}_{t+}\vee\mathcal{N}\right\}_{t\geq 0}$ where $\mathcal{N}$ is the collection of P – null subsets of Ω in $\mathcal{B}$. Then:

1. *$\left\{\hat{X}_t\right\}_{t\geq 0}$ is a $\left\{\bar{\mathcal{B}}_{t+}\right\}_{t\geq 0}$ – submartingale which is right continuous with left hand limits.*
2. *$\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right]\geq X_t$ a.s. for all $t\in\mathbb{R}_+$ with equality holding for all $t\in\mathbb{R}_+$ iff $t\to\mathbb{E}X_t$ is right continuous.*
3. *If $\{\mathcal{B}_t\}_{t\geq 0}$ is right continuous, then $\hat{X}_t\geq X_t$ a.s. for all $t\in\mathbb{R}_+$ with equality holding for all $t\in\mathbb{R}_+$ iff $t\to\mathbb{E}X_t$ is right continuous.*

In particular if $\{X_t\}_{t\geq 0}$ is right continuous in probability, then $\{X_t\}_{t\geq 0}$ has a right continuous modification possessing left hand limits, $\left\{\hat{X}_t\right\}_{t\geq 0}$, such that $\left\{\hat{X}_t\right\}_{t\geq 0}$ is a $\left\{\bar{\mathcal{B}}_{t+}\right\}_{t\geq 0}$ – submartingale.

Proof. 1. Since $\Omega\setminus\Omega_0\in\mathcal{N}$ and $\bar{X}_t$ is a $\mathcal{B}_{t+}$ – measurable, it follows that $\hat{X}_t$ is $\bar{\mathcal{B}}_{t+}$ – measurable. Hence $\left\{\hat{X}_t\right\}_{t\geq 0}$ is an adapted process. Since $\hat{X}$ is a modification of $\bar{X}$ which is already a $\left\{\bar{\mathcal{B}}_{t+}\right\}_{t\geq 0}$ – submartingale (see Lemma 28.15) it follows that $\hat{X}$ is also a $\left\{\bar{\mathcal{B}}_{t+}\right\}_{t\geq 0}$ – submartingale.

2. Since $\hat{X}_t=\bar{X}_t$, we may replace $\hat{X}$ by $\bar{X}$ in the statement 2. We now need only follow the proof of item 6. in Theorem 28.26. Indeed, if $t\in\mathbb{R}_+$, $A\in\mathcal{B}_t$, and $\{t_n\}\subset\mathbb{D}$ with $t_n\downarrow t\in\mathbb{D}$, then

$$\mathbb{E}\left[\bar{X}_t:A\right]=\lim_{n\to\infty}\mathbb{E}\left[X_{t_n}:A\right]\geq\lim_{n\to\infty}\mathbb{E}\left[X_t:A\right].$$

Since $A\in\mathcal{B}_t$ was arbitrary, it follows that

$$X_t\leq\mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]=\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right]\text{ a.s.}\tag{28.24}$$

If equality holds in Eq. (28.24), $\mathbb{E}X_t=\mathbb{E}\hat{X}_t$ which is right continuous by Exercise 28.2. Conversely if $t\to\mathbb{E}X_t$ is right continuous, it follows from item 6. of Theorem 28.26 that $\hat{X}_t=X_t$ a.s.

3. Since $\hat{X}_t=\bar{X}_t$ a.s. and $\bar{X}_t$ is $\mathcal{B}_{t+}=\mathcal{B}_t$ – measurable, we find

$$\hat{X}_t=\bar{X}_t=\mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]=\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right]\text{ a.s.}$$

With this observation (i.e. $\hat{X}_t=\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right]$ a.s.) the assertions in item 3. follow directly from those in item 2.

Now suppose that $\{X_t\}_{t\geq 0}$ is right continuous in probability. By Proposition 28.16 and Lemma 28.15, $\{X_t\}_{t\geq 0}$ is a $\left\{\bar{\mathcal{B}}_{t+}\right\}_{t\geq 0}$ – submartingale such that, by Exercise 28.3, $t\to\mathbb{E}X_t$ is right continuous. Therefore, by items 1. and 3. of the theorem, $\hat{X}$ defined above is the desired modification of X. ■

Example 28.27. Let be a Poisson process, $\{N_t\}_{t\geq 0}$, with parameter λ as described in Example 25.8. Since $\{N\}_{t\geq 0}$ has independent increments, it follows that $\{N_t\}_{t\geq 0}$ is a $\{\mathcal{B}_t:=\sigma(N_s:s\leq t)\}_{t\geq 0}$ – martingale. By example 22.14) we know that $\mathbb{E}|N_t-N_s|=\lambda|t-s|$ and in particular, $t\to N_t$ is continuous in probability. Hence it follows from Theorem 28.26 that there is a modification, $\hat{N}$ and N such that $\left\{\hat{N}_t\right\}_{t\geq 0}$ is a $\left\{\bar{\mathcal{B}}_{t+}\right\}_{t\geq 0}$ – martingale which has continuous sample paths possessing left hand limits.

The ideas of this example significantly generalize to produce good modifications of large classes of Markov processes, see for example [7, Theorem I.9.4 on p. 46], [20] and [35]. See [42, Chapter I] where this is carried out in the context of Lévy processes. We end this section with another version of the optional sampling theorem.

Theorem 28.28 (Optional sampling II). *Suppose $\{X_t\}_{t\geq 0}$ is a right continuous submartingale on a filtered probability space, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ such that $\{X_t^+\}_{t\geq 0}$ is uniformly integrable. Then for any two optional times, σ and τ, $X_\tau\in L^1(P)$ and*

$$X_{\sigma\wedge\tau}\leq\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma^+\right].\tag{28.25}$$

In particular if $\{M_t\}_{t\geq 0}$ is a right continuous uniformly integrable martingale, then

$$M_{\sigma\wedge\tau}=\mathbb{E}\left[M_\tau|\mathcal{B}_\sigma^+\right].\tag{28.26}$$

Proof. Let $X_\infty:=\lim_{\mathbb{D}\ni t\uparrow\infty}X_t\in L^1(P)$ as in Theorem 28.24 so that $X_t\leq\mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$ for all $t\in\mathbb{D}$. For $t\in\mathbb{R}_+$, let $\{t_n\}_{n=1}^{\infty}\subset\mathbb{D}\cap(t,\infty)$ be such that $t_n\downarrow t$, then by Corollary 18.77,

$$X_t=\lim_{n\to\infty}X_{t_n}\leq\lim_{n\to\infty}\mathbb{E}\left[X_\infty|\mathcal{B}_{t_n}\right]=\mathbb{E}\left[X_\infty|\mathcal{B}_t^+\right]\text{ a.s.}$$

Conditioning this inequality on $\mathcal{B}_t$ also allows us to conclude that $X_t \leq \mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$.[4] We may now reduce the inequality in Eq. (28.25) to the case in Theorem 28.9 where τ is a bounded stopping time by simply identifying $[0,\pi/2]$ with $[0,\infty]$ via the map, $t \to \tan t$. More precisely, let $\{Y_t\}_{0\leq t\leq \pi/2}$ be the right continuous $\left\{\tilde{\mathcal{B}}_t := \mathcal{B}_{\tan t}\right\}_{0\leq t\leq \pi/2}$ – submartingale defined by $Y_t := X_{\tan t}$. As $\tan^{-1}(\sigma)$ and $\tan^{-1}(\tau)$ are two bounded $\left\{\tilde{\mathcal{B}}_t\right\}_{0\leq t\leq \pi/2}$ – optional times, we may apply Theorem 28.9 to find;

$$X_{\sigma\wedge\tau} = Y_{\tan^{-1}(\sigma)\wedge\tan^{-1}(\tau)} \leq \mathbb{E}\left[Y_{\tan^{-1}(\tau)}|\tilde{\mathcal{B}}^+_{\tan^{-1}(\sigma)}\right] = \mathbb{E}\left[X_\tau|\mathcal{B}^+_\sigma\right] \text{ a.s.}$$

For the martingale assertions, simply apply Eq. (28.25) with $X_t = M_t$ and $X_t = -M_t$. ■

[4] According to Exercise 28.2, $\{X_t\}_{t\geq 0}$ is also a $\left\{\mathcal{B}^+_t\right\}$ – submartingale. Therefore for the purposes of this Theorem, there is no loss in generality in assuming that $\mathcal{B}^+_t = \mathcal{B}_t$.

Part VI

Markov Processes II

29

The Strong Markov Property

The main theme of this part of the book is that many Markov processes (including all the ones we have seen) satisfy a stronger version of the Markov property than we proved in Theorems 17.4 and 17.14. Our first goal is to describe these stronger Markov properties which involve stopping times. We will then show in different classes of examples how this strong Markov property implies a number of interesting properties of the corresponding Markov process.

Our immediate goal is to extend the Markov property in Theorem 17.14 to allow for restarting the processes at random times. As in Chapter 17 we will assume that $T = \mathbb{N}_0$ or $\mathbb{R}_+$. We will start with a simple extension of the Markov property associated to a stopping time with countable range. We will then use a limiting procedure in order to handle general optional times.

To keep the setup relatively general but not overly general, in this part of the book we will assume that (S, ρ) is a complete separable metric space and let $\mathcal{S}$ denote the Borel σ – algebra on S. Further suppose that $\{Q_t : S \times \mathcal{S} \to [0,1]\}_{t \in \mathbb{N}_0}$ are time homogeneous Markov transition kernels. As usual we let S^T denote the set of all functions, $\gamma : T \to S$ and $\pi_t : S^T \to S$ will denote the projection, $\pi_t(\gamma) = \gamma(t)$.

Definition 29.1 (RC). *Let $RC(T,S)$ denote those $\gamma \in S^T$ which are right continuous, i.e. $\gamma(t) = \gamma(t+) := \lim_{s \downarrow t} \gamma(s)$ for all $t \in T$. Similarly we let $C(T,S)$ denote those $\gamma \in S^T$ which are continuous.*

Notation 29.2 (Path Spaces) *Let Γ denote either of the three path spaces, S^T, $RC(T,S)$ or $C(T,S)$ and write in all cases, π_t for $\pi_t|_\Gamma$ and $\mathcal{F}_t := \sigma(\pi_s : s \le t)$, $\mathcal{F}_t^+ := \mathcal{F}_{t+}$, and $\mathcal{F} := \vee_{t \ge 0} \mathcal{F}_t = \sigma(\pi_s : 0 \le s < \infty)$.*

Assumption 3 *To each $x \in S$ we will assume there exists a probability measure, P_x, on $(\Gamma, \mathcal{F})$ such that $\left(\Gamma, \mathcal{F}, \{\mathcal{F}_t\}_{t \ge 0}, \{\pi_t\}_{t \ge 0}, P_x\right)$ is a time homogeneous Markov process with transition kernels $\{Q_t\}_{t \ge 0}$ such that $P_x(\{\pi_0 = x\}) = 1$.*

From Theorem 17.11, we know that Assumption 3 automatically holds when $\Gamma = S^T$. On the other hand if Γ is equal to $RC(T,S)$ or $C(T,S)$ and $T = \mathbb{R}_+$, the existence of the Markov – measures $\{P_x\}_{x \in S}$ requires additional restriction on the Markov – semi-group, $\{Q_t\}_{t \in T}$. Our starting point for this chapter is the following minor variant of Theorem 17.14.

Theorem 29.3 (The time homogeneous Markov property). *Suppose that $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \ge 0}, P\right)$ is a filtered probability space and $\{X_t : \Omega \to S\}_{t \ge 0}$ are adapted functions such that $\{X_t\}_{t \ge 0}$ is a time homogeneous Markov process with transition kernels $\{Q_t\}_{t \in T}$. If $F : \Gamma \to \mathbb{R}$ is bounded and $\mathcal{F}/\mathcal{B}_\mathbb{R}$ – measurable and $t \in T$, then $S \ni x \to \mathbb{E}_x F$ is $\mathcal{S}/\mathcal{B}_\mathbb{R}$ – measurable and*

$$\mathbb{E}_P\left[F(X_{t+\cdot}) | \mathcal{B}_t\right] = \mathbb{E}_P\left[F(X_{t+\cdot}) | X_t\right] = \mathbb{E}_{X_t}[F] \quad P\text{ – a.s.} \tag{29.1}$$

We omit the proof since it is virtually identical to the proof of Theorem 17.14.

29.1 The denumerable strong Markov property

In this section we are going to prove a strong form of the Markov property which is useful for discrete time Markov chains. It will be the stepping stone to the more general strong Markov property in the continuous time case considered in Theorem 29.6.

Theorem 29.4 (The denumerable strong Markov property). *Suppose again that $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \ge 0}, \{X_t\}_{t \ge 0}, P\right)$ is a time homogeneous Markov process with transition kernels $\{Q_t\}_{t \in T}$ and $\tau : \Omega \to [0, \infty]$ is a $\{\mathcal{B}_t\}_{t \ge 0}$ – stopping time with countable range, i.e. $\tau(\Omega)$ is finite or countable. If $F : \Gamma \to \mathbb{R}$ is bounded and $\mathcal{F}/\mathcal{B}_\mathbb{R}$ – measurable, then*

$$\mathbb{E}\left[F(X_{\tau+\cdot}) | \mathcal{B}_\tau\right] = \mathbb{E}_{X_\tau} F, \ P\text{ – a.s. on } \{\tau < \infty\}. \tag{29.2}$$

Proof. Let $T_0 := \tau(\Omega) \setminus \{\infty\} = \tau(\Omega) \cap \mathbb{R}_+$. If $B \in \mathcal{S}$ we have

$$\{X_\tau \in B\} \cap \{\tau = t\} = \{X_t \in B\} \cap \{\tau = t\} \in \mathcal{B}_t$$

for all $t \in T_0$ from which it follows that X_τ is $\mathcal{B}_\tau/\mathcal{S}$ – measurable on $\{\tau < \infty\}$ by item 7. of Lemma 27.16. Since $x \to \mathbb{E}_x F$ is $\mathcal{S}/\mathcal{B}_\mathbb{R}$ – measurable by Theorem 29.3 we may now conclude that $1_{\{\tau < \infty\}} \mathbb{E}_{X_\tau} F$ is $\mathcal{B}_\tau/\mathcal{B}_\mathbb{R}$ – measurable.

By Theorem 29.3, if $A \in \mathcal{B}_\tau$ we find,

$$\begin{aligned}
&\mathbb{E}\left[F\left(X_{\tau+\cdot}\right)\cdot 1_{\tau<\infty}\cdot 1_A\right] \\
&\quad=\mathbb{E}\left[\sum_{t\in T_0}F\left(X_{t+\cdot}\right)\cdot 1_{A\cap\{\tau=t\}}\right]=\sum_{t\in T_0}\mathbb{E}\left[F\left(X_{t+\cdot}\right)\cdot 1_{A\cap\{\tau=t\}}\right] \\
&\quad=\sum_{t\in T_0}\mathbb{E}\left[\mathbb{E}_{X_t}F\cdot 1_{A\cap\{\tau=t\}}\right] \\
&\quad=\mathbb{E}\left[\sum_{t\in T_0}1_{\tau=t}\left[\mathbb{E}_{X_t}F\right]\cdot 1_A\right]=\mathbb{E}\left[\left[\mathbb{E}_{X_\tau}F\right]\cdot 1_A 1_{\tau<\infty}\right],
\end{aligned}$$

where in we have used Lemma 27.16 again to see that $A\cap\{\tau=t\}\in\mathcal{B}_t$ for all t. Since $A\in\mathcal{B}_\tau$ is arbitrary and $1_{\tau<\infty}\mathbb{E}_{X_\tau}F$ is $\mathcal{B}_\tau$ – measurable it follows that

$$\mathbb{E}\left[F\left(X_{\tau+\cdot}\right)\cdot 1_{\tau<\infty}|\mathcal{B}_\tau\right]=1_{\tau<\infty}\mathbb{E}_{X_\tau}F,\ P-\text{a.s.}$$

which is precisely Eq. (29.2). ∎

Let us write out this theorem more explicitly in the case of a discrete time Markov chain.

Corollary 29.5. *Suppose $\{X_t\}_{t\in\mathbb{N}_0}$ is a discrete time Markov chain in a countable or finite state space S. If τ is a stopping time and $x\in S$, then, conditioned on $\{\tau<\infty$ and $X_\tau=x\}$, $\mathcal{B}_\tau$ and $\{X_{\tau+t}:t\geq 0\}$ are independent and $\{X_{\tau+t}:t\geq 0\}$ has the same distribution as $\{X_t\}_{t\geq 0}$ under P_x.*

Proof. Let $g:\Omega\to\mathbb{R}$ be a bounded $\mathcal{B}_\tau$ – measurable function and $f:S^{\mathbb{N}_0}\to\mathbb{R}$ be a bounded $\mathcal{S}^{\otimes\mathbb{N}_0}$ – measurable function. Then

$$\begin{aligned}
\mathbb{E}_\nu\left[g\cdot f\left(X_{\tau+\cdot}\right):\tau<\infty\ \&\ X_\tau=x\right]&=\mathbb{E}_\nu\left[\left(1_{X_\tau=x}g\right)\cdot f\left(X_{\tau+\cdot}\right):\tau<\infty\right] \\
&=\mathbb{E}\left[\left(1_{X_\tau=x}g\right)\cdot\mathbb{E}_{X_\tau}\left[f\left(X_\cdot\right)\right]:\tau<\infty\right] \\
&=\mathbb{E}\left[\left(1_{X_\tau=x}g\right)\cdot\mathbb{E}_x\left[f\left(X_\cdot\right)\right]:\tau<\infty\right] \\
&=\mathbb{E}_x\left[f\left(X_\cdot\right)\right]\cdot\mathbb{E}\left[g:\tau<\infty\ \&\ X_\tau=x\right]
\end{aligned}$$

and therefore,

$$\begin{aligned}
\mathbb{E}_\nu\left[g\cdot f\left(X_{\tau+\cdot}\right)|\tau<\infty\ \&\ X_\tau=x\right]&=\frac{\mathbb{E}\left[g:\tau<\infty\ \&\ X_\tau=x\right]}{\mathbb{E}_\nu\left[\tau<\infty\ \&\ X_\tau=x\right]}\cdot\mathbb{E}_x f \\
&=\mathbb{E}_\nu\left(g|\tau<\infty\ \&\ X_\tau=x\right)\cdot\mathbb{E}_x f.
\end{aligned}$$

Taking $g=1$ in this equation then shows,

$$\mathbb{E}_\nu\left[f\left(X_{\tau+\cdot}\right)|\tau<\infty\ \&\ X_\tau=x\right]=\mathbb{E}_x f$$

which combined with the previously displayed equation completes the proof of the corollary. ∎

In Chapter 30 we will see that Corollary 29.5 will allow us to say a fair bit more about the behavior of discrete time Markov chains. In fact you may jump directly to Chapter 30 now if you wish.

29.2 The strong Markov property in continuous time

Let us now assume that $T=\mathbb{R}_+$. We would like to remove the countable range restriction on the stopping time τ appearing in Theorem 29.4. In order to do this we are going to need to impose some continuity conditions on our Markov processes and their transition kernels.

Theorem 29.6 (Strong Markov Property). *Let Γ be either $RC(T,S)$ or $C(T,S)$, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ be a filtered probability space, $\{X_t\}_{t\geq 0}$ be a time homogeneous Markov with Markov transition kernels $\{Q_t\}_{t\geq 0}$ and suppose that τ is a $\{\mathcal{B}_t\}$ – optional time, see Definition 27.8. We further assume that that $X_\cdot\in\Gamma$ (i.e. $(t\to X_t(\omega))\in\Gamma$ for all $\omega\in\Omega$) and assume the existence of a multiplicative system, $\mathbb{M}\subset BC(S)$, such that $\sigma(\mathbb{M})=\mathcal{S}$ and $Q_t\mathbb{M}\subset\mathbb{M}$ for all $t\in T$. Then for any $F:\Gamma\to\mathbb{R}$ which is bounded and $\mathcal{F}$ – measurable, we have*

$$\mathbb{E}\left[F\left(X_{\tau+\cdot}\right)|\mathcal{B}_\tau^+\right]=\mathbb{E}_{X_\tau}[F],\ P\text{ – a.s. on }\{\tau<\infty\}. \tag{29.3}$$

Part of the assertion here is that $1_{\{\tau<\infty\}}\mathbb{E}_{X_\tau}[F]$ is $\mathcal{B}_\tau^+$ – measurable. Moreover if τ is a stopping time, then $1_{\{\tau<\infty\}}\mathbb{E}_{X_\tau}[F]$ is $\mathcal{B}_\tau$ – measurable and

$$\mathbb{E}\left[F\left(X_{\tau+\cdot}\right)|\mathcal{B}_\tau^+\right]=\mathbb{E}\left[F\left(X_{\tau+\cdot}\right)|\mathcal{B}_\tau\right]=\mathbb{E}_{X_\tau}[F],\ P\text{ – a.s. on }\{\tau<\infty\}. \tag{29.4}$$

Proof. The proof will be divided into five steps.

1. By Lemmas 27.6 and 27.14 we know that X_τ is $\mathcal{B}_\tau^+$ – measurable on $\{\tau<\infty\}$ which combined with the measurability of $x\to\mathbb{E}_x F$ (see Theorem 29.3) implies $1_{\{\tau<\infty\}}\mathbb{E}_{X_\tau}[F]$ is $\mathcal{B}_\tau^+$ – measurable. If we further assume that τ is a stopping time, the same reasoning imply that $1_{\{\tau<\infty\}}\mathbb{E}_{X_\tau}[F]$ is now $\mathcal{B}_\tau$ – measurable. With this measurability issue out of the way it follows by the tower property for conditional expectations that Eq. (29.3) implies Eq. (29.4). So we now embark on the proof of Eq. (29.3).
2. Let $\{\tau_n\}_{n=1}^\infty$ be the discrete stopping times defined in Eq. (27.5) of Lemma 27.18. Recall that $\{\tau=\infty\}=\{\tau_n=\infty\}$ and $\tau_n\downarrow\tau$ as $n\to\infty$ and $\mathcal{B}_\tau^+\subset\mathcal{B}_{\tau_n}$ for all n. Therefore if $A\in\mathcal{B}_\tau^+\subset\mathcal{B}_{\tau_n}$, it follows from Theorem 29.4 that

$$\begin{aligned}
\mathbb{E}\left[F\left(X_{\tau_n+\cdot}\right)\cdot 1_{\tau<\infty}\cdot 1_A\right]&=\mathbb{E}\left[F\left(X_{\tau_n+\cdot}\right)\cdot 1_{\tau_n<\infty}\cdot 1_A\right] \\
&=\mathbb{E}\left[1_{\tau_n<\infty}\left[\mathbb{E}_{X_{\tau_n}}F\right]\cdot 1_A\right] \\
&=\mathbb{E}\left[\left[\mathbb{E}_{X_{\tau_n}}F\right]\cdot 1_{\tau<\infty}\cdot 1_A\right].
\end{aligned} \tag{29.5}$$

Our goal now is to let $n\to\infty$ in Eq. (29.5). However, we are not going to be able to take this limit for general $F\in\mathcal{F}_b$ and hence the need for three more steps.

3. For the moment suppose that
$$F = \prod_{i=1}^{n} f_i \circ \pi_{t_i} \tag{29.6}$$
(i.e. $F(\gamma) = \prod_{i=1}^n f_i(\gamma(t_i)))$ for some $f_i \in \mathbb{M}$ and times $0 \le t_1 < t_2 < \cdots < t_n$. For $x \in S$ one shows by induction that
$$\mathbb{E}_x[F] = \left(Q_{t_1} M_{f_1} Q_{t_2 - t_1} M_{f_2} \dots M_{f_{n-1}} Q_{t_n - t_{n-1}} f_n\right)(x) \tag{29.7}$$
where M_f denotes the operation of multiplication by f on $\mathbb{M}$, i.e. $M_f g = fg$ for all $g \in \mathbb{M}$. Since $\mathbb{M}$ is multiplicative and $Q_t \mathbb{M} \subset \mathbb{M}$ for all t it follows from Eq. (29.7) that $(S \ni x \to \mathbb{E}_x F) \in \mathbb{M} \subset C(S)$. The continuity of $\mathbb{E}_{(\cdot)} F$ along with the right continuity of $t \to X_t$ then implies;
$$\mathbb{E}_{X_{\tau_n}} F \to \mathbb{E}_{X_\tau} F \text{ boundedly as } n \to \infty. \tag{29.8}$$
We also have that
$$F(X_{\tau_n + \cdot}) = \prod_{i=1}^{n} f_i(X_{\tau_n + t_i}) \to \prod_{i=1}^{n} f_i(X_{\tau + t_i}) = F(X_{\tau + \cdot}) \text{ as } n \to \infty. \tag{29.9}$$
Therefore for F as in Eq. (29.6) we may use the DCT to pass to the limit in Eq. (29.5) to find;
$$\mathbb{E}[F(X_{\tau+\cdot}) \cdot 1_{\tau<\infty} \cdot 1_A] = \mathbb{E}[[\mathbb{E}_{X_\tau} F] \cdot 1_{\tau<\infty} \cdot 1_A]. \tag{29.10}$$
4. Let $\mathbb{H}$ denote those $F \in \mathcal{F}_b$ such that Eq. (29.10) is valid and $\tilde{\mathbb{M}}$ denote those $F \in \mathcal{F}_b$ which are of the form in Eq. (29.6). We have just proved that $\tilde{\mathbb{M}} \subset \mathbb{H}$. The reader may now easily check that $1 \in \mathbb{H}$, $\mathbb{H}$ is a linear subspace which is closed under bounded convergence, and $\tilde{\mathbb{M}}$ is a multiplicative system. Therefore an application of the multiplicative systems Theorem 8.2 implies that $\mathbb{H}$ contains all $\sigma\left(\tilde{\mathbb{M}}\right)$ – bounded measurable functions.
5. To finish the proof we must now show $\sigma\left(\tilde{\mathbb{M}}\right) = \mathcal{F}$ and for this it suffices to show that π_t is $\sigma\left(\tilde{\mathbb{M}}\right)$ – measurable for all $t \in \mathbb{R}_+$. Since $f \circ \pi_t \in \tilde{\mathbb{M}}$ for all $f \in \mathbb{M}$ it follows that $\mathbb{M} \subset \mathbb{H}$ where $\mathbb{H}$ now denotes those $f \in \mathcal{S}_b$ such that $f \circ \pi_t$ is $\sigma\left(\tilde{\mathbb{M}}\right)$ – measurable. Again the reader may easily check that $1 \in \mathbb{H}$ and $\mathbb{H}$ is a linear subspace which is closed under bounded convergence. Therefore another application of the multiplicative systems Theorem 8.2 shows that $\mathbb{H} = \mathcal{S}_b$. In particular if $B \in \mathcal{S}$ then $1_B \circ \pi_t$ is $\sigma\left(\tilde{\mathbb{M}}\right)$ – measurable and therefore $\pi_t^{-1}(B) = \{1_B \circ \pi_t = 1\} \in \sigma\left(\tilde{\mathbb{M}}\right)$ for all $B \in \mathcal{S}$. This shows that π_t is $\sigma\left(\tilde{\mathbb{M}}\right) / \mathcal{S}$ – measurable for all $t \in \mathbb{R}_+$.

■

Remark 29.7. We can also see that X_τ is $\mathcal{B}_\tau^+ / \mathcal{S}$ – measurable on $\{\tau < \infty\}$ from the fact that $X_\tau = \lim_{n\to\infty} X_{\tau_n}$. In order to do this first observe that if $A \in \cap_{n=1}^\infty \mathcal{B}_{\tau_n}$ and $t > 0$ then
$$A \cap \{\tau < t\} = \cup_{n=1}^\infty A \cap \{\tau_n < t\} \in \mathcal{B}_t$$
from which it follows that $A \in \mathcal{B}_\tau^+$. Therefore for any $x \in S$ we have $\rho(x, X_\tau) = \lim_{n\to\infty} \rho(x, X_{\tau_n})$ on $\{\tau < \infty\}$ where the limit is necessarily $\cap_{n=1}^\infty \mathcal{B}_{\tau_n}$ – measurable on $\{\tau < \infty\}$. Therefore it follows that $\rho(x, X_\tau)$ is $\mathcal{B}_\tau^+$ – measurable on $\{\tau < \infty\}$ for all $x \in S$. Therefore by Exercise 27.1 it follows that X_τ is $\mathcal{B}_\tau^{+s}$ – measurable on $\{\tau < \infty\}$.

29.3 Examples

In this section we will verify that the examples of continuous time Markov processes we have considered so far all verify the hypotheses of Theorem 29.6 and therefore satisfy the strong Markov property.

Example 29.8 (Poisson Process). From the construction given in Theorem 11.15 we know that the Poisson process may be chosen to have right continuous paths in $S = \mathbb{N}_0$. From example 17.20, the conservative Markov semi-group of the Poisson process is given by
$$Q_t f(x) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} f(x+n)$$
for all bounded functions, $f : S \to \mathbb{R}$. We may now take $\mathbb{M} = BC(S)$ – the bounded continuous functions on S. (As continuity is no restriction we really have that $\mathbb{M}$ consists of all bounded functions on S.) Since
$$|Q_t f(x)| \le \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} |f(x+n)| \le \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \|f\|_u = \|f\|_u$$
it follows that $Q_t f$ is bounded whenever f is bounded. Therefore $Q_t \mathbb{M} \subset \mathbb{M}$ for all $t \ge 0$.

More generally the same argument works for any right continuous Markov chain with a discrete state space.

Proposition 29.9. *Suppose that S is a countable or finite state space, $\{Q_t : S \times S \to [0,1]\}_{t\ge0}$ is a time homogeneous Markov – semi-group. (In this case $\mathcal{S} = 2^S$.) If $\mathbb{M}$ denotes the bounded (necessarily continuous) functions on S then $Q_t \mathbb{M} \subset \mathbb{M}$ for all $t \ge 0$.*

Proof. As for all probability kernels we have $\|Q_t f\|_u \le \|f\|_u$ so that $Q_t\mathbb{M} \subset \mathbb{M}$ for all $t \ge 0$. ∎

The next results is the continuous time Markov chain analogue of Corollary 29.5 for discrete time Markov chains.

Corollary 29.10. *Let S and $\{Q_t\}_{t\ge 0}$ be as in Proposition 29.9, $\Gamma = RC(T,S)$, and suppose that $\{X_t\}_{t\ge 0}$ is an associated right continuous Markov chain. If τ is an optional time and $j \in S$, then, then conditioned on $\{\tau < \infty$ and $X_\tau = j\}$ we have that $\mathcal{B}_\tau^+$ and $\{X_{t+\tau} : t \ge 0\}$ are independent and $\{X_{t+\tau} : t \ge 0\}$ has the same distribution as $\{X_t\}_{t\ge 0}$ under P_j.*

Proof. Let g be a bounded $\mathcal{B}_\tau^+$ – measurable function and $f : \Gamma \to \mathbb{R}$ be a bounded $\mathcal{B}$ – measurable function. Then by Theorem 29.6,

$$\begin{aligned}\mathbb{E}\left[g\cdot f\left(X_{\tau+\cdot}\right) : \tau<\infty \ \&\ X_\tau = j\right] &= \mathbb{E}\left(\mathbb{E}\left[f\left(X_{\tau+\cdot}\right)\cdot g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}|\mathcal{B}_\tau^+\right]\right)\\ &= \mathbb{E}\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\cdot\mathbb{E}\left[f\left(X_{\tau+\cdot}\right)|\mathcal{B}_\tau^+\right]\right)\\ &= \mathbb{E}\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\cdot\mathbb{E}_{X_\tau}f\right)\\ &= \mathbb{E}\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\cdot\mathbb{E}_j f\right)\\ &= \mathbb{E}\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\right)\cdot\mathbb{E}_j f\end{aligned}$$

and therefore,

$$\begin{aligned}\mathbb{E}\left[g\cdot f\left(X_{\tau+\cdot}\right)|\tau<\infty \ \&\ X_\tau = j\right] &= \frac{\mathbb{E}\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\right)}{\mathbb{E}\left[\tau<\infty \ \&\ X_\tau=j\right]}\cdot\mathbb{E}_j f\\ &= \mathbb{E}\left(g|\tau<\infty \ \&\ X_\tau=j\right)\cdot\mathbb{E}_j f.\end{aligned}$$

Taking $g = 1$ in this equation then shows,

$$\mathbb{E}\left[f\left(X_{\tau+\cdot}\right)|\tau<\infty \ \&\ X_\tau = j\right] = \mathbb{E}_j f$$

which combined with the previously displayed equation completes the proof of the corollary. ∎

Example 29.11 (Bounded Rate Markov Chains). Suppose that S is countable set and $a : S\times S \to \mathbb{R}$ is a function such that $a(x,y) \ge 0$ for all $x \ne y$, and there exists $\lambda < \infty$ such that

$$a_x := \sum_{y\ne x} a(x,y) \le \lambda \text{ for all } x \in S.$$

Let us now define $a(x,x) = -a_x$ and $A : \mathcal{S}_b \to \mathcal{S}_b$ by

$$Af(x) := \sum_{y\in S} a(x,y) f(y) = \sum_{y\ne x} a(x,y)\left[f(y) - f(x)\right] \text{ for all } x \in S$$

and $Q_t = e^{tA}$. It was shown in Corollary 17.27 that $\{Q_t\}_{t\ge 0}$ defines a conservative Markov semi-group and that there is an associated right continuous Markov process associated to this semi-group. In light of Proposition 29.9 and Theorem 29.6 this Markov process has the strong Markov property.

Example 29.12 (Brownian Motion). Suppose that $S = \mathbb{R}^d$ and and

$$Q_t(x,dy) = p_t(x,y)\,dy = p_t(y-x)\,dy,$$

where

$$p_t(x,y) := \left(\frac{1}{2\pi t}\right)^{d/2}\exp\left(-\frac{1}{2t}|x-y|^2\right), \tag{29.11}$$

where $|x|^2 = \sum_{i=1}^d x_i^2$. As you showed in Exercise 17.8, $\{Q_t\}_{t\ge 0}$ is a time homogeneous Markov semi-group (the associated Markov process being Brownian motion) which may be described as;

$$(Q_t f)(x) = \mathbb{E}\left[f\left(x+\sqrt{t}Z\right)\right] \tag{29.12}$$

where $Z \overset{d}{=} N(0,I)$. In this case we may take $\mathbb{M} = BC(S)$ – the bounded continuous functions on S. Indeed if $f \in BC(S)$, then from Eq. (29.12) and the DCT we see that $Q_t f$ is still continuous and that $\|Q_t f\|_u \le \|f\|_u$ so that $Q_t f \in \mathbb{M}$ whenever $f \in \mathbb{M}$. Since we know Brownian motion has continuous paths (Theorem 26.3), it follows from Theorem 29.6 that Brownian motion satisfies the strong Markov property.

Alternatively we could take

$$\mathbb{M} = \left\{f_{a,\lambda}(x) := ae^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d \text{ and } a > 0\right\}.$$

In this case we have

$$\begin{aligned}(Q_t f_{a,\lambda})(x) &= a\mathbb{E}\left[e^{i\lambda\cdot\left(x+\sqrt{t}Z\right)}\right] = a\mathbb{E}\left[e^{i\sqrt{t}\lambda\cdot Z}\right]e^{i\lambda\cdot x}\\ &= ae^{-\frac{t}{2}|\lambda|^2}e^{i\lambda\cdot x} = f_{ae^{-\frac{t}{2}|\lambda|^2},\lambda}(x).\end{aligned}$$

In order to make use of this choice for $\mathbb{M}$, it is necessary to generalize Theorem 29.6 to the complex setting which we leave to the interested reader.

29.4 Applications

Throughout this section we will continue the notation and hypotheseses as in Theorem 29.6. To simplify the statments of the theorems we will further assume that $\Omega = \Gamma$, $\mathcal{B} = \mathcal{F}$, $\mathcal{B}_t = \mathcal{F}_t$, and $X_t = \pi_t$. That is we are going to work in the

canonical path space model for our Markov process. For the time being I will write $b\mathcal{B}$ for the bounded $\mathcal{B}$ – measurable functions on the path space Γ rather than $\mathcal{B}_b$. Let us introduce the **shift operator,** $\theta_t : \Gamma \to \Gamma$ defined by $\theta_t(\gamma) := \gamma(t + \cdot)$ for all $t \geq 0$. With this notation we may now write $Z \circ \theta_t$ for $Z(X_{t+\cdot})$.

Corollary 29.13. *For all $Z \in b\mathcal{B}$, and $t \geq 0$,*

$$\mathbb{E}[Z|\mathcal{B}_{t+}] = \mathbb{E}[Z|\mathcal{B}_t], \ P \text{ – a.s.} \tag{29.13}$$

(More precisely, if U is any version of $\mathbb{E}[Z|\mathcal{B}_{t+}]$ and V is any version of $\mathbb{E}[Z|\mathcal{B}_t]$, then $U = V$, P – a.s.)

Proof. First suppose that $Z = G \cdot F \circ \theta_t$ with $F \in b\mathcal{B}$ and $G \in b\mathcal{B}_t$. For such a Z we have, according to Theorem 29.6 with $\tau \equiv t$ (a stopping time),

$$\begin{aligned}\mathbb{E}[Z|\mathcal{B}_{t+}] &= \mathbb{E}[G \cdot F \circ \theta_t|\mathcal{B}_{t+}] = G \cdot \mathbb{E}[F \circ \theta_t|\mathcal{B}_{t+}] \\ &= G \cdot \mathbb{E}_{B_t}[F] = G \cdot \mathbb{E}[F \circ \theta_t|\mathcal{B}_t] \\ &= \mathbb{E}[G \cdot F \circ \theta_t|\mathcal{B}_t] = \mathbb{E}[Z|\mathcal{B}_t] \qquad P-a.s..\end{aligned}$$

An application of the multiplicative systems Theorem 8.2 (or Theorem 8.16) then shows this identity remains valid for all $Z \in b\mathcal{B}$. (In applying Theorem 8.2, you may want to make use of the conditional version of the DCT.) ∎

Corollary 29.14. *If $A \in \mathcal{B}_{t+}$, there exists $\tilde{A} \in \mathcal{B}_t$ such that $P\left(A \triangle \tilde{A}\right) = 0$. In short, $\mathcal{B}_{t+}$ is equal to $\mathcal{B}_t$ modulo P – null sets.*

Proof. Let $A \in \mathcal{B}_{t_+}$. Applying Corollary 29.13 to $Z = 1_A \in b\mathcal{B}_{t+}$ gives;

$$Z = \mathbb{E}[Z|\mathcal{B}_{t+}] = \mathbb{E}[Z|\mathcal{B}_t] \quad P-\text{a.s.}$$

Letting $\tilde{Z} \in b\mathcal{B}_t$ be a fixed version of $\mathbb{E}[Z|\mathcal{B}_t]$, we have $1_A = \tilde{Z}$ (P – a.s.). Composing this equation with the function, $\varphi(x) = (x \vee 0) \wedge 1$ shows that we may assume $\tilde{Z} = 1_{\tilde{A}}$ where $\tilde{A} = \left\{\tilde{Z} = 1\right\} \in \mathcal{B}_t$. This completes the proof since,

$$P\left(\tilde{A} \triangle A\right) = \mathbb{E}|1_A - 1_{\tilde{A}}| = \mathbb{E}\left|1_A - \tilde{Z}\right| = 0.$$

∎

Recall if $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ is a filtered probability space, the augmentation of a filtration, $\{\mathcal{B}_t\}_{t\geq 0}$, is the new filtration $\left\{\tilde{\mathcal{B}}_t := \mathcal{B}_t \vee \mathcal{N}(P)\right\}_{t\geq 0}$ where

$$\mathcal{N}(P) := \{A \in \mathcal{B} : P(A) = 0\}$$

is the collection of P – null sets in $\mathcal{B}$.

Corollary 29.15 (Augmented filtrations are right continuous). *The augmentation, $\left\{\tilde{\mathcal{B}}_t := \mathcal{B}_t \vee \mathcal{N}(P)\right\}_{t\geq 0}$, of a Brownian filtration, $\{\mathcal{B}_t\}_{t\geq 0}$, is automatically right continuous.*

Proof. By Corollary 29.13 we know that $\mathcal{B}_t \subset \mathcal{B}_{t+} \subset \tilde{\mathcal{B}}_t$ and therefore, $\tilde{\mathcal{B}}_t \subset (\mathcal{B}_{t+})^{\sim} \subset \tilde{\mathcal{B}}_t$, i.e. $\tilde{\mathcal{B}}_t = (\mathcal{B}_{t+})^{\sim}$. On the other hand by Lemma 27.23, $(\mathcal{B}_{t+})^{\sim} = \tilde{\mathcal{B}}_{t+}$. (Also see Example 29.16.) ∎

More generally we have the following result.

Example 29.16 (Augmented filtrations are right continuous). Let ν be a probability measure on S. As in Notation **??**, let $P_\nu := \int_S d\nu(x) \ P_x$ be the Wiener measure on $\Omega := C([0,\infty), S)$, $B_t : \Omega \to S$ be the projection map, $B_t(\omega) = \omega(t)$, $\mathcal{B}_t = \sigma(B_s : s \leq t)$, and $\mathcal{N}_{t+}(\nu) := \{N \in \mathcal{B}_{t+} : P_\nu(N) = 0\}$. Then by Corollary 29.13, $\mathcal{B}_{t+} = \mathcal{B}_t \vee \mathcal{N}_{t+}(\nu)$. Hence if we let

$$\mathcal{N}(\nu) := \{N \in \mathcal{B} : P_\nu(N) = 0\}$$

and

$$\bar{\mathcal{N}}(\nu) := \left\{B \subset 2^{\Omega} : B \subset N \text{ for some } N \in \mathcal{N}(\nu)\right\}$$

then $\mathcal{B}_{t+} \vee \mathcal{N}(\nu) = \mathcal{B}_t \vee \mathcal{N}(\nu) = \bar{\mathcal{B}}_t$ and $\mathcal{B}_{t+} \vee \bar{\mathcal{N}}(\nu) = \mathcal{B}_t \vee \bar{\mathcal{N}}(\nu)$ for all $t \in \mathbb{R}_+$. This shows that the augmented Brownian filtration, $\{\mathcal{B}_t \vee \mathcal{N}(\nu)\}_{t\geq 0}$, is already right continuous and hence satisfies the weak usual hypothesis, (see Definition 27.21). Similarly, the completed and augmented Brownian filtration, $\left\{\mathcal{B}_t \vee \bar{\mathcal{N}}(\nu)\right\}_{t\geq 0}$, satisfies the usual hypothesis.

Theorem 29.17 (Blumenthal $0-1$ Law). *The σ – field, $\mathcal{B}_{0+}$ is P_x – trivial for all $x \in S$.*

Proof. By Corollary 29.14, if $A \in \mathcal{B}_{0+}$ there exists $\tilde{A} \in \mathcal{B}_0$ such that $P_x\left(A \Delta \tilde{A}\right) = 0$. In particular it follows that $P_x(A) = P_x\left(\tilde{A}\right)$. On the other hand $1_{\tilde{A}}$ is $\mathcal{B}_0$ – measurable and hence $1_{\tilde{A}} = F(X_0) = F(x)$ P_x – a.s. for some $F \in \mathcal{S}_b$. Since $1_{\tilde{A}}$ is constant a.s. it must be either 0 or 1 a.s. and hence $P_x(A) = P_x\left(\tilde{A}\right) \in \{0, 1\}$.

∎

Long Run Behavior of Discrete Markov Chains

(This chapter needs more editing and in particular include restatements and proofs of theorems already covered.) In this chapter, X_n will be a Markov chain with a finite or countable state space, S. To each state $i \in S$, let

$$R_i := \min\{n \geq 1 : X_n = i\} \tag{30.1}$$

be the **first passage time of the chain to site** i, and

$$M_i := \sum_{n \geq 1} 1_{X_n = i} \tag{30.2}$$

be number of visits of $\{X_n\}_{n \geq 1}$ to site i.

Definition 30.1. *A state j is accessible from i (written $i \to j$) iff $P_i(R_j < \infty) > 0$ and $i \longleftrightarrow j$ (i communicates with j) iff $i \to j$ and $j \to i$. Notice that $i \to j$ iff there is a path, $i = x_0, x_1, \ldots, x_n = j \in S$ such that $p(x_0, x_1)\, p(x_1, x_2) \ldots p(x_{n-1}, x_n) > 0$.*

Definition 30.2. *For each $i \in S$, let $C_i := \{j \in S : i \longleftrightarrow j\}$ be the **communicating class of** i. The state space, S, is partitioned into a disjoint union of its communicating classes.*

Definition 30.3. *A communicating class $C \subset S$ is **closed** provided the probability that X_n leaves C given that it started in C is zero. In other words $P_{ij} = 0$ for all $i \in C$ and $j \notin C$. (Notice that if C is closed, then X_n restricted to C is a Markov chain.)*

Definition 30.4. *A state $i \in S$ is:*

1. ***transient** if $P_i(R_i < \infty) < 1$,*
2. ***recurrent** if $P_i(R_i < \infty) = 1$,*
   a) ***positive recurrent** if $1/(\mathbb{E}_i R_i) > 0$, i.e. $\mathbb{E}_i R_i < \infty$,*
   b) ***null recurrent** if it is recurrent ($P_i(R_i < \infty) = 1$) and $1/(\mathbb{E}_i R_i) = 0$, i.e. $\mathbb{E} R_i = \infty$.*

We let S_t, S_r, S_{pr}, and S_{nr} be the transient, recurrent, positive recurrent, and null recurrent states respectively.

The next two sections give the main results of this chapter along with some illustrative examples. The remaining sections are devoted to some of the more technical aspects of the proofs.

30.1 The Main Results

Bruce – see Kallenberg [30, pages 148-158]for more information along the lines of what is to follow.

Proposition 30.5 (Class properties). *The notions of being recurrent, positive recurrent, null recurrent, or transient are all class properties. Namely if $C \subset S$ is a communicating class then either all $i \in C$ are recurrent, positive recurrent, null recurrent, or transient. Hence it makes sense to refer to C as being either recurrent, positive recurrent, null recurrent, or transient.*

Proof. See Proposition 30.13 for the assertion that being recurrent or transient is a class property. For the fact that positive and null recurrence is a class property, see Proposition 30.39 below. ■

Lemma 30.6. *Let $C \subset S$ be a communicating class. Then*

$$C \text{ not closed} \implies C \text{ is transient}$$

or equivalently put,

$$C \text{ is recurrent} \implies C \text{ is closed.}$$

Proof. If C is not closed and $i \in C$, there is a $j \notin C$ such that $i \to j$, i.e. there is a path $i = x_0, x_1, \ldots, x_n = j$ with all of the $\{x_j\}_{j=0}^n$ being distinct such that

$$P_i(X_0 = i, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n = j) > 0.$$

Since $j \notin C$ we must have $j \nrightarrow C$ and therefore on the event,

$$A := \{X_0 = i, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n = j\},$$

$X_m \notin C$ for all $m \geq n$ and therefore $R_i = \infty$ on the event A which has positive probability. ■

Proposition 30.7. *Suppose that $C \subset S$ is a finite communicating class and $T = \inf\{n \geq 0 : X_n \notin C\}$ be the first exit time from C. If C is not closed, then not only is C transient but $\mathbb{E}_i T < \infty$ for all $i \in C$. We also have the equivalence of the following statements:*

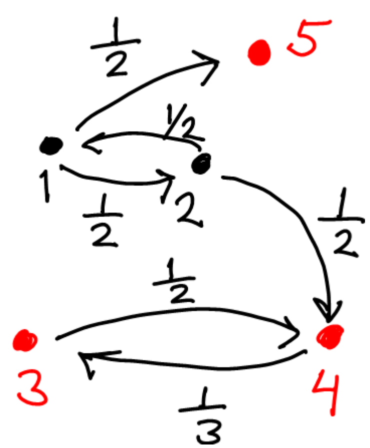
1. *C is closed.*
2. *C is positive recurrent.*
3. *C is recurrent.*

In particular if $\#(S) < \infty$, then the recurrent (= positively recurrent) states are precisely the union of the closed communication classes and the transient states are what is left over.

Proof. These results follow fairly easily from Proposition **??**. Also see Corollary 30.20 for another proof. ■

Remark 30.8. Let $\{X_n\}_{n=0}^\infty$ denote the fair random walk on $\{0,1,2,\dots\}$ with 0 being an absorbing state. The communication classes are $\{0\}$ and $\{1,2,\dots\}$ with the latter class not being closed and hence transient. Using Remark **??**, it follows that $\mathbb{E}_i T = \infty$ for all $i > 0$ which shows we can not drop the assumption that $\#(C) < \infty$ in the first statement in Proposition 30.7. Similarly, using the fair random walk example, we see that it is not possible to drop the condition that $\#(C) < \infty$ for the equivalence statements as well.

Example 30.9. Let P be the Markov matrix with jump diagram given in Figure 30.9. In this case the communication classes are $\{\{1,2\},\{3,4\},\{5\}\}$. The latter two are closed and hence positively recurrent while $\{1,2\}$ is transient.



Warning: if $C \subset S$ is closed and $\#(C) = \infty$, C could be recurrent or it could be transient. Transient in this case means the walk goes off to "infinity." The following proposition is a consequence of the strong Markov property in Corollary 29.5 (or Corollary 30.35 below).

Proposition 30.10. *If $j \in S$, $k \in \mathbb{N}$, and $\nu : S \to [0,1]$ is any probability on S, then*

$$P_\nu(M_j \ge k) = P_\nu(R_j < \infty) \cdot P_j(R_j < \infty)^{k-1}. \tag{30.3}$$

Proof. Intuitively, $M_j \ge k$ happens iff the chain first visits j with probability $P_\nu(R_j < \infty)$ and then revisits j again $k-1$ times which the probability of each revisit being $P_j(R_j < \infty)$. Since Markov chains are forgetful, these probabilities are all independent and hence we arrive at Eq. (30.3). See Proposition 30.36 below for the formal proof based on the strong Markov property in Corollary 29.5 (or Corollary 30.35 below). ■

Corollary 30.11. *If $j \in S$ and $\nu : S \to [0,1]$ is any probability on S, then*

$$P_\nu(M_j = \infty) = P_\nu(X_n = j \text{ i.o.}) = P_\nu(R_j < \infty) 1_{j \in S_r}, \tag{30.4}$$
$$P_j(M_j = \infty) = P_j(X_n = j \text{ i.o.}) = 1_{j \in S_r}, \tag{30.5}$$

$$\mathbb{E}_\nu M_j = \sum_{n=1}^\infty \sum_{i \in S} \nu(i) P_{ij}^n = \frac{P_\nu(R_j < \infty)}{1 - P_j(R_j < \infty)}, \tag{30.6}$$

and

$$\mathbb{E}_i M_j = \sum_{n=1}^\infty P_{ij}^n = \frac{P_i(R_j < \infty)}{1 - P_j(R_j < \infty)} \tag{30.7}$$

where the following conventions are used in interpreting the right hand side of Eqs. (30.6) and (30.7): $a/0 := \infty$ if $a > 0$ while $0/0 := 0$.

Proof. Since

$$\{M_j \ge k\} \downarrow \{M_j = \infty\} = \{X_n = j \text{ i.o. } n\} \text{ as } k \uparrow \infty,$$

it follows, using Eq. (30.3), that

$$P_\nu(X_n = j \text{ i.o. } n) = \lim_{k\to\infty} P_\nu(M_j \ge k) = P_\nu(R_j < \infty) \cdot \lim_{k\to\infty} P_j(R_j < \infty)^{k-1} \tag{30.8}$$

which gives Eq. (30.4). Equation (30.5) follows by taking $\nu = \delta_j$ in Eq. (30.4) and recalling that $j \in S_r$ iff $P_j(R_j < \infty) = 1$. Similarly Eq. (30.7) is a special case of Eq. (30.6) with $\nu = \delta_i$. We now prove Eq. (30.6).

Using the definition of M_j in Eq. (30.2),

$$\mathbb{E}_\nu M_j = \mathbb{E}_\nu \sum_{n\ge1} 1_{X_n = j} = \sum_{n\ge1} \mathbb{E}_\nu 1_{X_n=j}$$
$$= \sum_{n\ge1} P_\nu(X_n = j) = \sum_{n=1}^\infty \sum_{j\in S} \nu(j) P_{jj}^n$$

which is the first equality in Eq. (30.6). For the second, observe that

$$\sum_{k=1}^{\infty} P_\nu(M_j \geq k) = \sum_{k=1}^{\infty} \mathbb{E}_\nu 1_{M_j \geq k} = \mathbb{E}_\nu \sum_{k=1}^{\infty} 1_{k \leq M_j} = \mathbb{E}_\nu M_j.$$

On the other hand using Eq. (30.3) we have

$$\sum_{k=1}^{\infty} P_\nu(M_j \geq k) = \sum_{k=1}^{\infty} P_\nu(R_j < \infty) P_j(R_j < \infty)^{k-1} = \frac{P_\nu(R_j < \infty)}{1 - P_j(R_j < \infty)}$$

provided $a/0 := \infty$ if $a > 0$ while $0/0 := 0$. ■

It is worth remarking that if $j \in S_t$, then Eq. (30.6) asserts that

$$\mathbb{E}_\nu M_j = (\text{the expected number of visits to } j) < \infty$$

which then implies that M_j is a finite valued random variable almost surely. Hence, for almost all sample paths, X_n can visit j at most a finite number of times.

Theorem 30.12 (Recurrent States). *Let $j \in S$. Then the following are equivalent;*

1. *j is recurrent, i.e. $P_j(R_j < \infty) = 1$,*
2. *$P_j(X_n = j \text{ i.o. } n) = 1$,*
3. *$\mathbb{E}_j M_j = \sum_{n=1}^{\infty} P_{jj}^n = \infty$.*

Proof. The equivalence of the first two items follows directly from Eq. (30.5) and the equivalent of items 1. and 3. follows directly from Eq. (30.7) with $i = j$. ■

Proposition 30.13. *If $i \longleftrightarrow j$, then i is recurrent iff j is recurrent, i.e. the property of being recurrent or transient is a class property.*

Proof. Since i and j communicate, there exists α and β in $\mathbb{N}$ such that $P_{ij}^\alpha > 0$ and $P_{ji}^\beta > 0$. Therefore

$$\sum_{n \geq 1} P_{ii}^{n+\alpha+\beta} \geq \sum_{n \geq 1} P_{ij}^\alpha P_{jj}^n P_{ji}^\beta$$

which shows that $\sum_{n \geq 1} P_{jj}^n = \infty \implies \sum_{n \geq 1} P_{ii}^n = \infty$. Similarly $\sum_{n \geq 1} P_{ii}^n = \infty \implies \sum_{n \geq 1} P_{jj}^n = \infty$. Thus using item 3. of Theorem 30.12, it follows that i is recurrent iff j is recurrent. ■

Corollary 30.14. *If $C \subset S_r$ is a recurrent communication class, then*

$$P_i(R_j < \infty) = 1 \text{ for all } i, j \in C \tag{30.9}$$

and in fact

$$P_i(\cap_{j \in C} \{X_n = j \text{ i.o. } n\}) = 1 \text{ for all } i \in C. \tag{30.10}$$

More generally if $\nu : S \to [0,1]$ is a probability such that $\nu(i) = 0$ for $i \notin C$, then

$$P_\nu(\cap_{j \in C} \{X_n = j \text{ i.o. } n\}) = 1 \text{ for all } i \in C. \tag{30.11}$$

In words, if we start in C then every state in C is visited an infinite number of times. (Notice that $P_i(R_j < \infty) = P_i(\{X_n\}_{n \geq 1} \text{ hits } j)$.)

Proof. Let $i, j \in C \subset S_r$ and choose $m \in \mathbb{N}$ such that $P_{ji}^m > 0$. Since $P_j(M_j = \infty) = 1$ and

$$\begin{aligned}
&\{X_m = i \text{ and } X_n = j \text{ for some } n > m\} \\
&\quad = \sum_{n > m} \{X_m = i, X_{m+1} \neq j, \dots, X_{n-1} \neq j, X_n = j\},
\end{aligned}$$

we have

$$\begin{aligned}
P_{ji}^m &= P_j(X_m = i) = P_j(M_j = \infty, X_m = i) \\
&\leq P_j(X_m = i \text{ and } X_n = j \text{ for some } n > m) \\
&= \sum_{n > m} P_j(X_m = i, X_{m+1} \neq j, \dots, X_{n-1} \neq j, X_n = j) \\
&= \sum_{n > m} P_{ji}^m P_i(X_1 \neq j, \dots, X_{n-m-1} \neq j, X_{n-m} = j) \\
&= \sum_{n > m} P_{ji}^m P_i(R_j = n - m) = P_{ji}^m \sum_{k=1}^{\infty} P_i(R_j = k) \\
&= P_{ji}^m P_i(R_j < \infty).
\end{aligned} \tag{30.12}$$

Because $P_{ji}^m > 0$, we may conclude from Eq. (30.12) that $1 \leq P_i(R_j < \infty)$, i.e. that $P_i(R_j < \infty) = 1$ and Eq. (30.9) is proved. Feeding this result back into Eq. (30.4) with $\nu = \delta_i$ shows $P_i(M_j = \infty) = 1$ for all $i, j \in C$ and therefore, $P_i(\cap_{j \in C} \{M_j = \infty\}) = 1$ for all $i \in C$ which is Eq. (30.10). Equation (30.11) follows by multiplying Eq. (30.10) by $\nu(i)$ and then summing on $i \in C$. ■

Theorem 30.15 (Transient States). *Let $j \in S$. Then the following are equivalent;*

1. *j is transient, i.e. $P_j(R_j < \infty) < 1$,*
2. *$P_j(X_n = j \text{ i.o. } n) = 0$, and*

3. $\mathbb{E}_j M_j = \sum_{n=1}^{\infty} P_{jj}^n < \infty$.

Moreover, if $i \in S$ and $j \in S_t$, then

$$\sum_{n=1}^{\infty} P_{ij}^n = \mathbb{E}_i M_j < \infty \implies \begin{cases} \lim_{n\to\infty} P_{ij}^n = 0 \\ P_i(X_n = j \text{ i.o. } n) = 0. \end{cases} \tag{30.13}$$

and more generally if $\nu : S \to [0,1]$ is any probability, then

$$\sum_{n=1}^{\infty} P_\nu(X_n = j) = \mathbb{E}_\nu M_j < \infty \implies \begin{cases} \lim_{n\to\infty} P_\nu(X_n = j) = 0 \\ P_\nu(X_n = j \text{ i.o. } n) = 0. \end{cases} \tag{30.14}$$

Proof. The equivalence of the first two items follows directly from Eq. (30.5) and the equivalent of items 1. and 3. follows directly from Eq. (30.7) with $i = j$. The fact that $\mathbb{E}_i M_j < \infty$ and $\mathbb{E}_\nu M_j < \infty$ for all $j \in S_t$ are consequences of Eqs. (30.7) and (30.6) respectively. The remaining implication in Eqs. (30.13) and (30.6) follow from the first Borel Cantelli Lemma **??** and the fact that n^{th} – term in a convergent series tends to zero as $n \to \infty$. ■

Corollary 30.16. *1) If the state space, S, is a finite set, then $S_r \neq \emptyset$. 2) Any finite and closed communicating class $C \subset S$ is a recurrent.*

Proof. First suppose that $\#(S) < \infty$ and for the sake of contradiction, suppose $S_r = \emptyset$ or equivalently that $S = S_t$. Then by Theorem 30.15, $\lim_{n\to\infty} P_{ij}^n = 0$ for all $i, j \in S$. On the other hand, $\sum_{j\in S} P_{ij}^n = 1$ so that

$$1 = \lim_{n\to\infty} \sum_{j\in S} P_{ij}^n = \sum_{j\in S} \lim_{n\to\infty} P_{ij}^n = \sum_{j\in S} 0 = 0,$$

which is a contradiction. (Notice that if S were infinite, we could not interchange the limit and the above sum without some extra conditions.)

To prove the first statement, restrict X_n to C to get a Markov chain on a finite state space C. By what we have just proved, there is a recurrent state $i \in C$. Since recurrence is a class property, it follows that all states in C are recurrent. ■

Definition 30.17. *A function, $\pi : S \to [0,1]$ is a* ***sub-probability*** *if $\sum_{j\in S} \pi(j) \le 1$. We call $\sum_{j\in S} \pi(j)$ the* ***mass*** *of π. So a probability is a sub-probability with mass one.*

Definition 30.18. *We say a sub-probability, $\pi : S \to [0,1]$, is* ***invariant*** *if $\pi P = \pi$, i.e.*

$$\sum_{i\in S} \pi(i) p_{ij} = \pi(j) \text{ for all } j \in S. \tag{30.15}$$

An invariant probability, $\pi : S \to [0,1]$, is called an ***invariant distribution.***

Theorem 30.19. *Suppose that $P = (p_{ij})$ is an irreducible Markov kernel and $\pi_j := \frac{1}{\mathbb{E}_j R_j}$ for all $j \in S$. Then:*

1. For all $i, j \in S$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} 1_{X_n = j} = \pi_j \quad P_i - a.s. \tag{30.16}$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_i(X_n = j) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} P_{ij}^n = \pi_j. \tag{30.17}$$

2. If $\mu : S \to [0,1]$ is an invariant sub-probability, then either $\mu(i) > 0$ for all i or $\mu(i) = 0$ for all i.
3. P has at most one invariant distribution.
4. P has a (necessarily unique) invariant distribution, $\mu : S \to [0,1]$, iff P is positive recurrent in which case $\mu(i) = \pi(i) = \frac{1}{\mathbb{E}_i R_i} > 0$ for all $i \in S$.

(These results may of course be applied to the restriction of a general non-irreducible Markov chain to any one of its communication classes.)

Proof. These results are the contents of Theorem 30.38 and Propositions 30.39 and 30.40 below. ■

Using this result we can give another proof of Proposition 30.7.

Corollary 30.20. *If C is a closed finite communicating class then C is positive recurrent. (Recall that we already know that C is recurrent by Corollary 30.16.)*

Proof. For $i, j \in C$, let

$$\pi_j := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_i(X_n = j) = \frac{1}{\mathbb{E}_j R_j}$$

as in Theorem 30.21. Since C is closed,

$$\sum_{j\in C} P_i(X_n = j) = 1$$

1and therefore,

$$\sum_{j\in C} \pi_j = \lim_{N\to\infty} \frac{1}{N} \sum_{j\in C} \sum_{n=1}^{N} P_i(X_n = j) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{j\in C} P_i(X_n = j) = 1.$$

Therefore $\pi_j > 0$ for some $j \in C$ and hence all $j \in C$ by Theorem 30.19 with S replaced by C. Hence we have $\mathbb{E}_j R_j < \infty$, i.e. every $j \in C$ is a positive recurrent state. ■

Theorem 30.21 (General Convergence Theorem). *Let $\nu : S \to [0,1]$ be any probability, $i \in S$, C be the communicating class containing i,*

$$\{X_n \text{ hits } C\} := \{X_n \in C \text{ for some } n\},$$

and

$$\pi_i := \pi_i(\nu) = \frac{P_\nu(X_n \text{ hits } C)}{\mathbb{E}_i R_i}, \tag{30.18}$$

where $1/\infty := 0$. Then:

1. *P_ν – a.s.,*
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n = i} = \frac{1}{\mathbb{E}_i R_i} 1_{\{X_n \text{ hits } C\}}, \tag{30.19}$$
2.
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{j\in S} \nu(j) P^n_{ji} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_\nu(X_n = i) = \pi_i, \tag{30.20}$$
3. *π is an invariant sub-probability for P, and*
4. *the mass of π is*
$$\sum_{i\in S} \pi_i = \sum_{C:\text{ pos. recurrent}} P_\nu(X_n \text{ hits } C) \le 1. \tag{30.21}$$

Proof. If $i \in S$ is a transient site, then according to Eq. (30.14), $P_\nu(M_i < \infty) = 1$ and therefore $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n=i} = 0$ which agrees with Eq. (30.19) for $i \in S_t$.

So now suppose that $i \in S_r$ and let C be the communication class containing i and

$$T = \inf\{n \ge 0 : X_n \in C\}$$

be the first time when X_n enters C. It is clear that $\{R_i < \infty\} \subset \{T < \infty\}$. On the other hand, for any $j \in C$, it follows by the strong Markov property (Corollary 29.5 (or Corollary 30.35 below)) and Corollary 30.14 that, conditioned on $\{T < \infty, X_T = j\}$, $\{X_n\}$ hits i i.o. and hence $P(R_i < \infty | T < \infty, X_T = j) = 1$. Equivalently put,

$$P(R_i < \infty, T < \infty, X_T = j) = P(T < \infty, X_T = j) \text{ for all } j \in C.$$

Summing this last equation on $j \in C$ then shows

$$P(R_i < \infty) = P(R_i < \infty, T < \infty) = P(T < \infty)$$

and therefore $\{R_i < \infty\} = \{T < \infty\}$ modulo an event with P_ν – probability zero.

Another application of the strong Markov property (in Corollary 29.5 (or Corollary 30.35 below)), observing that $X_{R_i} = i$ on $\{R_i < \infty\}$, allows us to conclude that the $P_\nu(\cdot|R_i < \infty) = P_\nu(\cdot|T < \infty)$ – law of $(X_{R_i}, X_{R_i+1}, X_{R_i+2}, \dots)$ is the same as the P_i – law of $(X_0, X_1, X_2, \dots)$. Therefore, we may apply Theorem 30.19 to conclude that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n=i} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_{R_i+n}=i} = \frac{1}{\mathbb{E}_i R_i} \; P_\nu(\cdot|R_i < \infty) \text{ – a.s.}$$

On the other hand, on the event $\{R_i = \infty\}$ we have $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n=i} = 0$. Thus we have shown P_ν – a.s. that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n=i} = \frac{1}{\mathbb{E}_i R_i} 1_{R_i<\infty} = \frac{1}{\mathbb{E}_i R_i} 1_{T<\infty} = \frac{1}{\mathbb{E}_i R_i} 1_{\{X_n \text{ hits } C\}}$$

which is Eq. (30.19). Taking expectations of this equation, using the dominated convergence theorem, gives Eq. (30.20).

Since $1/\mathbb{E}_i R_i = \infty$ unless i is a positive recurrent site, it follows that

$$\sum_{i\in S} \pi_i P_{ij} = \sum_{i\in S_{\text{pr}}} \pi_i P_{ij} = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C) \sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} P_{ij}. \tag{30.22}$$

As each positive recurrent class, C, is closed; if $i \in C$ and $j \notin C$, then $P_{ij} = 0$. Therefore $\sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} P_{ij}$ is zero unless $j \in C$. So if $j \notin S_{\text{pr}}$ we have $\sum_{i\in S} \pi_i P_{ij} = 0 = \pi_j$ and if $j \in S_{\text{pr}}$, then by Theorem 30.19,

$$\sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} P_{ij} = 1_{j\in C} \cdot \frac{1}{\mathbb{E}_j R_j}.$$

Using this result in Eq. (30.22) shows that

$$\sum_{i\in S} \pi_i P_{ij} = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C) 1_{j\in C} \cdot \frac{1}{\mathbb{E}_j R_j} = \pi_j$$

so that π is an invariant distribution. Similarly, using Theorem 30.19 again,

$$\sum_{i\in S} \pi_i = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C) \sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C).$$

■

Definition 30.22. *A state $i \in S$ is* ***aperiodic*** *if $P^n_{ii} > 0$ for all n sufficiently large.*

Lemma 30.23. *If $i \in S$ is aperiodic and $j \longleftrightarrow i$, then j is aperiodic. So being aperiodic is a class property.*

Proof. We have

$$P_{jj}^{n+m+k} = \sum_{w,z \in S} P_{j,w}^{n} P_{w,z}^{m} P_{z,j}^{k} \geq P_{j,i}^{n} P_{i,i}^{m} P_{i,j}^{k}.$$

Since $j \longleftrightarrow i$, there exists $n, k \in \mathbb{N}$ such that $P_{j,i}^{n} > 0$ and $P_{i,j}^{k} > 0$. Since $P_{i,i}^{m} > 0$ for all large m, it follows that $P_{jj}^{n+m+k} > 0$ for all large m and therefore, j is aperiodic as well. ■

Lemma 30.24. *A state $i \in S$ is aperiodic iff 1 is the greatest common divisor of the set,*

$$\{n \in \mathbb{N} : P_i (X_n = i) = P_{ii}^{n} > 0\}.$$

Proof. Use the number theory Lemma 30.41 below. ■

Theorem 30.25. *If P is an irreducible, aperiodic, and recurrent Markov chain, then*

$$\lim_{n \to \infty} P_{ij}^{n} = \pi_j = \frac{1}{\mathbb{E}_j (R_j)}. \tag{30.23}$$

More generally, if C is an aperiodic communication class, then

$$\lim_{n \to \infty} P_\nu (X_n = i) := \lim_{n \to \infty} \sum_{j \in S} \nu (j) P_{ji}^{n} = P_\nu (R_i < \infty) \frac{1}{\mathbb{E}_j (R_j)} \text{ for all } i \in C.$$

Proof. I will not prove this theorem here but refer the reader to Norris [38, Theorem 1.8.3] or Kallenberg [30, Chapter 8]. The proof given there is by a "coupling argument" is given. ■

30.1.1 More finite state space examples

Example 30.26 (Analyzing a non-irreducible Markov chain). In this example we are going to analyze the limiting behavior of the non-irreducible Markov chain determined by the Markov matrix,

$$P = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \end{array}.$$

Here are the steps to follow.

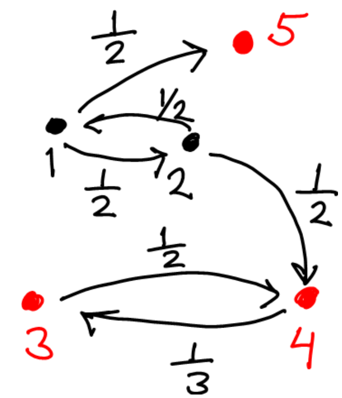


Fig. 30.1. The jump diagram for P above.

1. **Find the jump diagram for** P. In our case it is given in Figure 30.1.
2. **Identify the communication classes**. In our example they are $\{1, 2\}$, $\{5\}$, and $\{3, 4\}$. The first is not closed and hence transient while the second two are closed and finite sets and hence recurrent.
3. **Find the invariant distributions for the recurrent classes.** For $\{5\}$ it is simply $\pi'_{\{5\}} = [1]$ and for $\{3, 4\}$ we must find the invariant distribution for the 2×2 Markov matrix,

$$Q = \begin{array}{c} \begin{matrix} 3 & 4 \end{matrix} \\ \begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix} \begin{matrix} 3 \\ 4 \end{matrix} \end{array}.$$

We do this in the usual way, namely

$$\mathrm{Nul}\left(I - Q^{\mathrm{tr}}\right) = \mathrm{Nul}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{2}{3} \end{bmatrix}\right) = \mathbb{R} \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

so that $\pi'_{\{3,4\}} = \frac{1}{5} \left[2\ 3\right]$.

4. We can turn $\pi'_{\{3,4\}}$ and $\pi'_{\{5\}}$ into invariant distributions for P by padding the row vectors with zeros to get

$$\pi_{\{3,4\}} = \left[0\ 0\ 2/5\ 3/5\ 0\right]$$
$$\pi_{\{5\}} = \left[0\ 0\ 0\ 0\ 1\right].$$

The general invariant distribution may then be written as;

$$\pi = \alpha \pi_{\{5\}} + \beta \pi_{\{3,4\}} \text{ with } \alpha, \beta \geq 0 \text{ and } \alpha + \beta = 1.$$

5. We can now work out the $\lim_{n\to\infty} P^n$. If we start at site i we are considering the i^{th} – row of $\lim_{n\to\infty} P^n$. If we start in the recurrent class $\{3,4\}$ we will simply get $\pi_{\{3,4\}}$ for these rows and we start in the recurrent class $\{5\}$ we will get $\pi_{\{5\}}$. However if start in the non-closed transient class, $\{1,2\}$ we have
$$\text{first row of } \lim_{n\to\infty} P^n = P_1\left(X_n \text{ hits } 5\right)\pi_{\{5\}} + P_1\left(X_n \text{ hits } \{3,4\}\right)\pi_{\{3,4\}} \tag{30.24}$$
and
$$\text{second row of } \lim_{n\to\infty} P^n = P_2\left(X_n \text{ hits } 5\right)\pi_{\{5\}} + P_2\left(X_n \text{ hits } \{3,4\}\right)\pi_{\{3,4\}}. \tag{30.25}$$

6. **Compute the required hitting probabilities.** Let us begin by computing the fraction of one pound of sand put at site 1 will end up at site 5, i.e. we want to find $h_1 := P_1\left(X_n \text{ hits } 5\right)$. To do this let $h_i = P_i\left(X_n \text{ hits } 5\right)$ for $i = 1, 2, \ldots, 5$. It is clear that $h_5 = 1$, and $h_3 = h_4 = 0$. A first step analysis then shows
$$\begin{aligned} h_1 &= \frac{1}{2}\cdot P_2\left(X_n \text{ hits } 5\right) + \frac{1}{2}P_5\left(X_n \text{ hits } 5\right) \\ h_2 &= \frac{1}{2}\cdot P_1\left(X_n \text{ hits } 5\right) + \frac{1}{2}P_4\left(X_n \text{ hits } 5\right) \end{aligned}$$
which leads to[1]
$$\begin{aligned} h_1 &= \frac{1}{2}h_2 + \frac{1}{2} \\ h_2 &= \frac{1}{2}h_1 + \frac{1}{2}0. \end{aligned}$$
The solutions to these equations are
$$P_1\left(X_n \text{ hits } 5\right) = h_1 = \frac{2}{3} \text{ and } P_2\left(X_n \text{ hits } 5\right) = h_2 = \frac{1}{3}.$$
Since the process is either going to end up in $\{5\}$ or in $\{3,4\}$, we may also conclude that
$$P_1\left(X_n \text{ hits } \{3,4\}\right) = \frac{1}{3} \text{ and } P_2\left(X_n \text{ hits } \{3,4\}\right) = \frac{2}{3}.$$

7. Using these results in Eqs. (30.24) and (30.25) shows,
$$\begin{aligned} \text{first row of } \lim_{n\to\infty} P^n &= \frac{2}{3}\pi_{\{5\}} + \frac{1}{3}\pi_{\{3,4\}} \\ &= \left[0\ 0\ \tfrac{2}{15}\ \tfrac{1}{5}\ 2/3\right] \\ &= \left[0.0\ 0.0\ 0.133\,33\ 0.2\ 0.666\,67\right] \end{aligned}$$
and
$$\begin{aligned} \text{second row of } \lim_{n\to\infty} P^n &= \frac{1}{3}\pi_{\{5\}} + \frac{2}{3}\pi_{\{3,4\}} \\ &= \frac{1}{3}\left[0\ 0\ 0\ 0\ 1\right] + \frac{2}{3}\left[0\ 0\ 2/5\ 3/5\ 0\right] \\ &= \left[0\ 0\ \tfrac{4}{15}\ \tfrac{2}{5}\ \tfrac{1}{3}\right] \\ &= \left[0.0\ 0.0\ 0.266\,67\ 0.4\ 0.333\,33\right]. \end{aligned}$$
These answers already compare well with
$$P^{10} = \begin{bmatrix} 9.7656\times10^{-4} & 0.0 & 0.132\,76 & 0.200\,24 & 0.666\,02 \\ 0.0 & 9.7656\times10^{-4} & 0.266\,26 & 0.399\,76 & 0.333\,01 \\ 0.0 & 0.0 & 0.4 & 0.600\,00 & 0.0 \\ 0.0 & 0.0 & 0.400\,00 & 0.6 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

1

Example 30.27. **Note:** If we were to make use of Theorem **??** we would have not set $h_3 = h_4 = 0$ and we would have added the equations,
$$\begin{aligned} h_3 &= \frac{1}{2}h_3 + \frac{1}{2}h_4 \\ h_4 &= \frac{1}{3}h_3 + \frac{2}{3}h_4, \end{aligned}$$
to those above. The general solution to these equations is $c(1,1)$ for some $c \in \mathbb{R}$ and the non-negative minimal solution is the special case where $c = 0$, i.e. $h_3 = h_4 = 0$. The point is, since $\{3,4\}$ is a closed communication class there is no way to hit 5 starting in $\{3,4\}$ and therefore clearly $h_3 = h_4 = 0$.

30.2 The Strong Markov Property

In proving the results above, we are going to make essential use of a strong form of the Markov property which asserts that Theorem **??** continues to hold even when n is replaced by a random "stopping time."

Definition 30.28 (Stopping times). *Let τ be an $\mathbb{N}_0 \cup \{\infty\}$ - valued random variable which is a functional of a sequence of random variables, $\{X_n\}_{n=0}^{\infty}$ which we write by abuse of notation as, $\tau = \tau(X_0, X_1, \ldots)$. We say that τ is a stopping time if for all $n \in \mathbb{N}_0$, the indicator random variable, $1_{\tau=n}$ is a functional of $(X_0, \ldots, X_n)$. Thus for each $n \in \mathbb{N}_0$ there should exist a function, σ_n such that $1_{\tau=n} = \sigma_n(X_0, \ldots, X_n)$. In other words, the event $\{\tau = n\}$ may be described using only $(X_0, \ldots, X_n)$ for all $n \in \mathbb{N}$.*

Example 30.29. Here are some example of random times which are which are not stopping times. In these examples we will always use the convention that the minimum of the empty set is $+\infty$.

1. The random time, $\tau = \min\{k : |X_k| \geq 5\}$ (the first time, k, such that $|X_k| \geq 5$) is a stopping time since
$$\{\tau = k\} = \{|X_1| < 5, \dots, |X_{k-1}| < 5, |X_k| \geq 5\}.$$
2. Let $W_k := X_1 + \dots + X_k$, then the random time,
$$\tau = \min\{k : W_k \geq \pi\}$$
is a stopping time since,
$$\{\tau = k\} = \left\{ \begin{array}{c} W_j = X_1 + \dots + X_j < \pi \text{ for } j = 1, 2, \dots, k-1, \\ \& \ X_1 + \dots + X_{k-1} + X_k \geq \pi \end{array} \right\}.$$
3. For $t \geq 0$, let $N(t) = \#\{k : W_k \leq t\}$. Then
$$\{N(t) = k\} = \{X_1 + \dots + X_k \leq t, \ X_1 + \dots + X_{k+1} > t\}$$
which shows that $N(t)$ is **not** a stopping time. On the other hand, since
$$\begin{aligned} \{N(t) + 1 = k\} &= \{N(t) = k-1\} \\ &= \{X_1 + \dots + X_{k-1} \leq t, \ X_1 + \dots + X_k > t\}, \end{aligned}$$
we see that $N(t) + 1$ **is** a stopping time!
4. If τ is a stopping time then so is $\tau + 1$ because,
$$1_{\{\tau+1=k\}} = 1_{\{\tau=k-1\}} = \sigma_{k-1}(X_0, \dots, X_{k-1})$$
which is also a function of $(X_0, \dots, X_k)$ which happens not to depend on X_k.
5. On the other hand, if τ is a stopping time it is **not** necessarily true that $\tau - 1$ is still a stopping time as seen in item 3. above.
6. One can also see that the last time, k, such that $|X_k| \geq \pi$ is typically **not** a stopping time. (Think about this.)

Remark 30.30. If τ is an $\{X_n\}_{n=0}^{\infty}$ - stopping time then
$$1_{\tau \geq n} = 1 - 1_{\tau < n} = 1 - \sum_{k<n} \sigma_k(X_0, \dots, X_k) =: u_n(X_0, \dots, X_{n-1}).$$
That is for a stopping time τ, $1_{\tau \geq n}$ is a function of $(X_0, \dots, X_{n-1})$ only for all $n \in \mathbb{N}_0$.

The following presentation of Wald's equation is taken from Ross [45, p. 59-60].

Theorem 30.31 (Wald's Equation). *Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables, $f(x)$ is a non-negative function of $x \in \mathbb{R}$, and τ is a stopping time. Then*
$$\mathbb{E}\left[\sum_{n=1}^{\tau} f(X_n)\right] = \mathbb{E}f(X_1) \cdot \mathbb{E}\tau. \tag{30.26}$$
This identity also holds if $f(X_n)$ are real valued but integrable and τ is a stopping time such that $\mathbb{E}\tau < \infty$. (See Resnick for more identities along these lines.)

Proof. If $f(X_n) \geq 0$ for all n, then the the following computations need no justification,
$$\begin{aligned} \mathbb{E}\left[\sum_{n=1}^{\tau} f(X_n)\right] &= \mathbb{E}\left[\sum_{n=1}^{\infty} f(X_n) 1_{n\leq\tau}\right] = \sum_{n=1}^{\infty} \mathbb{E}[f(X_n) 1_{n\leq\tau}] \\ &= \sum_{n=1}^{\infty} \mathbb{E}[f(X_n) u_n(X_1, \dots, X_{n-1})] \\ &= \sum_{n=1}^{\infty} \mathbb{E}[f(X_n)] \cdot \mathbb{E}[u_n(X_1, \dots, X_{n-1})] \\ &= \sum_{n=1}^{\infty} \mathbb{E}[f(X_n)] \cdot \mathbb{E}[1_{n\leq\tau}] = \mathbb{E}f(X_1) \sum_{n=1}^{\infty} \mathbb{E}[1_{n\leq\tau}] \\ &= \mathbb{E}f(X_1) \cdot \mathbb{E}\left[\sum_{n=1}^{\infty} 1_{n\leq\tau}\right] = \mathbb{E}f(X_1) \cdot \mathbb{E}\tau. \end{aligned}$$

If $\mathbb{E}|f(X_n)| < \infty$ and $\mathbb{E}\tau < \infty$, the above computation with f replaced by $|f|$ shows all sums appearing above are equal $\mathbb{E}|f(X_1)| \cdot \mathbb{E}\tau < \infty$. Hence we may remove the absolute values to again arrive at Eq. (30.26). ■

Example 30.32. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. such that $P(X_n = 0) = P(X_n = 1) = 1/2$ and let
$$\tau := \min\{n : X_1 + \dots + X_n = 10\}.$$
For example τ is the first time we have flipped 10 heads of a fair coin. By Wald's equation (valid because $X_n \geq 0$ for all n) we find
$$10 = \mathbb{E}\left[\sum_{n=1}^{\tau} X_n\right] = \mathbb{E}X_1 \cdot \mathbb{E}\tau = \frac{1}{2}\mathbb{E}\tau$$
and therefore $\mathbb{E}\tau = 20 < \infty$.

Example 30.33 (Gambler's ruin). Let $\{X_n\}_{n=1}^\infty$ be i.i.d. such that $P(X_n = -1) = P(X_n = 1) = 1/2$ and let

$$\tau := \min\{n : X_1 + \cdots + X_n = 1\}.$$

So τ may represent the first time that a gambler is ahead by 1. Notice that $\mathbb{E}X_1 = 0$. If $\mathbb{E}\tau < \infty$, then we would have $\tau < \infty$ a.s. and by Wald's equation would give,

$$1 = \mathbb{E}\left[\sum_{n=1}^{\tau} X_n\right] = \mathbb{E}X_1 \cdot \mathbb{E}\tau = 0 \cdot \mathbb{E}\tau$$

which can not hold. Hence it must be that

$$\mathbb{E}\tau = \mathbb{E}\left[\text{first time that a gambler is ahead by 1}\right] = \infty.$$

Theorem 30.34 (Strong Markov Property). *Let $\left(\{X_n\}_{n=0}^\infty, \{P_x\}_{x\in S}, p\right)$ be Markov chain as above and $\tau : \Omega \to [0,\infty]$ be a stopping time as in Definition 30.28. Then*

$$\begin{aligned} &\mathbb{E}_\pi\left[f(X_\tau, X_{\tau+1}, \dots) g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right] \\ &\quad = \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau} f(X_0, X_1, \dots)\right] g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right]. \end{aligned} \tag{30.27}$$

for all $f, g = \{g_n\} \geq 0$ or f and g bounded.

Proof. The proof of this deep result is now rather easy to reduce to Theorem **??**. Indeed,

$$\begin{aligned} &\mathbb{E}_\pi\left[f(X_\tau, X_{\tau+1}, \dots) g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[f(X_n, X_{n+1}, \dots) g_n(X_0, \dots, X_n) 1_{\tau=n}\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[f(X_n, X_{n+1}, \dots) g_n(X_0, \dots, X_n) \sigma_n(X_0, \dots, X_n)\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_n} f(X_0, X_1, \dots)\right] g_n(X_0, \dots, X_n) \sigma_n(X_0, \dots, X_n)\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau} f(X_0, X_1, \dots)\right] g_\tau(X_0, \dots, X_n) 1_{\tau=n}\right] \\ &= \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau} f(X_0, X_1, \dots)\right] g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right] \end{aligned}$$

wherein we have used Theorem **??** in the third equality. ■

The analogue of Corollary **??** in this more general setting states; conditioned on $\tau < \infty$ and $X_\tau = x$, $X_\tau, X_{\tau+1}, X_{\tau+2}, \dots$ is independent of $X_0, \dots, X_\tau$ and is distributed as $X_0, X_1, \dots$ under P_x.

Corollary 30.35. *Let τ be a stopping time, $x \in S$ and π be any probability on S. Then relative to $P_\pi(\cdot|\tau < \infty,\ X_\tau = x)$, $\{X_{\tau+k}\}_{k\geq 0}$ is independent of $\{X_0, \dots, X_\tau\}$ and $\{X_{\tau+k}\}_{k\geq 0}$ has the same distribution as $\{X_k\}_{k=0}^\infty$ under P_x.*

Proof. According to Eq. (30.27),

$$\begin{aligned} &\mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) f(X_\tau, X_{\tau+1}, \dots) : \tau < \infty,\ X_\tau = x\right] \\ &\quad = \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) 1_{\tau<\infty} \delta_x(X_\tau) f(X_\tau, X_{\tau+1}, \dots)\right] \\ &\quad = \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) 1_{\tau<\infty} \delta_x(X_\tau) \mathbb{E}_{X_\tau}\left[f(X_0, X_1, \dots)\right]\right] \\ &\quad = \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) 1_{\tau<\infty} \delta_x(X_\tau) \mathbb{E}_x\left[f(X_0, X_1, \dots)\right]\right] \\ &\quad = \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) : \tau < \infty,\ X_\tau = x\right] \mathbb{E}_x\left[f(X_0, X_1, \dots)\right]. \end{aligned}$$

Dividing this equation by $P(\tau < \infty,\ X_\tau = x)$ shows,

$$\begin{aligned} &\mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) f(X_\tau, X_{\tau+1}, \dots) | \tau < \infty,\ X_\tau = x\right] \\ &\quad = \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) | \tau < \infty,\ X_\tau = x\right] \mathbb{E}_x\left[f(X_0, X_1, \dots)\right]. \end{aligned} \tag{30.28}$$

Taking $g = 1$ in this equation then shows,

$$\mathbb{E}_\pi\left[f(X_\tau, X_{\tau+1}, \dots) | \tau < \infty,\ X_\tau = x\right] = \mathbb{E}_x\left[f(X_0, X_1, \dots)\right]. \tag{30.29}$$

This shows that $\{X_{\tau+k}\}_{k\geq 0}$ under $P_\pi(\cdot|\tau < \infty,\ X_\tau = x)$ has the same distribution as $\{X_k\}_{k=0}^\infty$ under P_x and, in combination, Eqs. (30.28) and (30.29) shows $\{X_{\tau+k}\}_{k\geq 0}$ and $\{X_0, \dots, X_\tau\}$ are conditionally, on $\{\tau < \infty,\ X_\tau = x\}$, independent. ■

To match notation in the book, let

$$f_{ii}^{(n)} = P_i(R_i = n) = P_i(X_1 \neq i, \dots, X_{n-1} \neq i, X_n = i)$$

and $m_{ij} := \mathbb{E}_i(M_j)$ – the expected number of visits to j after $n = 0$.

Proposition 30.36. *Let $i \in S$ and $n \geq 1$. Then P_{ii}^n satisfies the "renewal equation,"*

$$P_{ii}^n = \sum_{k=1}^n P(R_i = k) P_{ii}^{n-k}. \tag{30.30}$$

Also if $j \in S$, $k \in \mathbb{N}$, and $\nu : S \to [0,1]$ is any probability on S, then Eq. (30.3) holds, i.e.

$$P_\nu(M_j \geq k) = P_\nu(R_j < \infty) \cdot P_j(R_j < \infty)^{k-1}. \tag{30.31}$$

Proof. To prove Eq. (30.30) we first observe for $n \geq 1$ that $\{X_n = i\}$ is the disjoint union of $\{X_n = i, R_i = k\}$ for $1 \leq k \leq n$ and therefore[2],

[2] Alternatively, we could use the Markov property to show,

$$\begin{aligned} P_{ii}^n = P_i(X_n = i) &= \sum_{k=1}^n P_i(R_i = k, X_n = i) \\ &= \sum_{k=1}^n P_i(X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i, X_n = i) \\ &= \sum_{k=1}^n P_i(X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i) P_{ii}^{n-k} \\ &= \sum_{k=1}^n P_{ii}^{n-k} P(R_i = k). \end{aligned}$$

For Eq. (30.31) we have $\{M_j \geq 1\} = \{R_j < \infty\}$ so that $P_i(M_j \geq 1) = P_i(R_j < \infty)$. For $k \geq 2$, since $R_j < \infty$ if $M_j \geq 1$, we have

$$P_i(M_j \geq k) = P_i(M_j \geq k | R_j < \infty) P_i(R_j < \infty).$$

Since, on $R_j < \infty$, $X_{R_j} = j$, it follows by the strong Markov property (Corollary 29.5 (or Corollary 30.35 below)) that;

$$\begin{aligned} P_i(M_j \geq k | R_j < \infty) &= P_i\left(M_j \geq k | R_j < \infty,\ X_{R_j} = j\right) \\ &= P_i\left(1 + \sum_{n \geq 1} 1_{X_{R_j + n} = j} \geq k | R_j < \infty,\ X_{R_j} = j\right) \\ &= P_j\left(1 + \sum_{n \geq 1} 1_{X_n = j} \geq k\right) = P_j(M_j \geq k - 1). \end{aligned}$$

By the last two displayed equations,

$$P_i(M_j \geq k) = P_j(M_j \geq k - 1) P_i(R_j < \infty) \tag{30.32}$$

Taking $i = j$ in this equation shows,

$$P_j(M_j \geq k) = P_j(M_j \geq k - 1) P_j(R_j < \infty)$$

$$\begin{aligned} P_{ii}^n = P_i(X_n = i) &= \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k} \cdot 1_{X_n = i}) = \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k} \cdot \mathbb{E}_i 1_{X_{n-k} = i}) \\ &= \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k}) \mathbb{E}_i\left(1_{X_{n-k} = i}\right) = \sum_{k=1}^n P_i(R_i = k) P_i(X_{n-k} = i) \\ &= \sum_{k=1}^n P_{ii}^{n-k} P(R_i = k). \end{aligned}$$

and so by induction,

$$P_j(M_j \geq k) = P_j(R_j < \infty)^k. \tag{30.33}$$

Equation (30.31) now follows from Eqs. (30.32) and (30.33). ∎

30.3 Irreducible Recurrent Chains

For this section we are going to assume that X_n is a irreducible recurrent Markov chain. Let us now fix a state, $j \in S$ and define,

$$\begin{aligned} \tau_1 &= R_j = \min\{n \geq 1 : X_n = j\}, \\ \tau_2 &= \min\{n \geq 1 : X_{n+\tau_1} = j\}, \\ &\vdots \\ \tau_n &= \min\{n \geq 1 : X_{n+\tau_{n-1}} = j\}, \end{aligned}$$

so that τ_n is the time it takes for the chain to visit j after the $(n-1)$'st visit to j. By Corollary 30.14 we know that $P_i(\tau_n < \infty) = 1$ for all $i \in S$ and $n \in \mathbb{N}$. We will use strong Markov property to prove the following key lemma in our development.

Lemma 30.37. *We continue to use the notation above and in particular assume that X_n is an irreducible recurrent Markov chain. Then relative to any P_i with $i \in S$, $\{\tau_n\}_{n=1}^\infty$ is a sequence of independent random variables, $\{\tau_n\}_{n=2}^\infty$ are identically distributed, and $P_i(\tau_n = k) = P_j(\tau_1 = k)$ for all $k \in \mathbb{N}_0$ and $n \geq 2$.*

Proof. Let $T_0 = 0$ and then define T_k inductively by, $T_{k+1} = \inf\{n > T_k : X_n = j\}$ so that T_n is the time of the n'th visit of $\{X_n\}_{n=1}^\infty$ to site j. Observe that $T_1 = \tau_1$,

$$\tau_{n+1}(X_0, X_1, \dots) = \tau_1\left(X_{T_n}, X_{T_n+1}, X_{T_n+2}, \dots\right),$$

and $(\tau_1, \dots, \tau_n)$ is a function of $(X_0, \dots, X_{T_n})$. Since $P_i(T_n < \infty) = 1$ (Corollary 30.14) and $X_{T_n} = j$, we may apply the strong Markov property in the form of Corollary 29.5 (or Corollary 30.35) to learn:

1. τ_{n+1} is independent of $(X_0, \dots, X_{T_n})$ and hence τ_{n+1} is independent of $(\tau_1, \dots, \tau_n)$, and
2. the distribution of τ_{n+1} under P_i is the same as the distribution of τ_1 under P_j.

The result now follows from these two observations and induction. ∎

Theorem 30.38. *Suppose that X_n is a irreducible recurrent Markov chain, and let $j \in S$ be a fixed state. Define*

$$\pi_j := \frac{1}{\mathbb{E}_j(R_j)}, \tag{30.34}$$

with the understanding that $\pi_j = 0$ if $\mathbb{E}_j(R_j) = \infty$. Then

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N} 1_{X_n=j} = \pi_j \quad P_i - a.s. \tag{30.35}$$

for all $i \in S$ and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N} P_{ij}^n = \pi_j. \tag{30.36}$$

Proof. Let us first note that Eq. (30.36) follows by taking expectations of Eq. (30.35). So we must prove Eq. (30.35).

By Lemma 30.37, the sequence $\{\tau_n\}_{n\geq 2}$ is i.i.d. relative to P_i and $\mathbb{E}_i\tau_n = \mathbb{E}_j\tau_j = \mathbb{E}_jR_j$ for all $i \in S$. We may now use the strong law of large numbers (Theorem **??**) to conclude that

$$\lim_{N\to\infty} \frac{\tau_1+\tau_2+\dots+\tau_N}{N} = \mathbb{E}_i\tau_2 = \mathbb{E}_j\tau_1 = \mathbb{E}_jR_j \qquad (P_i\text{– a.s.}). \tag{30.37}$$

This may be expressed as follows, let $R_j^{(N)} = \tau_1+\tau_2+\dots+\tau_N$, be the time when the chain first visits j for the N^{th} time, then

$$\lim_{N\to\infty} \frac{R_j^{(N)}}{N} = \mathbb{E}_jR_j \qquad (P_i\text{– a.s.}) \tag{30.38}$$

Let

$$\nu_N = \sum_{n=0}^{N} 1_{X_n} = j$$

be the number of time X_n visits j up to time N. Since j is visited infinitely often, $\nu_N \to \infty$ as $N \to \infty$ and therefore, $\lim_{N\to\infty} \frac{\nu_N+1}{\nu_N} = 1$. Since there were ν_N visits to j in the first N steps, the of the $\nu_N{}^{\text{th}}$ time j was hit is less than or equal to N, i.e. $R_j^{(\nu_N)} \leq N$. Similarly, the time, $R_j^{(\nu_N+1)}$, of the $(\nu_N+1)^{\text{st}}$ visit to j must be larger than N, so we have $R_j^{(\nu_N)} \leq N \leq R_j^{(\nu_N+1)}$. Putting these facts together along with Eq. (30.38) shows that

$$\begin{array}{ccccc} \frac{R_j^{(\nu_N)}}{\nu_N} \leq & \frac{N}{\nu_N} & \leq \frac{R_j^{(\nu_N+1)}}{\nu_N+1}\cdot\frac{\nu_N+1}{\nu_N} & & \\ \downarrow & \downarrow & \downarrow & N\to\infty, \\ \mathbb{E}_jR_j \leq & \lim_{N\to\infty}\frac{N}{\nu_N} \leq & \mathbb{E}_jR_j\cdot 1 & & \end{array}$$

i.e. $\lim_{N\to\infty}\frac{N}{\nu_N} = \mathbb{E}_jR_j$ for P_i – almost every sample path. Taking reciprocals of this last set of inequalities implies Eq. (30.35). ■

Proposition 30.39. *Suppose that X_n is a irreducible, recurrent Markov chain and let $\pi_j = \frac{1}{\mathbb{E}_j(R_j)}$ for all $j \in S$ as in Eq. (30.34). Then either $\pi_i = 0$ for all $i \in S$ (in which case X_n is null recurrent) or $\pi_i > 0$ for all $i \in S$ (in which case X_n is positive recurrent). Moreover if $\pi_i > 0$ then*

$$\sum_{i\in S}\pi_i = 1 \text{ and} \tag{30.39}$$

$$\sum_{i\in S}\pi_iP_{ij} = \pi_j \text{ for all } j \in S. \tag{30.40}$$

That is $\pi = (\pi_i)_{i\in S}$ is the unique stationary distribution for P.

Proof. Let us define

$$T_{ki}^n := \frac{1}{n}\sum_{l=1}^{n} P_{ki}^l \tag{30.41}$$

which, according to Theorem 30.38, satisfies,

$$\lim_{n\to\infty} T_{ki}^n = \pi_i \text{ for all } i,k \in S.$$

Observe that,

$$(T^nP)_{ki} = \frac{1}{n}\sum_{l=1}^{n} P_{ki}^{l+1} = \frac{1}{n}\sum_{l=1}^{n} P_{ki}^{l} + \frac{1}{n}\left[P_{ki}^{n+1} - P_{ki}\right] \to \pi_i \text{ as } n\to\infty.$$

Let $\alpha := \sum_{i\in S}\pi_i$. Since $\pi_i = \lim_{n\to\infty} T_{ki}^n$, Fatou's lemma implies for all $i,j \in S$ that

$$\alpha = \sum_{i\in S}\pi_i = \sum_{i\in S}\liminf_{n\to\infty} T_{ki}^n \leq \liminf_{n\to\infty}\sum_{i\in S} T_{ki}^n = 1$$

and

$$\sum_{i\in S}\pi_iP_{ij} = \sum_{i\in S}\lim_{n\to\infty} T_{li}^nP_{ij} \leq \liminf_{n\to\infty}\sum_{i\in S} T_{li}^nP_{ij} = \liminf_{n\to\infty} T_{lj}^{n+1} = \pi_j$$

where $l \in S$ is arbitrary. Thus

$$\sum_{i\in S}\pi_i =: \alpha \leq 1 \text{ and } \sum_{i\in S}\pi_iP_{ij} \leq \pi_j \text{ for all } j \in S. \tag{30.42}$$

By induction it also follows that

$$\sum_{i\in S}\pi_i P_{ij}^k \le \pi_j \text{ for all } j\in S. \tag{30.43}$$

So if $\pi_j = 0$ for some $j \in S$, then given any $i \in S$, there is a integer k such that $P_{ij}^k > 0$, and by Eq. (30.43) we learn that $\pi_i = 0$. This shows that either $\pi_i = 0$ for all $i \in S$ or $\pi_i > 0$ for all $i \in S$.

For the rest of the proof we assume that $\pi_i > 0$ for all $i \in S$. If there were some $j \in S$ such that $\sum_{i\in S}\pi_i P_{ij} < \pi_j$, we would have from Eq. (30.42) that

$$\alpha = \sum_{i\in S}\pi_i = \sum_{i\in S}\sum_{j\in S}\pi_i P_{ij} = \sum_{j\in S}\sum_{i\in S}\pi_i P_{ij} < \sum_{j\in S}\pi_j = \alpha,$$

which is a contradiction and Eq. (30.40) is proved.

From Eq. (30.40) and induction we also have

$$\sum_{i\in S}\pi_i P_{ij}^k = \pi_j \text{ for all } j\in S$$

for all $k \in \mathbb{N}$ and therefore,

$$\sum_{i\in S}\pi_i T_{ij}^k = \pi_j \text{ for all } j\in S. \tag{30.44}$$

Since $0 \le T_{ij} \le 1$ and $\sum_{i\in S}\pi_i = \alpha \le 1$, we may use the dominated convergence theorem to pass to the limit as $k \to \infty$ in Eq. (30.44) to find

$$\pi_j = \lim_{k\to\infty}\sum_{i\in S}\pi_i T_{ij}^k = \sum_{i\in S}\lim_{k\to\infty}\pi_i T_{ij}^k = \sum_{i\in S}\pi_i\pi_j = \alpha\pi_j.$$

Since $\pi_j > 0$, this implies that $\alpha = 1$ and hence Eq. (30.39) is now verified. ■

Proposition 30.40. *Suppose that P is an irreducible Markov kernel which admits a stationary distribution μ. Then P is positive recurrent and $\mu_j = \pi_j = \frac{1}{\mathbb{E}_j(R_j)}$ for all $j \in S$. In particular, an irreducible Markov kernel has at most one invariant distribution and it has exactly one iff P is positive recurrent.*

Proof. Suppose that $\mu = (\mu_i)$ is a stationary distribution for P, i.e. $\sum_{i\in S}\mu_i = 1$ and $\mu_j = \sum_{i\in S}\mu_i P_{ij}$ for all $j \in S$. Then we also have

$$\mu_j = \sum_{i\in S}\mu_i T_{ij}^k \text{ for all } k\in\mathbb{N} \tag{30.45}$$

where T_{ij}^k is defined above in Eq. (30.41). As in the proof of Proposition 30.39, we may use the dominated convergence theorem to find,

$$\mu_j = \lim_{k\to\infty}\sum_{i\in S}\mu_i T_{ij}^k = \sum_{i\in S}\lim_{k\to\infty}\mu_i T_{ij}^k = \sum_{i\in S}\mu_i\pi_j = \pi_j.$$

Alternative Proof. If P were not positive recurrent then P is either transient or null-recurrent in which case $\lim_{n\to\infty}T_{ij}^n = \frac{1}{\mathbb{E}_j(R_j)} = 0$ for all i, j. So letting $k \to \infty$, using the dominated convergence theorem, in Eq. (30.45) allows us to conclude that $\mu_j = 0$ for all j which contradicts the fact that μ was assumed to be a distribution. ■

Lemma 30.41 (A number theory lemma). *Suppose that 1 is the greatest common denominator of a set of positive integers, $\Gamma := \{n_1, \ldots, n_k\}$. Then there exists $N \in \mathbb{N}$ such that the set,*

$$A = \{m_1 n_1 + \cdots + m_k n_k : m_i \ge 0 \text{ for all } i\},$$

contains all $n \in \mathbb{N}$ with $n \ge N$.

Proof. (The following proof is from Durrett [14].) We first will show that A contains two consecutive positive integers, a and $a+1$. To prove this let,

$$k := \min\{|b-a| : a, b \in A \text{ with } a \ne b\}$$

and choose $a, b \in A$ with $b = a + k$. If $k > 1$, there exists $n \in \Gamma \subset A$ such that k does not divide n. Let us write $n = mk + r$ with $m \ge 0$ and $1 \le r < k$. It then follows that $(m+1)\,b$ and $(m+1)\,a + n$ are in A,

$$(m+1)\,b = (m+1)\,(a+k) > (m+1)\,a + mk + r = (m+1)\,a + n,$$

and

$$(m+1)\,b - (m+1)\,a + n = k - r < k.$$

This contradicts the definition of k and therefore, $k = 1$.

Let $N = a^2$. If $n \ge N$, then $n - a^2 = ma + r$ for some $m \ge 0$ and $0 \le r < a$. Therefore,

$$n = a^2 + ma + r = (a+m)\,a + r = (a+m-r)\,a + r\,(a+1) \in A.$$

■

31

Brownian Motion II (Markov Property)

Definition 31.1. *For $d \in \mathbb{N}$, we say a $\mathbb{R}^d$ – valued process, $\left\{X_t = \left(X_t^1, \dots, X_t^d\right)^{\mathrm{tr}}\right\}_{t\geq 0}$ is a d –* ***dimensional Brownian motion*** *provided $\left\{X^i_\cdot\right\}_{i=1}^d$ is an independent collection of one dimensional Brownian motions.*

Recall from Example 29.12 we know the hypothesis of Theorem 29.6 are verified for Brownian motion and therefore it satisfies the strong Markov property. In this chapter we will develop some properties of Brownian motion based on the strong Markov property. To keep the setup as clean as possible we will now work with the "canonical path space model" for Brownian motion.

Notation 31.2 (Canonical path space) *In what follows, let $\Omega := C\left(\mathbb{R}_+^d, \mathbb{R}^d\right)$ and let $\theta_t : \Omega \to \Omega$ and $B_t : \Omega \to \mathbb{R}^d$ be defined by,*

$$\begin{aligned} \theta_t(\omega) &= \omega(t+\cdot) \text{ and} \\ B_t(\omega) &:= \omega(t) \end{aligned}$$

respectively. Further let $\mathcal{B}_t := \sigma(B_s : s \leq t)$, $\mathcal{B}_{t+} := \cap_{s>t}\mathcal{B}_t$, and $\mathcal{B} := \sigma(B_s : s < \infty) = \vee_{t<\infty}\mathcal{B}_t$.

Definition 31.3. *Let $\{X_t\}_{t\geq 0}$ be a Brownian motion defined on some probability space, $(Y, \mathcal{M}, \nu)$. For $x \in \mathbb{R}^d$, let $\varphi_x : Y \to \Omega$ be defined by,*

$$\varphi_x(y) := ([0,\infty) \ni t \to x + X_t(y)) \in \Omega.$$

Then φ_x is a measurable map and we let $P_x := \nu \circ \varphi_x^{-1}$ for all $x \in \mathbb{R}^d$. When $x = 0$, measure, P_0, is called ***Wiener measure*** *on $(\Omega, \mathcal{B})$. More generally for any probability measure ν on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$, let*

$$P_\nu(A) := \int_{\mathbb{R}^d} P_x(A)\, d\nu(x) \text{ for all } A \in \mathcal{B}. \tag{31.1}$$

Exercise 31.1. For $T > 0$, let $\Omega_T := C([0,T], \mathbb{R})$ which relative to the sup-norm (or uniform norm) $\|\cdot\|_u$, is a Banach space. Show that the Borel σ – algebra, $\mathcal{B}$, on Ω_T is the same as $\mathcal{B}_T := \sigma(B_s : s \leq T)$.

Corollary 31.4. *Let ν be a probability measure on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$, $T > 0$, and let*

$$b_t := B_{t+T} - B_T \text{ for } t \geq 0.$$

Then b is a Brownian motion starting at $0 \in \mathbb{R}^d$ which is independent of $\mathcal{B}_T^+$ relative to the probability measure, P_ν in Eq. (31.1).

Proof. Let $F : \Omega \to \mathbb{R}$ be a bounded $\mathcal{B}$ – measurable function and $A \in \mathcal{B}_T^+$. Since $B_{t+T} = B_t \circ \theta_T$ we may write

$$b_t = (B_t - B_0) \circ \theta_T \text{ and } F(b) = F(B - B_0) \circ \theta_T.$$

So by the Markov property in Theorem 29.6 with $\tau = T$ (a stopping time) and the fact that the P_x – law of $B_\cdot - B_0$ is the P_0 – law of $B_\cdot$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned} \mathbb{E}_\nu\left[F(b)\,|\mathcal{B}_T^+\right] &= \mathbb{E}_\nu\left[F(B - B_0) \circ \theta_T |\mathcal{B}_T^+\right] \\ &= \mathbb{E}_{B_T}\left[F(B - B_0)\right] = \mathbb{E}_0\left[F(B)\right] = \mathbb{E}_0\left[F\right]. \end{aligned}$$

■

Corollary 31.5 (Rapid oscillation of B. M.). *Suppose that $d = 1$. Let $T_+ := \inf\{t > 0 : B_t > 0\}$, $T_- := \inf\{t > 0 : B_t < 0\}$, and $T_0 := \inf\{t > 0 : B_t = 0\}$. Then*

$$P_0(T_\pm = 0) = 1 = P_0(T_0 = 0).$$

In particular, the typical Brownian path revisits 0 for infinitely many $t \in (0, \varepsilon)$ for any $\varepsilon > 0$, see Figure 27.1.

Proof. From Proposition 27.11, we know that T_+, T_-, and T_0 are all optional times and therefore,

$$\{T_+ = 0\}, \ \{T_- = 0\}, \text{ and } \{T_0 = 0\}$$

are all $\mathcal{B}_{0+}$ – measurable. For any $t > 0$, $\{B_t > 0\} \subset \{T_+ \leq t\}$ and therefore,

$$P_0(T_+ \leq t) \geq P(B_t > 0) = \frac{1}{2}.$$

Letting $t \downarrow 0$ in this equation shows $P_0(T_+ = 0) \geq \frac{1}{2}$ and so by Blumenthal $0-1$ law (Theorem 29.17) we know $P_0(T_+ = 0) = 1$. Similarly, or by using $B \overset{d}{=} -B$ under P_0, $P_0(T_- = 0) = 1$ as well. Finally, if $T_+(\omega) = 0$ and $T_-(\omega) = 0$, it follows by the intermediate value theorem that $T_0(\omega) = 0$. Thus $\{T_+ = 0\} \cap \{T_- = 0\} \subset \{T_0 = 0\}$ and therefore $P_0(T_0 = 0) = 1$. ■

Exercise 31.2. Suppose $d = 1$ and $f : (0, 1) \to (0, \infty)$ is a continuous function. Show $X := \limsup_{t\downarrow 0} \frac{B_t}{f(t)}$ is $\mathcal{B}_{0+}$ measurable. Use Blumenthal $0-1$ law (Theorem 29.17), Corollary 31.5, and Lemma 10.49 to conclude that there exists $c \in [0, \infty]$ such that $\limsup_{t\downarrow 0} \frac{B_t}{f(t)} = c$, P_0 – a.s.

Exercise 31.3. Show $\limsup_{t\downarrow 0} \frac{B_t}{t^{1/2}} = \infty$, P_0 – a.s. **Hint:** use a scaling argument to show for any $M \in (0, \infty)$ and $n \in \mathbb{N}$ that

$$P_0\left(\sup_{t\in(0,1/n]} \frac{B_t}{t^{1/2}} \leq M\right) \leq P_0(B_1 \leq M) < 1.$$

The following corollary is the stopping time analogue of Corollary 31.4.

Corollary 31.6. *Let ν be a probability measure on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$, τ be an optional time with $P_\nu(\tau < \infty) > 0$, and let*

$$b_t := B_{t+\tau} - B_\tau \text{ on } \{\tau < \infty\}.$$

Then, conditioned on $\{\tau < \infty\}$, b is a Brownian motion starting at $0 \in \mathbb{R}^d$ which is independent of $\mathcal{B}_\tau^+$. To be more precise we are claiming, for all bounded measurable functions $F : \Omega \to \mathbb{R}^d$ and all $A \in \mathcal{B}_\tau^+$, that

$$\mathbb{E}_\nu[F(b)\,|\tau < \infty] = \mathbb{E}_0[F] \tag{31.2}$$

and

$$\mathbb{E}_\nu[F(b)\,1_A|\tau < \infty] = \mathbb{E}_\nu[F(b)\,|\tau < \infty] \cdot \mathbb{E}_\nu[1_A|\tau < \infty]. \tag{31.3}$$

Proof. On $\{\tau < \infty\}$, $B_{t+\tau} = B_t \circ \theta_\tau$ and therefore,

$$b_t = (B_t - B_0) \circ \theta_\tau \text{ and } F(b) = F(B - B_0) \circ \theta_\tau \text{ on } \{\tau < \infty\}.$$

So by the Markov property in Theorem 29.6 and the fact that the P_x – law of $B_\cdot - B_0$ is the P_0 – law of $B_\cdot$ for all $x \in \mathbb{R}^d$, we have

$$\begin{aligned}\mathbb{E}_\nu\left[1_{\tau<\infty}F(b)\,|\mathcal{B}_\tau^+\right] &= \mathbb{E}_\nu\left[1_{\tau<\infty}F(B - B_0) \circ \theta_\tau|\mathcal{B}_\tau^+\right] \\ &= 1_{\tau<\infty}\mathbb{E}_{B_\tau}[F(B - B_0)] \\ &= 1_{\tau<\infty}\mathbb{E}_0[F(B)] = 1_{\tau<\infty}\mathbb{E}_0[F].\end{aligned}$$

Therefore if $A \in \mathcal{B}_\tau^+$,

$$\begin{aligned}\mathbb{E}_\nu[F(b)\,1_A|\tau < \infty] &= \frac{1}{P_\nu(\tau < \infty)}\mathbb{E}_\nu[1_{\tau<\infty}F(b)\,1_A] \\ &= \mathbb{E}_0[F] \cdot \mathbb{E}_\nu[1_A|\tau < \infty].\end{aligned} \tag{31.4}$$

Taking $A = \Omega$ in this equation proves Eq. (31.2) and then combining Eq. (31.2) and (31.4) proves Eq. (31.3). ■

Proposition 31.7 (Stitching Lemma). *Suppose that $\{B_t\}_{t\geq 0}$ and $\{X_t\}_{t\geq 0}$ are Brownian motions, τ is an optional time, and $\{X_t\}_{t\geq 0}$ is independent of $\mathcal{B}_\tau^+$. Then $\tilde{B}_t := B_{t\wedge\tau} + X_{(t-\tau)_+}$ is another Brownian motion.*

Proof. Given any finite subset, $\Lambda := \{t_i\}_{i=1}^n \subset [0, \infty)$, we must show $(B_{t_1}, \ldots, B_{t_n})$ has the same distribution as $\left(\tilde{B}_{t_1}, \ldots, \tilde{B}_{t_n}\right)$. In checking this we may replace τ by $\tau \wedge M$ where $M := \max \Lambda$ and in this way we may assume without loss of generality that $\{\tau < \infty\} = \Omega$.

By Corollary 31.6, the process, $b_t := B_{t+\tau} - B_\tau$ is again a Brownian motion independent of $\mathcal{B}_\tau^+$ and $B_t = B_t^\tau + b_{(t-\tau)_+}$. As $X \overset{d}{=} b$ and both b and X are independent of $\mathcal{B}_\tau^+$ and therefore independent of (B^τ, τ), it follows that $(B^\tau, \tau, b) \overset{d}{=} (B^\tau, \tau, X)$. So we may conclude that $F(B^\tau, \tau, b) \overset{d}{=} F(B^\tau, \tau, X)$ for any measurable function F. Applying this fact with $F : \Omega \times [0, \infty] \times \Omega \to \mathbb{R}^n$ defined by[1]

$$F(x, t, y) := \left(x_{t_1\wedge t} + y_{(t_1-t)_+}, \ldots, x_{t_n\wedge t} + y_{(t_n-t)_+}\right)$$

shows

$$(B_{t_1}, \ldots, B_{t_n}) = F(B^\tau, \tau, b) \overset{d}{=} F(B^\tau, \tau, X) = \left(\tilde{B}_{t_1}, \ldots, \tilde{B}_{t_n}\right).$$

(In broad strokes the the three steps in the above proof may be summarized as; 1) $B = B^\tau + b_{(\cdot-\tau)}$, 2) $\tilde{B} = B^\tau + X_{(\cdot-\tau)}$, and 3) $(B^\tau, \tau, b) \overset{d}{=} (B^\tau, \tau, X)$.) ■

For the next couple of results we will follow [30, Chapter 13] (also see [31, Section 2.8]) where more results along this line may be found.

Theorem 31.8 (Reflection Principle). *Let τ be an optional time and $\{B_t\}_{t\geq 0}$ be a Brownian motion. Then the "reflected" process (see Figure* **??***),*

$$\tilde{B}_t := B_{t\wedge\tau} - (B_t - B_{t\wedge\tau}) = \begin{cases} B_t & \text{if } t \leq \tau \\ B_\tau - (B_t - B_\tau) & \text{if } t > \tau, \end{cases}$$

is again a Brownian motion.

[1] You should convince yourself that this function is measurable, see Lemma 27.6.

Proof. Again it suffices to show $(B_{t_1},\dots,B_{t_n}) \overset{d}{=} \left(\tilde{B}_{t_1},\dots,\tilde{B}_{t_n}\right)$ for all finite subsets, $\Lambda := \{t_i\}_{i=1}^n \subset [0,\infty)$. As in the proof of Proposition 31.7 we may assume the $\{\tau<\infty\}=\Omega$. By Corollary 31.6, the process, $b_t := B_{t+\tau}-B_\tau$ is a Brownian motion independent of $\mathcal{B}_\tau^+$. it then follows that $X_t := -b_t$ is another Brownian motion independent of $\mathcal{B}_\tau^+$ and since $\tilde{B}_t = B_t^\tau + X_{(t-\tau)_+}$ the proof is finished with an application of Proposition 31.7. ■

Lemma 31.9. *If $d=1$, then*

$$P_0\left(B_t>a\right) \le \min\left(\sqrt{\frac{t}{2\pi a^2}}e^{-a^2/2t}, \frac{1}{2}e^{-a^2/2t}\right)$$

for all $t>0$ and $a>0$.

Proof. This follows from Lemma 7.59 and the fact that $B_t \overset{d}{=} \sqrt{t}N$, where N is a standard normal random variable. ■

Lemma 31.10 (Running maximum). *If B_t is a 1 - dimensional Brownian motion and $z>0$ and $y\ge 0$, then*

$$P\left(\max_{s\le t} B_s \ge z,\ B_t < z-y\right) = P(B_t > z+y) \tag{31.5}$$

and

$$P\left(\max_{s\le t} B_s \ge z\right) = 2P(B_t>z) = P\left(|B_t|>z\right). \tag{31.6}$$

In particular we have $\max_{s\le t} B_s \overset{d}{=} |B_t|$.

Proof. For $z>0$, let $\tau := T_z = \inf\{t>0 : B_t = z\}$ and let $\tilde{B}_t := B_t^\tau - b_{(t-\tau)_+}$ where $b_t := (B_{t\wedge\tau}-B_\tau)\,1_{\tau<\infty}$, see Figure 31.1. (By Corollary 26.19 we actually know that $T_z<\infty$ a.s. but we will not use this fact in the proof.) Observe that $\tilde{\tau} := \inf\left\{t>0 : \tilde{B}_t = z\right\} = \tau$ and that

$$\left\{\max_{s\le t} B_s \ge z\right\} = \{\tau\le\tau\} = \{\tilde{\tau}\le t\}.$$

Furthermore $\{B_t < z-y\} = \left\{\tilde{B}_t > z+y\right\}$ and hence we have

$$\begin{aligned} P\left(\max_{s\le t} B_s \ge z,\ B_t<z-y\right) &= P\left(\tau\le t,\ B_t<z-y\right) = P\left(\tau=\tilde{\tau}\le t, \tilde{B}_t > z+y\right) \\ &= P\left(\tau\le t, B_t>z+y\right) = P\left(B_t>z+y\right) \end{aligned}$$

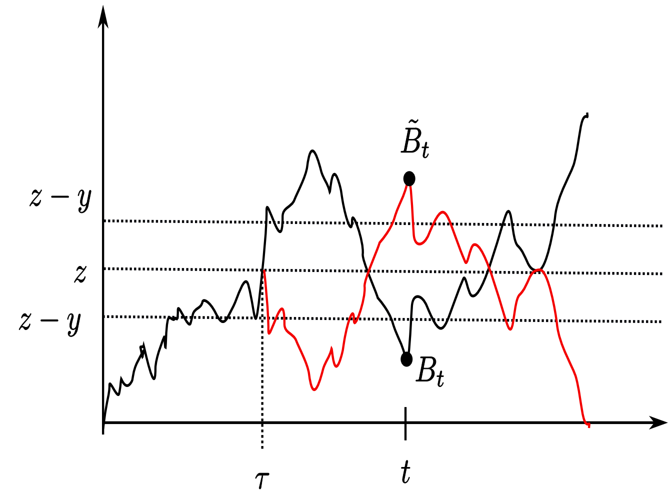


Fig. 31.1. A Brownian path B along with its reflection about the first time, τ, that B_t hits the level z.

wherein we have use $\{B_t>z+y\}\subset\{\tau\le t\}$ for the last equality.

Taking $y=0$ in this estimate shows

$$P\left(\max_{s\le t} B_s\ge z,\ B_t<z\right) = P(B_t>z)$$

and then using $P\left(B_t=z\right)=0$ we may add $P\left(B_t\ge z\right) = P\left(\max_{s\le t} B_s\ge z,\ B_t\ge z\right)$ to both sides of this equation to learn

$$P\left(\max_{s\le t} B_s\ge z\right) = 2P(B_t>z).$$

■

Remark 31.11. Notice that

$$\begin{aligned} P\left(\max_{s\le t} B_s\ge z\right) &= P\left(|B_t|>z\right) = P\left(\sqrt{t}\,|B_1|>z\right) \\ &= P\left(|B_1|>\frac{z}{\sqrt{t}}\right) \to 1 \text{ as } t\to\infty \end{aligned}$$

and therefore it follows that $\sup_{s<\infty} B_s = \infty$ a.s. In particular this shows that $T_z<\infty$ a.s. for all $z\ge 0$ which we have already seen in Corollary 26.19.

Corollary 31.12. *Suppose now that* $T = \inf\{t > 0 : |B_t| = a\}$, *i.e. the first time* B_t *leaves the strip* $(-a, a)$. *Then*

$$\begin{aligned} P_0(T < t) &\le 4P_0(B_t > a) = \frac{4}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t} dx \\ &\le \min\left(\sqrt{\frac{8t}{\pi a^2}} e^{-a^2/2t}, 1\right). \end{aligned} \tag{31.7}$$

Notice that $P_0(T < t) = P_0(B_t^* \ge a)$ *where* $B_t^* = \max\{|B_\tau| : \tau \le t\}$. *So Eq. (31.7) may be rewritten as*

$$P_0(B_t^* \ge a) \le 4P_0(B_t > a) \le \min\left(\sqrt{\frac{8t}{\pi a^2}} e^{-a^2/2t}, 1\right) \le 2e^{-a^2/2t}. \tag{31.8}$$

Proof. By definition $T = T_a \wedge T_{-a}$ so that $\{T < t\} = \{T_a < t\} \cup \{T_{-a} < t\}$ and therefore

$$\begin{aligned} P_0(T < t) &\le P_0(T_a < t) + P_0(T_{-a} < t) \\ &= 2P_0(T_a < t) = 4P_0(B_t > a) = \frac{4}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t} dx \\ &\le \frac{4}{\sqrt{2\pi t}} \int_a^\infty \frac{x}{a} e^{-x^2/2t} dx = \frac{4}{\sqrt{2\pi t}} \left(-\frac{t}{a} e^{-x^2/2t}\right)\Big|_a^\infty = \sqrt{\frac{8t}{\pi a^2}} e^{-a^2/2t}. \end{aligned}$$

This proves everything but the very last inequality in Eq. (31.8). To prove this inequality first observe the elementary calculus inequality:

$$\min\left(\frac{4}{\sqrt{2\pi} y} e^{-y^2/2}, 1\right) \le 2e^{-y^2/2}. \tag{31.9}$$

Indeed Eq. (31.9) holds $\frac{4}{\sqrt{2\pi} y} \le 2$, i.e. if $y \ge y_0 := 2/\sqrt{2\pi}$. The fact that Eq. (31.9) holds for $y \le y_0$ follows from the following trivial inequality

$$1 \le 1.4552 \cong 2e^{-\frac{1}{\pi}} = e^{-y_0^2/2}.$$

Finally letting $y = a/\sqrt{t}$ in Eq. (31.9) gives the last inequality in Eq. (31.8). ∎

Theorem 31.13 (The Dirichlet problem). *Suppose* D *is an open subset of* $\mathbb{R}^d$ *and* $\tau := \inf\{t \ge 0 : B_t \in D^c\}$ *is the first exit time form* D. *Given a bounded measurable function,* $f : \mathrm{bd}(D) \to \mathbb{R}$, *let* $u : D \to \mathbb{R}$ *be defined by (see Figure 31.2),*

$$u(x) := \mathbb{E}_x[f(B_\tau) : \tau < \infty] \text{ for } x \in D.$$

Then $u \in C^\infty(D)$ *and* $\Delta u = 0$ *on* D, *i.e.* u *is a harmonic function.*

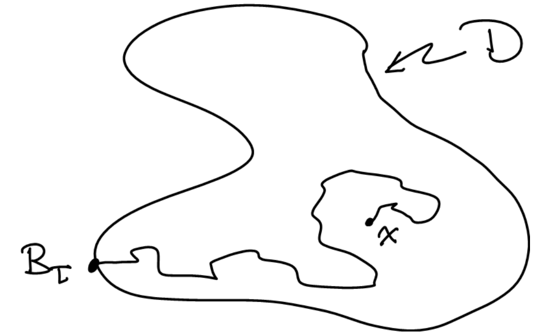


Fig. 31.2. Brownian motion starting at $x \in D$ and exiting on the boundary of D at B_τ.

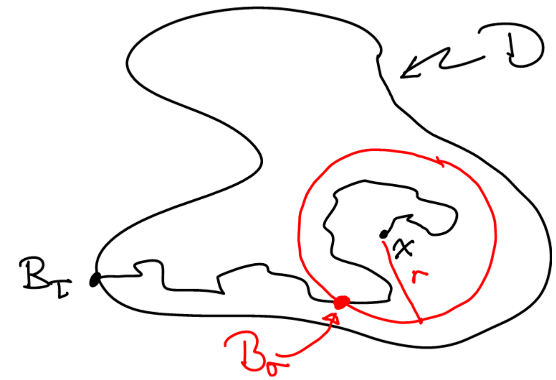


Fig. 31.3. Brownian motion starting at $x \in D$ and exiting the boundary of $B(x, r)$ at B_σ before exiting on the boundary of D at B_τ.

Proof. (Sketch.) Let $x \in D$ and $r > 0$ be such that $B(x, r) \subset D$ and let $\sigma := \inf\{t \ge 0 : B_t \notin B(x, r)\}$ as in Figure 31.3. Setting $F := f(B_\tau) 1_{\{\tau < \infty\}}$, we see that $F \circ \theta_\sigma = F$ on $\{\sigma < \infty\}$ and that $\{\sigma < \infty\}$, P_x - a.s. on $\{\tau < \infty\}$. Moreover, by either Corollary 26.19 or by Lemma 31.10 (see Remark 31.11), we know that $\{\sigma < \infty\}$, P_x – a.s. Therefore by the strong Markov property,

$$\begin{aligned} u(x) &= \mathbb{E}_x[F] = \mathbb{E}_x[F : \sigma < \infty] = \mathbb{E}_x[F \circ \theta_\sigma : \sigma < \infty] \\ &= \mathbb{E}_x[\mathbb{E}_{B_\sigma} F : \sigma < \infty] = \mathbb{E}_x[u(B_\sigma)]. \end{aligned}$$

Using the rotation invariance of Brownian motion, we may conclude that

$$\mathbb{E}_x[u(B_\sigma)] = \frac{1}{\rho(\mathrm{bd}(B(x, r)))} \int_{\mathrm{bd}(B(x,r))} u(y)\, d\rho(y)$$

where ρ denotes surface measure on $\mathrm{bd}(B(x,r))$. This shows that $u(x)$ satisfies the mean value property, i.e. $u(x)$ is equal to its average about in sphere centered at x which is contained in D. It is now a well known that this property, see for example [31, Proposition 4.2.5 on p. 242], that this implies $u \in C^\infty(D)$ and that $\Delta u = 0$. ∎

When the boundary of D is sufficiently regular and f is continuous on $\mathrm{bd}(D)$, it can be shown that, for $x \in \mathrm{bd}(D)$, that $u(y) \to f(x)$ as $y \in D$ tends to x. For more details in this direction, see [2], [3], [31, Section 4.2], [16], and [15].

31.1 Some "Brownian" Martingales

For this chapter, let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}, P\right)$ be a filtered probability space as described in Chapter 27. For the reader's convenience, let us repeat the definition of a (sub or super) martingale here that was given in Definition 28.1.

Definition 31.14. *Given a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, an adapted process, $X_t : \Omega \to \mathbb{R}$, is said to be a $(\{\mathcal{B}_t\})$ **martingale** provided, $\mathbb{E}|X_t| < \infty$ for all t and $\mathbb{E}[X_t - X_s|\mathcal{B}_s] = 0$ for all $0 \leq s \leq t < \infty$. If $\mathbb{E}[X_t - X_s|\mathcal{B}_s] \geq 0$ or $\mathbb{E}[X_t - X_s|\mathcal{B}_s] \leq 0$ for all $0 \leq s \leq t < \infty$, then X is said to be **submartingale or supermartingale** respectively.*

Let us also recall optional sampling Theorem 28.9. In this theorem we assumed that $\{M_t\}_{t\geq 0}$ is a right continuous $\{\mathcal{B}_t\}$ (or $\{\mathcal{B}_t^+\}$) – submartingale (martingale) and suppose that σ and τ are two $\{\mathcal{B}_t\}$ – optional times. (Equivalently put, we are assuming that σ and τ are two $\{\mathcal{B}_t^+\}$ – stopping times.) If there is a constant $K < \infty$ such that $\sigma \leq \tau \leq K$, then

$$M_\sigma \leq \mathbb{E}\left[M_\tau|\mathcal{B}_\sigma^+\right] \tag{31.10}$$

with equality when $\{M_t\}_{t\geq 0}$ is a martingale.

Theorem 31.15. *Suppose that $h : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ is a continuous function such that $\dot{h}(t,x) = \frac{\partial}{\partial t} h(t,x)$, $\nabla_x h(t,x)$, and $\Delta_x h(t,x)$ exist and are continuous on $[0,T]\times\mathbb{R}^d$ and satisfy*

$$\sup_{0\leq t\leq T} \mathbb{E}_\nu\left[|h(t,B_t)| + \left|\dot{h}(t,B_t)\right| + |\nabla_x h(t,B_t)| + |\Delta_x h(t,B_t)|\right] < \infty. \tag{31.11}$$

Then the process,

$$M_t := h(t,B_t) - \int_0^t \left[\dot{h}(\tau,B_\tau) + \frac{1}{2}\Delta_x h(\tau,B_\tau)\right] d\tau \tag{31.12}$$

is a $\{\mathcal{B}_t^+\}_{t\in\mathbb{R}_+}$ – martingale. In particular, if h also satisfies the heat equation in reverse time,

$$\dot{h}(t,x) + \frac{1}{2}\Delta_x h(t,x) = 0, \tag{31.13}$$

then $M_t = h(t,B_t)$ is a martingale.

Proof. Working formally for the moment,

$$\begin{aligned}
\frac{d}{d\tau}\mathbb{E}_\nu\left[h(\tau,B_\tau)|\mathcal{B}_{s+}\right] &= \frac{d}{d\tau}\left[e^{\frac{\tau-s}{2}\Delta} h(\tau,B_s)\right] \\
&= e^{\frac{\tau-s}{2}\Delta}\dot{h}(\tau,B_s) + \frac{1}{2}e^{\frac{\tau-s}{2}\Delta}\Delta h(\tau,B_s) \\
&= \mathbb{E}_\nu\left[\dot{h}(\tau,B_\tau) + \frac{1}{2}\Delta h(\tau,B_\tau)|\mathcal{B}_{s+}\right].
\end{aligned}$$

Integrating this equation on $\tau \in [s,t]$ then shows

$$\mathbb{E}_\nu\left[h(t,B_t) - h(s,B_s)|\mathcal{B}_{s+}\right] = \mathbb{E}_\nu\left[\int_s^t \left[\dot{h}(\tau,B_\tau) + \frac{1}{2}\Delta h(\tau,B_\tau)\right] d\tau|\mathcal{B}_{s+}\right].$$

This statement is equivalent to the statement that $\mathbb{E}_\nu\left[M_t - M_s|\mathcal{B}_{s+}\right] = 0$ i.e. to the assertion that $\{M_t\}_{t\geq 0}$ is a martingale. We now need to justify the above computations.

1. Let us first suppose there exists an $R < \infty$ such that $h(t,x) = 0$ if $|x| = \sqrt{\sum_{i=1}^d x_i^2} \geq R$. Using Corollary 7.30 and a couple of integration by parts, we find

$$\begin{aligned}
\frac{d}{d\tau}\left(e^{\frac{\tau-s}{2}\Delta} h\right)(\tau,x) &= \int_{\mathbb{R}^d} \frac{d}{d\tau}\left[p_{\tau-s}(x-y)\, h(\tau,y)\right] dy \\
&= \int_{\mathbb{R}^d} \left[\frac{1}{2}(\Delta p_{\tau-s})(x-y)\, h(\tau,y) + p_{\tau-s}(x-y)\,\dot{h}(\tau,y)\right] dy \\
&= \int_{\mathbb{R}^d} \left[\frac{1}{2}p_{\tau-s}(x-y)\,\Delta_y h(\tau,y) + p_{\tau-s}(x-y)\,\dot{h}(\tau,y)\right] dy \\
&= \int_{\mathbb{R}^d} p_{\tau-s}(x-y)\,\Box h(\tau,y)\, dy =: \left(e^{\frac{\tau-s}{2}\Delta}\Box h\right)(\tau,x),
\end{aligned} \tag{31.14}$$

where

$$\Box h(\tau,y) := \dot{h}(\tau,y) + \frac{1}{2}\Delta_y h(\tau,y). \tag{31.15}$$

Since

$$\left(e^{\frac{\tau-s}{2}\Delta} h\right)(\tau,x) := \int_{\mathbb{R}^d} p_{\tau-s}(x-y)\, h(\tau,y)\, dy = \mathbb{E}_\nu\left[h\left(\tau, x + \sqrt{\tau-s}N\right)\right]$$

where $N \overset{d}{=} N(0, I_{d\times d})$, we see that $\left(e^{\frac{\tau-s}{2}\Delta}h\right)(\tau,x)$ (and similarly that $\square h(\tau,x)$) is a continuous function in (τ,x) for $\tau \geq s$. Moreover, $\left(e^{\frac{\tau-s}{2}\Delta}h\right)(\tau,x)|_{\tau=s} = h(s,x)$ and so by the fundamental theorem of calculus (using Eq. (31.14

$$\left(e^{\frac{\tau-s}{2}\Delta}h\right)(\tau,x) = h(s,x) + \int_s^t \left(e^{\frac{\tau-s}{2}\Delta}\square h\right)(\tau,x)\,d\tau. \tag{31.16}$$

Hence for $A \in \mathcal{B}_{s+}$,

$$\begin{aligned}\mathbb{E}_\nu[h(t,B_t):A] &= \mathbb{E}_\nu[\mathbb{E}_\nu[h(t,B_t)|\mathcal{B}_{s+}]:A]\\ &= \mathbb{E}_\nu\left[\left(e^{\frac{\tau-s}{2}\Delta}h\right)(\tau,B_s):A\right]\\ &= \mathbb{E}_\nu\left[h(s,B_s) + \int_s^t \left(e^{\frac{\tau-s}{2}\Delta}\square h\right)(\tau,B_s)\,d\tau : A\right]\\ &= \mathbb{E}_\nu[h(s,B_s):A] + \int_s^t \mathbb{E}_\nu\left[\left(e^{\frac{\tau-s}{2}\Delta}\square h\right)(\tau,B_s):A\right]d\tau.\end{aligned} \tag{31.17}$$

Since

$$\left(e^{\frac{\tau-s}{2}\Delta}\square h\right)(\tau,B_s) = \mathbb{E}_\nu[(\square h)(\tau,B_\tau)|\mathcal{B}_{s+}],$$

we have

$$\mathbb{E}_\nu\left[\left(e^{\frac{\tau-s}{2}\Delta}\square h\right)(\tau,B_s):A\right] = \mathbb{E}_\nu[(\square h)(\tau,B_\tau):A]$$

which combined with Eq. (31.17) shows,

$$\begin{aligned}\mathbb{E}_\nu[h(t,B_t):A] &= \mathbb{E}_\nu[h(s,B_s):A] + \int_s^t \mathbb{E}_\nu[(\square h)(\tau,B_\tau):A]\,d\tau\\ &= \mathbb{E}_\nu\left[h(s,B_s) + \int_s^t (\square h)(\tau,B_\tau)\,d\tau : A\right].\end{aligned}$$

This proves $\{M_t\}_{t\geq 0}$ is a martingale when $h(t,x) = 0$ if $|x| \geq R$.

2. For the general case, let $\varphi \in C_c^\infty(\mathbb{R}^d,[0,1])$ such that $\varphi = 1$ in a neighborhood of $0 \in \mathbb{R}^d$ and for $n \in \mathbb{N}$, let $\varphi_n(x) := \varphi(x/n)$. Observe that $\varphi_n \to 1$, $\nabla\varphi_n(x) = \frac{1}{n}(\nabla\varphi)(x/n)$, and $\Delta\varphi_n(x) = \frac{1}{n^2}(\Delta\varphi)(x/n)$ all go to zero boundedly as $n \to \infty$. Applying case 1. to $h_n(t,x) = \varphi_n(x)h(t,x)$ we find that

$$M_t^n := \varphi_n(B_t)h(t,B_t) - \int_0^t \varphi_n(B_\tau)\left[\dot h(\tau,B_\tau) + \frac{1}{2}\Delta h(\tau,B_\tau)\right]d\tau + \varepsilon_n(t)$$

is a martingale, where

$$\varepsilon_n(t) := \int_0^t \left[\frac{1}{2}\Delta\varphi_n(B_\tau)h(\tau,B_\tau) + \nabla\varphi_n(B_\tau)\cdot\nabla h(\tau,B_\tau)\right]d\tau.$$

By DCT,

$$\begin{aligned}\mathbb{E}_\nu|\varepsilon_n(t)| \leq& \int_0^t \frac{1}{2}\mathbb{E}_\nu|\Delta\varphi_n(B_\tau)h(\tau,B_\tau)|\,d\tau\\ &+ \int_0^t \mathbb{E}_\nu[|\nabla\varphi_n(B_\tau)||\nabla h(\tau,B_\tau)|]\,d\tau \to 0 \text{ as } n\to\infty,\end{aligned}$$

and similarly,

$$\begin{aligned}\mathbb{E}_\nu|\varphi_n(B_t)h(t,B_t) - h(t,B_t)| &\to 0,\\ \int_0^t \mathbb{E}_\nu\left|\varphi_n(B_\tau)\dot h(\tau,B_\tau) - \dot h(\tau,B_\tau)\right|d\tau &\to 0, \text{ and}\\ \frac{1}{2}\int_0^t \mathbb{E}_\nu|\varphi_n(B_\tau)\Delta h(\tau,B_\tau) - \Delta h(\tau,B_\tau)|\,d\tau &\to 0\end{aligned}$$

as $n \to \infty$. From these comments one easily sees that $\mathbb{E}_\nu|M_t^n - M_t| \to 0$ as $n \to \infty$ which is sufficient to show $\{M_t\}_{t\in[0,T]}$ is still a martingale. ■

Rewriting Eq. (31.12) we have

$$h(t,B_t) = M_t + \int_0^t \left[\dot h(\tau,B_\tau) + \frac{1}{2}\Delta_x h(\tau,B_\tau)\right]d\tau.$$

Considering the simplest case where $h(t,x) = f(x)$ and $x \in \mathbb{R}$, i.e. $d = 1$, we have

$$f(B_t) = M_t + \frac{1}{2}\int_0^t f''(B_t)\,dt$$

where M_t is a martingale provided,

$$\sup_{0\leq t\leq T}\mathbb{E}_\nu[|f(B_t)| + |f'(B_t)| + |f''(B_t)|] < \infty.$$

From this it follows that if $f'' \geq 0$ (i.e. f is subharmonic) then $f(B_t)$ is a submartingale and if $f'' \leq 0$ (i.e. f is super harmonic) then $f(B_t)$ is a supermartingale. More precisely, if $f : \mathbb{R}^d \to \mathbb{R}$ is a C^2 function, then $f(B_t)$ is a "local" submartingale (supermartingale) iff $\Delta f \geq 0$ $(\Delta f \leq 0)$.

Corollary 31.16. *Suppose $h : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$ is a C^2 – function such that $\square h = 0$ and both h and h^2 satisfy the hypothesis of Theorem 31.15. If we let M_t denote the martingale, $M_t := h(t,B_t)$ and*

$$A_t := \int_0^t |\nabla_x h(\tau,B_\tau)|^2\,d\tau,$$

then $N_t := M_t^2 - A_t$ is a martingale. Thus the submartingale has the "Doob" decomposition,

$$M_t^2 = N_t + A_t.$$

and we call the increasing process, A_t, the ***compensator to*** *M_t^2.*

Proof. We need only apply Theorem 31.15 to h^2. In order to do this we need to compute $\Box h^2$. Since

$$\partial_i^2 h^2 = \partial_i (2h \cdot \partial_i h) = 2 (\partial_i h)^2 + 2h\partial_i h$$

we see that $\frac{1}{2}\Delta h^2 = h\Delta h + |\nabla h|^2$. Therefore,

$$\Box h^2 = 2h\dot{h} + h\Delta h + |\nabla h|^2 = |\nabla h|^2$$

and hence the proposition follows form Eq. (31.12) with h replaced by h^2. ■

Exercise 31.4 (h – transforms of B_t). Let $\{B_t\}_{t\in\mathbb{R}_+}$ be a d – dimensional Brownian motion. Show the following processes are martingales;

1. $M_t = u(B_t)$ where $u : \mathbb{R}^d \to \mathbb{R}$ is a Harmonic function, $\Delta u = 0$, such that $\sup_{t\le T} \mathbb{E}[|u(B_t)| + |\nabla u(B_t)|] < \infty$ for all $T < \infty$.
2. $M_t = \lambda \cdot B_t$ for all $\lambda \in \mathbb{R}^d$.
3. $M_t = e^{-\lambda Y_t} \cos(\lambda X_t)$ and $M_t = e^{-\lambda Y_t} \sin(\lambda X_t)$ where $B_t = (X_t, Y_t)$ is a two dimensional Brownian motion and $\lambda \in \mathbb{R}$.
4. $M_t = |B_t|^2 - d \cdot t$.
5. $M_t = (a \cdot B_t)(b \cdot B_t) - (a \cdot b)\, t$ for all $a, b \in \mathbb{R}^d$.
6. $M_t := e^{\lambda\cdot B_t - |\lambda|^2 t/2}$ for any $\lambda \in \mathbb{R}^d$.

Exercise 31.5 (Compensators). Let $\{B_t\}_{t\in\mathbb{R}_+}$ be a d – dimensional Brownian motion. Find the compensator, A_t, for each of the following square integrable martingales.

1. $M_t = u(B_t)$ where $u : \mathbb{R}^d \to \mathbb{R}$ is a harmonic function, $\Delta u = 0$, such that
$$\sup_{t\le T} \mathbb{E}\left[\left|u^2(B_t)\right| + |\nabla u(B_t)|^2\right] < \infty \text{ for all } T < \infty.$$
2. $M_t = \lambda \cdot B_t$ for all $\lambda \in \mathbb{R}^d$.
3. $M_t = |B_t|^2 - d \cdot t$.
4. $M_t := e^{\lambda\cdot B_t - |\lambda|^2 t/2}$ for any $\lambda \in \mathbb{R}^d$.

For the next three exercises let $a < 0 < b$,

$$T_y := \inf\{t > 0 : B_t = y\} \ \forall\, y \in \mathbb{R}, \tag{31.18}$$

and $\tau := T_a \wedge T_b$. Recall from the law of large numbers for Brownian motions (Corollary 26.19) that $P_0(\tau < \infty) = 1$. This may also be deduced from Lemma 31.10 – but this is using a rather large hammer to conclude a simple result.

Remark 31.17. In applying the optional sampling Theorem 28.9, observe that if τ is any **optional** time and $K < \infty$ is a constant, then $\tau \wedge K$ is a bounded optional time. Indeed,

$$\{\tau \wedge K < t\} = \begin{cases} \{\tau < t\} \in \mathcal{B}_t & \text{if } t \le K \\ \Omega \in \mathcal{B}_t & \text{if } t > K. \end{cases}$$

Therefore if σ and τ are any optional times with $\sigma \le \tau$, we may apply Theorem 28.9 with σ and τ replaced by $\sigma \wedge K$ and $\tau \wedge K$. One may then try to pass to limit as $K \uparrow \infty$ in the resulting identity.

Exercise 31.6 (Compare with Exercise 18.23). Let $-\infty < a < 0 < b < \infty$. Show

$$P_0(T_b < T_a) = \frac{-a}{b-a} = \frac{|a|}{b+|a|}.$$

Use this to conclude that $P_0(T_b < \infty) = 1$. **Hint:** consider the optional sampling Theorem 28.9 with $M_t = B_t$.

Exercise 31.7 (Compare with Exercise 18.24). Let $-\infty < a < 0 < b < \infty$. Apply the optional sampling Theorem 28.9 to the martingale, $M_t = B_t^2 - t$ to conclude that

$$\mathbb{E}_0[T_a \wedge T_b] = \mathbb{E}_0\tau = |a|\, b.$$

Use this to conclude that $\mathbb{E}_0[T_b] = \infty$ for all $b > 0$.

Exercise 31.8. By considering the martingale,

$$M_t := e^{\lambda B_t - \frac{1}{2}\lambda^2 t},$$

show

$$\mathbb{E}_0\left[e^{-\lambda T_a}\right] = e^{-a\sqrt{2\lambda}}. \tag{31.19}$$

Remark 31.18. Equation (31.19) may also be proved using Lemma 31.10 directly. Indeed, if

$$p_t(x) := \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t},$$

then

$$\begin{aligned}\mathbb{E}_0\left[e^{-\lambda T_a}\right] &= \mathbb{E}_0\left[\int_{T_a}^\infty \lambda e^{-\lambda t} dt\right] = \mathbb{E}_0\left[\int_0^\infty 1_{T_a<t}\lambda e^{-\lambda t} dt\right] \\ &= \mathbb{E}_0\left[\int_0^\infty P(T_a < t)\,\lambda e^{-\lambda t} dt\right] \\ &= -2\int_0^\infty \left[\int_a^\infty p_t(x)\, dx\right] \frac{d}{dt} e^{-\lambda t} dt.\end{aligned}$$

Integrating by parts in t, shows

$$\begin{aligned}-2\int_0^\infty\left[\int_a^\infty p_t(x)\,dx\right]\frac{d}{dt}e^{-\lambda t}dt &= 2\int_0^\infty\left[\int_a^\infty \frac{d}{dt}p_t(x)\,dx\right]e^{-\lambda t}dt\\ &= 2\int_0^\infty\left[\int_a^\infty \frac{1}{2}p_t''(x)\,dx\right]e^{-\lambda t}dt\\ &= -\int_0^\infty p_t'(a)\,e^{-\lambda t}dt\\ &= -\int_0^\infty \frac{1}{\sqrt{2\pi t}}\frac{-a}{t}e^{-a^2/2t}e^{-\lambda t}dt\\ &= \frac{a}{\sqrt{2\pi}}\int_0^\infty t^{-3/2}e^{-a^2/2t}e^{-\lambda t}dt\end{aligned}$$

where the last integral may be evaluated to be $e^{-a\sqrt{2\lambda}}$ using the Laplace transform function of Mathematica.

Alternatively, consider

$$\int_0^\infty p_t(x)\,e^{-\lambda t}dt=\int_0^\infty\frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}e^{-\lambda t}dt=\frac{1}{\sqrt{2\pi}}\int_0^\infty\frac{1}{\sqrt{t}}\exp\left(-\frac{x^2}{2t}-\lambda t\right)dt$$

which is evaluated in Exercise 34.1 in the real analysis notes. Here is a brief sketch.

Using the theory of the Fourier transform (or characteristic functions if you prefer),

$$\begin{aligned}\frac{\sqrt{2\pi}}{2m}e^{-m|x|} &= \int_{\mathbb{R}}\left(|\xi|^2+m^2\right)^{-1}e^{i\xi x}\mathbf{d}\xi\\ &= \int_{\mathbb{R}}\left[\int_0^\infty e^{-\lambda(|\xi|^2+m^2)}d\lambda\right]e^{i\xi x}\mathbf{d}\xi\\ &= \int_0^\infty d\lambda\left[e^{-\lambda m^2}\int_{\mathbb{R}}\mathbf{d}\xi e^{-\lambda|\xi|^2}e^{i\xi x}\right]\\ &= \int_0^\infty e^{-\lambda m^2}(2\lambda)^{-1/2}e^{-\frac{1}{4\lambda}x^2}d\lambda,\end{aligned}$$

where $\mathbf{d}\xi := (2\pi)^{-1/2}\,d\xi$. Now make appropriate change of variables and carry out the remaining x integral to arrive at the result.

The Feynman-Kac Formula

The goal of this section is to more deeply explore the connection between Brownian motion and certain heat type equations. We will start with some basic facts about the heat equation and give some heuristics indicating how one might solve these equations via Brownian motion. The resulting probabilistic representation is referred to as the Feynman-Kac formula. At the end we will see how to make the heuristics rigorous. We begin with the heat equation.

32.1 The Heat Equation

Suppose that $\Omega \subset \mathbb{R}^d$ is a region of space filled with a material, $\rho(x)$ is the density of the material at $x \in \Omega$ and $c(x)$ is the heat capacity. Let $u(x,t)$ denote the temperature at time $t \in [0,\infty)$ at the spatial point $x \in \Omega$. Now suppose that $B \subset \mathbb{R}^d$ is a subregion in Ω, ∂B is the boundary of B, and $E_B(t)$ is the heat energy contained in the volume B at time t. Then

$$E_B(t) = \int_B \rho(x)c(x)u(t,x)dx.$$

So on one hand (writing $\dot{f}(t)$ for $\frac{\partial f(t)}{\partial t}$),

$$\dot{E}_B(t) := \frac{d}{dt}E_B(t) = \int_B \rho(x)c(x)\dot{u}(t,x)dx \tag{32.1}$$

while on the other hand,

$$\dot{E}_B(t) = \int_{\partial B} \langle G(x)\nabla u(t,x), n(x)\rangle d\sigma(x), \tag{32.2}$$

where $G(x)$ is a $n \times n$–positive definite matrix representing the conduction properties of the material, $n(x)$ is the outward pointing normal to B at $x \in \partial B$, and $d\sigma$ denotes surface measure on ∂B. (We are using $\langle\cdot,\cdot\rangle$ to denote the standard dot product on $\mathbb{R}^d$ and $\nabla u(x,t) = \left(\frac{\partial u(x,t)}{\partial x_1}, \dots \frac{\partial u(x,t)}{\partial x_n}\right)^{\mathrm{tr}}$.)

In order to see that we have the sign correct in Eq. (32.2), suppose that $x \in \partial B$ and

$$\langle G(x)\nabla u(t,x), n(x)\rangle = \langle \nabla u(t,x), G(x)n(x)\rangle > 0.$$

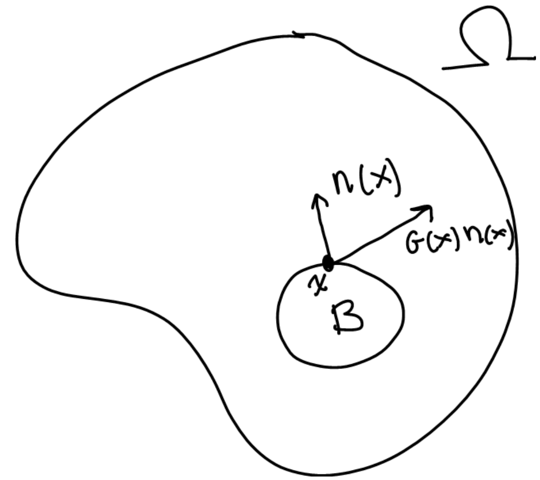


Fig. 32.1. The geometry of the test region, $B \subset \Omega$.

In this case temperature is increasing as one move from x in the direction of $G(x)n(x)$ and since $\langle n(x), G(x)n(x)\rangle > 0$, $G(x)n(x)$ is outward pointing to B. Thus if $\langle G(x)\nabla u(t,x), n(x)\rangle > 0$, near x, it is hotter outside B then inside and hence heat energy is flowing into B in a neighborhood of the ∂B near x, see Figure 32.1.

Comparing Eqs. (32.1) to (32.2) after an application of the divergence theorem shows that

$$\int_B \rho(x)c(x)\dot{u}(t,x)dx = \int_B \nabla\cdot(G(\cdot)\nabla u(t,\cdot))(x)\,dx. \tag{32.3}$$

Since this holds for all nice volumes $B \subset \Omega$, we conclude that the temperature functions should satisfy the following partial differential equation.

$$\rho(x)c(x)\dot{u}(t,x) = \nabla\cdot(G(\cdot)\nabla u(t,\cdot))(x)\,. \tag{32.4}$$

or equivalently that

$$\dot{u}(t,x) = \frac{1}{\rho(x)c(x)}\nabla\cdot(G(x)\nabla u(t,x)). \tag{32.5}$$

Setting $g^{ij}(x) := G_{ij}(x)/(\rho(x)c(x))$ and

$$z^j(x) := \sum_{i=1}^n \partial(G_{ij}(x)/(\rho(x)c(x)))/\partial x^i$$

the above equation may be written as:

$$\dot{u}(t,x) = Lu(t,x), \tag{32.6}$$

where

$$(Lf)(x) = \sum_{i,j} g^{ij}(x)\frac{\partial^2}{\partial x^i \partial x^j} f(x) + \sum_j z^j(x)\frac{\partial}{\partial x^j} f(x). \tag{32.7}$$

The operator L is a prototypical example of a second order "elliptic" differential operator. In the next section we will consider the special case of the standard Laplacian, Δ, on $\mathbb{R}^d$, i.e. $z^j \equiv 0$ and $g^{ij} = \delta_{ij}$ so that

$$\Delta := \sum_{i=1}^n \frac{\partial^2}{(\partial x^i)^2}. \tag{32.8}$$

32.2 Solving the heat equation on $\mathbb{R}^n$.

Let

$$\hat{f}(k) = (\mathcal{F}f)(k) = \left(\frac{1}{2\pi}\right)^n \int_{\mathbb{R}^d} f(x) e^{-ik\cdot x} dx$$

be the Fourier transform of f and

$$f^\vee(x) = \left(\mathcal{F}^{-1}f\right)(x) = \int_{\mathbb{R}^d} f(k) e^{ik\cdot x} dk$$

be the inverse Fourier transform. For f nice enough (or in the sense of tempered distributions), we know that

$$f(x) = \int_{\mathbb{R}^d} \hat{f}(k) e^{ik\cdot x} dk = \left(\mathcal{F}^{-1}\hat{f}\right)(x).$$

Also recall that the Fourier transform and the convolution operation are related by;

$$\begin{aligned}\widehat{f * g}(k) &= \left(\frac{1}{2\pi}\right)^n \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x-y)g(y)e^{-ik\cdot x} dx\ dy \\ &= \left(\frac{1}{2\pi}\right)^n \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x)g(y)e^{-ik\cdot(x+y)} dx\ dy = (2\pi)^n \hat{f}(k)\hat{g}(k).\end{aligned}$$

Inverting this relation gives the relation,

$$\mathcal{F}^{-1}\left(\hat{f}\cdot\hat{g}\right)(x) = \left(\frac{1}{2\pi}\right)^n (f*g)(x).$$

The heat equation for a function $u : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{C}$ is the partial differential equation

$$\left(\partial_t - \frac{1}{2}\Delta_x\right) u(t,x) = 0 \text{ with } u(0,x) = f(x), \tag{32.9}$$

where f is a given function on $\mathbb{R}^n$. By Fourier transforming Eq. (32.9) in the x – variables only, one finds (after some integration by parts) that (32.9) implies that

$$\left(\partial_t + \frac{1}{2}|k|^2\right)\hat{u}(t,k) = 0 \text{ with } \hat{u}(0,k) = \hat{f}(k). \tag{32.10}$$

Solving for $\hat{u}(t,k)$ gives

$$\hat{u}(t,k) = e^{-t|k|^2/2}\hat{f}(k).$$

Inverting the Fourier transform then shows that

$$u(t,x) = \mathcal{F}^{-1}\left(e^{-t|k|^2/2}\hat{f}(k)\right)(x) = \left(\frac{1}{2\pi}\right)^n \left(\mathcal{F}^{-1}\left(e^{-t|k|^2/2}\right) * f\right)(x). \tag{32.11}$$

Let

$$g(x) := \left(\mathcal{F}^{-1}e^{-\frac{t}{2}|k|^2}\right)(x) = \int_{\mathbb{R}^d} e^{-\frac{t}{2}|k|^2} e^{ik\cdot x} dk.$$

Making the change of variables, $k \to k/t^{1/2}$ and using a standard Gaussian integral formula gives

$$\begin{aligned} g(x) &= \left(\frac{1}{t}\right)^{n/2} \int_{\mathbb{R}^d} e^{-\frac{1}{2}|k|^2} e^{ik\cdot x/\sqrt{t}} dk \\ &= \left(\frac{2\pi}{t}\right)^{n/2}\left(\frac{1}{2\pi}\right)^{n/2} \int_{\mathbb{R}^d} e^{-\frac{1}{2}|k|^2} e^{ik\cdot x/\sqrt{t}} dk \\ &= \left(\frac{2\pi}{t}\right)^{n/2} \exp\left(-\frac{1}{2}\left|\frac{x}{\sqrt{t}}\right|^2\right) = \left(\frac{2\pi}{t}\right)^{n/2} \exp\left(-\frac{1}{2t}|x|^2\right). \end{aligned} \tag{32.12}$$

Using this result in Eq. (32.11) implies

$$u(t,x) = \int_{\mathbb{R}^n} p_t(x-y) f(y) dy$$

where

$$p_t(x) := \left(\frac{1}{2\pi t}\right)^{n/2} \exp\left(-\frac{1}{2t}|x|^2\right). \tag{32.13}$$

This suggests the following theorem.

Theorem 32.1. *Let $p_t(x)$ be the **heat kernel** on $\mathbb{R}^n$ defined in Eq. (32.13). Then*

$$\left(\partial_t - \frac{1}{2}\Delta_x\right) p_t(x-y) = 0 \text{ and } \lim_{t\downarrow 0} p_t(x-y) = \delta_x(y), \tag{32.14}$$

where δ_x is the δ – function at x in $\mathbb{R}^n$. More precisely, if f is a bounded continuous function on $\mathbb{R}^n$, then

$$u(t,x) = \int_{\mathbb{R}^n} p_t(x-y) f(y) dy \tag{32.15}$$

is a solution to Eq. (32.9) and $\lim_{t\downarrow 0} u(t,x) = f(x)$ uniformly for $x \in K$, where K is any compact subset of $\mathbb{R}^n$.

Proof. Direct computations show that $\left(\partial_t - \frac{1}{2}\Delta_x\right) p_t(x-y) = 0$ which coupled with a few applications of Corollary 7.30 shows $\left(\partial_t - \frac{1}{2}\Delta_x\right) u(t,x) = 0$ for $t>0$. After a making the changes of variables, $y \to x-y$ and then $y \to \sqrt{t}y$, Eq. (32.15) may be written as

$$u(t,x) = \int_{\mathbb{R}^n} p_t(y) f(x-y) dy = \int_{\mathbb{R}^n} p_1(y) f(x-\sqrt{t}y) dy$$

and therefore,

$$\begin{aligned} |u(t,x) - f(x)| &= \left| \int_{\mathbb{R}^n} p_1(y) \left[f(x-\sqrt{t}y) - f(x) \right] dy \right| \\ &\le \int_{\mathbb{R}^n} p_1(y) \left| f(x-\sqrt{t}y) - f(x) \right| dy. \end{aligned} \tag{32.16}$$

For $R>0$,

$$\sup_{x\in K} \int_{|y|\le R} p_1(y) \left| f(x-\sqrt{t}y) - f(x) \right| dy \le \sup_{x\in K} \sup_{|y|\le R} \left| f(x-\sqrt{t}y) - f(x) \right| \to 0$$

as $t \downarrow 0$ by the uniform continuity of f on compact sets. If $M = \sup_{x\in\mathbb{R}^n} |f(x)|$, then by Chebyschev's inequality,

$$\int_{|y|>R} p_1(y) \left| f(x-\sqrt{t}y) - f(x) \right| dy \le 2M \int_{|y|>R} p_1(y) dy \le C\frac{2M}{R}$$

where $C := \int_{\mathbb{R}^n} |y| p_1(y) dy$. Hence we have shown,

$$\limsup_{t\downarrow 0} \sup_{x\in K} |u(t,x) - f(x)| \le C\frac{2M}{R} \to 0 \text{ as } R \to \infty.$$

This shows that $\lim_{t\downarrow 0} u(t,x) = f(x)$ uniformly on compact subsets of $\mathbb{R}^n$. ■

Notation 32.2 *We will let $\left(e^{t\Delta/2} f\right)(x)$ be defined by*

$$\left(e^{t\Delta/2} f\right)(x) = \int_{\mathbb{R}^n} p_t(x-y) f(y) dy = (p_t * f)(x).$$

Hence for nice enough f (for example when f is bounded and continuous), $u(t,x) := \left(e^{t\Delta/2} f\right)(x)$ solves the heat equation in Eq. (32.9).

Exercise 32.1 (Semigroup Property). Verify the semi-group identity for p_t;

$$p_{t+s} = p_s * p_t \text{ for all } s,t>0. \tag{32.17}$$

Proposition 32.3 (Properties of $e^{t\Delta/2}$). *Let $t \in (0,\infty)$, then;*

1. *for $f \in L^p(\mathbb{R}^n, dx)$ with $1 \le p \le \infty$, the function*
$$\left(e^{t\Delta/2} f\right)(x) = \int_{\mathbb{R}^n} f(y) \frac{e^{-\frac{1}{2t}|x-y|^2}}{(2\pi t)^{n/2}} dy$$
is smooth[1] *in (t,x) for $t>0$ and $x \in \mathbb{R}^n$.*
2. *$e^{t\Delta/2}$ acts as a contraction on $L^p(\mathbb{R}^n, dx)$ for all $p \in [0,\infty]$ and $t>0$.*
3. *For $p \in [0,\infty)$, $e^{t\Delta/2} f = p_t * f \to f$ in $L^p(\mathbb{R}^n, dx)$ as $t \to 0$.*

Proof. Item 1. follows by multiple applications of Corollary 7.30.
Item 2.

$$|(p_t * f)(x)| \le \int_{\mathbb{R}^n} |f(y)| p_t(x-y) dy$$

and hence with the aid of Jensen's inequality we have,

$$\|p_t * f\|_{L^p}^p \le \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |f(y)|^p p_t(x-y) dy dx = \|f\|_{L^p}^p$$

So p_t is a contraction $\forall t>0$.

Item 3. First, let us suppose that $f \in C_c(\mathbb{R}^n)$. From Eq. (32.16) along with Jensen's inequality, we find

$$\begin{aligned} \int_{\mathbb{R}^n} |(p_t * f)(x) - f(x)|^p dx &\le \int_{\mathbb{R}^n} dx \int_{\mathbb{R}^n} p_1(y) \left| f(x-\sqrt{t}y) - f(x) \right|^p dy \\ &= \int_{\mathbb{R}^n} dy\, p_1(y) \int_{\mathbb{R}^n} dx \left| f(x-\sqrt{t}y) - f(x) \right|^p. \end{aligned} \tag{32.18}$$

[1] In fact, $u(t,x)$ is real analytic for $x \in \mathbb{R}^n$ and $t>0$. One just notices that $p_t(x-y)$ analytically continues to $\operatorname{Re} t > 0$ and $x \in \mathbb{C}^n$ and then shows that it is permissible to differentiate under the integral.

Since

$$g\left(t,y\right):=\int_{\mathbb{R}^n}\left|f(x-\sqrt{t}y)-f\left(x\right)\right|^p dx\leq 2^{p(p-1)^{-1}}\int_{\mathbb{R}^n}\left|f\left(x\right)\right|^p dx$$

and $\lim_{t\downarrow 0}g\left(t,y\right)=0$, we may pass to the limit (using the DCT) in Eq. (32.18) to find $\lim_{t\downarrow 0}\left\|p_t*f-f\right\|=0$.

Now suppose $g\in L^p\left(\mathbb{R}^n\right)$ and $f\in C_c\left(\mathbb{R}^n\right)$, then

$$\begin{aligned}\left\|p_t*g-g\right\|_p&\leq\left\|p_t*g-p_t*f\right\|_p+\left\|p_t*f-f\right\|_p+\left\|f-g\right\|_p\\&\leq 2\left\|f-g\right\|_p+\left\|p_t*f-f\right\|_p\end{aligned}$$

and therefore,

$$\limsup_{t\downarrow 0}\left\|p_t*g-g\right\|_p\leq 2\left\|f-g\right\|_p.$$

Since this inequality is valid for all $f\in C_c\left(\mathbb{R}^n\right)$ and, by Theorem 12.27, $C_c\left(\mathbb{R}^n\right)$ is dense in $L^p\left(\mathbb{R}^n\right)$, we may conclude that $\limsup_{t\downarrow 0}\left\|p_t*g-g\right\|_p=0$. ∎

Theorem 32.4 (Forced Heat Equation). *Suppose $g\in C_b(\mathbb{R}^d)$ and $f\in C_b^{1,2}([0,\infty)\times\mathbb{R}^d)$ then*

$$u(t,x):=p_t*g(x)+\int_0^t p_{t-\tau}*f(\tau,x)d\tau$$

solves

$$\frac{\partial u}{\partial t}=\frac{1}{2}\triangle u+f\ \text{ with } u(0,\cdot)=g.$$

Proof. Because of Theorem 32.1, we may without loss of generality assume $g=0$ in which case

$$u(t,x)=\int_0^t p_t*f(t-\tau,x)d\tau.$$

Therefore

$$\begin{aligned}\frac{\partial u}{\partial t}(t,x)&=p_t*f(0,x)+\int_0^t p_\tau*\frac{\partial}{\partial t}f(t-\tau,x)d\tau\\&=p_t*f_0(x)-\int_0^t p_\tau*\frac{\partial}{\partial\tau}f(t-\tau,x)d\tau\end{aligned}$$

and

$$\frac{\triangle}{2}u(t,x)=\int_0^t p_t*\frac{\triangle}{2}f(t-\tau,x)d\tau.$$

Hence we find, using integration by parts and approximate δ − function arguments, that

$$\begin{aligned}\left(\frac{\partial}{\partial t}-\frac{\triangle}{2}\right)u(t,x)&=p_t*f_0(x)+\int_0^t p_\tau*\left(-\frac{\partial}{\partial\tau}-\frac{1}{2}\triangle\right)f(t-\tau,x)d\tau\\&=p_t*f_0(x)\\&\quad+\lim_{\varepsilon\downarrow 0}\int_\varepsilon^t p_\tau*\left(-\frac{\partial}{\partial\tau}-\frac{1}{2}\triangle\right)f(t-\tau,x)d\tau\\&=p_t*f_0(x)-\lim_{\varepsilon\downarrow 0}p_\tau*f(t-\tau,x)\Big|_\varepsilon^t\\&\quad+\lim_{\varepsilon\downarrow 0}\int_\varepsilon^t\left(\frac{\partial}{\partial\tau}-\frac{1}{2}\triangle\right)p_\tau*f(t-\tau,x)d\tau\\&=p_t*f_0(x)-p_t*f_0(x)+\lim_{\varepsilon\downarrow 0}\ p_\varepsilon*f(t-\varepsilon,x)\\&=f(t,x).\end{aligned}$$

∎

32.3 Wiener Measure Heuristics and the Feynman-Kac formula

Theorem 32.5 (Trotter Product Formula). *Let A and B be $d\times d$ matrices. Then $e^{(A+B)}=\lim_{n\to\infty}\left(e^{\frac{A}{n}}e^{\frac{B}{n}}\right)^n$.*

Proof. By the chain rule,

$$\frac{d}{d\varepsilon}|_0\log(e^{\varepsilon A}e^{\varepsilon B})=A+B.$$

Hence by Taylor's theorem with remainder,

$$\log(e^{\varepsilon A}e^{\varepsilon B})=\varepsilon\left(A+B\right)+O\left(\varepsilon^2\right)$$

which is equivalent to

$$e^{\varepsilon A}e^{\varepsilon B}=e^{\varepsilon(A+B)+O\left(\varepsilon^2\right)}.$$

Taking $\varepsilon=1/n$ and raising the result to the n^{th} − power gives

$$\begin{aligned}(e^{n^{-1}A}e^{n^{-1}B})^n&=\left[e^{n^{-1}(A+B)+O\left(n^{-2}\right)}\right]^n\\&=e^{A+B+O\left(n^{-1}\right)}\to e^{(A+B)}\text{ as }n\to\infty.\end{aligned}$$

∎

Fact 32.6 (Trotter product formula) *For "nice enough" V,*

$$e^{T(\Delta/2-V)} = strong\text{-}\lim_{n\to\infty}[e^{\frac{T}{2n}\Delta}e^{-\frac{T}{n}V}]^n. \tag{32.19}$$

See [49] for a rigorous statement of this type.

Lemma 32.7. *Let $V:\mathbb{R}^d\to\mathbb{R}$ be a continuous function which is bounded from below, then*

$$\begin{aligned}
&\left(\left(e^{\frac{T}{n}\Delta/2}e^{-\frac{T}{n}V}\right)^n f\right)(x_0)\\
&= \int_{\mathbb{R}^{dn}} p_{\frac{T}{n}}(x_0,x_1)e^{-\frac{T}{n}V(x_1)}\dots p_{\frac{T}{n}}(x_{n-1},x_n)e^{-\frac{T}{n}V(x_n)}f(x_n)dx_1\dots dx_n\\
&= \left(\frac{1}{\sqrt{2\pi\frac{T}{n}}}\right)^{dn}\int_{(\mathbb{R}^d)^n} e^{-\frac{n}{2T}\sum_{i=1}^n|x_i-x_{i-1}|^2-\frac{T}{n}\sum_{i=1}^n V(x_i)}f(x_n)dx_1\dots dx_n.
\end{aligned} \tag{32.20}$$

Notation 32.8 *Given $T>0$, and $n\in\mathbb{N}$, let $W_{n,T}$ denote the set of piecewise C^1 – paths, $\omega:[0,T]\to\mathbb{R}^d$ such that $\omega(0)=0$ and $\omega''(\tau)=0$ if $\tau\notin\left\{\frac{i}{n}T\right\}_{i=0}^n =: \mathcal{P}_n(T)$ – see Figure 32.2. Further let dm_n denote the unique translation invariant measure on $W_{n,T}$ which is well defined up to a multiplicative constant.*

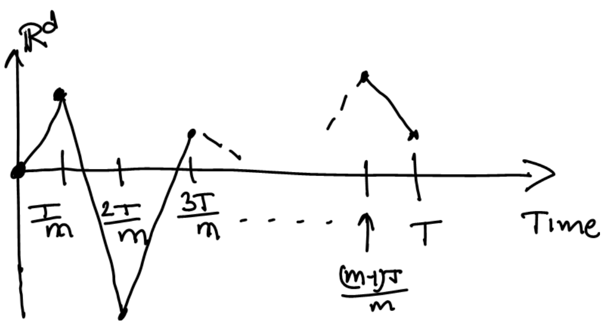


Fig. 32.2. A typical path in $W_{m,T}$.

With this notation we may rewrite Lemma 32.7 as follows.

Theorem 32.9. *Let $T>0$ and $n\in\mathbb{N}$ be given. For $\tau\in[0,T]$, let $\tau_+=\frac{i}{n}T$ if $\tau\in(\frac{i-1}{n}T,\frac{i}{n}T]$. Then Eq. (32.20) may be written as,*

$$\begin{aligned}
&\left(\left(e^{\frac{T}{n}\Delta/2}e^{-\frac{T}{n}V}\right)^n f\right)(x_0)\\
&\quad= \frac{1}{Z_n(T)}\int_{W_{n,T}} e^{-\int_0^T\left[\frac{1}{2}|\omega'(\tau)|^2+V(x_0+\omega(\tau_+))\right]d\tau}f(x_0+\omega(T))\,dm_n(\omega)
\end{aligned}$$

where

$$Z_n(T) := \int_{W_{n,T}} e^{-\frac{1}{2}\int_0^T|\omega'(\tau)|^2 d\tau}dm_n(\omega).$$

Moreover, by Trotter's product formula,

$$\begin{aligned}
&e^{T(\Delta/2-V)}f(x_0)\\
&\quad= \lim_{n\to\infty}\frac{1}{Z_n(T)}\int_{W_{n,T}} e^{-\int_0^T\left[\frac{1}{2}|\omega'(\tau)|^2+V(x_0+\omega(\tau_+))\right]d\tau}f(x_0+\omega(T))\,dm_n(\omega).
\end{aligned} \tag{32.21}$$

Following Feynman, at an informal level (see Figure 32.3), $W_{n,T}\to W_T$ as $n\to\infty$, where

$$W_T := \left\{\omega\in C\left([0,T]\to\mathbb{R}^d\right) : \omega(0)=0\right\}.$$

Moreover, formally passing to the limit in Eq. (32.21) leads us to the following

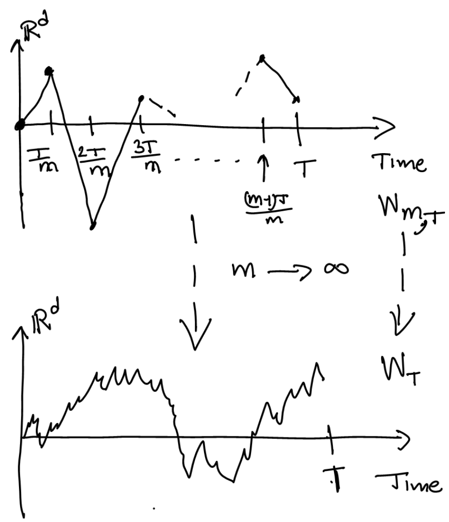


Fig. 32.3. A typical path in W_T may be approximated better and better by paths in $W_{m,T}$ as $m\to\infty$.

heuristic expression for $\left(e^{T(\Delta/2-V)}f\right)(x_0)$;

$$\left(e^{T(\Delta/2-V)}f\right)(x_0)=\text{“}\frac{1}{Z(T)}\int_{W_T}e^{-\int_0^T\left[\frac{1}{2}|\omega'(\tau)|^2+V(x_0+\omega(\tau))\right]d\tau}f(x_0+\omega(T))\,\mathcal{D}\omega\text{”}\tag{32.22}$$

where $\mathcal{D}\omega$ is the **non-existent** Lebesgue measure on W_T, and $Z(T)$ is the "normalization" constant (or partition function) given by

$$Z(T)=\text{“}\int_{W_T}e^{-\frac{1}{2}\int_0^T|\omega'(\tau)|^2d\tau}\mathcal{D}\omega.\text{”}$$

This expression may also be written in the **Feynman − Kac** form as

$$e^{T(\Delta/2-V)}f(x_0)=\int_{W_T}e^{-\int_0^T V(x_0+\omega(\tau))d\tau}f(x_0+\omega(T))\,d\mu(\omega),\tag{32.23}$$

where

$$d\mu(\omega)=\text{“}\frac{1}{Z(T)}e^{-\frac{1}{2}\int_0^T|\omega'(\tau)|^2d\tau}\mathcal{D}\omega.\text{”}\tag{32.24}$$

Thus our immediate goal is to make sense out of Eq. (32.24).

Let

$$H_T:=\left\{h\in W_T:\int_0^T|h'(\tau)|^2\,d\tau<\infty\right\}$$

with the convention that $\int_0^T|h'(\tau)|^2\,d\tau:=\infty$ if h is not absolutely continuous. Further let

$$\langle h,k\rangle_T:=\int_0^T h'(\tau)\cdot k'(\tau)\,d\tau\text{ for all }h,k\in H_T$$

and $X_h(\omega):=\langle h,\omega\rangle_T$ for $h\in H_T$. Since

$$d\mu(\omega)=\text{“}\frac{1}{Z(T)}e^{-\frac{1}{2}\|\omega\|^2_{H_T}}\mathcal{D}\omega,\text{”}\tag{32.25}$$

$d\mu(\omega)$ should be a Gaussian measure on H_T and hence we expect,

$$\mathbb{E}_\mu[X_hX_k]=\langle h,k\rangle_T\text{ for all }h,k\in H_T.\tag{32.26}$$

According to Proposition 24.6, there exists a Gaussian random field, $\{X_h\}_{h\in H_T}$, on some probability space, $(\Omega,\mathcal{B},P)$, such that Eq. (32.26) holds. We are applying this corollary with $T\to H_T$, and $Q(h,k):=\langle h,k\rangle_T$. Notice that if $\Lambda\subset_f H_T$ and $\lambda:\Lambda\to\mathbb{R}$ is a function, then

$$\sum_{h,k\in\Lambda}Q(h,k)\,\lambda(h)\,\lambda(k)=\left\|\sum_{h\in\Lambda}\lambda(h)\,h\right\|_T^2\geq 0.$$

Heuristically, we are thinking that Ω should be the Hilbert space, H_T, and P should be the "measure" in Eq. (32.25). In this hypothetical setting, we could define, $B_t:H_T\to\mathbb{R}^d$ to be the projection, $B_t(\omega)=\omega(t)$ for $t\in[0,T]$. Hence for $a\in\mathbb{R}^d$,

$$a\cdot B_t(\omega)=a\cdot\omega(t)=\int_0^T a1_{[0,t]}(\tau)\,\dot\omega(\tau)\,d\tau=\langle h_{a,t},\omega\rangle_T=X_{h_{a,t}}(\omega)$$

where

$$h_{a,t}(\tau):=\int_0^\tau a1_{[0,t]}(u)\,du=a\cdot(t\wedge\tau).$$

Since the B^j are independent processes, it follows from Eq. (32.26) that

$$\begin{aligned}\mathbb{E}_\mu[(a\cdot B_t)(b\cdot B_s)]=\langle h_{a,t},h_{b,s}\rangle_T&=\int_0^T a\cdot b1_{[0,t]}(\tau)\,1_{[0,s]}(\tau)\,d\tau\\&=a\cdot b\,(s\wedge t).\end{aligned}\tag{32.27}$$

Hence we recognize that the process $\{B_t\}_{t\geq0}$ should be our old friend – the multidimensional Brownian motion and so $\mu=\mathrm{Law}(B_\cdot)$. Assuming this to be the case, the informal expression in Eq. (32.23) leads the following conjecture;

$$\left(e^{T(\Delta/2-V)}f\right)(x_0)=\mathbb{E}\left[e^{-\int_0^T V(x_0+B_\tau)d\tau}f(x_0+B_T)\right]$$

which is the **Feynman − Kac formula.** We will discuss a rigorous proof of this formula in the next section.

32.4 Proving the Feynman Kac Formula

Suppose that $V:\mathbb{R}^d\to\mathbb{R}$ is a smooth function such that $k:=\inf_{x\in\mathbb{R}^d}V(x)>-\infty$ and for $f\geq0$ or f bounded and measurable, let[2]

$$T_tf(x):=\mathbb{E}_x\left[e^{-\int_0^t V(B_\tau)d\tau}f(B_t)\right]=\mathbb{E}_0\left[e^{-\int_0^t V(x+B_\tau)d\tau}f(x+B_t)\right].$$

Let us observe that for $f\geq0$ and $p,q\in(1,\infty)$ such that $p^{-1}+q^{-1}=1$

$$\begin{aligned}|(T_tf)(x)|\leq e^{-kt}\mathbb{E}_x[f(B_t)]&=e^{-kt}\int_{\mathbb{R}^d}|f(y)|\,p_t(x,y)\,dy\\&\leq e^{-kt}\|f\|_p\cdot\|p_t(x,\cdot)\|_q\end{aligned}\tag{32.28}$$

where

[2] In what follows, the reader feeling queasy about measurability issues, please have a look at Lemma 27.6 and Lemma 27.7 in the next section.

$$\|f\|_p := \left(\int_{\mathbb{R}^d} |f(y)|^p \, dy\right)^{1/p}.$$

In particular if $f = 0$, m – a.e., then $T_t f(x) = 0$ and we can use this to see that $(T_t f)(x)$ is well defined for all $f \in L^p$. Since

$$\begin{aligned}
\|p_t(x,\cdot)\|_q^q &= \int_{\mathbb{R}^d} p_t(y-x)^q \, dy = \int_{\mathbb{R}^d} p_t(y)^q \, dy \\
&= \frac{1}{(2\pi t)^{dq/2}} \int_{\mathbb{R}^d} e^{-\frac{1}{2(t/q)}|y|^2} dy = \frac{1}{(2\pi t)^{dq/2}} \left(\frac{2\pi t}{q}\right)^{d/2} \\
&= q^{-d/2} \frac{1}{(2\pi t)^{d(q-1)/2}},
\end{aligned}$$

Eq. (32.28) gives the quantitative estimate;

$$|T_t f(x)| \le C(p,t) \|f\|_{L^p(\mathbb{R}^d)}, \tag{32.29}$$

where

$$C(p,t) := q^{-\frac{d}{2q}} (2\pi t)^{-\frac{d}{2p}} e^{-kt}. \tag{32.30}$$

Theorem 32.10 (Feynman-Kac Formula). *Suppose $f \in L^2(\mathbb{R}^d, m)$ and $t \ge 0$. Then;*

1. *T_t is a bounded linear operator on $L^2(\mathbb{R}^d)$ with $\|T_t\|_{op} \le e^{-kt}$, i.e.*
$$\|T_t f\|_2 \le e^{-kt} \|f\|_2 \text{ for all } f \in L^2(\mathbb{R}^d, m).$$
2. *T_t is self-adjoint, i.e. $(T_t f, g) = (f, T_t g)$ for all $f, g \in L^2(\mathbb{R}^d, m)$ where*
$$(f,g) := \int_{\mathbb{R}^d} f(x) \bar{g}(x) \, dm(x).$$
3. *$\{T_t\}_{t\ge0}$ is a semi-group, i.e. $T_{t+s} = T_t T_s$ for all $t, s \ge 0$.*
4. *T_t is strongly continuous, i.e.*
$$\lim_{t\downarrow 0} \|T_t f - f\|_{L^2} = 0 \text{ for all } f \in L^2(\mathbb{R}^d, m).$$
5. *Let*
$$Af := \frac{d}{dt}|_{0+}(T_t f) := L^2\text{–}\lim_{t\downarrow0} \frac{T_t f - f}{t}$$
*for those f for which the limit exists. Then $Af = \left(\frac{1}{2}\Delta - V\right) f$ for all $f \in C_c^2(\mathbb{R}^d)$. The operator A with its natural domain[3] is called the **infinitesimal generator** of $\{T_t\}_{t\ge0}$.*

[3] The domain, $\mathcal{D}(A)$, of A consists of those $f \in L^2(\mathbb{R}^d, m)$ such that the limit defining Af exists in the $L^2(\mathbb{R}^d, m)$ sense. So we are asserting that $C_c^2(\mathbb{R}^d) \subset \mathcal{D}(A)$ and $Af = \left(\frac{1}{2}\Delta - V\right) f$ for all $f \in C_c^2(\mathbb{R}^d)$.

Remark 32.11. Some functional analysis along with basic "elliptic regularity" shows, that $u(t,x) = T_t f(x)$ solves the heat equation,

$$u_t(t,x) = \frac{1}{2}\Delta u(t,x) - V(x) u(t,x) \text{ with } \lim_{t\downarrow0} u(t,\cdot) = f(\cdot) \text{ in } L^2.$$

See Simon [49], for a proof of Theorem 32.10 (in more generality) using Trotter's product formula.

Proof. To simplify notation a bit we will assume $d = 1$ in the proof below and let $\|f\| := \|f\|_{L^2(\mathbb{R},m)}$.

1. By Fubini's theorem and simple estimates,
$$\begin{aligned}
\|T_t f\|^2 &= \int_{\mathbb{R}} dx \left| \mathbb{E}_0 \left[e^{-\int_0^t V(x+B_\tau)d\tau} f(x+B_t) \right] \right|^2 \\
&\le \int_{\mathbb{R}} \mathbb{E}_0 \left| e^{-\int_0^t V(x+B_\tau)d\tau} f(x+B_t) \right|^2 dx \\
&\le e^{-2kt} \mathbb{E}_0 \int_{\mathbb{R}} |f(x+B_t)|^2 dx = e^{-2kt} \|f\|_2^2.
\end{aligned}$$
2. We have
$$\begin{aligned}
(T_t f, g) &= \int_{\mathbb{R}} \mathbb{E}_0 \left[e^{-\int_0^t V(x+B_\tau)d\tau} f(x+B_t) \right] g(x) \, dx \\
&= \mathbb{E}_0 \int_{\mathbb{R}} e^{-\int_0^t V(x+B_\tau)d\tau} f(x+B_t) g(x) \, dx \\
&= \mathbb{E}_0 \int_{\mathbb{R}} e^{-\int_0^t V(x-B_t+B_\tau)d\tau} f(x) g(x-B_t) \, dx.
\end{aligned}$$
Now let $b_\tau := B_{t-\tau} - B_t$ so that b_τ is another Brownian motion on $[0,t]$ and observe that $b_t = -B_t$. Hence we have
$$\begin{aligned}
(T_t f, g) &= \mathbb{E}_0 \int_{\mathbb{R}} e^{-\int_0^t V(x+b_{t-\tau})d\tau} f(x) g(x+b_t) \, dx \\
&= \mathbb{E}_0 \int_{\mathbb{R}} e^{-\int_0^t V(x+b_\tau)d\tau} f(x) g(x+b_t) \, dx \\
&= \mathbb{E}_0 \int_{\mathbb{R}} e^{-\int_0^t V(x+B_\tau)d\tau} g(x+B_t) f(x) \, dx \\
&= \int_{\mathbb{R}} \mathbb{E}_0 \left[e^{-\int_0^t V(x+B_\tau)d\tau} g(x+B_t) \right] f(x) \, dx \\
&= (f, T_t g).
\end{aligned}$$

3. Using the Markov property in Theorem 29.3 we find,

$$\begin{aligned}(T_{t+s}f)(x) &= \mathbb{E}_x\left[e^{-\int_0^{t+s}V(B_\tau)d\tau}f(B_{t+s})\right] \\ &= \mathbb{E}_x\left[e^{-\int_0^{t}V(B_\tau)d\tau}e^{-\int_s^{t+s}V(B_\tau)d\tau}f(B_{t+s})\right] \\ &= \mathbb{E}_x\left[e^{-\int_0^{t}V(B_\tau)d\tau}\left[e^{-\int_0^{s}V(B_\tau)d\tau}f(B_s)\right]\circ\theta_t\right] \\ &= \mathbb{E}_x\left[e^{-\int_0^{t}V(B_\tau)d\tau}\mathbb{E}_{B_t}\left[e^{-\int_0^{s}V(B_\tau)d\tau}f(B_s)\right]\right] \\ &= \mathbb{E}_x\left[e^{-\int_0^{t}V(B_\tau)d\tau}(T_sf)(B_t)\right] = (T_tT_sf)(x).\end{aligned}$$

4. From the estimate,

$$\begin{aligned}\|T_tf-f\|^2 &= \int_{\mathbb{R}}dx\left|\mathbb{E}_0\left[e^{-\int_0^t V(x+B_\tau)d\tau}f(x+B_t)-f(x)\right]\right|^2 \\ &\le \int_{\mathbb{R}}\mathbb{E}_0\left|e^{-\int_0^t V(x+B_\tau)d\tau}f(x+B_t)-f(x)\right|^2dx \\ &= \int_{\mathbb{R}}\mathbb{E}_0\left|\left(e^{-\int_0^t V(x+B_\tau)d\tau}-1\right)f(x+B_t)+f(x+B_t)-f(x)\right|^2dx,\end{aligned}$$

it follows that

$$\limsup_{t\downarrow 0}\|T_tf-f\|^2 \le \limsup_{t\downarrow 0}D_t+\limsup_{t\downarrow 0}E_t$$

where

$$D_t = 2\int_{\mathbb{R}}\mathbb{E}_0\left|\left(e^{-\int_0^t V(x+B_\tau)d\tau}-1\right)f(x+B_t)\right|^2dx$$

and

$$E_t := 2\mathbb{E}_0\int_{\mathbb{R}}|f(x+B_t)-f(x)|^2dx.$$

Let us now assume for the moment that $f\in C_c(\mathbb{R})$. In this case, using DCT twice, we learn that $\int_{\mathbb{R}}|f(x+B_t)-f(x)|^2dx\to 0$ boundedly and hence that $\limsup_{t\downarrow 0}E_t=0$. Similarly, if M is a bound on $|f|$, then

$$D_t \le 2M^2\int_{\mathbb{R}}\mathbb{E}_0\left|e^{-\int_0^t V(x+B_\tau)d\tau}-1\right|^2dx \to 0 \text{ as } t\downarrow 0.$$

So in this special case we have shown $\lim_{t\downarrow 0}\|T_tf-f\|=0$. For general $f\in L^2$ and $g\in C_c(\mathbb{R})$ we have

$$\begin{aligned}\limsup_{t\downarrow 0}\|T_tf-f\| &\le \limsup_{t\downarrow 0}\left[\|T_tf-T_tg\|+\|T_tg-g\|+\|g-f\|\right] \\ &\le \limsup_{t\downarrow 0}\left[\|T_tg-g\|+\left(1+e^{-kt}\right)\|g-f\|\right] \\ &\le 2\|g-f\|.\end{aligned}$$

This completes the proof because $C_c(\mathbb{R})$ is dense in $L^2(\mathbb{R},m)$ (see Example 12.28) and hence we may make $\|f-g\|$ as small as we please.

5. Finally we **sketch** the computation of the infinitesimal generator, A, on $C_c^2(\mathbb{R})$. By the chain rule,

$$\begin{aligned}\frac{d}{dt}|_{0+}T_tf(x) &= \frac{d}{dt}|_{0+}\mathbb{E}_x\left[e^{-\int_0^t V(B_\tau)d\tau}f(B_t)\right] \\ &= \frac{d}{dt}|_{0+}\mathbb{E}_x\left[e^{-\int_0^t V(B_\tau)d\tau}f(B_0)\right]+\frac{d}{dt}|_{0+}\mathbb{E}_x[f(B_t)] \\ &= \mathbb{E}_x[-V(B_0)f(B_0)]+\frac{d}{dt}|_{0+}(p_t*f)(x) \\ &= -V(x)f(x)+\frac{1}{2}\Delta f(x).\end{aligned}$$

■

Exercise 32.2 (Ultracontractivity of T_t). Let $BC(\mathbb{R}^d)$ denote the bounded continuous functions on $\mathbb{R}^d$ and define

$$\|g\|_\infty = \sup_{x\in\mathbb{R}^d}|g(x)|$$

for $g\in BC(\mathbb{R}^d)$. Suppose $1<p<\infty$ and $f\in L^p(\mathbb{R}^d,m)$. Show $u(t,x):=(T_tf)(x)$ is continuous for $(t,x)\in(0,\infty)\times\mathbb{R}^d$ and is bounded in x for fixed $t>0$. In particular, for any $t>0$, show T_t maps $L^p(\mathbb{R}^d,m)$ into $BC(\mathbb{R}^d)$ and

$$\|T_tf\|_\infty \le C(p,t)\|f\|_{L^p(\mathbb{R}^d)},$$

where $C(p,t)$ is define as in Eq. (32.30). **Hint:** first verify the continuity of $u(t,x)$ under the additional assumption that $f\in C_c(\mathbb{R}^d)$.

32.5 Appendix: Extensions of Theorem 32.1

Proposition 32.12. *Suppose $f:\mathbb{R}^d\to\mathbb{R}$ is a measurable function and there exists constants $c,C<\infty$ such that*

$$|f(x)| \le Ce^{\frac{c}{2}|x|^2}.$$

Then $u(t,x) := p_t * f(x)$ is smooth for $(t,x) \in (0,c^{-1}) \times \mathbb{R}^d$ and for all $k \in \mathbb{N}$ and all multi-indices α,

$$D^\alpha \left(\frac{\partial}{\partial t}\right)^k u(t,x) = \left(D^\alpha \left(\frac{\partial}{\partial t}\right)^k p_t\right) * f(x). \tag{32.31}$$

In particular u satisfies the heat equation $u_t = \Delta u/2$ on $(0,c^{-1}) \times \mathbb{R}^d$.

Proof. The reader may check that

$$D^\alpha \left(\frac{\partial}{\partial t}\right)^k p_t(x) = q(t^{-1},x)p_t(x)$$

where q is a polynomial in its variables. Let $x_0 \in \mathbb{R}^d$ and $\varepsilon > 0$ be small, then for $x \in B(x_0,\varepsilon)$ and any $\beta > 0$,

$$\begin{aligned}|x-y|^2 &= |x|^2 - 2|x||y| + |y|^2 \geq |y|^2 + |x|^2 - \left(\beta^{-2}|x|^2 + \beta^2|y|^2\right) \\ &\geq \left(1-\beta^2\right)|y|^2 - \left(\beta^{-2}-1\right)\left(|x_0|^2 + \varepsilon\right).\end{aligned}$$

Hence

$$\begin{aligned}g(y) &:= \sup\left\{\left|D^\alpha \left(\frac{\partial}{\partial t}\right)^k p_t(x-y)f(y)\right| : \varepsilon \leq t \leq c-\varepsilon \;\&\; x \in B(x_0,\varepsilon)\right\} \\ &\leq \sup\left\{\left|q(t^{-1},x-y)\frac{e^{-\frac{1}{2t}|x-y|^2}}{(2\pi t)^{n/2}}Ce^{\frac{c}{2}|y|^2}\right| : \varepsilon \leq t \leq c-\varepsilon \;\&\; x \in B(x_0,\varepsilon)\right\} \\ &\leq C(\beta,x_0,\varepsilon)\sup\left\{\left|q(t^{-1},x-y)\frac{e^{\left[-\frac{1}{2t}\left(1-\beta^2\right)+\frac{c}{2}\right]|y|^2}}{(2\pi t)^{n/2}}\right| : \begin{array}{c}\varepsilon \leq t \leq c-\varepsilon \text{ and} \\ x \in B(x_0,\varepsilon)\end{array}\right\}.\end{aligned}$$

By choosing β close to 0, the reader should check using the above expression that for any $0 < \delta < (1/t - c)/2$ there is a $\tilde{C} < \infty$ such that $g(y) \leq \tilde{C}e^{-\delta|y|^2}$. In particular $g \in L^1\left(\mathbb{R}^d\right)$. Hence one is justified in differentiating past the integrals in $p_t * f$ and this proves Eq. (32.31). ■

Lemma 32.13. *There exists a polynomial $q_n(x)$ such that for any $\beta > 0$ and $\delta > 0$,*

$$\int_{\mathbb{R}^d} 1_{|y|\geq\delta} e^{-\beta|y|^2} dy \leq \delta^n q_n\left(\frac{1}{\beta\delta^2}\right)e^{-\beta\delta^2}$$

Proof. Making the change of variables $y \to \delta y$ and then passing to polar coordinates shows

$$\int_{\mathbb{R}^d} 1_{|y|\geq\delta} e^{-\beta|y|^2} dy = \delta^n \int_{\mathbb{R}^d} 1_{|y|\geq 1} e^{-\beta\delta^2|y|^2} dy = \sigma\left(S^{n-1}\right)\delta^n \int_1^\infty e^{-\beta\delta^2 r^2} r^{n-1} dr.$$

Letting $\lambda = \beta\delta^2$ and $\phi_n(\lambda) := \int_{r=1}^\infty e^{-\lambda r^2} r^n dr$, integration by parts shows

$$\begin{aligned}\phi_n(\lambda) &= \int_{r=1}^\infty r^{n-1} d\left(\frac{e^{-\lambda r^2}}{-2\lambda}\right) = \frac{1}{2\lambda}e^{-\lambda} + \frac{1}{2}\int_{r=1}^\infty (n-1)r^{(n-2)}\frac{e^{-\lambda r^2}}{\lambda}dr \\ &= \frac{1}{2\lambda}e^{-\lambda} + \frac{n-1}{2\lambda}\phi_{n-2}(\lambda).\end{aligned}$$

Iterating this equation implies

$$\phi_n(\lambda) = \frac{1}{2\lambda}e^{-\lambda} + \frac{n-1}{2\lambda}\left(\frac{1}{2\lambda}e^{-\lambda} + \frac{n-3}{2\lambda}\phi_{n-4}(\lambda)\right)$$

and continuing in this way shows

$$\phi_n(\lambda) = e^{-\lambda}r_n(\lambda^{-1}) + \frac{(n-1)!!}{2^\delta\lambda^\delta}\phi_i(\lambda)$$

where δ is the integer part of $n/2$, $i = 0$ if n is even and $i = 1$ if n is odd and r_n is a polynomial. Since

$$\phi_0(\lambda) = \int_{r=1}^\infty e^{-\lambda r^2} dr \leq \phi_1(\lambda) = \int_{r=1}^\infty re^{-\lambda r^2} dr = \frac{e^{-\lambda}}{2\lambda},$$

it follows that

$$\phi_n(\lambda) \leq e^{-\lambda}q_n(\lambda^{-1})$$

for some polynomial q_n. ■

Proposition 32.14. *Suppose $f \in C(\mathbb{R}^d,\mathbb{R})$ such that $|f(x)| \leq Ce^{\frac{c}{2}|x|^2}$ then $p_t * f \to f$ uniformly on compact subsets as $t \downarrow 0$. In particular in view of Proposition 32.12, $u(t,x) := p_t * f(x)$ is a solution to the heat equation with $u(0,x) = f(x)$.*

Proof. Let $M > 0$ be fixed and assume $|x| \leq M$ throughout. By uniform continuity of f on compact set, given $\varepsilon > 0$ there exists $\delta = \delta(t) > 0$ such that $|f(x) - f(y)| \leq \varepsilon$ if $|x-y| \leq \delta$ and $|x| \leq M$. Therefore, choosing $a > c/2$ sufficiently small,

$$\begin{aligned}|p_t * f(x) - f(x)| &= \left|\int p_t(y)\left[f(x-y) - f(x)\right]dy\right| \\ &\leq \int p_t(y)\left|f(x-y) - f(x)\right|dy \\ &\leq \varepsilon\int_{|y|\leq\delta} p_t(y)dy + \frac{C}{(2\pi t)^{n/2}}\int_{|y|\geq\delta}\left[e^{\frac{c}{2}|x-y|^2} + e^{\frac{c}{2}|x|^2}\right]e^{-\frac{1}{2t}|y|^2}dy \\ &\leq \varepsilon + \tilde{C}\left(2\pi t\right)^{-n/2}\int_{|y|\geq\delta} e^{-\left(\frac{1}{2t}-a\right)|y|^2}dy.\end{aligned}$$

So by Lemma 32.13, it follows that

$$|p_t * f(x) - f(x)| \le \varepsilon + \tilde{C}\,(2\pi t)^{-n/2}\,\delta^n q_n\left(\frac{1}{\beta\left(\frac{1}{2t}-a\right)^2}\right) e^{-\left(\frac{1}{2t}-a\right)\delta^2}$$

and therefore

$$\limsup_{t\downarrow 0}\ \sup_{|x|\le M}\ |p_t * f(x) - f(x)| \le \varepsilon \to 0 \text{ as } \varepsilon \downarrow 0.$$

■

Lemma 32.15. *If $q(x)$ is a polynomial on $\mathbb{R}^d$, then*

$$\int_{\mathbb{R}^d} p_t(x-y)q(y)dy = \sum_{n=0}^{\infty} \frac{t^n}{n!}\frac{\Delta^n}{2^n} q(x).$$

Proof. Since

$$f(t,x) := \int_{\mathbb{R}^d} p_t(x-y)q(y)dy = \int_{\mathbb{R}^d} p_t(y)\sum a_{\alpha\beta}x^\alpha y^\beta dy = \sum C_\alpha(t)x^\alpha,$$

$f(t,x)$ is a polynomial in x of degree no larger than that of q. Moreover $f(t,x)$ solves the heat equation and $f(t,x)\to q(x)$ as $t\downarrow 0$. Since $g(t,x) := \sum_{n=0}^{\infty}\frac{t^n}{n!}\frac{\Delta^n}{2^n}q(x)$ has the same properties of f and Δ is a bounded operator when acting on polynomials of a fixed degree we conclude $f(t,x)=g(t,x)$. ■

Example 32.16. Suppose $q(x) = x_1x_2 + x_3^4$, then

$$\begin{aligned} e^{t\Delta/2}q(x) &= x_1x_2 + x_3^4 + \frac{t}{2}\Delta\left(x_1x_2+x_3^4\right) + \frac{t^2}{2!\cdot 4}\Delta^2\left(x_1x_2+x_3^4\right) \\ &= x_1x_2 + x_3^4 + \frac{t}{2}12x_3^2 + \frac{t^2}{2!\cdot 4}4! \\ &= x_1x_2 + x_3^4 + 6tx_3^2 + 3t^2. \end{aligned}$$

Proposition 32.17. *Suppose $f\in C^\infty(\mathbb{R}^d)$ and there exists a constant $C<\infty$ such that*

$$\sum_{|\alpha|=2N+2} |D^\alpha f(x)| \le Ce^{C|x|^2},$$

then

$$(p_t * f)(x) = \text{“}e^{t\Delta/2}f(x)\text{”} = \sum_{k=0}^{N}\frac{t^k}{k!}\Delta^k f(x) + O(t^{N+1}) \text{ as } t\downarrow 0$$

Proof. Fix $x\in\mathbb{R}^d$ and let

$$f_N(y) := \sum_{|\alpha|\le 2N+1}\frac{1}{\alpha!}D^\alpha f(x)y^\alpha.$$

Then by Taylor's theorem with remainder

$$\begin{aligned} |f(x+y)-f_N(y)| &\le C\,|y|^{2N+2}\sup_{t\in[0,1]} e^{C|x+ty|^2} \\ &\le C\,|y|^{2N+2}\,e^{2C\left[|x|^2+|y|^2\right]} \le \tilde{C}\,|y|^{2N+2}\,e^{2C|y|^2} \end{aligned}$$

and thus

$$\begin{aligned} &\left|\int_{\mathbb{R}^d} p_t(y)f(x+y)dy - \int_{\mathbb{R}^d} p_t(y)f_N(y)dy\right| \\ &\quad\le \tilde{C}\int_{\mathbb{R}^d} p_t(y)\,|y|^{2N+2}\,e^{2C|y|^2}dy \\ &\quad= \tilde{C}t^{N+1}\int_{\mathbb{R}^d} p_1(y)\,|y|^{2N+2}\,e^{2t^2C|y|^2}dy = O(t^{N+1}). \end{aligned}$$

Since $f(x+y)$ and $f_N(y)$ agree to order $2N+1$ for y near zero, it follows that

$$\begin{aligned} \int_{\mathbb{R}^d} p_t(y)f_N(y)dy &= \sum_{k=0}^{N}\frac{t^k}{k!}\Delta^k f_N(0) = \sum_{k=0}^{N}\frac{t^k}{k!}\Delta_y^k f(x+y)|_{y=0} \\ &= \sum_{k=0}^{N}\frac{t^k}{k!}\Delta^k f(x) \end{aligned}$$

which completes the proof. ■

33

Feller Processes

In this chapter we are going to introduce the class of Feller Markov processes. This class of processes contains most of the examples of continuous time Markov processes that are studied in this book.

Throughout this part of the book, let (S, ρ) be a locally compact, separable metric space and recall that for $x \in S$ and $\varepsilon > 0$ we let

$$B(x, \varepsilon) = \{y \in S : \rho(x, y) < \varepsilon\} \text{ and}$$
$$C(x, \varepsilon) = \{y \in S : \rho(x, y) \le \varepsilon\}$$

be the open and closed ε – balls about x – respectively. Let us now recall that (S, ρ) is σ – compact, i.e. there exists compact subsets $\{K_n\}_{n=1}^{\infty}$ such that $K_n \uparrow S$ as $n \uparrow \infty$. To prove this let $\{x_n\}_{n=1}^{\infty}$ be a countable dense subset of S and let

$$\varepsilon_n := \sup\{\varepsilon > 0 : C(x_n, 2\varepsilon) \text{ is compact}\}.$$

Then define $C_n := C(x_n, \varepsilon_n \wedge 1)$is compact for each n. We may then take

$$K_n = \cup_{l=1}^{n} C(x_l, \varepsilon_l \wedge 1) \text{ for all } n \in \mathbb{N}.$$

By construction each K_n is compact and I will leave it to the reader to verify that $\cup_{n=1}^{\infty} K_n = S$.

Definition 33.1. *A function $f : S \to \mathbb{R}$ is said to **vanish at infinity** if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset S$ such that $|f(x)| \le \varepsilon$ for all $x \in S \setminus K_\varepsilon$. We denote those $f \in C(S)$ which vanish at infinity by $C_0(S)$.*

In terms of an exhausting sequence of compact sets $\{K_n\}_{n=1}^{\infty}$ as constructed above we can rephrase the definition that f vanishes at infinity as the

$$\limsup_{n\to\infty} \sup_{x \in S \setminus K_n} |f(x)| = 0.$$

Finally when S is not compact, let $\hat{S} := S \cup \{\Delta\}$ denote the one point compactification of S where $V \subset \hat{S}$ is an open neighborhood of Δ iff $\hat{S} \setminus V$ is compact in S. To any $f \in C_0(S)$ we will define $f(\Delta) = 0$ and view f as a function on $\hat{S}$. This map is inverse to the restriction map, $C_\Delta\left(\hat{S}\right) \ni f \to f|_S \in C(S)$ where $C_\Delta\left(\hat{S}\right)$ consists of those $f \in C\left(\hat{S}\right)$ such that $f(\Delta) = 0$.

We are now going extend the notion introduced in Definition 17.12.

Definition 33.2 (sub-Markov transition kernels). *We say a collection of sub-probabilty kernels, $\{Q_t\}_{t\ge 0}$ on $S \times S$ are time homogeneous **sub-Markov transition kernels** Q if $Q_0(x, dy) = \delta_x(dy)$ for all $s \in T$ and the **Chapmann-Kolmogorov equations** hold;*

$$Q_{s+t} = Q_s Q_t \text{ for all } 0 \le s, t < \infty. \tag{33.1}$$

The point is that we are now allowing for the possibility that $Q_t(x, 1) = Q_t(x, S) < 1$ and hence the terminology of being sub-Markovian. For $t \ge 0$ and $x \in S$ let

$$\gamma_t(x) := 1 - Q_t(x, S) = 1 - (Q_t 1)(x)$$

and let $\gamma_t(\Delta) = 1$. Then we may defined probability kernels on $\hat{S} \times \hat{S}$ by

$$\hat{Q}_t(x, A) := Q_t(x, A \setminus \{\Delta\}) + \gamma_t(x)\,\delta_\Delta(A)$$

for all $x \in \hat{S}$ and $A \in \mathcal{B}_{\hat{S}}$ where by convention $Q_t(\Delta, \cdot) \equiv 0$. Alternatively stated, if $f : \hat{S} \to \mathbb{R}$ is a bounded measurable function, then

$$\left(\hat{Q}_t f\right)(x) = \hat{Q}_t(x, f) = Q_t(x, f|_S) + \gamma_t(x)\, f(\Delta).$$

Lemma 33.3. *The probability kernels $\left\{\hat{Q}_t\right\}_{t\ge 0}$ defined above are Markov transition kernels as in Definition 17.12.*

Proof. The point is to verify the Chapmann-Kolmogorov equations hold. If $f : \hat{S} \to \mathbb{R}$ is a bounded measurable function, then for $x \in S$ we have

$$\begin{aligned}
\left(\hat{Q}_s \hat{Q}_t f\right)(x) &= Q_s\left(x, \left(\hat{Q}_t f\right)|_S\right) + \gamma_s(x)\left(\hat{Q}_t f\right)(\Delta) \\
&= Q_s(x, Q_t(\cdot, f|_S) + \gamma_t(\cdot) f(\Delta)) + \gamma_s(x) f(\Delta) \\
&= Q_{s+t}(x, f|_S) + [(Q_s\gamma_t)(x) + \gamma_s(x)] f(\Delta).
\end{aligned}$$

Since

$$\begin{aligned}
\gamma_s + (Q_s\gamma_t) &= \gamma_s + Q_s(1 - Q_t(\cdot, S)) \\
&= 1 - Q_s 1 + Q_s 1 - Q_{s+t} 1 = 1 - Q_{s+t} 1 = \gamma_{s+t}
\end{aligned}$$

we have shown that $\left(\hat{Q}_s\hat{Q}_tf\right)(x)=\left(\hat{Q}_{s+t}f\right)(x)$ for all $x\in S$. The case where $x=\Delta$ is simple to check since $\left(\hat{Q}_tf\right)(\Delta)=f(\Delta)$ and therefore,

$$\left(\hat{Q}_s\hat{Q}_tf\right)(\Delta)=\left(\hat{Q}_tf\right)(\Delta)=f(\Delta)=\left(\hat{Q}_{s+t}f\right)(\Delta).$$

■

According to Theorem 17.11 we may now associate an $\hat{S}$ – valued Markov process to the Markov transition kernels, $\left\{\hat{Q}_t\right\}_{t\geq 0}$. However, Theorem 17.11 makes no guarantee that this process will have any reasonable sample path properties, e.g. measurability or continuity properties. To overcome this problem we are going to require more restrictions on $\{Q_t\}$.

Definition 33.4 (Feller Semigroup). *We say that $\{Q_t\}_{t\geq 0}$ is a Feller semi-group if $\{Q_t\}_{t\geq 0}$ is are sub-Markov transition kernels such that;*

1. *$Q_t(C_0(S))\subset C_0(S)$ for all $t\geq 0$ and*
2. *$Q_tf\to f$ in the uniform norm as $t\downarrow 0$ for all $f\in C_0(S)$.*

It is possible to associate to these Feller semi-groups Markov processes with nice sample space properties.

Definition 33.5 (Canonical path space). *Let $\Omega=\Omega\left(\hat{S}\right)$ denote the collection of functions, $\omega:\mathbb{R}_+\to\hat{S}$ which are right continuous and possess left hand limits (**rcll** for short) and are absorbing at Δ. Here we way that ω is **absorbing** at Δ if either $\omega(t-):=\lim_{\tau\uparrow t}\omega(\tau)=\Delta$ or $\omega(t)=\Delta$ then $\omega(s)=\Delta$ for all $s>t$. For $\omega\in\Omega$ we let*

$$\zeta(\omega):=\inf\{t\geq 0:\omega(t)=\Delta\}$$

and call $\zeta(\omega)$ the life time of ω.

For example if $S=\mathbb{R}$ and

$$\omega(t):=\begin{cases}\frac{1}{1-t} & \text{if } 0\leq t<1\\ \Delta & \text{if } \quad t\geq 1\end{cases},$$

then $\omega\in\Omega\left(\hat{S}\right)$.

Notation 33.6 *For $t\geq 0$, let $X_t:\Omega\to\hat{S}$ be defined by $X_t(\omega)=\omega(t)$ for all $\omega\in\Omega$, $\mathcal{B}_t:=\mathcal{B}_t^X:=\sigma(X_s:s\leq t)$, $\mathcal{B}_t^+:=\mathcal{B}_{t+}$, and $\mathcal{B}:=\vee_{t\geq 0}\mathcal{B}_t=\sigma(X_s:0\leq s<\infty)$.*

Theorem 33.7. *Suppose that $\{Q_t\}_{t\geq 0}$ is Feller semi-group. Then for each probability measure, ν, on $(S,\mathcal{B}_S)$, there is a probability measure P_ν on $(\Omega,\mathcal{B})$ such that; 1) $\mathrm{Law}_{P_\nu}(X_0)=\nu$ and 2) $\{X_t\}_{t\geq 0}$ is a time homogeneous Markov process having $\{Q_t\}_{t\geq 0}$ as its Markov transition kernels. Moreover if Q_t is conservative for all $t\geq 0$ (i.e. $Q_t1=1$ for all t) then we may replace Ω by the rcll paths on S.*

Rather than give a proof[1] of this theorem we will remind the reader of a few examples which we already know satisfy the hypothesis and the conclusion of the theorem.

Example 33.8 (Poisson Process). Let us continue the set up in Example 29.8 so that $S=\mathbb{N}_0$ and

$$Q_tf(x)=\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}e^{-\lambda t}f(x+n)$$

for all bounded functions, $f:S\to\mathbb{R}$. In this case $f\in C_0(S)$ iff $\lim_{x\to\infty}f(x)=0$ and if this is the case then for all $t>0$ we have by DCT that,

$$\lim_{x\to\infty}Q_tf(x)=\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}e^{-\lambda t}\lim_{x\to\infty}f(x+n)=0$$

so that $Q_t(C_0(S))\subset C_0(S)$. Moreover if $f:S\to\mathbb{R}$ is a bounded functions, then

$$\begin{aligned}|Q_tf(x)-f(x)|&=\left|\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}e^{-\lambda t}f(x+n)-f(x)\right|\\&\leq\left(1-e^{-\lambda t}\right)|f(x)|+e^{-\lambda t}\sum_{n=1}^{\infty}\frac{(\lambda t)^n}{n!}|f(x+n)|\\&\leq\left[\left(1-e^{-\lambda t}\right)+e^{-\lambda t}\sum_{n=1}^{\infty}\frac{(\lambda t)^n}{n!}\right]\|f\|_u.\end{aligned}$$

It now easily follows that $\lim_{t\downarrow 0}\|Q_tf-f\|_u=0$ for all bounded functions on S and hence for all $f\in C_0(S)$. This shows that $\{Q_t\}_{t\geq 0}$ is a Feller semi-group and so by Theorem 33.7 there we have a rccl Markov process on S which we have already constructed in Section 11.3.

[1] The idea of the proof is to start with the Markov processes constructed in Theorem 17.11. One then shows using martingale regularization techniques (see Section 28.2) in order to show that this Markov process admits a version with sample paths in $\Omega\left(\hat{S}\right)$. The interested reader may find the missing details in Kallenberg [30, Theorem 19.15 on p. 379].

Example 33.9 (Bounded Rate Markov Chains). We now re-examine Example 29.11. Recall that S is countable set and $a : S \times S \to \mathbb{R}$ is a function such that $a(x,y) \geq 0$ for all $x \neq y$, and there exists $\lambda < \infty$ such that

$$a_x := \sum_{y \neq x} a(x,y) \leq \lambda \text{ for all } x \in S.$$

We then set $A : \mathcal{S}_b \to \mathcal{S}_b$ by

$$Af(x) := \sum_{y \neq x} a(x,y)\,[f(y) - f(x)] \text{ for all } x \in S$$

and $Q_t = e^{tA}$ so that $Q_t f = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n f$ where

$$A^n f(x) = \sum_{y_1,\dots,y_n \in S} a(x,y_1)\,a(y_1,y_2)\dots a(y_{n-1},y_n)\,f(y_n)$$

where $a(x,x) := -\sum_{y \neq x} a(x,y)$ in these formulas. It was shown in Corollary 17.27 that $\{Q_t\}_{t \geq 0}$ defines a conservative Markov semi-group and that there is an associated rcll – Markov process associated to this semi-group. However in this case it is not actually true that Q_t is necessarily Feller. For example if we take $S = \mathbb{Z}$ and define $a(x,y) = 1_{x \neq 0} \cdot 1_{y=0}$ then $\sum_{y \neq x} a(x,y) = 1$ for $x \neq 0$ and 0 if $x = 0$. Thus we extend a via

$$a(x,y) = 1_{x \neq 0} \cdot (1_{y=0} - 1_{x=y}).$$

Taking $f = 1_{\{0\}} \in C_0(S)$ we find

$$(Af)(x) := \sum_{y \in S} a(x,y)\,f(y) = a(x,0) = 1_{x \neq 0} \cdot (1 - 1_{x=0}) = 1_{x \neq 0}.$$

So if $e^{tA} f \in C_0(S)$ then $e^{t(A+I)} f = e^t e^{tA} f \in C_0(S)$ and since $A + I$ has a positive matrix entries this would imply that $(A+I)f \in C_0(S)$ for $f \geq 0$. But this is not the case for $f = 1_{\{0\}}$. Nevertheless we still have a nice Markov process associated to these Markov kernels.

It is worth observing that $e^{tA} C(S) \subset C(S)$ as $C(S)$ is a Banach space and $e^{tA} f$ is convergent in the uniform norm topology. In fact we have

$$|Af(x)| \leq \sum_{y \in S} |a(x,y)|\,|f(y)| \leq 2\lambda \cdot \|f\|_u$$

so that $\|Af\|_u \leq 2\lambda \|f\|_u$. Therefore A is bounded in the operator norm and therefore e^{tA} is convergent in $\mathrm{End}(C_0(S))$ with $e^{tA} \to I$ as $t \downarrow 0$ in the operator norm.

Exercise 33.1. Suppose that $a : S \times S \to [-\lambda, \lambda]$ is as described in Example 33.9. Show that the associated semi-group is Feller provided $\lim_{x \to \infty} a(x,y) = 0$ for all $y \in S$, i.e. $a(\cdot, y) \in C_0(S)$ for all $y \in S$.

Solution to Exercise (33.1). Suppose that $f \in C_0(S)$ and $\varepsilon > 0$ are given. Choose $\Lambda_\varepsilon \subset_f S$ such that $|f| \leq \varepsilon$ on Λ_ε^c, then

$$\begin{aligned}|Af(x)| &\leq \sum_{y \in \Lambda_\varepsilon} |a(x,y)|\,|f(y)| + \varepsilon \cdot \sum_{y \notin \Lambda_\varepsilon} |a(x,y)| \\ &\leq \sum_{y \in \Lambda_\varepsilon} |a(x,y)|\,|f(y)| + 2\lambda\varepsilon.\end{aligned}$$

Therefore

$$\limsup_{x \to \infty} |Af(x)| \leq 2\lambda\varepsilon \to 0 \text{ as } \varepsilon \downarrow 0.$$

This shows that $Af \in C_0(S)$. As $C_0(S)$ is a closed subspace of $C(S)$ with the uniform norm and $e^{tA} f$ is convergent in $C(S)$ and each partial sum is in $C_0(S)$, it follows that $e^{tA} f \in C_0(S)$ as well. We have already seen that $e^{tA} \to I$ in the operator norm on $C(S)$ and therefore $e^{tA} f \to f$ for all $f \in C(S)$ as $t \downarrow 0$.

Example 33.10 (Brownian Motion). We now reconsider Example 29.12 so that $S = \mathbb{R}^d$ and

$$Q_t(x, dy) = \left(\frac{1}{2\pi t}\right)^{d/2} \exp\left(-\frac{1}{2t}\|y - x\|^2\right) dy.$$

In this case $\{Q_t\}_{t \geq 0}$ is a Feller semi-group. Indeed let $Z \overset{d}{=} N(0, I)$. If $f \in C_0(\mathbb{R}^d)$ and $t > 0$, then by DCT

$$\lim_{x \to \infty} Q_t(x; f) = \lim_{x \to \infty} \mathbb{E}\left[f\left(x + \sqrt{t} Z\right)\right] = \mathbb{E}\left[\lim_{x \to \infty} f\left(x + \sqrt{t} Z\right)\right] = 0$$

and

$$\begin{aligned}|Q_t(x; f) - f(x)| &= \left|\mathbb{E}\left[f\left(x + \sqrt{t} Z\right) - f(x)\right]\right| \\ &\leq \mathbb{E}\left|f\left(x + \sqrt{t} Z\right) - f(x)\right| \\ &\leq \mathbb{E}\left[\left|f\left(x + \sqrt{t} Z\right) - f(x)\right| : \|Z\| \leq M\right] + \mathbb{E}\left[\left|f\left(x + \sqrt{t} Z\right) - f(x)\right|\right.\end{aligned}$$

Thus it follows that

$$\|Q_t f - f\|_u \leq \sup_x \sup_{\|y\| \leq \sqrt{t} M} |f(x+y) - f(x)| + 2\|f\|_u P(\|Z\| > M).$$

Since f is uniformly continuous it follows that

$$\limsup_{t \downarrow 0} \|Q_t f - f\|_u \leq 2 \|f\|_u P(\|Z\| > M) \to 0 \text{ as } M \to \infty.$$

As we have already seen in Theorem 26.3, there is a corresponding rcll Markov process (called Brownian motion) and in fact we have seen that this process may be taken to be continuous.

*Nelson's Continuity Criteria

This chapter is devoted to giving Nelson's convenient criteria for verifying that the sample paths of a Markov process may be chosen to be continuous. Let (S, ρ) be a metric space and suppose that $\left(Q_t, X_t, \Omega, \mathcal{B}_t, \mathcal{B}, P, \{P_x\}_{x\in S}\right)$ is a time homogeneous Markov process as in Definition 17.13 and Theorem 17.14. For $\varepsilon, \delta > 0$ let

$$c(\varepsilon,\delta) := \sup\left\{Q_s\left(x, C(x,\varepsilon)^c\right) : 0 \le s \le \delta \text{ and } x \in S\right\} \tag{34.1}$$
$$= \sup_{0\le s\le\delta, x\in S} P_x\left(\rho(x, X_s) > \varepsilon\right) \tag{34.2}$$

where

$$C(x,\varepsilon) := \{y : \rho(y,x) \le \varepsilon\}$$

is the closed ε – ball in S which is centered at x. We are going to give a version of Nelson's continuity criteria, see [36, Theorem 2] and [37, p. 339-340]. Roughly speaking Theorem 34.4 below states if the probabilities of jumps of any fixed size over small time intervals is sufficiently small then the process "wants" to have continuous sample paths. More precisely we will prove if $c(\varepsilon,\delta) = o(\delta)$ for all $\varepsilon > 0$ then $\{X_t\}$ has a continuous modification. We begin with the following preliminary result.

Lemma 34.1. *For all $s, t \in \mathbb{R}_+$ and $\varepsilon > 0$ we have*

$$P\left(\rho(X_t, X_s) > \varepsilon\right) \le c\left(\varepsilon, |t-s|\right). \tag{34.3}$$

Hence if $c(\varepsilon,\delta) \to 0$ as $\delta \downarrow 0$, then $X_s \xrightarrow{P} X_t$ as $s \to t$.

Proof. Let $\nu := \mathrm{Law}_P(X_0)$ and suppose that $0 \le s < t$ for definiteness. We then have,

$$\begin{aligned}
P\left(\rho(X_t, X_s) > \varepsilon\right) &= \int_S d\nu(x)\, Q_{t-s}\left(x, 1_{\rho(x,\cdot)>\varepsilon}\right) \\
&= \int_S d\nu(x)\, Q_{t-s}\left(x, C(x,\varepsilon)^c\right) \\
&\le \int_S d\nu(x)\, c\left(\varepsilon, |t-s|\right) = c\left(\varepsilon, |t-s|\right).
\end{aligned}$$

■

We can in fact greatly improve on this estimate as we see in the next theorem.

Theorem 34.2. *For all countable subsets $\Lambda \subset \mathbb{R}_+$ with $\sup\Lambda - \inf\Lambda \le \delta$, we have*

$$P\left(\sup_{s,t\in\Lambda} \rho(X_s, X_t) > 4\varepsilon\right) \le 2c(\varepsilon,\delta) \text{ and} \tag{34.4}$$

$$P\left(\sup_{s,t\in\Lambda} \rho(X_s, X_t) > 4\varepsilon\right) \le \frac{c(\varepsilon,\delta)}{1 - c(\varepsilon,\delta)} \tag{34.5}$$

where P is any measure on $(\Omega, \mathcal{B})$ such that $\{X_t\}_{t\ge0}$ is a Markov process with transition kernels, $\{Q_t\}_{t\ge0}$. (We will refer the first inequality as Nelson's inequality. The second inequality is a variant of Skorohod's inequality in Theorem 20.45.)

Proof. We may with out loss of generality assume that Λ is a finite set. To see this let $\{\Lambda_n\}_{n=1}^\infty$ be a sequence of finite sets such that $\Lambda_n \uparrow \Lambda$ as $n \uparrow \infty$, then

$$\left\{\max_{s,t\in\Lambda_n} \rho(X_s, X_t) > 2\varepsilon\right\} \uparrow \left\{\sup_{s,t\in\Lambda} \rho(X_s, X_t) > 2\varepsilon\right\}$$

and so the estimates for finite Λ will give the estimates for countable Λ. So we now let $\Lambda = \{t_1 < t_2 < \dots < t_n\} \subset \mathbb{R}_+$ with $t_n - t_1 \le \delta$ and to simplify notation let $Y_j := X_{t_j}$ for all $1 \le j \le n$ and

$$\tau = \min\left\{k \in \{1, \dots, n\} : \rho(Y_k, Y_1) > 2\varepsilon\right\}$$

where $\min \emptyset := \infty$. Notice that

$$\{\tau \le n\} = \left\{\max_{t\in\Lambda} \rho(X_{t_1}, X_t) > 2\varepsilon\right\}.$$

Let

$$B := \left\{\rho(X_{t_n}, X_{t_1}) \ge \varepsilon\right\} = \left\{\rho(Y_n, Y_1) > \varepsilon\right\}.$$

For $1 \le k \le n$ on the event $\{\tau = k\} \setminus B$ we have

$$2\varepsilon < \rho(Y_k, Y_1) \le \rho(Y_k, Y_n) + \rho(Y_n, Y_1) < \rho(Y_k, Y_n) + \varepsilon$$

and so $\{\tau = k\} \setminus B \subset \left\{\rho(Y_k, Y_n) > \varepsilon\right\}$. So we have shown,

$$\{\tau = k\} \subset B \cup \{\tau = k \;\&\; \rho(Y_k, Y_n) > \varepsilon\}$$

which implies that

$$\{\tau \le n\} \subset B \cup \cup_{k=1}^{n} \{\tau = k \;\&\; \rho(Y_k, Y_n) > \varepsilon\}.$$

This leads to the estimate;

$$\begin{aligned} P(\tau \le n) &\le P(B) + \sum_{k=1}^{n} P(\tau = k \;\&\; \rho(Y_k, Y_n) > \varepsilon) \\ &= P(\rho(Y_n, Y_1) > \varepsilon) + \sum_{k=1}^{n} P(\tau = k \;\&\; \rho(Y_k, Y_n) > \varepsilon) \\ &\le c(\varepsilon, |t_n - t_1|) + \sum_{k=1}^{n} P(\tau = k \;\&\; \rho(Y_k, Y_n) > \varepsilon), \qquad (34.6) \end{aligned}$$

wherein we used Lemma 34.1 in the last inequality.

Since $\{\tau = k\}$ is $\mathcal{B}_{t_k}$ – measurable we may use the Markov property in Theorem 17.14 to find;

$$\begin{aligned} P(\tau = k \;\&\; \rho(Y_k, Y_n) > \varepsilon) &= \mathbb{E}\left[1_{\tau=k} \cdot P_{Y_k}\left[\rho\left(X_0, X_{t_n - t_k}\right) > \varepsilon\right]\right] \\ &= \mathbb{E}\left[1_{\tau=k} \cdot Q_{t_n - t_k}\left(Y_k, \{y \in S : \rho(Y_k, y) > \varepsilon\}\right)\right] \\ &\le \mathbb{E}\left[1_{\tau=k} \cdot c(\varepsilon, \delta)\right] = c(\varepsilon, \delta) P(\tau = k). \end{aligned}$$

Using this estimate back in Eq. (34.6) gives the estimate

$$P\left(\max_{t \in \Lambda} \rho(X_{t_1}, X_t) \ge 2\varepsilon\right) = P(\tau \le n) \le 2c(\varepsilon, \delta).$$

This inequality along with the observation;

$$\max_{s,t \in \Lambda} \rho(X_t, X_s) \le \max_{s,t \in \Lambda} \left[\rho(X_t, X_{t_1}) + \rho(X_{t_1}, X_s)\right] \le 2 \max_{t \in \Lambda} \rho(X_t, X_{t_1}) \quad (34.7)$$

proves Eq. (34.4).

The proof of Eq. (34.5) is similar. Working on the event $\{\tau \le n\} \cap \{\rho(Y_\tau, Y_n) \le \varepsilon\}$ we have

$$2\varepsilon < \rho(Y_\tau, Y_1) \le \rho(Y_\tau, Y_n) + \rho(Y_n, Y_1) \le \varepsilon + \rho(Y_n, Y_1)$$

and so we have shown

$$\{\tau \le n \;\&\; \rho(Y_\tau, Y_n) \le \varepsilon\} \subset \{\rho(Y_n, Y_1) > \varepsilon\}.$$

Hence it follows (again using the Markov property in Theorem 17.14) that

$$\begin{aligned} P(\rho(Y_n, Y_1) > \varepsilon) &\ge P(\tau \le n \;\&\; \rho(Y_\tau, Y_n) \le \varepsilon) \\ &= \sum_{k=1}^{n} P(\tau = k \;\&\; \rho(Y_k, Y_n) \le \varepsilon) \\ &= \sum_{k=1}^{n} \mathbb{E}\left[1_{\tau=k} \cdot P_{Y_k}\left(\rho\left(X_0, X_{t_n - t_k}\right)\right) \le \varepsilon\right] \\ &= \sum_{k=1}^{n} \mathbb{E}\left[1_{\tau=k} \cdot Q_{t_n - t_k}\left(Y_k, C(Y_k, \varepsilon)\right)\right] \\ &\ge \sum_{k=1}^{n} \mathbb{E}\left[1_{\tau=k} \cdot \inf_{x \in S} Q_{t_n - t_k}\left(x, C(x, \varepsilon)\right)\right] \\ &\ge \inf_{x \in S} \inf_{0 \le s \le \delta} Q_s(x, C(x, \varepsilon)) \cdot \sum_{k=1}^{n} P(\tau = k) \\ &= \inf_{x \in S} \inf_{0 \le s \le \delta} Q_s(x, C(x, \varepsilon)) \cdot P(\tau \le n). \end{aligned}$$

Thus we have shown

$$\begin{aligned} &P\left(\max_{t \in \Lambda} \rho(X_{t_1}, X_t) > 2\varepsilon\right) \\ &\le \frac{1}{\inf_{x \in S} \inf_{0 \le s \le \delta} Q_s(x, C(x, \varepsilon))} P(\rho(Y_n, Y_1) > \varepsilon) \\ &= \frac{1}{1 - \sup_{x \in S} \sup_{0 \le s \le \delta} Q_s(x, \{y : \rho(y, x) > \varepsilon\})} P(\rho(X_{t_n}, X_{t_1}) > \varepsilon) \\ &= \frac{1}{1 - c(\varepsilon, \delta)} P(\rho(X_{t_n}, X_{t_1}) > \varepsilon) \le \frac{c(\varepsilon, \delta)}{1 - c(\varepsilon, \delta)} \end{aligned}$$

wherein we have used Lemma 34.1 in the last inequality. Equation (34.5) now follows from this inequality and Eq. (34.7). ■

Corollary 34.3. *Continuing the notation used above and further assume that* $1/\delta \in \mathbb{N}$. *Then for all* $N \in \mathbb{N}$;

$$P\left(\sup\{\rho(X_s, X_t) : s, t \in [0, N) \cap \mathbb{Q} \;\&\; |t - s| \le \delta\} \ge 4\varepsilon\right) \le 2N \frac{c(\varepsilon, 2\delta)}{\delta}.$$

Proof. Let $J_k := [(k-1)\delta, k\delta)$ so that $[0, N) = \sum_{k=1}^{N/\delta} J_k$. If $s, t \in [0, N)$ with $s < t \le s + \delta$, then $s \in J_k$ for some unique k and consequently $t \in J_k \cup J_{k+1}$. From this observation it follows that

$$\begin{aligned} &\sup\{\rho(X_s, X_t) : s, t \in [0, N) \cap \mathbb{Q} \;\&\; |t - s| \le \delta\} \\ &\quad = \max_{1 \le k \le N/\delta} \sup\{\rho(X_s, X_t) : s, t \in [J_k \cup J_{k+1}] \cap \mathbb{Q}\} \end{aligned}$$

and therefore,

$$\begin{aligned} P\left(\sup\left\{\rho\left(X_s, X_t\right) : s,t \in [0,N)\cap\mathbb{Q} \;\&\; |t-s|\le\delta\right\}\ge 4\varepsilon\right) \\ \le \sum_{k=1}^{N/\delta} P\left(\sup\left\{\rho\left(X_s,X_t\right) : s,t\in [J_k\cup J_{k+1}]\cap\mathbb{Q}\right\}\ge 4\varepsilon\right) \\ \le \sum_{k=1}^{N/\delta} 2c\left(\varepsilon,2\delta\right) = 2N\frac{c\left(\varepsilon,2\delta\right)}{\delta}. \end{aligned}$$

■

Theorem 34.4 (Nelson's continuity theorem). *Suppose now that (S,ρ) is a complete metric space and $(Q_t, X_t, \Omega, \mathcal{B}_t, \mathcal{B}, P)$ are as in Theorem 17.14 and suppose for each $\varepsilon > 0$ that $c(\varepsilon,\delta) = o(\delta)$ (i.e. $\lim_{\delta\downarrow 0}[c(\varepsilon,\delta)/\delta] = 0$) where $c(\varepsilon,\delta)$ is defined in Eq. (34.1). Then $\{X_t\}_{t\ge 0}$ admits a continuous version.*

Proof. Let Ω_0 denote those $\omega\in\Omega$ such that $\mathbb{Q}_+ \ni t \to X_t(\omega)$ is uniformly continuous on bounded subsets. Since $\omega\notin\Omega_0$ iff there exists $N\in\mathbb{N}$ and an $\varepsilon = 1/k > 0$ such that for all $\delta = 1/n$ we will have $\sup\{\rho(X_s,X_t) : s,t\in[0,N)\cap\mathbb{Q} \;\&\; |t-s|\le\delta\}\ge 4\varepsilon$ it follows that

$$\Omega_0^c = \cup_{N\in\mathbb{N}}\cup_{k\in\mathbb{N}}\cap_{n\in\mathbb{N}} E_{N,k,n}$$

where

$$E_{N,k,n} := \left\{\sup\left\{\rho\left(X_s,X_t\right) : s,t\in[0,N)\cap\mathbb{Q} \;\&\; |t-s|\le 1/n\right\}\ge 4/k\right\}$$

However making use of the continuity properties of P and Corollary 34.3 we learn that

$$P\left(\Omega_0^c\right) = \lim_{N\to\infty}\lim_{k\to\infty}\lim_{n\to\infty} P\left(E_{N,k,n}\right) = 0.$$

Thus for $\omega\in\Omega_0$ (as set of full measure), $\mathbb{Q}_+\ni t\to X_t(\omega)$ is uniformly continuous on bounded subsets and therefore extends uniquely to a continuous function, $\tilde{X}_t(\omega)$ for all $t\in\mathbb{R}_+$. (This is where we use the completeness of (S,ρ).) For $\omega\in\Omega_0^c$ let us define $\tilde{X}_t(\omega) = *$ for some fixed point $*\in S$. To finish the proof it suffices to show that $\tilde{X}_t = X_t$ a.s. for every $t\in\mathbb{R}_+$. But this follows from the facts; 1) $X_s = \tilde{X}_s$ a.s. when $s\in\mathbb{Q}_+$, 2) $X_s \xrightarrow{P} X_t$ as $s\to t$ by Lemma 34.1, and $\tilde{X}_t = \lim_{\mathbb{Q}\ni s\to t} X_s$ a.s. and therefore $X_t = \tilde{X}_t$ a.s. ■

As an example we give yet another proof for the existence of Brownian motion.

Theorem 34.5 (Wiener). *Associated to the homogeneous Markov semi-group, $\{Q_t\}_{t\ge 0}$ on $\mathbb{R}$ defined by*

$$Q_t(x,dy) := \frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{1}{2}(y-x)^2\right)dy,$$

is a Markov process, $\{B_t\}_{t\ge 0}$, with continuous sample paths. (Of course when we start the process at 0, $\{B_t\}_{t\ge 0}$ is a standard Brownian motion as in Definition 17.21.)

Proof. In Exercise 17.8 you showed that is a time homogeneous Markov kernel and therefore by Theorem 17.11 there exists a corresponding Markov process. Recall from Exercise 17.8 that

$$Q_t(x,f) = \mathbb{E}\left[f\left(x+\sqrt{t}Z\right)\right]$$

where Z is standard normal random variable, we find that

$$\begin{aligned} Q_t\left(x,[x-\varepsilon,x+\varepsilon]^c\right) &= P\left(x+\sqrt{t}Z\in[x-\varepsilon,x+\varepsilon]^c\right) \\ &= P\left(|Z|>\varepsilon/\sqrt{t}\right) \end{aligned}$$

and therefore,

$$c(\varepsilon,\delta) = \sup_{0\le t\le\delta} P\left(|Z|>\varepsilon/\sqrt{t}\right) = P\left(|Z|>\varepsilon/\sqrt{\delta}\right).$$

By Chebyshev's inequality we have for all $p>0$ that $c(\varepsilon,\delta)\le\frac{\delta^{p/2}}{\varepsilon^p}\mathbb{E}|Z|^p$ which is clearly $o(\delta)$ when $p>2$. Alternatively we may use the Gaussian tail estimates in Lemma 7.59 in order to conclude that

$$c(\varepsilon,\delta)\sim\sqrt{\frac{2}{\pi}}\frac{\sqrt{\delta}}{\varepsilon}e^{-\delta/2\varepsilon^2} \text{ as } \delta\downarrow 0.$$

The result now follows from Theorem 34.4. ■

Part VII

Continuous Time Markov Chains

This part of the document needs significant editing! Please read at your own risk. Strictly speaking the material here might logically come before Brownian motion. But as this section is still needs so much work, I thought it was best to isolate it by itself for now.

Basics of continuous time chains

(This chapter needs serious editing! It was originally written for Math 180 and as such is not properly integrated into the material above. For example there is a lot of redundancy with Chapters 17 and 29. Moreover the notation is not consistent with that chapter.)

In this chapter we are going to begin out study continuous time homogeneous Markov chains on discrete state spaces S. In more detail we will assume that $\{X_t\}_{t\geq 0}$ is a stochastic process whose sample paths are right continuous, see Figures 35.1 and 35.2. (These processes need not have left hand limits if there are an infinite number of jumps in a finite time interval. For the most part we will assume that this does not happen almost surely.) Recall from Theorem 17.26 and Corollary 17.27 above that we have seen such Markov chains exist.

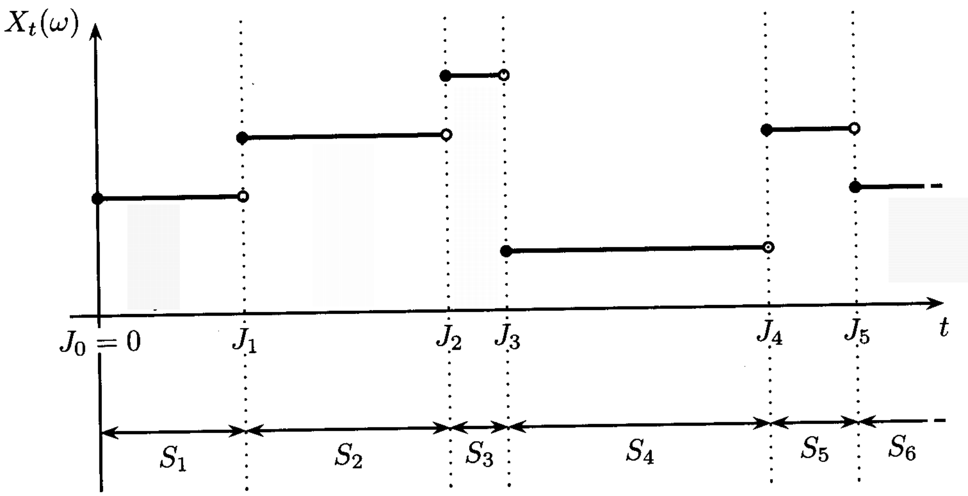


Fig. 35.1. Typical sample paths of a continuous time Markov chain in a discrete state space.

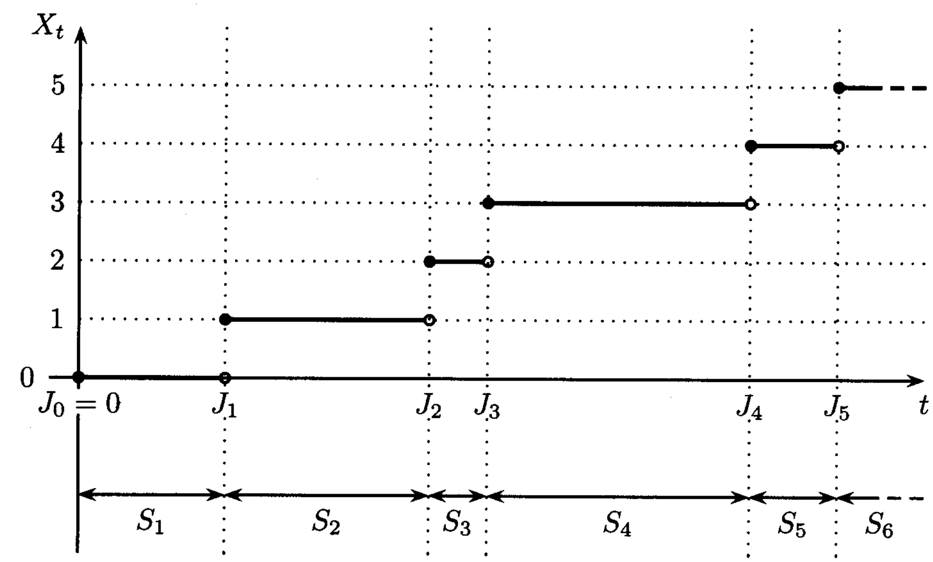


Fig. 35.2. A sample path of a birth process. Here the state space is $\{0,1,2,\dots\}$ to be thought of the possible population size.

As in the discrete time Markov chain setting, to each $i \in S$, we will write $P_i(A) := P(A|X_0 = i)$. That is P_i is the probability associated to the scenario where the chain is forced to start at site i. We now define, for $i, j \in S$,

$$P_{ij}(t) := P_i(X(t) = j) \tag{35.1}$$

which is the probability of finding the chain at time t at site j given the chain starts at i.

Definition 35.1. *The* ***time homogeneous Markov property*** *states for every* $0 \leq s < t < \infty$ *and any choices of* $0 = t_0 < t_1 < \cdots < t_n = s < t$ *and* $i_1, \dots, i_n \in S$ *that*

$$P_i(X(t) = j | X(t_1) = i_1, \dots, X(t_n) = i_n) = P_{i_n, j}(t - s), \tag{35.2}$$

and consequently,

$$P_i\left(X\left(t\right)=j|X\left(s\right)=i_n\right)=P_{i_n,j}\left(t-s\right). \tag{35.3}$$

Roughly speaking the Markov property may be stated as follows; the probability that $X\left(t\right)=j$ given knowledge of the process up to time s is $P_{X(s),j}\left(t-s\right)$. In symbols we might express this last sentence as

$$P_i\left(X\left(t\right)=j|\left\{X\left(\tau\right)\right\}_{\tau\le s}\right)=P_i\left(X\left(t\right)=j|X\left(s\right)\right)=P_{X(s),j}\left(t-s\right).$$

So again a continuous time Markov process is forgetful in the sense what the chain does for $t\ge s$ depend only on where the chain is located, $X\left(s\right)$, at time s and not how it got there. See Fact 35.3 below for a more general statement of this property.

Definition 35.2 (Informal). *A stopping time, T, for $\left\{X\left(t\right)\right\}$, is a random variable with the property that the event $\left\{T\le t\right\}$ is determined from the knowledge of $\left\{X\left(s\right):0\le s\le t\right\}$. Alternatively put, for each $t\ge0$, there is a functional, f_t, such that*

$$1_{T\le t}=f_t\left(\left\{X\left(s\right):0\le s\le t\right\}\right).$$

As in the discrete state space setting, the first time the chain hits some subset of states, $A\subset S$, is a typical example of a stopping time whereas the last time the chain hits a set $A\subset S$ is typically **not** a stopping time. Similar the discrete time setting, the Markov property leads to a strong form of forgetfulness of the chain. This property is again called the **strong Markov property** which we take for granted here.

Fact 35.3 (Strong Markov Property) *If $\left\{X\left(t\right)\right\}_{t\ge0}$ is a Markov chain, T is a stopping time, and $j\in S$, then, conditioned on $\left\{T<\infty \text{ and } X_T=j\right\}$,*

$$\left\{X\left(s\right):0\le s\le T\right\} \text{ and } \left\{X\left(t+T\right):t\ge0\right\} \text{ are independent}$$

and $\left\{X\left(t+T\right):t\ge0\right\}$ has the same distribution as $\left\{X\left(t\right)\right\}_{t\ge0}$ under P_j. (See Theorem 35.19 and Corollary 35.20.)

We will use the above fact later in our discussions. For the moment, let us go back to more elementary considerations.

Theorem 35.4 (Finite dimensional distributions). *Let $0<t_1<t_2<\dots<t_n$ and $i_0,i_1,i_2,\dots,i_n\in S$. Then*

$$\begin{aligned}P_{i_0}(X_{t_1}&=i_1,X_{t_2}=i_2,\dots,X_{t_n}=i_n)\\&=P_{i_0,i_1}(t_1)P_{i_1,i_2}(t_2-t_1)\dots P_{i_{n-1},i_n}(t_n-t_{n-1}).\end{aligned} \tag{35.4}$$

Proof. The proof is similar to that of Proposition **??**. For notational simplicity let us suppose that $n=3$. We then have

$$\begin{aligned}P_{i_0}(X_{t_1}=i_1,X_{t_2}=i_2,X_{t_3}=i_3)&=P_{i_0}(X_{t_3}=i_3|X_{t_1}=i_1,X_{t_2}=i_2)P_{i_0}\left(X_{t_1}=i_1,X_{t_2}=\right.\\&=P_{i_2,i_3}\left(t_3-t_2\right)P_{i_0}\left(X_{t_1}=i_1,X_{t_2}=i_2\right)\\&=P_{i_2,i_3}\left(t_3-t_2\right)P_{i_0}\left(X_{t_2}=i_2|X_{t_1}=i_1\right)P_{i_0}\left(X_{t_1}=i_1\right)\\&=P_{i_2,i_3}\left(t_3-t_2\right)P_{i_1,i_2}\left(t_2-t_1\right)P_{i_0,i_1}\left(t_1\right)\end{aligned}$$

wherein we have used the Markov property once in line 2 and twice in line 4. ■

Proposition 35.5 (Properties of P). *Let $P_{ij}\left(t\right):=P_i\left(X\left(t\right)=j\right)$ be as above. Then:*

1. *For each $t\ge0$, $P\left(t\right)$ is a Markov matrix, i.e.*
$$\sum_{j\in S}P_{ij}\left(t\right)=1 \text{ for all } i\in S \text{ and}$$
$$P_{ij}\left(t\right)\ge0 \text{ for all } i,j\in S.$$
2. *$\lim_{t\downarrow0}P_{ij}\left(t\right)=\delta_{ij}$ for all $i,j\in S$.*
3. *The **Chapman – Kolmogorov equation** holds:*
$$P(t+s)=P(t)P(s) \text{ for all } s,t\ge0, \tag{35.5}$$
i.e.
$$P_{ij}\left(t+s\right)=\sum_{k\in S}P_{ik}\left(s\right)P_{kj}\left(t\right) \text{ for all } s,t\ge0. \tag{35.6}$$

*We will call a matrix $\left\{P\left(t\right)\right\}_{t\ge0}$ satisfying items 1. – 3. a **continuous time Markov semigroup.***

Proof. Most of the assertions follow from the basic properties of conditional probabilities. The assumed right continuity of X_t implies that $\lim_{t\downarrow0}P(t)=P(0)=I$. From Equation (35.4) with $n=2$ we learn that

$$\begin{aligned}P_{i_0,i_2}(t_2)&=\sum_{i_1\in S}P_{i_0}(X_{t_1}=i_1,X_{t_2}=i_2)\\&=\sum_{i_1\in S}P_{i_0,i_1}(t_1)P_{i_1,i_2}(t_2-t_1)\\&=[P(t_1)P(t_2-t_1)]_{i_0,i_2}.\end{aligned}$$

■

At this point it is not so clear how to find a non-trivial (i.e. $P\left(t\right)\ne I$ for all t) example of a continuous time Markov semi-group. It turns out the Poisson process provides such an example.

Example 35.6. In this example we will take $S = \{0,1,2,\dots\}$ and then define, for $\lambda > 0$,

$$\begin{array}{c} \\ P(t) = e^{-\lambda t} \end{array} \begin{array}{c} \begin{array}{ccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & \dots \end{array} \\ \left[\begin{array}{cccccccc} 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \frac{(\lambda t)^4}{4!} & \frac{(\lambda t)^5}{5!} & \dots \\ 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \frac{(\lambda t)^4}{4!} & \dots \\ 0 & 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \dots \\ 0 & 0 & 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{array}\right] \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}.$$

In components this may be expressed as,

$$P_{ij}(t) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} 1_{i \le j}$$

with the convention that $0! = 1$. (See Exercise **??** to see where this example is coming from.)

If $i, j \in S$, then $P_{ik}(t) P_{kj}(s)$ will be zero unless $i \le k \le j$, therefore we have

$$\begin{aligned} \sum_{k \in S} P_{ik}(t) P_{kj}(s) &= 1_{i \le j} \sum_{i \le k \le j} P_{ik}(t) P_{kj}(s) \\ &= 1_{i \le j} e^{-\lambda(t+s)} \sum_{i \le k \le j} \frac{(\lambda t)^{k-i}}{(k-i)!} \frac{(\lambda s)^{j-k}}{(j-k)!}. \end{aligned} \tag{35.7}$$

Let $k = i + m$ with $0 \le m \le j - i$, then the above sum may be written as

$$\sum_{m=0}^{j-i} \frac{(\lambda t)^m}{m!} \frac{(\lambda s)^{j-i-m}}{(j-i-m)!} = \frac{1}{(j-i)!} \sum_{m=0}^{j-i} \binom{j-i}{m} (\lambda t)^m (\lambda s)^{j-i-m}$$

and hence by the Binomial formula we find,

$$\sum_{i \le k \le j} \frac{(\lambda t)^{k-i}}{(k-i)!} \frac{(\lambda s)^{j-k}}{(j-k)!} = \frac{1}{(j-i)!} (\lambda t + \lambda s)^{j-i}.$$

Combining this with Eq. (35.7) shows that

$$\sum_{k \in S} P_{ik}(t) P_{kj}(s) = P_{ij}(s+t).$$

Proposition 35.7. *Let $\{X_t\}_{t \ge 0}$ is the Markov chain determined by $P(t)$ of Example 35.6. Then relative to P_0, $\{X_t\}_{t \ge 0}$ is precisely the Poisson process on $[0, \infty)$ with intensity λ.*

Proof. Let $0 \le s < t$. Since $P_0(X_t = n | X_s = k) = P_{kn}(t-s) = 0$ if $n < k$, $\{X_t\}_{t \ge 0}$ is a non-decreasing integer value process. Suppose that $0 = s_0 < s_1 < s_2 < \dots < s_n = s$ and $i_k \in S$ for $k = 0,1,2,\dots,n$, then

$$\begin{aligned} P_0 &\left(X_t - X_s = i_0 | X_{s_j} = i_j \text{ for } 1 \le j \le n\right) \\ &= P_0\left(X_t = i_n + i_0 | X_{s_j} = i_j \text{ for } 1 \le j \le n\right) \\ &= P_0\left(X_t = i_n + i_0 | X_{s_n} = i_n\right) \\ &= e^{-\lambda(t-s)} \frac{(\lambda t)^{i_0}}{i_0!}. \end{aligned}$$

Since this answer is independent of $i_1, \dots, i_n$ we also have

$$\begin{aligned} &P_0(X_t - X_s = i_0) \\ &\quad = \sum_{i_1,\dots,i_n \in S} P_0\left(X_t - X_s = i_0 | X_{s_j} = i_j \text{ for } 1 \le j \le n\right) P_0\left(X_{s_j} = i_j \text{ for } 1 \le j \le n\right) \\ &\quad = \sum_{i_1,\dots,i_n \in S} e^{-\lambda(t-s)} \frac{(\lambda t)^{i_0}}{i_0!} P_0\left(X_{s_j} = i_j \text{ for } 1 \le j \le n\right) = e^{-\lambda(t-s)} \frac{(\lambda t)^{i_0}}{i_0!}. \end{aligned}$$

Thus we may conclude that $X_t - X_s$ is Poisson random variable with intensity λ which is independent of $\{X_r\}_{r \le s}$. That is $\{X_t\}_{t \ge 0}$ is a Poisson process with rate λ. ■

The next example is generalization of the Poisson process example above. You will be asked to work this example out on a future homework set.

Example 35.8. In problems VI.6.P1 on p. 406, you will be asked to consider a discrete time Markov matrix, ρ_{ij}, on some discrete state space, S,with associate Markov chain $\{Y_n\}$. It is claimed in this problem that if $\{N(t)\}_{t \ge 0}$ is Poisson process which is independent of $\{Y_n\}$, then $X_t := Y_{N(t)}$ is a continuous time Markov chain. More precisely the claim is that Eq. (35.2) holds with

$$P(t) = e^{-t} \sum_{m=0}^{\infty} \frac{t^m}{m!} \rho^m =: e^{t(\rho - I)},$$

i.e.

$$P_{ij}(t) = e^{-t} \sum_{m=0}^{\infty} \frac{t^m}{m!} (\rho^m)_{ij}.$$

(We will see a little later, that this example can be used to construct all finite state continuous time Markov chains.)

Notice that in each of these examples, $P(t) = I + Qt + O(t^2)$ for some matrix Q. In the first example,

$$Q_{ij} = -\lambda\delta_{ij} + \lambda\delta_{i,i+1}$$

while in the second example, $Q = \rho - I$.

For a general Markov semigroup, $P(t)$, we are going to show (at least when $\#(S) < \infty$) that $P(t) = I + Qt + O(t^2)$ for some matrix Q which we call the **infinitesimal generator (or Markov generator)** of P. We will see that every infinitesimal generator must satisfy:

$$Q_{ij} \le 0 \text{ for all } i \ne j, \text{ and} \tag{35.8}$$

$$\sum_j Q_{ij} = 0, \text{ i.e. } -Q_{ii} = \sum_{j\ne i} Q_{ij} \text{ for all } i. \tag{35.9}$$

Moreover, to any such Q, the matrix

$$P(t) = e^{tQ} := \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n = I + tQ + \frac{t^2}{2!}Q^2 + \frac{t^3}{3!}Q^3 + \dots$$

will be a Markov semigroup.

One useful way to understand what is going on here is to choose an initial distribution, π on S and then define $\pi(t) := \pi P(t)$. We are going to interpret π_j as the amount of sand we have placed at each of the sites, $j \in S$. We are going to interpret $\pi_j(t)$ as the mass at site j at a later time t under the assumption that π satisfies, $\dot\pi(t) = \pi(t)Q$, i.e.

$$\dot\pi_j(t) = \sum_{i\ne j} \pi_i(t) Qij - q_j\pi_j(t), \tag{35.10}$$

where $q_j = -Q_{j\,j}$. (See Example 36.19 below.) Here is how to interpret each term in this equation:

$$\begin{aligned} \dot\pi_j(t) &= \text{ rate of change of the amount of sand at } j \text{ at time } t,\\ \pi_i(t)Qij &= \text{ rate at which sand is shoveled from site } i \text{ to } j,\\ q_j\pi_j(t) &= \text{ rate at which sand is shoveled out of site } i \text{ to all other sites.} \end{aligned}$$

With this interpretation Eq. 35.10 has the clear meaning: namely the rate of change of the mass of sand at j at time t should be equal to the rate at which sand is shoveled into site j form all other sites minus the rate at which sand is shoveled out of site i. With this interpretation, the condition,

$$-Q_{j\,j} := q_j = \sum_{k\ne j} Q_{j,k},$$

just states the total sand in the system should be conserved, i.e. this guarantees the rate of sand leaving j should equal the total rate of sand being sent to all of the other sites from j.

Warning: the book denotes Q by A but then denotes the entries of A by q_{ij}. I have just decided to write $A = Q$ and identify, Q_{ij} and q_{ij}. To avoid some technical details, in the next chapter we are mostly going to restrict ourselves to the case where $\#(S) < \infty$. Later we will consider examples in more detail where $\#(S) = \infty$.

35.1 Construction of continuous time Markov processes

(Also see Section 37.3.)

Proposition 35.9 (Measure Theoretic). *Let Ω be a set, $\{\Omega_n\}$ be a finite or countable partition of Ω, and for each suppose that $\mathcal{B}_n$ is a σ – algebra on Ω_n. Then*

$$\mathcal{B} := \left\{\sum_n A_n : A_n \in \mathcal{B}_n\right\}$$

is a σ – algebra on Ω. Moreover, $f : \Omega \to \mathbb{R}$ is $\mathcal{B}$ – measurable iff $f|_{\Omega_n} : \Omega_n \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable for all n. We will write $\oplus_n \mathcal{B}_n$ for $\mathcal{B}$ in this case.

Proof. It is clear that $\emptyset \in \mathcal{B}$ and that $\mathcal{B}$ is closed under countable and finite unions. Thus it only is necessary to observe that $\mathcal{B}$ is closed under complementation which is a consequence of the following identity;

$$\begin{aligned} \Omega \setminus \left(\sum_n A_n\right) &= \left(\sum_n \Omega_n\right) \setminus \left(\sum_n A_n\right)\\ &= \sum_n \Omega_n \setminus \left(\sum_m A_m\right) = \sum_n \Omega_n \setminus A_n. \end{aligned}$$

For the measurability assertion regarding f just observe that $\mathcal{B}_{\Omega_n} = \mathcal{B}_n$ for all n. ■

Theorem 35.10 (Same as Theorem 17.26). *Let $\{\rho_{ij}\}_{i,j\in S}$ be a discrete time Markov matrix over a discrete state space, S and $\{Y_n\}_{n=0}^{\infty}$ be the corresponding Markov chain. Also let $\{N_t\}_{t\ge 0}$ be a Poisson process with rate $\lambda > 0$ which is independent of $\{Y_n\}$. Then $X_t := Y_{N_t}$ is a continuous time Markov chain with transition semi-group given by,*

$$P(t) = e^{t\lambda(\rho - I)} = e^{-\lambda t}e^{t\lambda\rho}.$$

Proof. Let us begin by computing,

$$\begin{aligned}
\mathbb{E}_x f(X_t) = \mathbb{E}_x f(Y_{N_t}) &= \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathbb{E}_x f(Y_n) \\
&= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \sum_{y \in S} \rho^n(x,y) f(y) \\
&= e^{-\lambda t} \sum_{y \in S} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \rho^n(x,y) f(y) \\
&= e^{-\lambda t} \sum_{y \in S} e^{\lambda t\ \rho}(x,y) f(y) = \left(e^{\lambda t(\rho - I)} f\right)(x) \\
&= P(t) f(x).
\end{aligned}$$

Thus if show that $\{X_t\}_{t\geq 0}$ is Markov it will follow that $P(t)$ is the desired semi-group.

For each $t \in \mathbb{R}_+$, notice that $\{N_t = n\}_{n=0}^{\infty}$ is a partition of Ω and so we may define

$$\mathcal{B}_t := \oplus_{n=0}^{\infty} \sigma(N_s : s \leq t,\ Y_0, \dots, Y_n)_{\{N_t = n\}}.$$

Observe that $\mathcal{B}_t$ is an increasing filtration. Indeed, if $s \leq t$, then $\mathcal{B}_s$ is generated by functions of the form,

$$g := 1_{N_s = k} f(N_{\sigma_1}, \dots, N_{\sigma_n}, Y_0, \dots, Y_k)$$

where $0 \leq \sigma_1 < \sigma_2 < \dots < \sigma_n \leq s$. Since $1_{N_t=n} 1_{N_s=k} = 0$ unless $k \leq n$, it follows that

$$1_{N_t = n} g = 1_{N_t = n} 1_{N_s = k} f(N_{\sigma_1}, \dots, N_{\sigma_n}, Y_0, \dots, Y_k)$$

is $\sigma(N_s : s \leq t,\ Y_0, \dots, Y_n)_{\{N_t=n\}}$ – measurable for each $n \in \mathbb{N}_0$ and therefore g is $\mathcal{B}_t$ – measurable. This shows that $\mathcal{B}_s \subset \mathcal{B}_t$ as required.

We are going to now show that $\{X_t\}$ is $\mathcal{B}_t$ – Markov. To this end we have for g as above (making use of the independence of N and Y and the Markov – property for Y),

$$\begin{aligned}
\mathbb{E}[g \cdot u(X_t)] &= \mathbb{E}[g \cdot u(Y_{N_t})] = \mathbb{E}[g \cdot u(Y_{N_s + N_t - N_s})] \\
&= \mathbb{E}[1_{N_s = k} f(N_{\sigma_1}, \dots, N_{\sigma_n}, Y_0, \dots, Y_k) \cdot u(Y_{k + N_t - N_s})] \\
&= e^{-(t-s)\lambda} \sum_{l=0}^{\infty} \frac{[(t-s)\lambda]^l}{l!} \mathbb{E}[1_{N_s=k} f(N_{\sigma_1}, \dots, N_{\sigma_n}, Y_0, \dots, Y_k) \cdot u(Y_{k+l})] \\
&= e^{-(t-s)\lambda} \sum_{l=0}^{\infty} \frac{[(t-s)\lambda]^l}{l!} \mathbb{E}\left[1_{N_s=k} f(N_{\sigma_1}, \dots, N_{\sigma_n}, Y_0, \dots, Y_k) \cdot (\rho^l u)(Y_k)\right] \\
&= \mathbb{E}\left[1_{N_s=k} f(N_{\sigma_1}, \dots, N_{\sigma_n}, Y_0, \dots, Y_k) \cdot e^{-(t-s)\lambda} \sum_{l=0}^{\infty} \frac{[(t-s)\lambda]^l}{l!} (\rho^l u)(Y_k)\right] \\
&= \mathbb{E}[g \cdot (P(t-s) u)(Y_{N_s})] = \mathbb{E}[g \cdot (P(t-s)u)(X_s)].
\end{aligned}$$

Thus it follows that $\mathbb{E}[u(X_t) | \mathcal{B}_t] = (P(t-s)u)(X_s)$ as we wished to show. ∎

Lemma 35.11. *If $q(x,y)$ is a rate matrix on S such that $q(x) := \sum_{y \neq x} q(x,y)$ is a bounded function on S, then there exists a Markov matrix $\rho(x,y)$ on S and $\lambda > 0$ such that $q(x,y) = \lambda(\rho(x,y) - \delta_{x,y})$.*

Proof. Let $\lambda := \sup_{x \in S} q(x)$ and let us look for $\rho(x,y)$ such that $q(x,y) = \lambda(\rho(x,y) - \delta_{xy})$. Clearly we must take $\rho(x,y) = \frac{1}{\lambda} q(x,y)$ for $x \neq y$ and for $x = y$ we will have $-q(x) = \lambda(\rho(x,x) - 1)$ so that $\rho(x,x) = 1 - \frac{1}{\lambda} q(x)$, i.e.

$$\rho(x,y) = \frac{1}{\lambda} q(x,y) 1_{x \neq y} + 1_{x=y}\left(1 - \frac{1}{\lambda} q(x)\right).$$

Notice that $\rho(x,y) \geq 0$ for all $x, y \in S$ and that

$$\sum_{y \in S} \rho(x,y) = \frac{1}{\lambda} q(x) + \left(1 - \frac{1}{\lambda} q(x)\right) = 1$$

so that $\rho(x,y)$ is indeed a Markov matrix. ∎

Corollary 35.12. *If $q(x,y)$ is a rate matrix on S such that $q(x) := \sum_{y \neq x} q(x,y)$ is a bounded function on S, then there exists a time homogeneous Markov process $\{X_t\}_{t \geq 0}$ with values in S such that $\mathbb{E}_x f(X_t) = \left(e^{tQ} f\right)(x)$ for all $f \in b\mathcal{S}$ and $t \geq 0$.*

Proof. Let λ and ρ be as in Lemma 35.11. Then define $X_t = Y_{N_t}$ as in Theorem 35.10. ∎

35.2 Markov Properties in more detail

This section is pretty much redundant now as it has all been done in Chapter 29.

Notation 35.13 *In what follows, let Ω denote those paths, $\omega : \mathbb{R}_+ \to S$, which are right continuous with left hand limits where S is equipped with the discrete topology. Let $\theta_t : \Omega \to \Omega$ and $X_t : \Omega \to S$ be defined by,*

$$\theta_t(\omega) = \omega(\cdot + t) \text{ and}$$
$$X_t(\omega) := \omega(t)$$

respectively. Further let $\mathcal{B}_t := \sigma(X_s : s \leq t)$, $\mathcal{B}_{t+} := \cap_{s > t} \mathcal{B}_t$, and $\mathcal{B} := \sigma(X_s : s < \infty) = \vee_{t < \infty} \mathcal{B}_t$.

We now assume for each $x \in S$ there is a probability measure, P_x, on $(\Omega, \mathcal{B})$ such that

$$P_x\left(X_0 = y_0, X_{t_1} = y_1, \dots, X_{t_n = y_n}\right) = \delta_{xy_0} \prod_{i=1}^{n} P_{y_{i-1},y_i}\left(t_i - t_{i-1}\right) \qquad (35.11)$$

for all $0 = t_0 < t_1 < \dots < t_n$, $\{y_i\}_{i=0}^n \subset S$, and $n \in \mathbb{N}$.

Definition 35.14. *A function, $F : \Omega \to \mathbb{R}$ is said to be a* ***cylinder function*** *if there exists times,*

$$0 = t_0 < t_1 < t_2 < \dots < t_n < \infty,$$

and a measurable function, $f : S^{n+1} \to \mathbb{R}$ such that

$$F = f\left(X_{t_0}, \dots, X_{t_n}\right). \qquad (35.12)$$

Notation 35.15 *Suppose that ν is a measure on S, let*

$$P_\nu := \int_S d\nu(x) P_x = \sum_{x \in S} \nu(x) P_x.$$

(Some of what follows is already done in Chapter 17.)

Theorem 35.16 (Markov Property I). *Let ν be a probability measure on S and $F \in b\mathcal{B}$ (the space of bounded $\mathcal{B}$ – measurable functions), then*

$$\mathbb{E}_\nu\left[F \circ \theta_t | \mathcal{B}_t\right] = \mathbb{E}_{X_t}[F] \;\; P_\nu - a.s., \qquad (35.13)$$

where

$$\left(\mathbb{E}_{X_t}[F]\right)(\omega) := \mathbb{E}_x[F]|_{x = X_t(\omega)} = \int_\Omega F(\omega') P_{X_t(\omega)}(d\omega') \;\; \forall\, \omega \in \Omega.$$

Proof. Let F be a bounded cylinder function of the form, $F = f\left(X_{t_0}, \dots, X_{t_n}\right)$, and G be a bounded cylinder function of the form, $G = g\left(X_{s_0}, \dots, X_{s_m}\right)$ where

$$0 = s_0 < s_1 < \dots < s_m = t.$$

Then

$$F \circ \theta_t = f\left(X_{t+t_0}, \dots, X_{t+t_n}\right)$$

and

$$\mathbb{E}_\nu\left[F \circ \theta_t \cdot G\right] = \sum_{x \in S^{m+1}, y \in S^n} g(x) f(x_m, y) q(x) q'(x_m; y) \qquad (35.14)$$

where $x \in S^{m+1}$, $y \in S^n$,

$$q(x) = \nu(x_0) p_{s_1 - s_0}(x_0, x_1) \cdot \dots \cdot p_{s_m - s_{m-1}}(x_{m-1}, x_m),$$

and

$$q'(x_m; y) = p_{t_1 - t_0}(x_m, y_1) \cdot \dots \cdot p_{t_n - t_{n-1}}(y_{n-1}, y_n). \qquad (35.15)$$

According to Eq. (35.11

$$\sum_{y \in S^n} f(x_m, y) q'(x_m; y) = \mathbb{E}_{x_m}[F]$$

we may rewrite Eq. (35.14) as,

$$\begin{aligned} \mathbb{E}_\nu\left[F \circ \theta_t \cdot G\right] &= \sum_{x \in S^{m+1}} g(x) \mathbb{E}_{x_m}[F] q(x) \\ &= \mathbb{E}_\nu\left[g\left(X_{s_0}, \dots, X_{s_m}\right) \mathbb{E}_{X_{s_m}}[F]\right] = \mathbb{E}_\nu\left[G \cdot \mathbb{E}_{X_t}[F]\right]. \qquad (35.16) \end{aligned}$$

An application of the multiplicative systems Theorem 8.2 (or Theorem 8.16) shows Eq. (35.16) holds for all $F \in b\mathcal{B}$ and $G \in b\mathcal{B}_t$ which is sufficient to prove Eq. (35.13). ∎

Here is another seemingly mild improvement on the previous theorem.

Theorem 35.17 (Markov Property II). *If $F \in b\mathcal{B}$ and ν is a probability measure on $\mathbb{R}$, then*

$$\mathbb{E}_\nu\left[F \circ \theta_t | \mathcal{B}_{t+}\right] = \mathbb{E}_{X_t}[F] \quad P_\nu \text{ – a.s.}$$

and in particular there is a version of $\mathbb{E}_\nu\left[F \circ \theta_t | \mathcal{B}_{t+}\right]$ which is $\sigma(B_t) \subset \mathcal{B}_t$ – measurable.

Proof. Again let F be a bounded cylinder function of the form, $F = f\left(X_{t_0}, \dots, X_{t_n}\right)$ with $f : S^{n+1} \to \mathbb{R}$ being a bounded function. Recall that

$$\mathbb{E}_x F = \sum_{y \in S^n} f(x, y_1, \dots, y_n) q'(x; y)$$

where

$$q'(x; y) = p_{t_1 - t_0}(x, y_1) \cdot \dots \cdot p_{t_n - t_{n-1}}(y_{n-1}, y_n).$$

Also observe that

$$\tau \to F \circ \theta_\tau(\omega) = f\left(\omega(t_0 + \tau), \dots, \omega(t_n + \tau)\right)$$

is right continuous in $\tau \geq 0$ for all $\omega \in \Omega$. Therefore if $G \in b\mathcal{B}_{t+}$, then by the DCT,

$$\mathbb{E}_\nu [F \circ \theta_t \cdot G] = \lim_{\tau \downarrow t} \mathbb{E}_\nu [F \circ \theta_\tau \cdot G].$$

By Theorem 35.16, for $\tau > t$,

$$\mathbb{E}_\nu [F \circ \theta_\tau \cdot G] = \mathbb{E}_\nu [\mathbb{E}_\nu [F \circ \theta_\tau | \mathcal{B}_\tau] \cdot G] = \mathbb{E}_\nu [\mathbb{E}_{X_\tau} [F] \cdot G].$$

Therefore, by another application of the DCT,

$$\mathbb{E}_\nu [F \circ \theta_t \cdot G] = \lim_{\tau \downarrow t} \mathbb{E}_\nu [F \circ \theta_\tau \cdot G] = \lim_{\tau \downarrow t} \mathbb{E}_\nu [\mathbb{E}_{X_\tau} [F] \cdot G] = \mathbb{E}_\nu [\mathbb{E}_{X_t} [F] \cdot G].$$

It now follows by an application of the multiplicative systems Theorem 8.2 (or Theorem 8.16) that

$$\mathbb{E}_\nu [F \circ \theta_t \cdot G] = \mathbb{E}_\nu [\mathbb{E}_{X_t} [F] \cdot G]$$

for all $F \in b\mathcal{B}$ which completes the proof. ■

Lemma 35.18 (Optional time approximation lemma). *Let τ be a $\{\mathcal{B}_t\}_{t \geq 0}$ – optional time (i.e. $\{\tau < t\} \in \mathcal{B}_t$ for all t, see Definition 27.8), and for $n \in \mathbb{N}$, let $\tau_n : \Omega \to [0, \infty]$ be defined by*

$$\tau_n := \infty 1_{\tau = \infty} + \sum_{k=1}^{\infty} \frac{k}{2^n} 1_{\frac{k-1}{2^n} \leq \tau < \frac{k}{2^n}}. \tag{35.17}$$

If $A \in \mathcal{B}_\tau^+$, then $A \cap \{\tau_n = k 2^{-n}\} \in \mathcal{B}_{k2^{-n}}$ for all $k \in \mathbb{N}$, τ_n is a stopping time, and $\tau_n \downarrow \tau$ as $n \to \infty$.

Proof. For $A \in \mathcal{B}_\tau^+$

$$\begin{aligned} A \cap \left\{\tau_n = k2^{-n}\right\} &= A \cap \left\{(k-1) 2^{-n} \leq \tau < k2^{-n}\right\} \\ &= \left[A \cap \left\{\tau < k2^{-n}\right\}\right] \setminus \left\{\tau < (k-1) 2^{-n}\right\} \in \mathcal{B}_{k2^{-n}}. \end{aligned}$$

Taking $A = \Omega$ in this equation shows $\{\tau_n = k2^{-n}\} \in \mathcal{B}_{k2^{-n}}$ for all $k \in \mathbb{N}$ and therefore

$$\left\{\tau_n \leq k2^{-n}\right\} = \cup_{l=1}^k \left\{\tau_n = l2^{-n}\right\} \in \mathcal{B}_{k2^{-n}}$$

for $k \in \mathbb{N}$. From this it follows that $\{\tau_n \leq t\} \in \mathcal{B}_t$ for all $t \leq \infty$ and hence τ_n is a stopping time. The fact that $\tau_n \downarrow \tau$ as $n \to \infty$ should be clear. ■

See Section 27.2 for more on optional and stopping times and in particular see Notation 27.1.

Theorem 35.19 (Strong Markov Property). *Suppose that $\tau : \Omega \to [0, \infty]$ is an optional time (i.e. $\{\tau < t\} \in \mathcal{B}_t$ for all t, see Definition 27.8), and $F : \Omega \to \mathbb{R}$ is a bounded measurable function. Then for any probability measure, ν, on $(S, \mathcal{B}_S)$ we have*

$$\mathbb{E}_\nu \left[F \circ \theta_\tau | \mathcal{B}_\tau^+\right] = \mathbb{E}_{X_\tau} F, \ P_\nu \text{ – a.s. on } \{\tau < \infty\}. \tag{35.18}$$

Proof. Suppose that $A \in \mathcal{B}_\tau^+$, $F := f(X_{t_0}, \ldots, X_{t_n})$ where $f : S^{n+1} \to \mathbb{R}$ is a bounded function, and $\{\tau_n\}_{n=1}^\infty$ are the discrete stopping times defined in Eq. (35.17) of Lemma 35.18. Then for each n, we may use Theorem 35.17, to conclude,

$$\begin{aligned} &\mathbb{E}_\nu [F \circ \theta_{\tau_n} \cdot 1_{\tau < \infty} \cdot 1_A] \\ &\quad = \mathbb{E}_\nu [F \circ \theta_{\tau_n} \cdot 1_{\tau_n < \infty} \cdot 1_A] \\ &\quad = \mathbb{E}_\nu \left[\sum_{k=1}^\infty F \circ \theta_{k2^{-n}} \cdot 1_{A \cap \{\tau_n = k2^{-n}\}}\right] \\ &\quad = \sum_{k=1}^\infty \mathbb{E}_\nu \left[F \circ \theta_{k2^{-n}} \cdot 1_{A \cap \{\tau_n = k2^{-n}\}}\right] \\ &\quad = \sum_{k=1}^\infty \mathbb{E}_\nu \left[\left[\mathbb{E}_{X_{k2^{-n}}} F\right] \cdot 1_{A \cap \{\tau_n = k2^{-n}\}}\right] \\ &\quad = \mathbb{E}_\nu \left[\sum_{k=1}^\infty 1_{\tau_n = k2^{-n}} \left[\mathbb{E}_{X_{k2^{-n}}} F\right] \cdot 1_A\right] \\ &\quad = \mathbb{E}_\nu \left[1_{\tau_n < \infty} \left[\mathbb{E}_{X_{\tau_n}} F\right] \cdot 1_A\right] = \mathbb{E}_\nu \left[1_{\tau < \infty} \left[\mathbb{E}_{X_{\tau_n}} F\right] \cdot 1_A\right]. \end{aligned} \tag{35.19}$$

Observe that $F \circ \theta_{\tau_n} \to F$ and $\mathbb{E}_{X_\tau} F \to \mathbb{E}_{X_{\tau_n}} F$ boundedly on $\{\tau < \infty\}$ because $t \to \mathbb{E}_{X_{t(\omega)}} F$ is right continuous and bounded. Using these observations and the DCT, we may pass to the limit $(n \to \infty)$ in Eq. (35.19) to arrive at the identity;

$$\mathbb{E}_\nu [F \circ \theta_\tau \cdot 1_{\tau < \infty} \cdot 1_A] = \mathbb{E}_\nu [1_{\tau < \infty} [\mathbb{E}_{X\tau} F] \cdot 1_A]. \tag{35.20}$$

By the multiplicative systems theorem, Eq. (35.20) is seen to be valid for all bounded measurable functions, $F : \Omega \to \mathbb{R}$. Since $A \in \mathcal{B}_\tau^+$ was arbitrary, Eq. (35.18) is proved. ■

Corollary 35.20. *If $\{X(t)\}_{t \geq 0}$ is a Markov chain, τ is an optional time (i.e. $\{\tau < t\} \in \mathcal{B}_t$ for all t, see Definition 27.8), and $j \in S$, then, conditioned on $\{\tau < \infty$ and $X_\tau = j\}$, $\mathcal{B}_\tau^+$ and $\{X(t + \tau) : t \geq 0\}$ are independent and $\{X(t + \tau) : t \geq 0\}$ has the same distribution as $\{X(t)\}_{t \geq 0}$ under P_j.*

Proof. Let $g : \Omega \to \mathbb{R}$ be a bounded $\mathcal{B}_\tau$ – measurable function and $f : \Omega \to \mathbb{R}$ be a bounded $\mathcal{B}$ – measurable function. Then

$$\begin{aligned}
\mathbb{E}_\nu\left[g\cdot f\left(X_{\tau+\cdot}\right):\tau<\infty\ \&\ X_\tau=j\right] &= \mathbb{E}_\nu\left[g\cdot f\circ\theta_\tau:\tau<\infty\ \&\ X_\tau=j\right]\\
&= \mathbb{E}_\nu\left(\mathbb{E}_\nu\left[f\circ\theta_\tau\cdot g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}|\mathcal{B}_\tau\right]\right)\\
&= \mathbb{E}_\nu\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\cdot\mathbb{E}_\nu\left[f\circ\theta_\tau|\mathcal{B}_\tau\right]\right)\\
&= \mathbb{E}_\nu\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\cdot\mathbb{E}_{X_\tau}f\right)\\
&= \mathbb{E}_\nu\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\cdot\mathbb{E}_j f\right)\\
&= \mathbb{E}_\nu\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\right)\cdot\mathbb{E}_j f
\end{aligned}$$

and therefore,

$$\begin{aligned}
\mathbb{E}_\nu\left[g\cdot f\left(X_{\tau+\cdot}\right)|\tau<\infty\ \&\ X_\tau=j\right] &= \frac{\mathbb{E}_\nu\left(g\cdot 1_{\tau<\infty}1_{X_{\tau=j}}\right)}{\mathbb{E}_\nu\left[\tau<\infty\ \&\ X_\tau=j\right]}\cdot\mathbb{E}_j f\\
&= \mathbb{E}_\nu\left(g|\tau<\infty\ \&\ X_\tau=j\right)\cdot\mathbb{E}_j f.
\end{aligned}$$

Taking $g=1$ in this equation then shows,

$$\mathbb{E}_\nu\left[f\left(X_{\tau+\cdot}\right)|\tau<\infty\ \&\ X_\tau=j\right]=\mathbb{E}_j f$$

which combined with the previously displayed equation completes the proof of the corollary. ■

36

Continuous Time M.C. Finite State Space Theory

For simplicity we will begin our study in the case where the state space is finite, say $S = \{1, 2, 3, \ldots, N\}$ for some $N < \infty$. It will be convenient to define,

$$\mathbf{1} := \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

be the column vector with all entries being 1.

Definition 36.1. *An $N \times N$ matrix function $P(t)$ for $t \geq 0$ is* ***Markov semi-group*** *if*

1. *$P(t)$ is Markov matrix for all $t \geq 0$, i.e. $P_{ij}(t) \geq 0$ for all i, j and*

$$\sum_{j \in S} P_{ij}(t) = 1 \text{ for all } i \in S. \tag{36.1}$$

The condition in Eq. (36.1) may be written in matrix notation as,

$$P(t)\mathbf{1} = \mathbf{1} \text{ for all } t \geq 0. \tag{36.2}$$

2. *$P(0) = I_{N \times N}$,*
3. *$P(t+s) = P(t)P(s)$ for all $s, t \geq 0$* ***(Chapman - Kolmogorov)****,*
4. *$\lim_{t \downarrow 0} P(t) = I$, i.e. P is continuous at $t = 0$.*

Definition 36.2. *An $N \times N$ matrix, Q, is an* ***infinitesimal generator*** *if $Q_{ij} \geq 0$ for all $i \neq j$ and*

$$\sum_{j \in S} Q_{ij} = 0 \text{ for all } i \in S. \tag{36.3}$$

The condition in Eq. (36.3) may be written in matrix notation as,

$$Q\mathbf{1} = 0. \tag{36.4}$$

36.1 Matrix Exponentials

In this section we are going to make use of the following facts from the theory of linear ordinary differential equations.

Theorem 36.3. *Let A and B be any $N \times N$ (real) matrices. Then there exists a unique $N \times N$ matrix function $P(t)$ solving the differential equation,*

$$\dot{P}(t) = AP(t) \text{ with } P(0) = B \tag{36.5}$$

which is in fact given by

$$P(t) = e^{tA}B \tag{36.6}$$

where

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \ldots \tag{36.7}$$

The matrix function e^{tA} may be characterized as the unique solution Eq. (36.5) with $B = I$ and it is also the unique solution to

$$\dot{P}(t) = AP(t) \text{ with } P(0) = I.$$

Moreover, e^{tA} satisfies the semi-group property (Chapman Kolmogorov equation),

$$e^{(t+s)A} = e^{tA}e^{sA} \text{ for all } s, t \geq 0. \tag{36.8}$$

Proof. We will only prove Eq. (36.8) here assuming the first part of the theorem. Fix $s > 0$ and let $R(t) := e^{(t+s)A}$, then

$$\dot{R}(t) = Ae^{(t+s)A} = AR(t) \text{ with } R(0) = P(s).$$

Therefore by the first part of the theorem

$$e^{(t+s)A} = R(t) = e^{tA}R(0) = e^{tA}e^{sA}.$$

■

Example 36.4 (Thanks to Mike Gao!). If $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $A^n = 0$ for $n \geq 2$, so that

$$e^{tA} = I + tA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

Similarly if $B = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$, then $B^n = 0$ for $n \geq 2$ and

$$e^{tB} = I + tB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -t & 1 \end{bmatrix}.$$

Now let $C = A + B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. In this case $C^2 = -I$, $C^3 = -C$, $C^4 = I$, $C^5 = C$ etc., so that

$$C^{2n} = (-1)^n I \text{ and } C^{2n+1} = (-1)^n C.$$

Therefore,

$$\begin{aligned} e^{tC} &= \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} C^{2n} + \sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)!} C^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} (-1)^n I + \sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)!} (-1)^n C \\ &= \cos(t) I + \sin(t) C = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \end{aligned}$$

which is the matrix representing rotation in the plan by t degrees.

Here is another way to compute e^{tC} in this example. Since $C^2 = -I$, we find

$$\begin{aligned} \frac{d^2}{dt^2} e^{tC} &= C^2 e^{tC} = -e^{tC} \text{ with} \\ e^{0C} &= I \text{ and } \frac{d}{dt} e^{tC}|_{t=0} = C. \end{aligned}$$

It is now easy to verify the solution to this second order equation is given by,

$$e^{tC} = \cos t \cdot I + \sin t \cdot C$$

which agrees with our previous answer.

Remark 36.5. **Warning:** if A and B are two $N \times N$ matrices it is not generally true that

$$e^{(A+B)} = e^A e^B \tag{36.9}$$

as can be seen from Example 36.4.

However we have the following lemma.

Lemma 36.6. *If A and B commute, i.e. $AB = BA$, then Eq. (36.9) holds. In particular, taking $B = -A$, shows that $e^{-A} = \left[e^A\right]^{-1}$.*

Proof. First proof. Simply verify Eq. (36.9) using explicit manipulations with the infinite series expansion. The point is, because A and B compute, we may use the binomial formula to find;

$$(A+B)^n = \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k}.$$

(Notice that if A and B do not compute we will have

$$(A+B) = A^2 + AB + BA + B^2 \neq A^2 + 2AB + B^2.)$$

Therefore,

$$\begin{aligned} e^{(A+B)} &= \sum_{n=0}^{\infty} \frac{1}{n!} (A+B)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} \\ &= \sum_{0 \leq k \leq n < \infty} \frac{1}{k!} \frac{1}{(n-k)!} A^k B^{n-k} \qquad (\text{let } n-k=l) \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{k!} \frac{1}{l!} A^k B^l = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \cdot \sum_{l=0}^{\infty} \frac{1}{l!} B^l = e^A e^B. \end{aligned}$$

Second proof. Here is another proof which uses the ODE interpretation of e^{tA}. We will carry it out in a number of steps.

1. By Theorem 36.3 and the product rule
$$\frac{d}{dt} e^{-tA} B e^{tA} = e^{-tA} (-A) B e^{tA} + e^{-tA} B A e^{tA} = e^{-tA} (BA - AB) e^{tA} = 0$$
since A and B commute. This shows that $e^{-tA} B e^{tA} = B$ for all $t \in \mathbb{R}$.
2. Taking $B = I$ in 1. then shows $e^{-tA} e^{tA} = I$ for all t ,i.e. $e^{-tA} = \left[e^{tA}\right]^{-1}$. Hence we now conclude from Item 1. that $e^{-tA} B = B e^{-tA}$ for all t.
3. Using Theorem 36.3, Item 2., and the product rule implies
$$\begin{aligned} &\frac{d}{dt} \left[e^{-tB} e^{-tA} e^{t(A+B)}\right] \\ &= e^{-tB} (-B) e^{-tA} e^{t(A+B)} + e^{-tB} e^{-tA} (-A) e^{t(A+B)} \\ &\quad + e^{-tB} e^{-tA} (A+B) e^{t(A+B)} \\ &= e^{-tB} e^{-tA} (-B) e^{t(A+B)} + e^{-tB} e^{-tA} (-A) e^{t(A+B)} \\ &\quad + e^{-tB} e^{-tA} (A+B) e^{t(A+B)} = 0. \end{aligned}$$

Therefore,

$$e^{-tB}e^{-tA}e^{t(A+B)} = e^{-tB}e^{-tA}e^{t(A+B)}|_{t=0} = I \text{ for all } t,$$

and hence taking $t = 1$, shows

$$e^{-B}e^{-A}e^{(A+B)} = I. \tag{36.10}$$

Multiplying Eq. (36.10) on the left by $e^A e^B$ gives Eq. (36.9). ■

The next two results gives a practical method for computing e^{tQ} in many situations.

Proposition 36.7. *If Λ is a diagonal matrix,*

$$\Lambda := \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

then

$$e^{t\Lambda} = \begin{bmatrix} e^{t\lambda_1} & & & \\ & e^{t\lambda_2} & & \\ & & \ddots & \\ & & & e^{t\lambda_n} \end{bmatrix}.$$

Proof. One easily shows that

$$\Lambda^n := \begin{bmatrix} \lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_m^n \end{bmatrix}$$

for all n and therefore,

$$\begin{aligned} e^{t\Lambda} = \sum_{n=0}^{\infty} \frac{t^n}{n!}\Lambda^n &= \begin{bmatrix} \sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_1^n & & & \\ & \sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_2^n & & \\ & & \ddots & \\ & & & \sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_m^n \end{bmatrix} \\ &= \begin{bmatrix} e^{t\lambda_1} & & & \\ & e^{t\lambda_2} & & \\ & & \ddots & \\ & & & e^{t\lambda_n} \end{bmatrix}. \end{aligned}$$

■

Theorem 36.8. *Suppose that Q is a diagonalizable matrix, i.e. there exists an invertible matrix, S, such that $S^{-1}QS = \Lambda$ with Λ being a diagonal matrix. In this case we have,*

$$e^{tQ} = Se^{t\Lambda}S^{-1} \tag{36.11}$$

Proof. We begin by observing that

$$\begin{aligned} \left(S^{-1}QS\right)^2 &= S^{-1}QSS^{-1}QS = S^{-1}Q^2S, \\ \left(S^{-1}QS\right)^3 &= S^{-1}Q^2SS^{-1}QS = S^{-1}Q^3S \\ &\vdots \\ \left(S^{-1}QS\right)^n &= S^{-1}Q^nS \quad \text{for all } n \geq 0. \end{aligned}$$

Therefore we find that

$$\begin{aligned} S^{-1}e^{tQ}S &= S^{-1}IS + \sum_{n=0}^{\infty} \frac{t^n}{n!}S^{-1}Q^nS \\ &= I + \sum_{n=0}^{\infty} \frac{t^n}{n!}\left(S^{-1}QS\right)^n \\ &= I + \sum_{n=0}^{\infty} \frac{t^n}{n!}\Lambda^n = e^{t\Lambda}. \end{aligned}$$

Solving this equation for e^{tQ} gives the desired result. ■

36.2 Characterizing Markov Semi-Groups

We now come to the main theorem of this chapter.

Theorem 36.9. *The collection of Markov semi-groups is in one to one correspondence with the collection of infinitesimal generators. More precisely we have;*

1. *$P(t) = e^{tQ}$ is Markov semi-group iff Q is an infinitesimal generator.*
2. *If $P(t)$ is a Markov semi-group, then $Q := \frac{d}{dt}|_{0+}P(t)$ exists, Q is an infinitesimal generator, and $P(t) = e^{tQ}$.*

Proof. The proof is completed by Propositions 36.10 – 36.13 below. (You might look at Example 36.4 to see what goes wrong if Q does not satisfy the properties of a Markov generator.) ■

We are now going to prove a number of results which in total will complete the proof of Theorem 36.9. The first result is technical and you may safely skip its proof.

Proposition 36.10 (Techinical proposition). *Every Markov semi-group, $\{P(t)\}_{t\geq 0}$ is continuously differentiable.*

Proof. First we want to show that $P(t)$ is continuous. For $t,h\geq 0$, we have

$$P(t+h)-P(t)=P(t)P(h)-P(t)=P(t)(P(h)-I)\to 0 \text{ as } h\downarrow 0.$$

Similarly if $t>0$ and $0\leq h<t$, we have

$$\begin{aligned}P(t)-P(t-h)&=P(t-h+h)-P(t-h)=P(t-h)P(h)-P(t-h)\\&=P(t-h)[P(h)-I]\to 0 \text{ as } h\downarrow 0\end{aligned}$$

where we use the fact that $P(t-h)$ has entries all bounded by 1 and therefore

$$\begin{aligned}\left|(P(t-h)[P(h)-I])_{ij}\right|&\leq\sum_k P_{ik}(t-h)\left|(P(h)-I)_{kj}\right|\\&\leq\sum_k\left|(P(h)-I)_{kj}\right|\to 0 \text{ as } h\downarrow 0.\end{aligned}$$

Thus we have shown that $P(t)$ is continuous.

To prove the differentiability of $P(t)$ we use a trick due to Gärding. Choose $\varepsilon>0$ such that

$$\Pi:=\frac{1}{\varepsilon}\int_0^\varepsilon P(s)\,ds$$

is invertible. To see this is possible, observe that by the continuity of P, $\frac{1}{\varepsilon}\int_0^\varepsilon P(s)\,ds\to I$ as $\varepsilon\downarrow 0$. Therefore, by the continuity of the determinant function,

$$\det\left(\frac{1}{\varepsilon}\int_0^\varepsilon P(s)\,ds\right)\to\det(I)=1 \text{ as } \varepsilon\downarrow 0.$$

With this definition of Π, we have

$$P(t)\Pi=\frac{1}{\varepsilon}\int_0^\varepsilon P(t)P(s)\,ds=\frac{1}{\varepsilon}\int_0^\varepsilon P(t+s)\,ds=\frac{1}{\varepsilon}\int_t^{t+\varepsilon}P(s)\,ds.$$

So by the fundamental theorem of calculus, $P(t)\Pi$ is differentiable and

$$\frac{d}{dt}[P(t)\Pi]=\frac{1}{\varepsilon}(P(t+\varepsilon)-P(t)).$$

As Π is invertible, we may conclude that $P(t)$ is differentiable and that

$$\dot{P}(t):=\frac{1}{\varepsilon}(P(t+\varepsilon)-P(t))\Pi^{-1}.$$

Since the right hand side of this equation is continuous in t it follows that $\dot{P}(t)$ is continuous as well. ■

Proposition 36.11. *If $\{P(t)\}_{t\geq 0}$ is a Markov semi-group and $Q:=\frac{d}{dt}|_{0+}P(t)$, then*

1. *$P(t)$ satisfies $P(0)=I$ and both,*
$$\dot{P}(t)=P(t)Q \quad \textit{(Kolmogorov's forward Eq.)}$$
and
$$\dot{P}(t)=QP(t) \quad \textit{(Kolmogorov's backwards Eq.)}$$
hold.
2. *$P(t)=e^{tQ}$.*
3. *Q is an infinitesimal generator.*

Proof. 1.-2. We may compute $\dot{P}(t)$ using

$$\dot{P}(t)=\frac{d}{ds}|_0 P(t+s).$$

We then may write $P(t+s)$ as $P(t)P(s)$ or as $P(s)P(t)$ and hence

$$\begin{aligned}\dot{P}(t)&=\frac{d}{ds}|_0[P(t)P(s)]=P(t)Q \text{ and}\\\dot{P}(t)&=\frac{d}{ds}|_0[P(s)P(t)]=QP(t).\end{aligned}$$

This proves Item 1. and Item 2. now follows from Theorem 36.3.

3. Since $P(t)$ is continuously differentiable, $P(t)=I+tQ+O(t^2)$, and so for $i\neq j$,

$$0\leq P_{ij}(t)=\delta_{ij}+tQ_{ij}+O(t^2)=tQ_{ij}+O(t^2).$$

Dividing this inequality by t and then letting $t\downarrow 0$ shows $Q_{ij}\geq 0$. Differentiating the Eq. (36.2), $P(t)\mathbf{1}=\mathbf{1}$, at $t=0_+$ to show $Q\mathbf{1}=0$. ■

Proposition 36.12. *Let Q be any matrix such that $Q_{ij}\geq 0$ for all $i\neq j$. Then $\left(e^{tQ}\right)_{ij}\geq 0$ for all $t\geq 0$ and $i,j\in S$.*

Proof. Choose $\lambda\in\mathbb{R}$ such that $\lambda\geq -Q_{ii}$ for all $i\in S$. Then $\lambda I+Q$ has all non-negative entries and therefore $e^{t(\lambda I+Q)}$ has non-negative entries for all $t\geq 0$. (Think about the power series expansion for $e^{t(\lambda I+Q)}$.) By Lemma 36.6 we know that $e^{t(\lambda I+Q)}=e^{t\lambda I}e^{tQ}$ and since $e^{t\lambda I}=e^{t\lambda}I$ (you verify), we have[1]

$$e^{t(\lambda I+Q)}=e^{t\lambda}e^{tQ}.$$

Therefore, $e^{tQ}=e^{-t\lambda}e^{t(\lambda I+Q)}$ again has all non-negative entries and the proof is complete. ■

[1] Actually if you do not want to use Lemma 36.6, you may check that $e^{t(\lambda I+Q)}=e^{t\lambda}e^{tQ}$ by simply showing both sides of this equation satisfy the same ordinary differential equation.

Proposition 36.13. *Suppose that Q is any matrix such that $\sum_{j\in S} Q_{ij} = 0$ for all $i \in S$, i.e. $Q\mathbf{1} = 0$. Then $e^{tQ}\mathbf{1} = \mathbf{1}$.*

Proof. Since
$$\frac{d}{dt}e^{tQ}\mathbf{1} = e^{tQ}Q\mathbf{1} = 0,$$
it follows that $e^{tQ}\mathbf{1} = e^{tQ}\mathbf{1}|_{t=0} = \mathbf{1}$. ∎

Lemma 36.14 (ODE Lemma). *If $h(t)$ is a given function and $\lambda \in \mathbb{R}$, then the solution to the differential equation,*
$$\dot{\pi}(t) = \lambda\pi(t) + h(t) \tag{36.12}$$
is
$$\pi(t) = e^{\lambda t}\left(\pi(0) + \int_0^t e^{-\lambda s}h(s)\,ds\right) \tag{36.13}$$
$$= e^{\lambda t}\pi(0) + \int_0^t e^{\lambda(t-s)}h(s)\,ds. \tag{36.14}$$

Proof. If $\pi(t)$ satisfies Eq. (36.12), then
$$\frac{d}{dt}\left(e^{-\lambda t}\pi(t)\right) = e^{-\lambda t}\left(-\lambda\pi(t) + \dot{\pi}(t)\right) = e^{-\lambda t}h(t).$$
Integrating this equation implies,
$$e^{-\lambda t}\pi(t) - \pi(0) = \int_0^t e^{-\lambda s}h(s)\,ds.$$
Solving this equation for $\pi(t)$ gives
$$\pi(t) = e^{\lambda t}\pi(0) + e^{\lambda t}\int_0^t e^{-\lambda s}h(s)\,ds \tag{36.15}$$
which is the same as Eq. (36.13). A direct check shows that $\pi(t)$ so defined solves Eq. (36.12). Indeed using Eq. (36.15) and the fundamental theorem of calculus shows,
$$\dot{\pi}(t) = \lambda e^{\lambda t}\pi(0) + \lambda e^{\lambda t}\int_0^t e^{-\lambda s}h(s)\,ds + e^{\lambda t}e^{-\lambda t}h(t)$$
$$= \lambda\pi(t) + h(t).$$
∎

Corollary 36.15. *Suppose $\lambda \in \mathbb{R}$ and $\pi(t)$ is a function which satisfies, $\dot{\pi}(t) \geq \lambda\pi(t)$, then*
$$\pi(t) \geq e^{\lambda t}\pi(0) \text{ for all } t \geq 0. \tag{36.16}$$
In particular if $\pi(0) \geq 0$ then $\pi(t) \geq 0$ for all t. In particular if Q is a Markov generator and $P(t) = e^{tQ}$, then
$$P_{ii}(t) \geq e^{-q_i t} \text{ for all } t > 0$$
where $q_i := -Q_{ii}$. (If we put all of the sand at site i at time 0, $e^{-q_i t}$ represents the amount of sand at a later time t in the worst case scenario where no one else shovels sand back to site i.)

Proof. Let $h(t) := \dot{\pi}(t) - \lambda\pi(t) \geq 0$ and then apply Lemma 36.14 to conclude that
$$\pi(t) = e^{\lambda t}\pi(0) + \int_0^t e^{\lambda(t-s)}h(s)\,ds. \tag{36.17}$$
Since $e^{\lambda(t-s)}h(s) \geq 0$, it follows that $\int_0^t e^{\lambda(t-s)}h(s)\,ds \geq 0$ and therefore if we ignore this term in Eq. (36.17) leads to the estimate in Eq. (36.16). ∎

36.3 Examples

Example 36.16 (2×2 case I). The most general 2×2 rate matrix Q is of the form
$$Q = \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ & \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$
with rate diagram being given in Figure 36.1. We now find e^{tQ} using Theorem

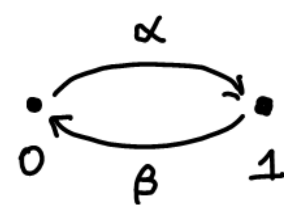


Fig. 36.1. Two state Markov chain rate diagram.

36.8. To do this we start by observing that

$$\det(Q-\lambda I) = \det\left(\begin{bmatrix} -\alpha-\lambda & \alpha \\ \beta & -\beta-\lambda \end{bmatrix}\right) = (\alpha+\lambda)(\beta+\lambda) - \alpha\beta$$
$$= \lambda^2 + \tau\lambda = \lambda(\lambda+\tau).$$

Thus the eigenvalues of Q are $\{0, -\tau\}$. The eigenvector for 0 is $\begin{bmatrix} 1 & 1 \end{bmatrix}^{\mathrm{tr}}$. Moreover,

$$Q - (-\tau) I = \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}$$

which has $\begin{bmatrix} \alpha & -\beta \end{bmatrix}^{\mathrm{tr}}$ and therefore we let

$$S = \begin{bmatrix} 1 & \alpha \\ 1 & -\beta \end{bmatrix} \text{ and } S^{-1} = \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \\ 1 & -1 \end{bmatrix}.$$

We then have

$$S^{-1}QS = \begin{bmatrix} 0 & 0 \\ 0 & -\tau \end{bmatrix} =: \Lambda.$$

So in our case

$$S^{-1}e^{tQ}S = e^{t\Lambda} = \begin{bmatrix} e^{0t} & 0 \\ 0 & e^{-\tau t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-\tau t} \end{bmatrix}.$$

Hence we must have,

$$\begin{aligned} e^{tQ} &= S\begin{bmatrix} 1 & 0 \\ 0 & e^{-\tau t} \end{bmatrix} S^{-1} \\ &= \frac{1}{\tau}\begin{bmatrix} 1 & \alpha \\ 1 & -\beta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & e^{-\tau t} \end{bmatrix}\begin{bmatrix} \beta & \alpha \\ 1 & -1 \end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix} \beta+\alpha e^{-\tau t} & \alpha-\alpha e^{-\tau t} \\ \beta-\beta e^{-\tau t} & \alpha+\beta e^{-\tau t} \end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix} \beta+\alpha e^{-\tau t} & \alpha\left(1-e^{-\tau t}\right) \\ \beta\left(1-e^{-\tau t}\right) & \alpha+\beta e^{-\tau t} \end{bmatrix}. \end{aligned}$$

Example 36.17 (2×2 case II). If $P(t) = e^{tQ}$ and $\pi(t) = \pi(0)P(t)$, then

$$\dot\pi(t) = \pi(t)Q = [\pi_0(t), \pi_1(t)]\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$
$$\begin{bmatrix} -\alpha\pi_0(t) + \beta\pi_1(t) & \alpha\pi_0(t) - \beta\pi_1(t) \end{bmatrix},$$

i.e

$$\dot\pi_0(t) = -\alpha\pi_0(t) + \beta\pi_1(t) \tag{36.18}$$
$$\dot\pi_1(t) = \alpha\pi_0(t) - \beta\pi_1(t). \tag{36.19}$$

The latter pair of equations is easy to write down using the jump diagram and the movement of sand interpretation. If we assume that $\pi_0(0) + \pi_1(0) = 1$ then we know $\pi_0(t) + \pi_1(t) = 1$ for all later times and therefore we may rewrite Eq. (36.18) as

$$\begin{aligned} \dot\pi_0(t) &= -\alpha\pi_0(t) + \beta(1-\pi_0(t)) \\ &= -\tau\pi_0(t) + \beta \end{aligned}$$

where $\tau := \alpha + \beta$. We may use Lemma 36.14 below to find

$$\begin{aligned} \pi_0(t) &= e^{-\tau t}\pi_0(0) + \int_0^t e^{-\tau(t-s)}\beta ds \\ &= e^{-\tau t}\pi_0(0) + \frac{\beta}{\tau}\left(1-e^{-\tau t}\right). \end{aligned}$$

We may also conclude that

$$\begin{aligned} \pi_1(t) = 1-\pi_0(t) &= 1 - e^{-\tau t}\pi_0(0) - \frac{\beta}{\tau}\left(1-e^{-\tau t}\right) \\ &= 1 - e^{-\tau t}(1-\pi_1(0)) - \frac{\beta}{\tau}\left(1-e^{-\tau t}\right) \\ &= e^{-\tau t}\pi_1(0) + \left(1-e^{-\tau t}\right) - \frac{\beta}{\tau}\left(1-e^{-\tau t}\right) \\ &= e^{-\tau t}\pi_1(0) + \frac{\alpha}{\tau}\left(1-e^{-\tau t}\right). \end{aligned}$$

By taking $\pi_0(0) = 1$ and $\pi_1(0) = 0$ we get the first row of $P(t)$ is equal to

$$\begin{bmatrix} e^{-\tau t}1 + \frac{\beta}{\tau}\left(1-e^{-\tau t}\right) & \frac{\alpha}{\tau}\left(1-e^{-\tau t}\right) \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} e^{-\tau t}\alpha + \beta & \alpha\left(1-e^{-\tau t}\right) \end{bmatrix}$$

and similarly the second row of $P(t)$ is found by taking $\pi_0(0) = 0$ and $\pi_1(0) = 1$ to find

$$\begin{bmatrix} \frac{\beta}{\tau}\left(1-e^{-\tau t}\right) & e^{-\tau t} + \frac{\alpha}{\tau}\left(1-e^{-\tau t}\right) \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} \beta\left(1-e^{-\tau t}\right) & \beta e^{-\tau t} + \alpha \end{bmatrix}.$$

Hence we have found

$$\begin{aligned} P(t) &= \frac{1}{\tau}\begin{bmatrix} e^{-\tau t}\alpha + \beta & \alpha\left(1-e^{-\tau t}\right) \\ \beta\left(1-e^{-\tau t}\right) & \beta e^{-\tau t} + \alpha \end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix} \left(e^{-\tau t}-1\right)\alpha + \beta + \alpha & \alpha\left(1-e^{-\tau t}\right) \\ \beta\left(1-e^{-\tau t}\right) & \beta\left(e^{-\tau t}-1\right) + \alpha + \beta \end{bmatrix} \\ &= I + \frac{1}{\tau}\left(1-e^{-\tau t}\right)\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \\ &= I + \frac{1}{\tau}\left(1-e^{-\tau t}\right)Q. \end{aligned}$$

Let us verify that this is indeed the correct solution. It is clear that $P(0) = I$,

$$\dot{P}(t) = e^{-\tau t}\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$

while on the other hand,

$$Q^2 = \begin{bmatrix} \alpha\beta + \alpha^2 & -\alpha\beta - \alpha^2 \\ -\alpha\beta - \beta^2 & \alpha\beta + \beta^2 \end{bmatrix} = \tau \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} = -\tau Q$$

and therefore,

$$P(t)\,Q = Q - \left(1 - e^{-\tau t}\right) Q = e^{-\tau t} Q$$

as desired.

We also have

$$\begin{aligned} P(s)\,P(t) &= \left(I + \frac{1}{\tau}\left(1 - e^{-\tau s}\right) Q\right)\left(I + \frac{1}{\tau}\left(1 - e^{-\tau t}\right) Q\right) \\ &= I + \frac{1}{\tau}\left(2 - e^{-\tau s} - e^{-\tau t}\right) Q + \frac{1}{\tau}\left(1 - e^{-\tau s}\right)\frac{1}{\tau}\left(1 - e^{-\tau t}\right)(-\tau)\,Q \\ &= I + \frac{1}{\tau}\left[\left(2 - e^{-\tau s} - e^{-\tau t}\right) - \left(1 - e^{-\tau s}\right)\left(1 - e^{-\tau t}\right)\right] Q \\ &= I + \frac{1}{\tau}\left[1 - e^{-\tau(s+t)}\right] Q = P(s+t) \end{aligned}$$

as it should be. Lastly let us observe that

$$\begin{aligned} \lim_{t\to\infty} P(t) &= I + \frac{1}{\tau}\lim_{t\to\infty}\left(1 - e^{-\tau t}\right)\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \\ &= I - \frac{1}{\tau}\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}. \end{aligned}$$

Moreover we have

$$\lim_{t\to\infty}\dot{P}(t) = \lim_{t\to\infty} e^{-\tau t}\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} = 0.$$

Suppose that π is any distribution, then

$$\lim_{t\to\infty}\pi P(t) = \frac{1}{\tau}\begin{bmatrix} \pi_0 & \pi_1 \end{bmatrix}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \end{bmatrix}$$

independent of π. Moreover, since

$$\begin{aligned} \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \end{bmatrix} P(s) &= \lim_{t\to\infty}\pi P(t)\,P(s) = \lim_{t\to\infty}\pi P(t+s) \\ &= \lim_{t\to\infty}\pi P(t) = \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \end{bmatrix} \end{aligned}$$

which shows that the limiting distribution is also an invariant distribution. If π is any invariant distribution for P, we must have

$$\pi = \lim_{t\to\infty}\pi P(t) = \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \end{bmatrix} = \begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix} \tag{36.20}$$

and moreover,

$$0 = \frac{d}{dt}|_0 \pi = \frac{d}{dt}|_0 \pi P(t) = \pi Q.$$

The solutions of $\pi Q = 0$ correspond to the null space of Q^{tr} which implies

$$\operatorname{Nul} Q^{\mathrm{tr}} = \operatorname{Nul}\begin{bmatrix} -\alpha & \beta \\ \alpha & -\beta \end{bmatrix} = \mathbb{R}\cdot\begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$

and hence we have again recovered $\pi = \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \end{bmatrix}$.

Example 36.18 (2×2 case III). We now compute e^{tQ} by the power series method as follows. A simple computation shows that

$$Q^2 = \begin{bmatrix} \alpha\beta + \alpha^2 & -\alpha\beta - \alpha^2 \\ -\alpha\beta - \beta^2 & \alpha\beta + \beta^2 \end{bmatrix} = \tau \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} = -\tau Q.$$

Hence it follows by induction that $Q^n = (-\tau)^{n-1} Q$ and therefore,

$$\begin{aligned} P(t) = e^{tQ} &= I + \sum_{n=1}^{\infty}\frac{t^n}{n!}(-\tau)^{n-1} Q \\ &= I - \frac{1}{\tau}\sum_{n=1}^{\infty}\frac{t^n}{n!}(-\tau)^n Q = I - \frac{1}{\tau}\left(e^{-\tau t} - 1\right) Q \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{\tau}\left(e^{-\tau t} - 1\right)\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \\ &= \begin{bmatrix} \frac{\alpha}{\tau}\left(e^{-t\tau} - 1\right) + 1 & -\frac{\alpha}{\tau}\left(e^{-t\tau} - 1\right) \\ -\frac{\beta}{\tau}\left(e^{-t\tau} - 1\right) & \frac{\beta}{\tau}\left(e^{-t\tau} - 1\right) + 1 \end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix} \alpha e^{-t\tau} + \beta & \alpha\left(1 - e^{-t\tau}\right) \\ \beta\left(1 - e^{-t\tau}\right) & \beta e^{-t\tau} + \alpha \end{bmatrix} \end{aligned}$$

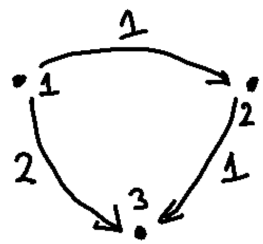
Let us again verify that this answer is correct;

$$\begin{aligned} &\dot{P}(t) = e^{-\tau t} Q \text{ while} \\ &P(t)\,Q = Q - \frac{1}{\tau}\left(e^{-\tau t} - 1\right)(-\tau)\,Q = Q + \left(e^{-\tau t} - 1\right) Q = \dot{P}(t). \end{aligned}$$

Example 36.19. Let $S = \{1, 2, 3\}$ and

$$Q = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left[\begin{array}{rrr} -3 & 1 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$

which we represent by Figure 36.19. Let $\pi = (\pi_1, \pi_2, \pi_3)$ be a given initial (at



$t = 0$) distribution (of sand say) on S and let $\pi(t) := \pi e^{tQ}$ be the distribution at time t. Then

$$\dot{\pi}(t) = \pi e^{tQ} Q = \pi(t) Q.$$

In this particular example this gives,

$$\begin{aligned} \left[\dot{\pi}_1 \; \dot{\pi}_2 \; \dot{\pi}_3\right] &= \left[\pi_1 \; \pi_2 \; \pi_3\right] \begin{bmatrix} -3 & 1 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \left[-3\pi_1 \;\; \pi_1 - \pi_2 \;\; 2\pi_1 + \pi_2\right], \end{aligned}$$

or equivalently,

$$\dot{\pi}_1 = -3\pi_1 \tag{36.21}$$

$$\dot{\pi}_2 = \pi_1 - \pi_2 \tag{36.22}$$

$$\dot{\pi}_3 = 2\pi_1 + \pi_2. \tag{36.23}$$

Notice that these equations are easy to read off from Figure 36.19. For example, the second equation represents the fact that rate of change of sand at site 2 is equal to the rate which sand is entering site 2 (in this case from 1 with rate $1\pi_1$) minus the rate at which sand is leaving site 2 (in this case $1\pi_2$ is the rate that sand is being transported to 3). Similarly, site 3 is greedy and never gives up any of its sand while happily receiving sand from site 1 at rate $2\pi_1$ and from site 2 are rate $1\pi_2$. Solving Eq. (36.21) gives,

$$\pi_1(t) = e^{-3t}\pi_1(0)$$

and therefore Eq. (36.22) becomes

$$\dot{\pi}_2 = e^{-3t}\pi_1(0) - \pi_2$$

which, by Lemma 36.14 below, has solution,

$$\begin{aligned} \pi_2(t) &= e^{-t}\pi_2(0) + e^{-t}\int_0^t e^{\tau}e^{-3\tau}\pi_1(0)\,d\tau \\ &= \frac{1}{2}\left(e^{-t} - e^{-3t}\right)\pi_1(0) + e^{-t}\pi_2(0). \end{aligned}$$

Using this back in Eq. (36.23) then shows

$$\begin{aligned} \dot{\pi}_3 &= 2e^{-3t}\pi_1(0) + \frac{1}{2}\left(e^{-t} - e^{-3t}\right)\pi_1(0) + e^{-t}\pi_2(0) \\ &= \left(\frac{1}{2}e^{-t} + \frac{3}{2}e^{-3t}\right)\pi_1(0) + e^{-t}\pi_2(0) \end{aligned}$$

which integrates to

$$\begin{aligned} \pi_3(t) &= \left(\frac{1}{2}\left[1 - e^{-t}\right] + \frac{1}{2}\left(1 - e^{-3t}\right)\right)\pi_1(0) + \left(1 - e^{-t}\right)\pi_2(0) + \pi_3(0) \\ &= \left(1 - \frac{1}{2}\left[e^{-t} + e^{-3t}\right]\right)\pi_1(0) + \left(1 - e^{-t}\right)\pi_2(0) + \pi_3(0). \end{aligned}$$

Thus we have

$$\begin{aligned} \begin{bmatrix} \pi_1(t) \\ \pi_2(t) \\ \pi_3(t) \end{bmatrix} &= \begin{bmatrix} e^{-3t}\pi_1(0) \\ \frac{1}{2}\left(e^{-t} - e^{-3t}\right)\pi_1(0) + e^{-t}\pi_2(0) \\ \left(1 - \frac{1}{2}\left[e^{-t} + e^{-3t}\right]\right)\pi_1(0) + \left(1 - e^{-t}\right)\pi_2(0) + \pi_3(0) \end{bmatrix} \\ &= \begin{bmatrix} e^{-3t} & 0 & 0 \\ \frac{1}{2}\left(e^{-t} - e^{-3t}\right) & e^{-t} & 0 \\ 1 - \frac{1}{2}\left[e^{-t} + e^{-3t}\right] & 1 - e^{-t} & 1 \end{bmatrix} \begin{bmatrix} \pi_1(0) \\ \pi_2(0) \\ \pi_3(0) \end{bmatrix}. \end{aligned}$$

From this we may conclude that

$$\begin{aligned} P(t) = e^{tQ} &= \begin{bmatrix} e^{-3t} & 0 & 0 \\ \frac{1}{2}\left(e^{-t} - e^{-3t}\right) & e^{-t} & 0 \\ 1 - \frac{1}{2}\left[e^{-t} + e^{-3t}\right] & 1 - e^{-t} & 1 \end{bmatrix}^{\mathrm{tr}} \\ &= \begin{bmatrix} e^{-3t} & \left(\frac{1}{2}e^{-t} - \frac{1}{2}e^{-3t}\right) & \left(1 - \frac{1}{2}e^{-t} - \frac{1}{2}e^{-3t}\right) \\ 0 & e^{-t} & -e^{-t} + 1 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

37

Jump and Hold Description

We would now like to make a direct connection between Q and the Markov process X_t. To this end, let τ denote the first time the process makes a jump between two states. In this section we are going to write x and y for typical element in the state space, S. (BRUCE: there is a notational problem here, namely below we are writing $\tau = S$ and also using S for the state space.) In this section we will describe the theory a bit informally saving the formal proofs for the next Section 37.3.

Theorem 37.1. *Let $Q_x := -Q_{x,x} \geq 0$. Then $P_x(S > t) = e^{-Q_x t}$, which shows that relative P_x, S is exponentially distributed with parameter Q_x. Moreover, X_S is independent of S and*

$$P_x(X_S = y) = Q_{x,y}/Q_x.$$

Proof. We give a number of proofs for this theorem. Perhaps the tightest proof appears in Section 37.3 below.

First proof. For the first assertion we let

$$A_n := \left\{ X\left(\frac{i}{2^n}t\right) = x \text{ for } i = 1, 2, \ldots, 2^n - 1, 2^n \right\}.$$

Then

$$A_n \downarrow \{X(s) = x \text{ for } s \leq t\} = \{S > t\}$$

and therefore, $P_x(A_n) \downarrow P_x(S > t)$. Since,

$$\begin{aligned} P(A_n) &= [P_{x,x}(t/2^n)]^{2^n} = \left[1 - \frac{tQ_x}{2^n} + O\left((1/2^n)^2\right)\right]^{2^n} \\ &\to e^{-tQ_x} \text{ as } n \to \infty, \end{aligned}$$

we have shown $P_x(S > t) = e^{-tQ_x}$.

For the second assertion. Let T be the time between the second and first jump of the process. Then by the strong Markov property (Corollary 35.20), for any $t \geq 0$ and $\Delta > 0$ small, we have,

$$\begin{aligned} P_x(t < S \leq t+\Delta,\ T \leq \Delta) &= \sum_{y \in S} P_x(t < S \leq t+\Delta,\ T \leq \Delta,\ X_S = y) \\ &= \sum_{y \in S} P_x(t < S \leq t+\Delta,\ S \circ \theta_S \leq \Delta,\ X_S = y) \\ &= \sum_{y \in S} P_x(t < S \leq t+\Delta,\ X_S = y) \cdot P_y(S \leq \Delta) \\ &= \sum_{y \in S} P_x(t < S \leq t+\Delta,\ X_S = y) \cdot \left(1 - e^{-Q_y \Delta}\right) \\ &\leq \max_{y \in S}\left(1 - e^{-Q_y \Delta}\right) \sum_{y \in S} P_x(t < S \leq t+\Delta,\ X_S = y) \\ &= \max_{y \in S}\left(1 - e^{-Q_y \Delta}\right) P_x(t < S \leq t+\Delta) \\ &= \max_{y \in S}\left(1 - e^{-Q_y \Delta}\right) \int_t^{t+\Delta} Q_x e^{-Q_x \tau} d\tau = O\left(\Delta^2\right). \end{aligned}$$

(Here we have used that the rates, $\{Q_y\}_{y \in S}$ are bounded which is certainly the case when $\#(S) < \infty$.) Therefore the probability of two jumps occurring in the time interval, $[t, t+\Delta]$, may be ignored and we have,

$$\begin{aligned} P_x(X_S = y, t < S \leq t+\Delta) &= P_x(X_{t+\Delta} = y, S > t) + o(\Delta) \\ &= P_x(X_{t+\Delta} = y, X_t = x, S > t) + o(\Delta) \\ &= \lim_{n \to \infty}\left[1 - \frac{tQ_x}{n} + O(n^{-2})\right]^n P_{x,y}(\Delta) + o(\Delta) \\ &= e^{-tQ_x} P_{x,y}(\Delta) + o(\Delta). \end{aligned}$$

Also

$$P_x(t < S \leq t+\Delta) = \int_t^{t+\Delta} Q_x e^{-Q_x s} ds = e^{-Q_x t} - e^{-Q_x(t+\Delta)} = Q_x e^{-Q_x t}\Delta + o(\Delta).$$

Therefore,

$$\begin{aligned} P_x(X_S = y | S = t) &= \lim_{\Delta \downarrow 0} \frac{P_x(X_S = y, t < S \leq t+\Delta)}{P_x(t < S \leq t+\Delta)} \\ &= \lim_{\Delta \downarrow 0} \frac{e^{-tQ_x} P_{x,y}(\Delta) + o(\Delta)}{Q_x e^{-Q_x t}\Delta + o(\Delta)} = \frac{1}{Q_x} \lim_{\Delta \downarrow 0} \frac{P_{x,y}(\Delta)}{\Delta} = Q_{x,y}/Q_x. \end{aligned}$$

This shows that S and X_S are independent and that $P_x(X_S = y) = Q_{x,y}/Q_x$.

Second Proof. For $t > 0$ and $\delta > 0$, we have that

$$\begin{aligned}
P_x(S > t, X_{t+\delta} = y) &= \lim_{n\to\infty} P_x\left(X_{t+\delta} = y \text{ and } X\left(\frac{i}{2^n}t\right) = x \text{ for } i = 1, 2, \dots, 2^n\right) \\
&= \lim_{n\to\infty} \left[P_{x,x}(t/2^n)\right]^{2^n} P_{xy}(\delta) \\
&= P_{xy}(\delta) \lim_{n\to\infty} \left[1 - \frac{tQ_x}{2^n} + O\left(2^{-2n}\right)\right]^{2^n} = P_{xy}(\delta) e^{-tQ_x}.
\end{aligned}$$

With this computation in hand, we may now compute $P_x(X_S = y, t < S \le t + \Delta)$ using the Figure 37.1 as our guide

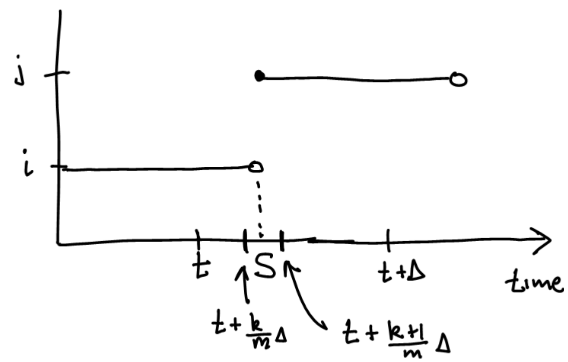


Fig. 37.1. Depicting the first jump of the chain staring at i.

So according Figure 37.1, we must have $X_S = y$ & $t < S \le t + \Delta$ iff for all large n there exists $0 \le k < n$ such that $S > t + k\Delta/n$ & $X_{t+(k+1)\Delta/n} = y$ and therefore

$$\begin{aligned}
&P_x(X_S = y \ \&\ t < S \le t + \Delta) \\
&\quad = \lim_{n\to\infty} P_x\begin{pmatrix} S > t + k\Delta/n \ \&\ X_{t+(k+1)\Delta/n} = y \\ \text{for some } 0 \le k < n \end{pmatrix} \\
&\quad = \lim_{n\to\infty} \sum_{k=0}^{n-1} P_x(S > t + k\Delta/n \ \&\ X_{t+(k+1)\Delta/n} = y) \\
&\quad = \lim_{n\to\infty} \sum_{k=0}^{n-1} P_{xy}(\Delta/n) e^{-(t+k\Delta/n)Q_x} \\
&\quad = \lim_{n\to\infty} \sum_{k=0}^{n-1} e^{-(t+k\Delta/n)Q_x} \left(\frac{\Delta}{n} Q_{xy} + o(n^{-1})\right) \\
&\quad = Q_{xy} \int_t^{t+\Delta} e^{-Q_x s} ds = \frac{Q_{x,y}}{Q_x} \int_t^{t+\Delta} Q_x e^{-Q_x s} ds \\
&\quad = \frac{Q_{x,y}}{Q_x} P_x(t < S \le t + \Delta).
\end{aligned}$$

Letting $t \downarrow 0$ and $\Delta \uparrow \infty$ in this equation we learn that

$$P_x(X_S = y) = \frac{Q_{x,y}}{Q_x}$$

and hence

$$P_x(X_S = y, t < S \le t + \Delta) = P_x(X_S = y) \cdot P_x(t < S \le t + \Delta).$$

This proves also that X_S and S are independent random variables. ■

Remark 37.2. Technically in the proof above, we have used the identity,

$$\begin{aligned}
&\{X_S = y \ \&\ t < S \le t + \Delta\} \\
&\quad = \cup_{N=1}^{\infty} \cap_{n \ge N} \cup_{0 \le m < n} \left\{S > t + k\Delta/n \ \&\ X_{t+(k+1)\Delta/n} = y\right\}.
\end{aligned}$$

Using Theorem 37.1 along with the strong Markov property in Corollary 35.20 leads to the following description of the Markov process associated to Q. Define a Markov matrix, $\tilde{P}$, by

$$\tilde{P}_{xy} := \begin{cases} \frac{Q_{x,y}}{-Q_{x,x}} & \text{if } x \ne y \\ 0 & \text{if } x = y \end{cases} \quad \text{for all } x, y \in S. \tag{37.1}$$

The process X starting at x may be described as follows: 1) stay at x for an $\exp(Q_x)$ – amount of time, S_1, then jump to x_1 with probability $\tilde{P}_{x,x_1}$. Stay at x_1 for an $\exp(Q_{x_1})$ amount of time, S_2, independent of S_1 and then jump to x_2 with probability $\tilde{P}_{x_1,x_2}$. Stay at x_2 for an $\exp(Q_{x_2})$ amount of time, S_3, independent of S_1 and S_2 and then jump to x_3 with probability $\tilde{P}_{x_2,x_3}$, etc. etc. etc. etc. The next corollary formalizes these rules.

Corollary 37.3. *Let Q be the infinitesimal generator of a Markov semigroup $P(t)$. Then the Markov chain, $\{X_t\}$, associated to $P(t)$ may be described as follows. Let $\{Y_k\}_{k=0}^{\infty}$ denote the discrete time Markov chain with Markov matrix $\tilde{P}$ as in Eq. (37.1). Let $\{S_j\}_{j=1}^{\infty}$ be random times such that given $\{Y_j = x_j : j \le n\}$, $S_j \overset{d}{=} \exp\left(Q_{x_{j-1}}\right)$ and the $\{S_j\}_{j=1}^{n}$ are independent for $1 \le j \le n$.*[1] *Now let $N_t = \max\{j : S_1 + \dots + S_j \le t\}$ (see Figure 37.2) and $X_t := Y_{N_t}$. Then $\{X_t\}_t \ge 0$ is the Markov process starting at x with Markov semi-group, $P(t) = e^t Q$.*

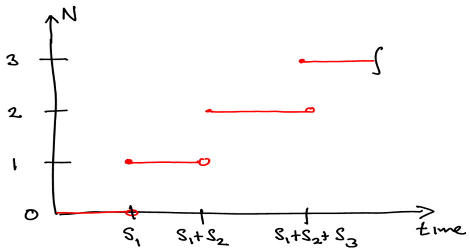


Fig. 37.2. Defining N_t.

In a manner somewhat similar to the proof of Example 35.8 one shows the description in Corollary 37.3 defines a Markov process with the correct semi-group, $P(t)$. For the details the reader is referred to Norris [38, See Theorems 2.8.2 and 2.8.4] and section 37.3 below.

37.1 Hitting and Expected Return times and Probabilities

Let $\{X(t)\}_{t \ge 0}$ be a continuous time Markov chain described by its infinitesimal generator, $Q = (Q_{ij})_{i,j \in S}$ where S is the state space. Further let

$$S_1 = \inf\{t > 0 : X(t) \ne X(0)\}$$

be the first jump time of the chain and $q_j := -Q_{j\,j}$ for all $j \in S$. Recall $P(S_1 > t | X(0) = j) = e^{-q_j t}$ for all $t > 0$ and $\mathbb{E}[S_1 | X(0) = j] = 1/q_j$. Given a subset, A, of the state space, S, let

$$T_A := \inf\{t \ge 0 : X(t) \in A\}$$

[1] A concrete way to chose the $\{S_j\}_{j=1}^{\infty}$ is as follows. Given a sequence, $\{T_j\}_{j=1}^{\infty}$, of i.i.d. $\exp(1)$ – random variables which are independent of $\{Y\}$, define $S_j := T_j / Q_{Y_{j-1}}$.

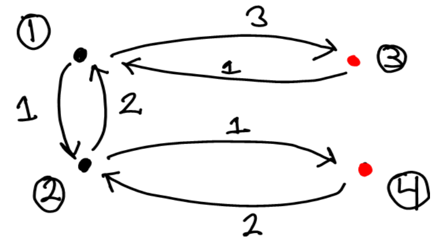


Fig. 37.3. A rate diagram for a four state Markov chain.

be the first time the process, $X(t)$, hits A. By convention, $T_A = \infty$ if $X(t) \notin A$ for all t, i.e. if $X(t)$ does not hit A.

Example 37.4. Let $S = \{1, 2, 3, 4\}$ and $X(t)$ be the continuous time Markov chain determined by the rate diagram, Further let $A = \{3, 4\}$. We would like to compute, $h_i = P_i(X(t) \text{ hits } A)$ for $i = 1, 2$. If $\{Y_n\}_{n=0}^{\infty}$ is the embedded discrete time chain, this is the same as computing, $h_i = P_i(Y_n \text{ hits } A)$ which we know how to do. We now carry out the details. First off the infinitesimal generator, Q, is given by

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} -4 & 1 & 3 & 0 \\ 2 & -3 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 2 & 0 & -2 \end{bmatrix} \end{array}$$

and hence the Markov matrix for $\{Y_n\}$ is given by,

$$\tilde{P} := \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1/4 & 3/4 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}.$$

The first step analysis for the hitting probabilities then implies,

$$\begin{aligned} h_1 = & P_1(X(t) \text{ hits } A | X_{S_1} = 3) P_1(X_{S_1} = 3) \\ & + P_1(X(t) \text{ hits } A | X_{S_1} = 2) P_1(X_{S_1} = 2) \\ = & \frac{3}{4} + h_2 \frac{1}{4} \end{aligned}$$

and

$$\begin{aligned}h_2 =& P_2 (X(t) \text{ hits } A|X_{S_1} = 1) P_2 (X_{S_1} = 1) \\ &+ P_2 (X(t) \text{ hits } A|X_{S_1} = 4) P_2 (X_{S_1} = 4) \\ =& \frac{2}{3} h_1 + \frac{1}{3}\end{aligned}$$

which have solutions, $h_1 = h_2 = 1$ as we know should be the case since this is an irreducible Markov chain.

Example 37.5. Continuing the set up in Example 37.4, we are going to compute $w_i = \mathbb{E}_i T_A$ for $i = 1, 2$. Again by a first step analysis we have,

$$\begin{aligned}w_1 &= \mathbb{E}_1 (T_A|X_{S_1} = 3) P_1 (X_{S_1} = 3) + \mathbb{E}_1 (T_A|X_{S_1} = 2) P_1 (X_{S_1} = 2) \\ &= \frac{1}{4}\frac{3}{4} + \left(\frac{1}{4} + w_2\right)\frac{1}{4} = \frac{1}{4} + \frac{1}{4} w_2\end{aligned}$$

and

$$\begin{aligned}w_2 &= \mathbb{E}_2 (T_A|X_{S_1} = 1) P_2 (X_{S_1} = 1) + \mathbb{E}_2 (T_A|X_{S_1} = 4) P_2 (X_{S_1} = 4) \\ &= \left(\frac{1}{3} + w_1\right)\frac{2}{3} + \frac{1}{3}\frac{1}{3} = \frac{1}{3} + \frac{2}{3} w_1,\end{aligned}$$

where $\frac{1}{4} = \mathbb{E}_1 (S_1)$ and $\frac{1}{3} = \mathbb{E}_2 (S_1)$. The solutions to these equations are:

$$\mathbb{E}_1 (T_A) = w_1 = \frac{2}{5} \text{ and } \mathbb{E}_2 (T_A) = w_2 = \frac{3}{5}.$$

With this example as background, let us now work out the general formula for these hitting times.

Proposition 37.6. *Let Q be the infinitesimal generator of a continuous time Markov chain, $\{X(t)\}_{t\geq 0}$, with state space, S. Suppose that $A \subset S$ and $T_A := \inf\{t \geq 0 : X(t) \in A\}$. If we let $w_i := \mathbb{E}_i T_A$ for all $i \notin A$, then $\{w_i\}_{i\in A^c}$ satisfy the system of linear equations,*

$$w_i = \frac{1}{q_i} + \sum_{j\notin A} \tilde{P}_{ij} w_j = \frac{1}{q_i} + \sum_{j\notin A} \frac{Q_{ij}}{q_i} w_j$$

where as usual, $q_i = -Q_{ii} = \sum_{j\neq i} Q_{ij}$.

Proof. By the first step analysis we have, for $i \notin A$,

$$\begin{aligned}w_i &= \sum_{j\neq i} \mathbb{E}_i [T_A|X_{S_1} = j] P_i (X_{S_1} = j) \\ &= \sum_{j\neq i} \tilde{P}_{ij} \ \mathbb{E}_i [T_A|X_{S_1} = j].\end{aligned}$$

By the strong Markov property,

$$\mathbb{E}_i [T_A|X_{S_1} = j] = \mathbb{E}_i S_1 + \mathbb{E}_j T_A = \frac{1}{q_i} + w_j$$

where $w_j := \mathbb{E}_j T_A = 0$ if $j \in A$. Therefore we have,

$$\begin{aligned}w_i &= \sum_{j\neq i} \tilde{P}_{ij} \ \left(\frac{1}{q_i} + w_j\right) = \frac{1}{q_i} + \sum_{j\neq i} \tilde{P}_{ij} \ w_j \\ &= \frac{1}{q_i} + \sum_{j\notin A} \tilde{P}_{ij} \ w_j\end{aligned}$$

as claimed. ■

Notation 37.7 *Now let*

$$R_j := \inf\{t > S_1 : X_t = j\}$$

be the first return time to j.

Our next goal is to find a formula for $\mathbb{E}_i R_j$ for all $i, j \in S$. Before going to the general case, let us work out an example.

Example 37.8. Let us do an example of a two state Markov chain. Say

$$0 \xrightarrow{\alpha} 1 \xrightarrow{\beta} 0.$$

Let $m_0 = \mathbb{E}_0 R_0$ and $m_1 = \mathbb{E}_1 R_0$, then

$$\begin{aligned}m_0 &= \mathbb{E}_0 [R_0|X_{S_1} = 1] P (X_{S_1} = 1) = \frac{1}{\alpha} + m_1 \\ m_1 &= \mathbb{E}_0 [R_0|X_{S_1} = 0] P (X_{S_1} = 0) = \frac{1}{\beta}\end{aligned}$$

and therefore, $m_0 = \frac{1}{\alpha} + \frac{1}{\beta}$ which is clearly the correct answer in this case. The long run fraction of the time we are in state 0 is therefore

$$\frac{1/\alpha}{m_0} = \frac{\beta}{\alpha + \beta}.$$

This is the same as computing $\lim_{t\to\infty} P (X(t) = 0) = \pi_0$. Indeed for this case,

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

has invariant distribution, $\pi = (\beta, \alpha) / (\alpha + \beta)$. Therefore,

$$\pi_0 = \frac{\beta}{\alpha + \beta} \text{ and } \pi_1 = \frac{\alpha}{\alpha + \beta}. \tag{37.2}$$

as argued above.

Proposition 37.9 (Expected return times). *If $m_{ij} := \mathbb{E}_i R_j$ for all $j \in S$, then*

$$m_{ij} = \frac{1}{q_i} + \sum_{k \neq i \text{ or } j} \frac{Q_{ik}}{q_i} m_{kj}. \tag{37.3}$$

Proof. By a first step analysis we have,

$$\begin{aligned} m_{ij} = \mathbb{E}_i R_j &= \sum_{k \neq i} \mathbb{E}_i \left[R_j | X_{S_1} = k\right] P\left(X_{S_1} = k\right) \\ &= \sum_{k \neq i} \mathbb{E}_i \left[R_j | X_{S_1} = k\right] \frac{Q_{ik}}{q_i}. \end{aligned}$$

Since

$$\begin{aligned} \mathbb{E}_i \left[R_j | X_{S_1} = k\right] &= \begin{cases} \mathbb{E}_i S_1 + \mathbb{E}_k R_j & \text{if } k \neq j \\ \mathbb{E}_i S_1 & \text{if } k = j \end{cases} \\ &= \begin{cases} \frac{1}{q_i} + m_{kj} & \text{if } k \neq j \\ \frac{1}{q_i} & \text{if } k = j \end{cases}. \end{aligned}$$

we arrive at the

$$\begin{aligned} m_{ij} &= \sum_{k \neq i} \mathbb{E}_i \left[R_j | X_{S_1} = k\right] \frac{Q_{ik}}{q_i} \\ &= \frac{1}{q_i^2} Q_{ij} + \sum_{k \neq i \text{ and } k \neq j} \left(\frac{1}{q_i} + m_{kj}\right) \frac{Q_{ik}}{q_i} \\ &= \sum_{k \neq i} \frac{Q_{ik}}{q_i^2} + \sum_{k \neq i \text{ and } k \neq j} m_{kj} \frac{Q_{ik}}{q_i} \\ &= \frac{1}{q_i} + \sum_k 1_{k \neq i, k \neq j} m_{kj} \frac{Q_{ik}}{q_i}. \end{aligned}$$

∎

Corollary 37.10. *Let $\{X(t)\}_{t \geq 0}$ be a finite state irreducible Markov chain with generator, $Q = (Q_{ij})_{i,j \in S}$. If $\pi = (\pi_i)$ is an invariant distribution, then*

$$\pi_i = \frac{1}{q_i m_{ii}} = \frac{1}{q_i \mathbb{E}_i \left[R_i\right]}. \tag{37.4}$$

Proof. Suppose that π_j is an invariant distribution for the chain, so that $\sum_i \pi_i Q_{ik} = 0$ or equivalently,

$$\sum_{i \neq k} \pi_i Q_{ik} = -\pi_k Q_{kk} = \pi_k q_k.$$

It follows from Eq. (37.3) that

$$\begin{aligned} \sum_i \pi_i q_i m_{ij} &= \sum_i \pi_i q_i \frac{1}{q_i} + \sum_i \pi_i q_i \sum_k 1_{k \neq i, k \neq j} \frac{Q_{ik}}{q_i} m_{kj} \\ &= 1 + \sum_{i,k} \pi_i 1_{i \neq k, k \neq j} Q_{ik} m_{kj} \\ &= 1 + \sum_k 1_{k \neq j} \pi_k q_k m_{kj} \\ &= 1 + \sum_i 1_{i \neq j} \pi_i q_i m_{ij}. \end{aligned}$$

Hence it follows that $\pi_i q_i m_{ii} = 1$ which proves Eq. (37.4). ∎

37.2 Long time behavior

In this section, suppose that $\{X(t)\}_{t \geq 0}$ is a continuous time Markov chain with infinitesimal generator, Q, so that

$$P\left(X(t+h) = j | X(t) = i\right) = \delta_{ij} + Q_{ij} h + o(h).$$

We further assume that Q completely determines the chain.

Definition 37.11. *The chain, $\{X(t)\}$, is irreducible iff the underlying discrete time jump chain, $\{Y_n\}$, determined by the Markov matrix, $\tilde{P}_{ij} := \frac{Q_{ij}}{q_i} 1_{i \neq j}$, is irreducible, where*

$$q_i := -Q_{ii} = \sum_{j \neq i} Q_{ij}.$$

Remark 37.12. Using the Sojourn time description of $X(t)$ it is easy to see that $P_{ij}(t) = \left(e^{tQ}\right)_{ij} > 0$ for all $t > 0$ and $i, j \in S$ if $X(t)$ is irreducible. Moreover, if for all $i, j \in S$, $P_{ij}(t) > 0$ for some $t > 0$ then, for the chain $\{Y_n\}$, $i \to j$ and hence $X(t)$ is irreducible. In short the following are equivalent:

1. $\{X(t)\}$ is irreducible,
2. or all $i, j \in S$, $P_{ij}(t) > 0$ for some $t > 0$, and
3. $P_{ij}(t) > 0$ for all $t > 0$ and $i, j \in S$.

In particular, in continuous time all chains are “aperiodic.”

The next theorem gives the basic limiting behavior of irreducible Markov chains. Before stating the theorem we need to introduce a little more notation.

Notation 37.13 *Let S_1 be the time of the first jump of $X(t)$, and*

$$R_i := \min\{t \geq S_1 : X(t) = i\},$$

is the first time hitting the site i after the first jump, and set

$$\pi_i = \frac{1}{q_i \cdot \mathbb{E}_i R_i} \text{ where } q_i := -Q_{ii}.$$

Theorem 37.14 (Limiting behavior). *Let $\{X(t)\}$ be an irreducible Markov chain. Then*

1. *for all initial staring distributions, $\nu(j) := P(X(0) = j)$ for all $j \in S$, and all $j \in S$,*
$$P_\nu\left(\lim_{T\to\infty} \frac{1}{T}\int_0^T 1_{X(t)=j}dt = \pi_j\right) = 1. \tag{37.5}$$
2. *$\lim_{t\to\infty} P_{ij}(t) = \pi_j$ independent of i.*
3. *$\pi = (\pi_j)_{j\in S}$ is stationary, i.e. $0 = \pi Q$, i.e.*
$$\sum_{i\in S} \pi_i Q_{ij} = 0 \text{ for all } j \in S,$$
which is equivalent to $\pi P(t) = \pi$ for all t and to $P_\pi(X(t) = j) = \pi(j)$ for all $t > 0$ and $j \in S$.
4. *If $\pi_i > 0$ for some $i \in S$, then $\pi_i > 0$ for all $i \in S$ and $\sum_{i\in S} \pi_i = 1$.*
5. *The π_i are all positive iff there exists a solution, $\nu_i \geq 0$ to*
$$\sum_{i\in S} \nu_i Q_{ij} = 0 \text{ for all } j \in S \text{ with } \sum_{i\in S} \nu_i = 1.$$
If such a solution exists it is unique and $\nu = \pi$.

Proof. We refer the reader to [38, Theorems 3.8.1.] for the full proof. Let us make a few comments on the proof taking for granted that $\lim_{t\to\infty} P_{ij}(t) =: \pi_j$ exists.

1. Suppose we assume that and that ν is a stationary distribution, i.e. $\nu P(t) = \nu$, then (by dominated convergence theorem),

$$\nu_j = \lim_{t\to\infty} \sum_i \nu_i P_{ij}(t) = \sum_i \lim_{t\to\infty} \nu_i P_{ij}(t) = \left(\sum_i \nu_i\right)\pi_j = \pi_j.$$

Thus $\nu_j = \pi_j$. If $\pi_j = 0$ for all j we must conclude there is not stationary distribution.

2. If we are in the finite state setting, the following computation is justified:

$$\sum_{j\in S} \pi_j P_{jk}(s) = \sum_{j\in S} \lim_{t\to\infty} P_{ij}(t) P_{jk}(s) = \lim_{t\to\infty} \sum_{j\in S} P_{ij}(t) P_{jk}(s)$$
$$= \lim_{t\to\infty} P_{ik}(t+s) = \pi_k.$$

This show that $\pi P(s) = \pi$ for all s and differentiating this equation at $s = 0$ then shows, $\pi Q = 0$.

3. Let us now explain why

$$\frac{1}{T}\int_0^T 1_{X(t)=j}dt \to \frac{1}{q_j \cdot \mathbb{E}_j R_j}.$$

The idea is that, because the chain is irreducible, no matter how we start the chain we will eventually hit the site j. Once we hit j, the (strong) Markov property implies the chain forgets how it got there and behaves as if it started at j. Since what happens for the initial time interval of hitting j in computing the average time spent at j, namely $\lim_{T\to\infty} \frac{1}{T}\int_0^T 1_{X(t)=j}dt$, we may as well have started our chain at j in the first place.

Now consider one typical cycle in the chain staring at j jumping away at time S_1 and then returning to j at time R_j. The average first jump time is $\mathbb{E}S_1 = 1/q_j$ while the average length of such as cycle is $\mathbb{E}R_j$. As the chain repeats this procedure over and over again with the same statistics, we expect (by a law of large numbers) that the average time spent at site j is given by

$$\frac{\mathbb{E}S_1}{\mathbb{E}R_j} = \frac{1/q_j}{\mathbb{E}_j R_j} = \frac{1}{q_j \cdot \mathbb{E}_j R_j}.$$

■

37.3 Formal Proofs

Let $S(\omega) = \tau(\omega) = \inf\{t > 0 : \omega(t) \neq \omega(0)\}$. Notice that τ is an optional time by Proposition 27.11 or by first principles. Notice that $\tau(\omega) < t$ iff $\omega(s) \neq \omega(0)$ for some $s \in \mathbb{Q} \cap (0,t)$, i.e.

$$\{\tau < t\} = \cup_{s\in\mathbb{Q}\cap(0,t)} \{X_s \neq X_0\} \in \mathcal{B}_t.$$

Lemma 37.15. *Suppose that $f : \mathbb{R}_+ \to (0,1]$ is a decreasing function such that $f(s+t) = f(s) f(t)$ for all $s, t > 0$, then $f(t) = e^{-ct}$ for some $c \geq 0$.*

Proof. Let $g(t) := \ln f(t)$, then $g(t)$ is decreasing and $g(t+s) = g(t) + g(s)$ for all $s, t > 0$. So for $n \in \mathbb{N}$ it follows that

$$g(1) = g\left(n \cdot \frac{1}{n}\right) = g\left(\frac{1}{n} + \dots + \frac{1}{n}\right) = n \cdot g\left(\frac{1}{n}\right)$$

and therefore $g(1/n) = \frac{1}{n}g(1)$. Similarly it follows that $g(k/n) = \frac{k}{n}g(1)$ for all $k, n \in \mathbb{N}$.

Let $c := -g(1) = -\ln f(1) \geq 0$. If $t \in \mathbb{R}_+$ and $a, b \in \mathbb{Q}_+$ with $a < t < b$ we have,

$$-bc = g(b) \leq g(t) \leq g(a) = -ac.$$

Letting $a \uparrow t$ and $b \downarrow t$ in these inequalities then shows that $g(t) = -ct$ and therefore that $f(t) = e^{-ct}$ as claimed. ∎

Let us begin with a better proof of Theorem 37.1 following Kallenberg, [30, Lemma 12.16]. For the reader's convenience we will restate the theorem here.

Theorem 37.16 (Theorem 37.1 restated). *Let $S_1 = \tau = \inf\{t > 0 : X_t \neq X_0\}$ and suppose that $x \in S$ is a **non-absorbing state**, i.e. $P_x(S_1 = \infty) < 1$. (That is there is a chance that we will leave the site x.) Then $\mathrm{Law}_{P_x}(S_1) \overset{d}{=} \exp(c(x))$ for some $c(x) > 0$ and $\theta_{S_1}^{-1}(\mathcal{B})$ and S_1 are independent.*

Proof. We then have, for $s, t \in \mathbb{R}_+$ that

$$\begin{aligned} P_x(S_1 > s + t) &= P_x(S_1 > s \;\&\; S_1 - s > t) = P_x(S_1 > s \;\&\; S_1 \circ \theta_s > t) \\ &= \mathbb{E}_x\left(1_{S_1 > s} \;\; \mathbb{E}_x\left(1_{S_1 \circ \theta_s > t} | \mathcal{B}_s\right)\right) = \mathbb{E}_x\left(1_{S_1 > s} \;\; \mathbb{E}_{X_s}\left[1_{S_1 > t}\right]\right) \\ &= \mathbb{E}_x\left(1_{S_1 > s} \;\; \mathbb{E}_x\left[1_{S_1 > t}\right]\right) = P_x(S_1 > s) P_x(S_1 > t). \end{aligned}$$

Since $P(S_1 > t) > 0$ for t small as our paths are right continuous, we may use Lemma 37.15 to conclude that $P_x(S_1 > t) = e^{-ct}$ for some $c = c(x) \geq 0$. Since $P_x(S_1 = \infty) < 1$ we must conclude that $c > 0$ and therefore $P_x(S_1 > t) = e^{-ct}$ for some $c > 0$. In particular, $P_x(S_1 = \infty) = 0$.

Next suppose that $f : \Omega \to \mathbb{R}$ is a bounded $\mathcal{B}$ – measurable function. Then

$$\begin{aligned} \mathbb{E}_x\left[f \circ \theta_{S_1} 1_{S_1 > t}\right] &= \mathbb{E}_x\left[1_{S_1 > t} \cdot \mathbb{E}_x\left(f \circ \theta_{S_1} | \mathcal{B}_{S_1}\right)\right] = \mathbb{E}_x\left[1_{S_1 > t} \cdot \mathbb{E}_{X_{S_1}} f\right] \\ &= \mathbb{E}_x\left[1_{S_1 > t} \cdot \mathbb{E}_x f\right] = \mathbb{E}_x\left[1_{S_1 > t}\right] \cdot \mathbb{E}_x f. \end{aligned}$$

Taking $t = 0$ in this identity then shows $\mathbb{E}_x\left[f \circ \theta_{S_1}\right] = \mathbb{E}_x f$ and therefore we have shown

$$\mathbb{E}_x\left[f \circ \theta_{S_1} 1_{S_1 > t}\right] = \mathbb{E}_x\left[1_{S_1 > t}\right] \cdot \mathbb{E}_x f = \mathbb{E}_x\left[1_{S_1 > t}\right] \cdot \mathbb{E}_x\left[f \circ \theta_{S_1}\right].$$

Since f and $t \geq 0$ were arbitrary, it follows that $\theta_{S_1}^{-1}(\mathcal{B})$ and S_1 are independent. ∎

Corollary 37.17. *Now suppose that all of the states are non-absorbing. Let S_n be the n^{th} – jump time of the chain. We now define $\pi(x, y) := P_x(X_{S_1} = y)$ and let $Y_n := X_{S_n}$ and $\tau_n := S_n - S_{n-1}$ for all n and let $\xi := \lim_{n\to\infty} S_n$. Then the joint distributions of the jumps and holds for this Markov chain are determined by*

$$\begin{aligned} &P_x(Y_1 = y_1, \dots, Y_n = y_n, \tau_1 > t_1, \dots, \tau_n > t_n) \\ &\quad = \pi(x, y_1)\pi(y_1, y_2)\dots\pi(y_{n-1}, y_n) e^{-c(x)t_1} e^{-c(y_1)t_2} \dots e^{-c(y_{n-1})t_n}. \end{aligned} \tag{37.6}$$

Proof. By the Strong Markov property we know that given $Y_1 = X_{S_1} = y_1$ that $\{X_{S_1+t}\}_{t\geq 0}$ is independent of $\mathcal{B}_{S_1}$ and has distribution of $\{X_t\}_{t\geq 0}$ under P_{y_1}. Therefore $S_2 - S_1$ is independent of $\mathcal{B}_{S_1}$ and given $Y_1 = X_{S_1} = y_1$ and $Y_2 = X_{S_2} = y_2$, $\{X_{S_1+S_2+t}\}_{t\geq 0}$ is independent of $\mathcal{B}_{S_2}$ and has law given by P_{y_2}. Continuing this way inductively leads to Eq. (37.6). ∎

One of our next goals is to identify $\pi(x, y)$ and $c(x)$ for all $x, y \in S$ in terms of the infinitesimal generators of the process. It is convenient to define $q : S \times S \to \mathbb{R}$ by

$$q(x, y) = \begin{cases} c(x)\pi(x, y) & \text{if } x \neq y \\ -c(x) & \text{if } x = y \end{cases} \tag{37.7}$$

and

$$T_t f(x) := \mathbb{E}_x f(X_t) \tag{37.8}$$

for all $f : S \to \mathbb{R}$ bounded. Using the Markov property we find,

$$T_{t+s} f(x) = \mathbb{E}_x f(X_{t+s}) = \mathbb{E}_x f(X_t \circ \theta_s) = \mathbb{E}_x \mathbb{E}_{X_s} f(X_t) = \mathbb{E}_x T_t f(X_s) = (T_s T_t f)(x)$$

which is to say that T_t is a semi-group of operators. The semi-group property is a strong indication that $T_t f$ satisfies a differential equation as we now show to be the case.

Theorem 37.18 (Backward's Kolmogorov's equation). *The Markov transition operators $\{T_t\}_{t\geq 0}$ satisfy the **backward's Kolmogorov equation**;*

$$\frac{d}{dt} T_t f(x) = \sum_{y \in S} q(x, y) T_t f(y) \tag{37.9}$$

for all $f : S \to \mathbb{R}$ bounded Alternatively put the transition functions, $p_t(x, y) := P_x(X_t = y)$ satisfy

$$\frac{d}{dt} p_t(x, z) = \sum_{y \in S} q(x, y) T_t \delta_z(y) = \sum_{y \in S} q(x, y) p_t(y, z) \tag{37.10}$$

with initial conditions, $p_0(x, y) = \delta_{x,y}$.

Proof. First note that Eq. (37.10) follows from Eq. (37.9) by taking $f(x) = \delta_z(x)$ in which case we have

$$T_t \delta_z(x) = \mathbb{E}_x \delta_z(X_t) = P_x(X_t = z) = p_t(x, z).$$

Our proof of Eq. (37.9) will follow [30, Theorem 12.22, p. 242].

Let $\tau = S_1$ and $\sigma := \tau \wedge t$, then

$$\begin{aligned} T_t f(x) &= \mathbb{E}_x f(X_t) = \mathbb{E}_x [f(X_t) : \tau > t] + \mathbb{E}_x [f(X_t) : \tau \le t] \\ &= \mathbb{E}_x [f(x) : \tau > t] + \mathbb{E}_x [f(X_t) : \tau \le t] \\ &= f(x) P(\tau > t) + \mathbb{E}_x [f(X_t) : \tau \le t]. \end{aligned}$$

In order to evaluate $\mathbb{E}_x [f(X_t) : \tau \le t]$, let us set $X_t = X_0$ if $t < 0$. We then have, for $t \ge \tau$, that

$$f(X_t) = [f(X_{t-s}) \circ \theta_\tau]_{s=\tau}.$$

Note well that it is **not** true that $f(X_t) = f(X_{t-\tau}) \circ \theta_\tau$ since,

$$\begin{aligned} f(X_{t-\tau}) \circ \theta_\tau(\omega) &= f(X_{t-\tau})(\omega(\tau(\omega) + \cdot)) \\ &= f(\omega(\tau(\omega) + t - \tau(\omega(\tau(\omega) + \cdot)))) \end{aligned}$$

while on the other hand,

$$\begin{aligned} ([f(X_{t-s}) \circ \theta_\tau]_{s=\tau})(\omega) &= [f(X_{t-s}) \circ \theta_\tau(\omega)]_{s=\tau(\omega)} = [f(X_{t-s})(\omega(\tau(\omega) + \cdot))]_{s=\tau(\omega)} \\ &= [f(\omega(\tau(\omega) + t - s))]_{s=\tau(\omega)} = f(\omega(t)) = f(X_t)(\omega) \end{aligned}$$

as desired. See Lemma 27.6 which shows that X is progressively measurable which we will use below.

With the above comments working informally at first,

$$\begin{aligned} \mathbb{E}_x [f(X_t) : \tau \le t] &= \mathbb{E}_x \left[[f(X_{t-s}) \circ \theta_\tau]_{s=\tau} : \tau \le t\right] \\ &= \mathbb{E}_x \left[\mathbb{E}_x \left([f(X_{t-s}) \circ \theta_\tau]_{s=\tau} | \mathcal{B}_\tau^+\right) : \tau \le t\right] \\ &= \mathbb{E}_x \left[\mathbb{E}_x \left([f(X_{t-s}) \circ \theta_\tau] | \mathcal{B}_\tau^+\right)|_{s=\tau} : \tau \le t\right] \\ &= \mathbb{E}_x \left[[\mathbb{E}_{X_\tau} f(X_{t-s})]|_{s=\tau} : \tau \le t\right] \\ &= \mathbb{E}_x \left[[\mathbb{E}_{Y_1} f(X_{t-s})]|_{s=\tau} : \tau \le t\right] \\ &= \sum_{y \ne x} \pi(x,y) \mathbb{E}_x \left([\mathbb{E}_y f(X_{t-s})]|_{s=\tau} : \tau \le t\right) \\ &= \sum_{y \ne x} \pi(x,y) \int_0^t c(x) e^{-c(x)u} [\mathbb{E}_y f(X_{t-s})]|_{s=u} du \\ &= \sum_{y \ne x} \pi(x,y) \int_0^t c(x) e^{-c(x)u} (T_{t-u} f)(y)\, du. \end{aligned} \tag{37.11}$$

We will now fill in the missing details in the above argument. (The general idea going on here is the notion of regular conditional expectations which we have available to us because of the Markov property.)

Let $F(s,\omega) := f(X_{t-s})(\omega) = f(\omega(t-s))$ which is jointly measurable by Lemma 27.6. To finish the proof of Eq. (37.11) it suffices to show

$$\mathbb{E}_x \left[[F(s,\cdot) \circ \theta_\tau]_{s=\tau} : \tau \le t\right] = \sum_{y \ne x} \pi(x,y) \int_0^t [\mathbb{E}_y F(u,\cdot)] c(x) e^{-c(x)u} du$$

for all bounded measurable, $F : \mathbb{R}_+ \times \Omega \to \mathbb{R}$. This is done using the multiplicative system theorem in the usual way. The point is that if $F(s,\omega) = h(s) f(\omega)$, then

$$\begin{aligned} \mathbb{E}_x \left[[F(s,\cdot) \circ \theta_\tau]_{s=\tau} : \tau \le t\right] &= \mathbb{E}_x \left[h(\tau) 1_{\tau \le t} f \circ \theta_\tau\right] = \mathbb{E}_x \left[h(\tau) 1_{\tau \le t} \mathbb{E}_{X_\tau} f\right] \\ &= \sum_{y \ne x} \pi(x,y) \mathbb{E}_y f \mathbb{E}_x [h(\tau) 1_{\tau \le t}] \\ &= \sum_{y \ne x} \pi(x,y) \mathbb{E}_y f \cdot \int_0^t c(x) e^{-c(x)u} h(u)\, du \\ &= \sum_{y \ne x} \pi(x,y) \int_0^t [\mathbb{E}_y F(u,\cdot)] c(x) e^{-c(x)u} du. \end{aligned}$$

Thus we have shown,

$$T_t f(x) = f(x) e^{-c(x)t} + c(x) \int_0^t \left(\sum_{y \ne x} \pi(x,y) (T_{t-u} f)(y)\right) e^{-c(x)u} du.$$

and therefore $T_t f(x)$ is differentiable in t and

$$\begin{aligned} \frac{d}{dt}\left[e^{c(x)t} T_t f(x)\right] &= \frac{d}{dt} c(x) \int_0^t \left(\sum_{y \ne x} \pi(x,y) (T_{t-u} f)(y)\right) e^{c(x)(t-u)} du \\ &= \frac{d}{dt} c(x) \int_0^t \left(\sum_{y \ne x} \pi(x,y) (T_s f)(y)\right) e^{c(x)s} ds \\ &= c(x) \left(\sum_{y \ne x} \pi(x,y) (T_t f)(y)\right) e^{c(x)t} \end{aligned}$$

and hence,

$$\frac{d}{dt} T_t f(x) + c(x) T_t f(x) = c(x) \left(\sum_{y \ne x} \pi(x,y) (T_t f)(y)\right)$$

or equivalently that

$$\frac{d}{dt}T_t f(x) = c(x)\left(\sum_{y\neq x}\pi(x,y)(T_t f)(y)\right) - c(x)T_t f(x)$$
$$= \sum_{y\in S} q(x,y)\,T_t f(y)$$

where $q(x,y) = c(x)\pi(x,y)$ for $y \neq x$ and $q(x,x) = -c(x)$ for all $x \in S$. ■

Our next goal is to show directly that we can use the jump hold description of the Markov chain in order to give a full definition of the process. The reader should compare the next theorem with Theorem 17.26.

Theorem 37.19. *Let $\{Y_n\}_{n=0}^{\infty}$ be a discrete time Markov chain in some countable state space, S. Also let $\{\gamma_n\}_{n=1}^{\infty}$ be an independent collection of i.i.d. exponential random variables with parameter 1 and suppose that $c : S \to (0,\infty)$ is a given function. As usual we let $S_n := \gamma_n / c(Y_{n-1})$ and following the notation in [38], $J_n = S_1 + \cdots + S_n$. Let $\zeta = \lim_{n\to\infty} J_n$ which will be the life-time of the process that we now describe. Let*

$$N_t := \#\{n : J_n \leq t\} = \sum_{n=1}^{\infty} 1_{(0,t]}(J_n) \text{ for } t < \zeta$$

and define

$$X_t = \begin{cases} Y_{N_t} & \text{if } t < \zeta \\ \partial & \text{if } t \geq \zeta \end{cases}$$

where ∂ is a new point we adjoin to S which we refer to as the cemetery. (Our jumping particle whose position is described by $\{X_t\}$ is sent to the cemetery at the first time it has made an infinite number of jumps which is the same a.s. as the first time it goes off to infinity[2].) Let Ω denote those paths from $\mathbb{R}_+ \to S \cup \{\partial\}$ which are right continuous with left limits such that $\omega(t) = \partial$ for $t > \zeta(\omega)$ – the first time that ω has made an infinite number of jumps. Thus $X_\cdot \in \Omega$. The process $\{X_t\}_{t\geq 0}$ is a Markov process, i.e. for all $T > 0$ and $f : \Omega \to \mathbb{R}$ bounded and measurable we have,

$$\mathbb{E}\left[f(X_{T+\cdot})\,|\mathcal{B}_T\right] = \mathbb{E}_{X_T} f(X) \text{ and } T < \zeta. \tag{37.12}$$

Proof. Let $g \in b\mathcal{B}_T$ and $f = f(X)$. Since $X_\cdot$ is a function of $(Y_0, Y_1, \dots; S_1, S_2, \dots)$, i.e. $X := \Gamma(Y_0, Y_1, \dots; S_1, S_2, \dots)$ as described in the statement of the theorem it follows that there is a function F such that

$$f(X) = f(\Gamma(Y_0, Y_1, \dots; S_1, S_2, \dots)) =: F(Y_0, Y_1, \dots; S_1, S_2, \dots)$$

[2] This is because if X_t keeps revisiting a compact (i.e. finite) subset K of S, then the accumulated jump times associated with jumps starting in K will be inifinite.

We have

$$\theta_T X = \Gamma(Y_m, Y_{m+1}, \dots; S_{m+1} - (T - J_m), S_{m+2}, \dots) \text{ on } \{J_m \leq T < J_{m+1}\}$$

and therefore

$$f(\theta_T X) = F(Y_m, Y_{m+1}, \dots; S_{m+1} - (T - J_m), S_{m+2}, \dots) \text{ on } \{J_m \leq T < J_{m+1}\}.$$

Similarly

$$g = G(Y_0, Y_1, \dots Y_m; S_1, S_2, \dots S_m) \text{ on } \{T < J_{m+1}\}$$

and define

$$g_m := G(Y_0, Y_1, \dots Y_m; S_1, S_2, \dots S_m).$$

(We need some more precision here to assert there exists a globally defined function, G, such that the above equation holds, see Norris [38, Lemma 6.5.3]) Using this notation we find,

$$\begin{aligned}
&\mathbb{E}\left[f(\theta_T X)\,g : J_m \leq T < J_{m+1}\right] \\
&= \mathbb{E}\left[F(Y_m, Y_{m+1}, \dots; S_{m+1} - (T - J_m), S_{m+2}, \dots)\,g_m : J_m \leq T < J_{m+1}\right] \\
&= \mathbb{E}\left[F(Y_m, Y_{m+1}, \dots; S_{m+1} - (T - J_m), S_{m+2}, \dots)\,g_m : J_m \leq T < J_m + S_{m+1}\right] \\
&= \mathbb{E}\left[F\left(Y_m, Y_{m+1}, \dots; \frac{\gamma_{m+1}}{c(Y_m)} - (T - J_m), S_{m+2}, \dots\right) g_m : J_m \leq T < J_m + \frac{\gamma_{m+1}}{c(Y_m)}\right] \\
&= \mathbb{E}\left[\int_{T-J_m}^{\infty} d\tau\, c(Y_m)\, e^{-c(Y_m)\tau} F(Y_m, Y_{m+1}, \dots; \tau - (T - J_m), S_{m+2}, \dots)\,g_m : J_m \leq T\right] \\
&= \mathbb{E}\left[e^{-c(Y_m)(T-J_m)} \int_0^{\infty} d\tau\, F(Y_m, Y_{m+1}, \dots; \tau, S_{m+2}, \dots)\,g_m c(Y_m)\, e^{-c(Y_m)\tau} : J_m \leq T\right] \\
&= \mathbb{E}\left[e^{-c(Y_m)(T-J_m)} \int_0^{\infty} d\tau\, \mathbb{E}_{Y_m}\left[F(Y_0, Y_1, \dots; \tau, S_2, \dots)\right] \cdot g_m c(Y_m)\, e^{-c(Y_m)\tau} : J_m \leq T\right] \\
&= \mathbb{E}\left[\mathbb{E}_{Y_m}\left[F(Y_0, Y_1, \dots; S_1, S_2, \dots)\right] \cdot g_m : J_m \leq T < J_{m+1}\right] \\
&= \mathbb{E}\left[\mathbb{E}_{X_T}\left[F(Y_0, Y_1, \dots; S_1, S_2, \dots)\right] \cdot g : J_m \leq T < J_{m+1}\right] \\
&= \mathbb{E}\left[\mathbb{E}_{X_T}[f] \cdot g : J_m \leq T < J_{m+1}\right]
\end{aligned}$$

Summing this last equation on m then shows,

$$\mathbb{E}\left[f(\theta_T X)\,g : T < \zeta\right] = \mathbb{E}\left[\mathbb{E}_{X_T}[f] \cdot g : T < \zeta\right]$$

which is equivalent to Eq. (37.12).

Observe that, for all $s < t$, that

$$\mathbb{E}\left[f(X_t)\,|\mathcal{B}_s\right] = \mathbb{E}\left[f(\theta_s X_{t-s})\,|\mathcal{B}_s\right] = \mathbb{E}_{X_s} f(X_{t-s}) = (T_{t-s}f)(X_s)$$

which shows, in light of Theorem 37.18, that the chains we have been constructing are time homogeneous with the correct Markovian semi-group. ■

38

Continuous Time M.C. Examples

38.1 Birth and Death Process basics

A **birth and death process** is a continuous time Markov chain with state space being $S = \{0, 1, 2, \ldots\}$ and transitions rates of the form;

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2 \underset{\mu_3}{\overset{\lambda_2}{\rightleftarrows}} 3 \ldots \underset{\mu_{n-1}}{\overset{\lambda_{n-2}}{\rightleftarrows}} (n-1) \underset{\mu_n}{\overset{\lambda_{n-1}}{\rightleftarrows}} n \underset{\mu_{n+1}}{\overset{\lambda_n}{\rightleftarrows}} (n+1) \ldots$$

The associated Q matrix for this chain is given by

$$Q = \begin{array}{c|cccccc} & 0 & 1 & 2 & 3 & 4 & \ldots \\ \hline 0 & -\lambda_0 & \lambda_0 & & & & \\ 1 & \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & & \\ 2 & & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & & \\ 3 & & & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & \\ 4 & & & & \ddots & \ddots & \ddots \\ \vdots & & & & & & \end{array}.$$

If $\pi_n(t) = P(X(t) = n)$, then $\pi(t) = (\pi_n(t))_{n\geq 0}$ satisfies, $\dot{\pi}(t) = \pi(t) Q$ which written out in components is the system of differential equations;

$$\begin{aligned} \dot{\pi}_0(t) &= -\lambda_0 \pi_0(t) + \mu_1 \pi_1(t) \\ \dot{\pi}_1(t) &= \lambda_0 \pi_0(t) - (\lambda_1 + \mu_1) \pi_1(t) + \mu_2 \pi_2(t) \\ &\vdots \\ \dot{\pi}_n(t) &= \lambda_{n-1} \pi_{n-1}(t) - (\lambda_n + \mu_n) \pi_n(t) + \mu_{n+1} \pi_{n+1}(t). \\ &\vdots \end{aligned}$$

The associated discrete time chain is described by the jump diagram,

$$0 \underset{\frac{\mu_1}{\lambda_1+\mu_1}}{\overset{1}{\rightleftarrows}} 1 \underset{\frac{\mu_2}{\lambda_2+\mu_2}}{\overset{\frac{\lambda_1}{\lambda_1+\mu_1}}{\rightleftarrows}} 2 \underset{\frac{\mu_3}{\lambda_3+\mu_3}}{\overset{\frac{\lambda_2}{\lambda_2+\mu_2}}{\rightleftarrows}} 3 \cdots \rightleftarrows (n-1) \underset{\frac{\mu_n}{\lambda_n+\mu_n}}{\overset{\frac{\lambda_{n-1}}{\lambda_{n-1}+\mu_{n-1}}}{\rightleftarrows}} n \underset{\frac{\mu_{n+1}}{\lambda_{n+1}+\mu_{n+1}}}{\overset{\frac{\lambda_n}{\lambda_n+\mu_n}}{\rightleftarrows}} \ldots$$

In the jump hold description, a particle follows this discrete time chain. When it arrives at a site, say n, it stays there for an $\exp(\lambda_n + \mu_\nu)$ – time and then jumps to either $n-1$ or n with probability $\frac{\mu_n}{\lambda_n+\mu_n}$ or $\frac{\lambda_n}{\lambda_n+\mu_n}$ respectively. Given your homework problem we may also describe these transitions by assuming at each site we have a death clock $D_n = \exp(\mu_n)$ and a Birth clock $B_n = \exp(\lambda_n)$ with B_n and D_n being independent. We then stay at site n until either B_n or D_n rings, i.e. for $\min(B_n, D_n) = \exp(\lambda_n + \mu_n)$ – amount of time. If B_n rings first we go to $n+1$ while if D_n rings first we go to $n-1$. When we are at 0 we go to 1 after waiting $\exp(\lambda_0)$ – amount of time.

38.2 Pure Birth Process:

The infinitesimal generator for a pure Birth process is described by the following rate diagram

$$0 \xrightarrow{\lambda_0} 1 \xrightarrow{\lambda_1} 2 \xrightarrow{\lambda_2} \ldots \xrightarrow{\lambda_{n-1}} (n-1) \xrightarrow{\lambda_{n-1}} n \xrightarrow{\lambda_n} \ldots.$$

For simplicity we are going to assume that we start at state 0. We will examine this model is both the sojourn description and the infinitesimal description. The typical sample path is shown in Figure 38.1.

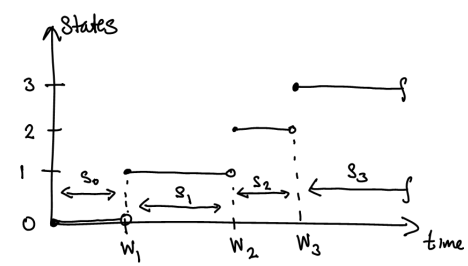


Fig. 38.1. A typical sample path for a pure birth process.

38.2.1 Infinitesimal description

The matrix Q is this case is given by

$$Q_{i,i+1} = \lambda_i \text{ and } Q_{ii} = -\lambda_i \text{ for all } i = 0, 1, 2, \dots$$

with all other entries being zero. Thus we have

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \dots & \end{array} \\ \begin{bmatrix} -\lambda_0 & \lambda_0 & & & & \\ & -\lambda_1 & \lambda_1 & & & \\ & & -\lambda_2 & \lambda_2 & & \\ & & & -\lambda_3 & \lambda_3 & \\ & & & & \ddots & \ddots \end{bmatrix} \end{array}.$$

If we now let

$$\pi_j(t) = P_0(X(t) = j) = \left[\pi(0) e^{tQ}\right]_j$$

then $\pi_j(t)$ satisfies the system of differential equations;

$$\begin{aligned} \dot{\pi}_0(t) &= -\lambda_0 \pi_0(t) \\ \dot{\pi}_1(t) &= \lambda_0 \pi_0(t) - \lambda_1 \pi_1(t) \\ &\vdots \\ \dot{\pi}_n(t) &= \lambda_{n-1}\pi_{n-1}(t) - \lambda_n \pi_n(t) \\ &\vdots \end{aligned}$$

The solution to the first equation is given by

$$\pi_0(t) = e^{-\lambda_0 t}\pi(0) = e^{-\lambda_0 t}$$

and the remaining may now be obtained inductively, see the ODE Lemma 36.14, using

$$\pi_n(t) = \lambda_{n-1} e^{-\lambda_n t} \int_0^t e^{\lambda_n \tau} \pi_{n-1}(\tau)\, d\tau. \tag{38.1}$$

So for example

$$\begin{aligned} \pi_1(t) &= \lambda_0 e^{-\lambda_1 t} \int_0^t e^{\lambda_1 \tau} \pi_0(\tau)\, d\tau = \lambda_0 e^{-\lambda_1 t} \int_0^t e^{\lambda_1 \tau} e^{-\lambda_0 \tau} d\tau \\ &= \frac{\lambda_0}{\lambda_1 - \lambda_0} e^{-\lambda_1 t} e^{(\lambda_1 - \lambda_0)\tau}|_{\tau=0}^{\tau=t} = \frac{\lambda_0}{\lambda_1 - \lambda_0}\left[e^{-\lambda_1 t} e^{(\lambda_1 - \lambda_0)t} - e^{-\lambda_1 t}\right] \\ &= \frac{\lambda_0}{\lambda_1 - \lambda_0}\left[e^{-\lambda_0 t} - e^{-\lambda_1 t}\right]. \end{aligned}$$

If $\lambda_1 = \lambda_0$, this becomes, $\pi_1(t) = (\lambda_0 t) e^{-\lambda_0 t}$ instead. In principle one can compute all of these integrals (you have already done the case where $\lambda_j = \lambda$ for all j) to find all of the $\pi_n(t)$. The formula for the solution is given as

$$\pi_n(t) = P(X(t) = n | X(0) = 0) = \lambda_0 \dots \lambda_{n-1}\left[\sum_{k=0}^n B_{k,n} e^{-\lambda_k t}\right]$$

where the $B_{k,n}$ are given on p. 338 of the book.

To see that this form of the answer is reasonable, if we look at the equations for $n = 0, 1, 2, 3$, we have

$$\begin{aligned} \dot{\pi}_0(t) &= -\lambda_0 \pi_0(t) \\ \dot{\pi}_1(t) &= \lambda_0 \pi_0(t) - \lambda_1 \pi_1(t) \\ \dot{\pi}_2(t) &= \lambda_1 \pi_1(t) - \lambda_2 \pi_2(t) \\ \dot{\pi}_3(t) &= \lambda_2 \pi_2(t) - \lambda_3 \pi_3(t) \end{aligned}$$

and the matrix associated to this system is

$$Q' = \begin{bmatrix} -\lambda_0 & \lambda_0 & & \\ & -\lambda_1 & \lambda_1 & \\ & & -\lambda_2 & \lambda_2 \\ & & & -\lambda_3 \end{bmatrix}$$

so that $(\pi_0(t), \dots, \pi_3(t)) = (1, 0, 0, 0) e^{tQ'}$. If all of the λ_j are distinct, then Q' has $\{\lambda_j\}_{j=0}^3$ as its distinct eigenvalues and hence is diagonalizable. Therefore we will have

$$(\pi_0(t), \dots, \pi_3(t)) = (1, 0, 0, 0) S \begin{bmatrix} e^{-t\lambda_0} & & & \\ & e^{-t\lambda_1} & & \\ & & e^{-t\lambda_2} & \\ & & & e^{-t\lambda_3} \end{bmatrix} S^{-1}$$

for some invertible matrix S. In particular it follows that $\pi_3(t)$ must be a linear combination of $\left\{e^{-t\lambda_j}\right\}_{j=0}^3$. Generalizing this argument shows that there must be constants, $\{C_{k,n}\}_{k=0}^n$ such that

$$\pi_n(t) = \sum_{k=0}^n C_{kn} e^{-t\lambda_k}.$$

We may now plug these expressions into the differential equations,

$$\dot{\pi}_n(t) = \lambda_{n-1}\pi_{n-1}(t) - \lambda_n \pi_n(t),$$

to learn

$$-\sum_{k=0}^n \lambda_k C_{kn} e^{-t\lambda_k} = \lambda_{n-1}\sum_{k=0}^{n-1} C_{k,n-1} e^{-t\lambda_k} - \lambda_n \sum_{k=0}^n C_{kn} e^{-t\lambda_k}.$$

Since one may show $\left\{e^{-t\lambda_k}\right\}_{k=0}^n$ are linearly independent, we conclude that

$$-\lambda_k C_{kn} = \lambda_{n-1} C_{k,n-1} \cdot 1_{k \leq n-1} - \lambda_n C_{kn} \text{ for } k = 0, 1, 2, \dots, n.$$

This equation gives no information for $k = n$, but for $k < n$ it implies,

$$C_{k,n} = \frac{\lambda_{n-1}}{\lambda_n - \lambda_k} C_{k,n-1} \text{ for } k \leq n-1.$$

To discover the value of $C_{n,n}$ we use the fact that $\sum_{k=0}^n C_{kn} = \pi_n(0) = 0$ for $n \geq 1$ to learn,

$$C_{n,n} = -\sum_{k=0}^{n-1} C_{k,n} = -\sum_{k=0}^{n-1} \frac{\lambda_{n-1}}{\lambda_n - \lambda_k} C_{k,n-1}.$$

One may determine all of the coefficients from these equations. For example, we know that $C_{00} = 1$ and therefore,

$$C_{0,1} = \frac{\lambda_0}{\lambda_1 - \lambda_0} \text{ and } C_{1,1} = -C_{0,1} = -\frac{\lambda_0}{\lambda_1 - \lambda_0}.$$

Thus we learn that

$$\pi_1(t) = \frac{\lambda_0}{\lambda_1 - \lambda_0}\left(e^{-\lambda_0 t} - e^{-\lambda_1 t}\right)$$

as we have seen from above.

Remark 38.1. It is interesting to observe that

$$\begin{aligned} \frac{d}{dt}(\pi_0(t), \dots, \pi_3(t)) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} &= \frac{d}{dt}(1,0,0,0)\, e^{tQ'} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= (1,0,0,0)\, e^{tQ'} Q' \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

where

$$Q' \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\lambda_0 & \lambda_0 & & \\ & -\lambda_1 & \lambda_1 & \\ & & -\lambda_2 & \lambda_2 \\ & & & -\lambda_3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\lambda_3 \end{bmatrix}$$

and therefore,

$$\frac{d}{dt}(\pi_0(t), \dots, \pi_3(t)) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \leq 0.$$

This shows that $\sum_{j=0}^3 \pi_j(t) \leq \sum_{j=0}^3 \pi_j(0) = 1$. Similarly one shows that

$$\sum_{j=0}^n \pi_j(t) \leq 1 \text{ for all } t \geq 0 \text{ and } n.$$

Letting $n \to \infty$ in this estimate then implies

$$\sum_{j=0}^\infty \pi_j(t) \leq 1.$$

It is possible that we have a strict inequality here! We will discuss this below.

Remark 38.2. We may iterate Eq. (38.1) to find,

$$\begin{aligned} \pi_1(t) &= \lambda_0 e^{-\lambda_1 t} \int_0^t e^{\lambda_1 \tau} \pi_0(\tau)\, d\tau = \lambda_0 e^{-\lambda_1 t} \int_0^t e^{\lambda_1 \tau} e^{-\lambda_0 \tau} d\tau \\ \pi_2(t) &= \lambda_1 e^{-\lambda_2 t} \int_0^t e^{\lambda_2 \sigma} \pi_1(\sigma)\, d\sigma \\ &= \lambda_1 e^{-\lambda_2 t} \int_0^t e^{\lambda_2 \sigma} \left[\lambda_0 e^{-\lambda_1 \sigma} \int_0^\sigma e^{\lambda_1 \tau} e^{-\lambda_0 \tau} d\tau\right] d\sigma \\ &= \lambda_0 \lambda_1 e^{-\lambda_2 t} \int_0^t d\sigma e^{(\lambda_2 - \lambda_1)\sigma} \int_0^\sigma e^{(\lambda_1 - \lambda_0)\tau} d\tau \\ &= \lambda_0 \lambda_1 e^{-\lambda_2 t} \int_{0 \leq \tau \leq \sigma \leq t} e^{(\lambda_2 - \lambda_1)\sigma + (\lambda_1 - \lambda_0)\tau} d\sigma d\tau \end{aligned}$$

and continuing on this way we find,

$$\pi_n(t) = \lambda_0 \lambda_1 \dots \lambda_{n-1} e^{-\lambda_n t} \int_{0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq t} e^{\sum_{j=1}^n (\lambda_j - \lambda_{j-1}) s_j} ds_1 \dots ds_n. \tag{38.2}$$

In the special case where $\lambda_j = \lambda$ for all j, this gives, by Lemma 38.3 below with $f(s) = 1$,

$$\pi_n(t) = \lambda^n e^{-\lambda t} \int_{0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq t} ds_1 \dots ds_n = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \tag{38.3}$$

Another special case of interest is when $\lambda_j = \beta(j+1)$ for all $j \geq 0$. This will be the Yule process discussed below. In this case,

$$\begin{aligned}\pi_n(t) &= n!\beta^n e^{-(n+1)\beta t}\int_{0\le s_1\le s_2\le\dots\le s_n\le t} e^{\beta\sum_{j=1}^n s_j} ds_1\dots ds_n \\ &= n!\beta^n e^{-(n+1)\beta t}\frac{1}{n!}\left(\int_0^t e^{\beta s}ds\right)^n = \beta^n e^{-(n+1)\beta t}\left(\frac{e^{\beta t}-1}{\beta}\right)^n \\ &= e^{-\beta t}\left(1-e^{-\beta t}\right)^n, \end{aligned}\tag{38.4}$$

wherein we have used Lemma 38.3 below for the the second equality.

Lemma 38.3. *Let $f(t)$ be a continuous function, then for all $n\in\mathbb{N}$ we have*

$$\int_{0\le s_1\le s_2\le\dots\le s_n\le t} f(s_1)\dots f(s_n)\,ds_1\dots ds_n = \frac{1}{n!}\left(\int_0^t f(s)\,ds\right)^n.$$

Proof. Let $F(t) := \int_0^t f(s)\,ds$. The proof goes by induction on n. The statement is clearly true when $n=1$ and if it holds at level n, then

$$\begin{aligned}&\int_{0\le s_1\le s_2\le\dots\le s_n\le s_{n+1}\le t} f(s_1)\dots f(s_n)\,f(s_{n+1})\,ds_1\dots ds_n ds_{n+1} \\ &= \int_0^t\left(\int_{0\le s_1\le s_2\le\dots\le s_n\le s_{n+1}} f(s_1)\dots f(s_n)\,ds_1\dots ds_n\right) f(s_{n+1})\,ds_{n+1} \\ &= \int_0^t\left(\frac{1}{n!}\left(F(s_{n+1})\right)^n\right)F'(s_{n+1})\,ds_{n+1} = \int_0^{F(t)}\left(\frac{1}{n!}u^n\right)du \\ &= \frac{F(t)^{n+1}}{(n+1)!}\end{aligned}$$

as required. ■

38.2.2 Yule Process

Suppose that each member of a population gives birth independently to one offspring at an exponential time with rate β. If there are k members of the population with birth times, $T_1,\dots,T_k$, then the time of the birth for this population is $\min(T_1,\dots,T_k) = S_k$ where S_k is now an exponential random variable with parameter, βk. This description gives rise to a pure Birth process with parameters $\lambda_k = \beta k$. In this case we start with initial distribution, $\pi_j(0) = \delta_{j,1}$. We have already solved for $\pi_k(t)$ in this case. Indeed from from Eq. (38.4) after a shift of the index by 1, we find,

$$\pi_n(t) = e^{-\beta t}\left(1-e^{-\beta t}\right)^{n-1} \text{ for } n\ge 1.$$

38.2.3 Sojourn description

Let $\{S_n\}_{n=0}^\infty$ be independent exponential random variables with $P(S_n>t) = e^{-\lambda_n t}$ for all n and let

$$W_k := S_0+\dots+S_{k-1}$$

be the time of the k^{th} – birth, see Figure 38.1 where the graph of $X(t)$ is shown as determined by the sequence $\{S_n\}_{n=0}^\infty$. With this notation we have

$$\begin{aligned}P(X(t)=0) &= P(S_0>t) = e^{-\lambda_0 t} \\ P(X(t)=1) &= P(S_0\le t<S_0+S_1) = P(W_1\le t<W_2) \\ P(X(t)=2) &= P(W_2\le t<W_3) \\ &\ \ \vdots \\ P(X(t)=j) &= P(W_j\le t<W_{j+1})\end{aligned}$$

where $\{W_j\le t<W_{j+1}\}$ represents the event where the j^{th} – birth has occurred by time t but the j^{th} – birth as not. Consider,

$$P(W_1\le t<W_2) = \lambda_0\lambda_1\int_{0\le x_0\le t<x_0+x_1} e^{-\lambda_0 x_0}e^{-\lambda_1 x_1}dx_0dx_1.$$

Doing the x_1 -integral first gives,

$$\begin{aligned}P(X(t)=1) &= P(W_1\le t<W_2) \\ &= \lambda_0\int_{0\le x_0\le t<x_0+x_1} e^{-\lambda_0 x_0}\left[-e^{-\lambda_1 x_1}\right]_{x_1=t-x_0}^\infty dx_0 \\ &= \lambda_0\int_{0\le x_0\le t} e^{-\lambda_0 x_0}e^{-\lambda_1(t-x_0)}dx_0 \\ &= \lambda_0 e^{-\lambda_1 t}\int_{0\le x_0\le t} e^{(\lambda_1-\lambda_0)x_0}dx_0 \\ &= \frac{\lambda_0}{\lambda_1-\lambda_0}e^{-\lambda_1 t}\left[e^{(\lambda_1-\lambda_0)t}-1\right] \\ &= \frac{\lambda_0}{\lambda_1-\lambda_0}\left[e^{-\lambda_0 t}-e^{-\lambda_1 t}\right].\end{aligned}$$

There is one point which we have not yet addressed in this model, namely does it make sense without further information. In terms of the Sojourn description this comes down to the issue as to whether $P\left(\sum_{j=1}^\infty S_j=\infty\right) = 1$. Indeed, if this is not the case, we will only have $X(t)$ defined for $t<\sum_{j=1}^\infty S_j$ which may be less than infinity. The next theorem tells us precisely when this phenomenon can happen.

Theorem 38.4. *Let $\{S_j\}_{j=1}^{\infty}$ be independent random variables such that $S_j \overset{d}{=} \exp(\lambda_j)$ with $0 < \lambda_j < \infty$ for all j. Then:*

1. *If $\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$ then $P\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1$.*
2. *If $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ then $P\left(\sum_{n=1}^{\infty} S_n = \infty\right) = 1$.*

Proof. 1. Since

$$\mathbb{E}\left[\sum_{n=1}^{\infty} S_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[S_n] = \sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$$

it follows that $\sum_{n=1}^{\infty} S_n < \infty$ a.s.

2. By the DCT, independence, and Eq. (**??**),

$$\begin{aligned}
\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} S_n}\right] &= \lim_{N\to\infty} \mathbb{E}\left[e^{-\sum_{n=1}^{N} S_n}\right] = \lim_{N\to\infty} \prod_{n=1}^{N} \mathbb{E}\left[e^{-S_n}\right] \\
&= \lim_{N\to\infty} \prod_{n=1}^{N} \left(\frac{1}{1+\lambda_n^{-1}}\right) = \lim_{N\to\infty} \exp\left(-\sum_{n=1}^{N} \ln\left(1+\lambda_n^{-1}\right)\right) \\
&= \exp\left(-\sum_{n=1}^{\infty} \ln\left(1+\lambda_n^{-1}\right)\right).
\end{aligned}$$

If λ_n does not go to infinity, then the latter sum is infinite and $\lambda_n \to \infty$ and $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ then $\sum_{n=1}^{\infty} \ln\left(1+\lambda_n^{-1}\right) = \infty$ as $\ln\left(1+\lambda_n^{-1}\right) \cong \lambda_n^{-1}$ for large n. In any case we have shown that $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} S_n}\right] = 0$ which can happen iff $e^{-\sum_{n=1}^{\infty} S_n} = 0$ a.s. or equivalently $\sum_{n=1}^{\infty} S_n = \infty$ a.s. ■

Remark 38.5. If $\sum_{k=1}^{\infty} 1/\lambda_k < \infty$ so that $P\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1$, one may define $X(t) = \infty$ on $\left\{t \geq \sum_{n=1}^{\infty} S_n\right\}$. With this definition, $\{X(t)\}_{t\geq 0}$ is again a Markov process. However, most of the examples we study will satisfy $\sum_{k=1}^{\infty} 1/\lambda_k = \infty$.

38.3 Pure Death Process

A pure death process is described by the following rate diagram,

$$0 \underset{\mu_1}{\longleftarrow} 1 \underset{\mu_2}{\longleftarrow} 2 \underset{\mu_3}{\longleftarrow} 3 \dots \underset{\mu_{N-1}}{\longleftarrow} (N-1) \underset{\mu_N}{\longleftarrow} N.$$

If $\pi_j(t) = P(X(t) = j | X(0) = \pi_j(0))$, we have that

$$\begin{aligned}
\dot{\pi}_N(t) &= -\mu_N \pi_N(t) \\
\dot{\pi}_{N-1}(t) &= \mu_N \pi_N(t) - \mu_{N-1}\pi_{N-1}(t) \\
&\vdots \\
\dot{\pi}_n(t) &= \mu_{n+1}\pi_{n+1}(t) - \mu_n \pi_n(t) \\
&\vdots \\
\dot{\pi}_1(t) &= \mu_2 \pi_2(t) - \mu_1 \pi_1(t) \\
\dot{\pi}_0(t) &= -\mu_1 \pi_1(t).
\end{aligned}$$

Let us now suppose that $\pi_j(t) = P(X(t) = j | X(0) = N)$. A little thought shows that we may find $\pi_j(t)$ for $j = 1, 2, \dots, N$ by using the solutions for the pure Birth process with $0 \to N$, $1 \to (N-1)$, $2 \to (N-2)$, …, and $(N-1) \to 1$. We may then compute

$$\pi_0(t) := 1 - \sum_{j=1}^{N} \pi_j(t).$$

The explicit formula for these solutions may be found in the book on p. 346 in the special case where all of the death parameters are distinct.

38.3.1 Cable Failure Model

Suppose that a cable is made up of N individual strands with the life time of each strand being a $\exp(K(l))$ – random variable where $K(l) > 0$ is some function of the load, l, on the strand. We suppose that the cable starts with N – fibers and is put under a total load of NL that L is the load applied per fiber when all N fibers are unbroken. If there are k – fibers in tact, the load per fiber is NL/k and the exponential life time of each fiber is now $K(NL/k)$. Thus when k - fibers are in tact the time to the next fiber breaking is $\exp(kK(NL/k))$. So if $\{S_j\}_{j=N}^{1}$ are the Sojourn times at state j, the time to failure of the cable is $T = \sum_{j=1}^{N} S_j$ and the expected time to failure is

$$\mathbb{E}T = \sum_{j=1}^{N} \mathbb{E}S_j = \sum_{j=1}^{N} \frac{1}{kK(NL/k)} = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{\frac{k}{N}K\left(\frac{N}{k}L\right)} \cong \int_0^1 \frac{1}{xK(L/x)}dx$$

if K is a nice enough function and N is large. For example, if $K(l) = l^{\beta}/A$ for some $\beta > 0$ and $A > 0$, we find

$$\mathbb{E}T = \int_0^1 \frac{A}{x(L/x)^{\beta}}dx = \frac{A}{L^{\beta}} \int_0^1 x^{\beta-1}dx = \frac{A}{L^{\beta}\beta}.$$

Where as the expected life, at the start, of any one strand is $1/K(L) = A/L^{\beta}$. Thus the cable last only $\frac{1}{\beta}$ times the average strand life. It is actually better to let L_0 be the total load applied so that $L = L_0/N$, then the above formula becomes,

$$\mathbb{E}T = \frac{A}{L_0^{\beta}} \frac{N^{\beta}}{\beta}.$$

38.3.2 Linear Death Process basics

Similar to the Yule process, suppose that each individual in a population has a life expectancy, $T \stackrel{d}{=} \exp(\alpha)$. Thus if there are k members in the population at time t, using the memoryless property of the exponential distribution, we the time of the next death is has distribution, $\exp(k\alpha)$. Thus the $\mu_k = \alpha k$ in this case. Using the formula in the book on p. 346, we then learn that if we start with an population of size N, then

$$\begin{aligned}\pi_n(t) &= P(X(t) = n | X(0) = N) \\ &= \binom{N}{n} e^{-n\alpha t}\left(1 - e^{-\alpha t}\right)^{N-n} \text{ for } n = 0, 1, 2, \dots, N. \end{aligned} \tag{38.5}$$

So $\{\pi_n(t)\}_{n=0}^N$ is the binomial distribution with parameter $e^{-\alpha t}$. This may be understood as follows. We have $\{X(t) = n\}$ iff there are exactly n members out of the original N still alive. Let ξ_j be the life time of the j^{th} member of the population, so that $\{\xi_j\}_{j=1}^N$ are i.i.d. $\exp(\mu)$ – distributed random variables. We then have the probability that a particular choice, $A \subset \{1, 2, \dots, N\}$ of n - members are alive with the others being dead is given by

$$P\left(\left(\cap_{j\in A}\{\xi_j > t\}\right) \cap \left(\cap_{j\notin A}\{\xi_j \le t\}\right)\right) = \left(e^{-\alpha t}\right)^n \left(1 - e^{-\alpha t}\right)^{N-n}.$$

As there are $\binom{N}{n}$ – ways to choose such subsets, $A \subset \{1, 2, \dots, N\}$, with n – members, we arrive at Eq. (38.5).

38.3.3 Linear death process in more detail

(You may safely skip this subsection.) In this subsection, we suppose that we start with a population of size N with ξ_j being the life time of the j^{th} member of the population. We assume that $\{\xi_j\}_{j=1}^N$ are i.i.d. $\exp(\mu)$ – distributed random variables and let $X(t)$ denote the number of people alive at time t, i.e.

$$X(t) = \#\{j : \xi_j > t\}.$$

Theorem 38.6. *The process, $\{X(t)\}_{t\ge 0}$ is the linear death Markov process with parameter, α.*

We will begin with the following lemma.

Lemma 38.7. *Suppose that B and $\{A_j\}_{j=1}^n$ are events such that: 1) $\{A_j\}_{j=1}^n$ are pairwise disjoint, 2) $P(A_j) = P(A_1)$ for all j, and 3) $P(B \cap A_j) = P(B \cap A_1)$ for all j. Then*

$$P\left(B | \cup_{j=1}^N A_j\right) = P(B|A_1). \tag{38.6}$$

We also use the identity, that

$$P(B|A\cap C) = P(B|A) \tag{38.7}$$

whenever C is independent of $\{A, B\}$.

Proof. The proof is the following simple computation,

$$\begin{aligned} P\left(B|\cup_{j=1}^n A_j\right) &= \frac{P\left(B \cap \left(\cup_{j=1}^n A_j\right)\right)}{P\left(\cup_{j=1}^n A_j\right)} = \frac{P\left(\cup_{j=1}^n B \cap A_j\right)}{P\left(\cup_{j=1}^n A_j\right)} \\ &= \frac{\sum_{j=1}^n P(B\cap A_j)}{\sum_{j=1}^n P(A_j)} = \frac{nP(B\cap A_1)}{nP(A_1)} = P(B|A_1). \end{aligned}$$

For the second assertion, we have

$$\begin{aligned} P(B|A\cap C) &= \frac{P(B\cap A\cap C)}{P(A\cap C)} = \frac{P(B\cap A)\cdot P(C)}{P(A)\cdot P(C)} \\ &= \frac{P(B\cap A)}{P(A)} = P(B|A). \end{aligned}$$

■

Proof. Sketch of the proof of Theorem 38.6. Let $0 < u < v < t$ and $k \ge l \ge m$ as in Figure 38.2. Given $V \subset U \subset \{1, 2, \dots, N\}$ with $\#V = l$ and

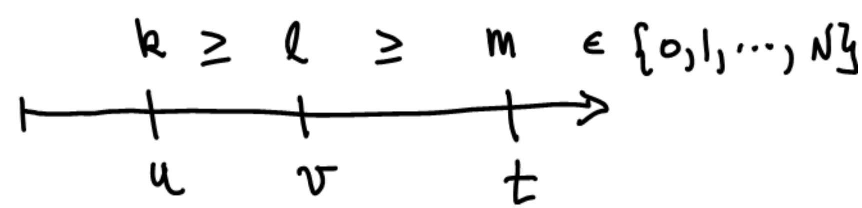


Fig. 38.2. The arangment of times.

$\#U = k$, let

$$A_{U,V} = \cap_{j\in U}\{\xi_j > u\} \cap \cap_{j\notin U}\{\xi_j \le u\} \cap \cap_{j\in V}\{\xi_j > v\} \cap \cap_{j\notin V}\{\xi_j \le v\}$$

so that $\{X_u = k, X_v = l\}$ is the disjoint union of $\{A_{U,V}\}$ over all such choices of $V \subset U$ as above. Notice that $P(A_{U,V})$ is independent of how $U \subset V$ as is $P(\{X_t = m\} \cap A_{U,V})$. Therefore by Lemma 38.7, we have, with $V = \{1, 2, \ldots, l\} \subset U = \{1, 2, \ldots, k\}$, that

$$\begin{aligned}
&P(X_t = m | X_u = k, X_v = l) = P(X_t = m | A_{U,V}) \\
&\quad = P(\text{Exactly } m \text{ of } \xi_1, \ldots, \xi_l \ > t | \xi_1 > v, \ldots, \xi_l > v, v \ge \xi_{l+1} > u, \ldots, v \ge \xi_k > u) \\
&\quad = P(\text{Exactly } m \text{ of } \xi_1, \ldots, \xi_l \ > t | \xi_1 > v, \ldots, \xi_l > v) \\
&\quad = \binom{l}{m} P(\xi_1 > t, \ldots, \xi_m > t, \xi_{m+1} \le t, \ldots \xi_l \le t | \xi_1 > v, \ldots, \xi_l > v) \\
&\quad = \binom{l}{m} \frac{P(\xi_1 > t)^m \cdot P(v < \xi_1 \le t)^{l-m}}{P(v < \xi_1)^l} \\
&\quad = \binom{l}{m} \frac{(e^{-\alpha t})^m \cdot (e^{-vt} - e^{-\alpha t})^{l-m}}{e^{-\alpha v l}} = \binom{l}{m} e^{-\alpha m(t-v)} \left(1 - e^{-\alpha(t-v)}\right)^{l-m}.
\end{aligned}$$

Similar considerations show that X_t has the Markov property and we have just found the transition matrix for this process to be,

$$P(X_t = m | X_v = l) = 1_{l \ge m} \binom{l}{m} e^{-\alpha m(t-v)} \left(1 - e^{-\alpha(t-v)}\right)^{l-m}.$$

So

$$P_{lm}(t) := P(X_t = m | X_0 = l) = 1_{\ge m} \binom{l}{m} e^{-\alpha m t} \left(1 - e^{-\alpha t}\right)^{l-m}.$$

Differentiating this equation at $t = 0$ implies $\frac{d}{dt}|_{0+} P_{lm}(t) = 0$ unless $m = l$ or $m = l - 1$ and

$$\begin{aligned}
\frac{d}{dt}|_{0+} P_{l\,l}(t) &= -\alpha l \text{ and} \\
\frac{d}{dt}|_{0+} P_{l\,,l-1}(t) &= \binom{l}{l-1}\alpha = \alpha l.
\end{aligned}$$

These are precisely the transition rate of the linear death process with parameter α. ■

Let us now also work out the Sojourn description in this model.

Theorem 38.8. *Suppose that $\{\xi_j\}_{j=1}^N$ are independent exponential random variables with parameter, α as in the above model for the life times of a population. Let $W_1 < W_2 < \cdots < W_N$ be the order statistics of $\{\xi_j\}_{j=1}^N$, i.e. $\{W_1 < W_2 < \cdots < W_N\} = \{\xi_j\}_{j=1}^N$. Hence W_j is the time of the j^{th} – death. Further let $S_1 = W_1$, $S_2 = W_2 - W_1, \ldots, S_N = W_N - W_{N-1}$ are times between successive deaths. Then $\{S_j\}_{j=1}^N$ are exponential random variables with $S_j \overset{d}{=} \exp((N-j)\alpha)$.*

Proof. Since $W_1 = S_1 = \min(\xi_1, \ldots, \xi_N)$, by a homework problem, $S_1 \overset{d}{=} \exp(N\alpha)$. Let

$$A_j := \left\{\xi_j < \min(\xi_k)_{k\neq j}\right\} \cap \{\xi_j = t\}.$$

We then have

$$\{W_1 = t\} = \cup_{j=1}^N A_j$$

and

$$A_j \cap \{W_2 > s + t\} = \left\{s + t < \min(\xi_k)_{k\neq j}\right\} \cap \{\xi_j = t\}.$$

By symmetry we have (this is the informal part)

$$\begin{gathered}
P(A_j) = P(A_1) \text{ and} \\
P(A_j \cap \{W_2 > s+t\}) = P(A_1 \cap \{W_2 > s+t\}),
\end{gathered}$$

and hence by Lemma 38.7,

$$P(W_2 > s + t | W_1 = t) = P$$

Now consider

$$\begin{aligned}
W_2 &= P(A_1 \cap \{W_2 > s+t\} | A_1) \\
&= P\left(\{\xi_1 = t\} \cap \left\{\min(\xi_k)_{k\neq 1} > s + t\right\} | \min(\xi_k)_{k\neq 1} > \xi_1 = t\right) \\
&= \frac{P\left(\min(\xi_k)_{k\neq 1} > s+t,\ \xi_1 = t\right)}{P\left(\min(\xi_k)_{k\neq 1} > t,\ \xi_1 = t\right)} \\
&= \frac{P\left(\min(\xi_k)_{k\neq 1} > s+t\right)}{P\left(\min(\xi_k)_{k\neq 1} > t\right)} = e^{-(N-1)\alpha s}
\end{aligned}$$

since $\min(\xi_k)_{k\neq 1} \overset{d}{=} \exp((N-1)\alpha)$ and the memoryless property of exponential random variables. This shows that $S_2 := W_2 - W_1 \overset{d}{=} \exp((N-1)\alpha)$.

Let us consider the next case, namely $P(W_3 - W_2 > t | W_1 = a, W_2 = a + b)$. In this case we argue as above that

$$\begin{aligned}
&P\left(W_3 - W_2 > t | W_1 = a, W_2 = a + b\right) \\
&= P\left(\min\left(\xi_3, \dots, \xi_N\right) - \xi_2 > t | \xi_1 = a,\ \xi_2 = a + b,\ \ \min\left(\xi_3, \dots, \xi_N\right) > \xi_2\right) \\
&= \frac{P\left(\min\left(\xi_3, \dots, \xi_N\right) > t + a + b,\ \xi_1 = a,\ \xi_2 = a + b,\ \ \min\left(\xi_3, \dots, \xi_N\right) > \xi_2\right)}{P\left(\xi_1 = a,\ \xi_2 = a + b,\ \ \min\left(\xi_3, \dots, \xi_N\right) > a + b\right)} \\
&= \frac{P\left(\min\left(\xi_3, \dots, \xi_N\right) > t + a + b\right)}{P\left(\min\left(\xi_3, \dots, \xi_N\right) > a + b\right)} = e^{-(N-2)\alpha t}.
\end{aligned}$$

We continue on this way to get the result. This proof is not rigorous, since $P\left(\xi_j = t\right) = 0$ but the spirit is correct.

Rigorous Proof. (Probably should be skipped.) In this proof, let g be a bounded function and $T_k := \min\left(\xi_l : l \neq k\right)$. We then have that T_k and ξ_k are independent, $T_k \overset{d}{=} \exp\left(\left(N-1\right)\alpha\right)$, and hence

$$\begin{aligned}
\mathbb{E}\left[1_{W_2 - W_1 > t}\ g\left(W_1\right)\right] &= \sum_k \mathbb{E}\left[1_{W_2 - W_1 > t}\ g\left(W_1\right) : \xi_k < T_k\right] \\
&= \sum_k \mathbb{E}\left[1_{T_k - \xi_k > t}\ g\left(\xi_k\right) : \xi_k < T_k\right] \\
&= \sum_k \mathbb{E}\left[1_{T_k - \xi_k > t}\ g\left(\xi_k\right)\right] \\
&= \sum_k \mathbb{E}\left[\exp\left(-\left(N-1\right)\alpha\left(t + \xi_k\right)\right) g\left(\xi_k\right)\right] \\
&= \exp\left(-\left(N-1\right)\alpha t\right) \sum_k \mathbb{E}\left[\exp\left(-\left(N-1\right)\alpha\xi_k\right) g\left(\xi_k\right)\right] \\
&= \exp\left(-\left(N-1\right)\alpha t\right) \sum_k \mathbb{E}\left[1_{T_k - \xi_k > 0}\ g\left(\xi_k\right)\right] \\
&= \exp\left(-\left(N-1\right)\alpha t\right) \sum_k \mathbb{E}\left[1_{T_k - \xi_k > 0}\ g\left(W_1\right)\right] \\
&= \exp\left(-\left(N-1\right)\alpha t\right) \cdot \mathbb{E}\left[g\left(W_1\right)\right].
\end{aligned}$$

It follows from this calculation that $W_2 - W_1$ and W_1 are independent, $W_2 - W_1 = \exp\left(\alpha\left(N-1\right)\right)$.

The general case may be done similarly. To see how this goes, let us show that $W_3 - W_2 \overset{d}{=} \exp\left((N-2)\alpha\right)$ and is independent of W_1 and W_2. To this end, let $T_{jk} := \min\left\{\xi_l : l \neq j \text{ or } k\right\}$ for $j \neq k$ in which case $T_{jk} \overset{d}{=} \exp\left(\left(N-2\right)\alpha\right)$ and is independent of $\left\{\xi_j, \xi_k\right\}$. We then have

$$\begin{aligned}
\mathbb{E}\left[1_{W_3 - W_2 > t}\ g\left(W_1, W_2\right)\right] &= \sum_{j \neq k} \mathbb{E}\left[1_{W_3 - W_2 > t}\ g\left(W_1, W_2\right) : \xi_j < \xi_k < T_{jk}\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > t}\ g\left(\xi_j, \xi_k\right) : \xi_j < \xi_k < T_{jk}\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > t}\ g\left(\xi_j, \xi_k\right) : \xi_j < \xi_k\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[\exp\left(-\left(N-2\right)\alpha\left(t + \xi_k\right)\right) g\left(\xi_j, \xi_k\right) : \xi_j < \xi_k\right] \\
&= \exp\left(-\left(N-2\right)\alpha t\right) \sum_{j \neq k} \mathbb{E}\left[\exp\left(-\left(N-2\right)\alpha\xi_k\right) g\left(\xi_j, \xi_k\right) : \xi_j <\right. \\
&= \exp\left(-\left(N-2\right)\alpha t\right) \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > 0}\ g\left(\xi_j, \xi_k\right) : \xi_j < \xi_k\right] \\
&= \exp\left(-\left(N-2\right)\alpha t\right) \sum_{j \neq k} \mathbb{E}\left[g\left(W_1, W_2\right) : \xi_j < \xi_k < T_{jk}\right] \\
&= \exp\left(-\left(N-2\right)\alpha t\right) \cdot \mathbb{E}\left[g\left(W_1, W_2\right)\right].
\end{aligned}$$

This again shows that $W_3 - W_2$ is independent of $\left\{W_1, W_2\right\}$ and $W_3 - W_2 \overset{d}{=} \exp\left((N-2)\alpha\right)$. We leave the general argument to the reader. ■

38.4 Birth and Death Processes

We have already discussed the basics of the Birth and death processes. To have the existence of the process requires some restrictions on the Birth and Death parameters which are discussed on p. 359 of the book. In general, we are not able to find solve for the transition semi-group, e^{tQ}, in this case. We will therefore have to ask more limited questions about more limited models. This is what we will consider in the rest of this section. We will also consider some interesting situations which one might model by a Birth and Death process.

Recall that the functions, $\pi_j\left(t\right) = P\left(X\left(t\right) = j\right)$, satisfy the differential equations

$$\begin{aligned}
\dot{\pi}_0\left(t\right) &= -\lambda_0\pi_0\left(t\right) + \mu_1\pi_1\left(t\right) \\
\dot{\pi}_1\left(t\right) &= \lambda_0\pi_0\left(t\right) - \left(\lambda_1 + \mu_1\right)\pi_1\left(t\right) + \mu_2\pi_2\left(t\right) \\
\dot{\pi}_2\left(t\right) &= \lambda_1\pi_1\left(t\right) - \left(\lambda_2 + \mu_2\right)\pi_2\left(t\right) + \mu_3\pi_3\left(t\right) \\
&\vdots \\
\dot{\pi}_n\left(t\right) &= \lambda_{n-1}\pi_{n-1}\left(t\right) - \left(\lambda_n + \mu_n\right)\pi_n\left(t\right) + \mu_{n+1}\pi_{n+1}\left(t\right). \\
&\vdots
\end{aligned}$$

Hence if are going to look for a stationary distribution, we must set $\dot{\pi}_j(t) = 0$ for all t and solve the system of algebraic equations:

$$\begin{aligned}
0 &= -\lambda_0\pi_0 + \mu_1\pi_1 \\
0 &= \lambda_0\pi_0 - (\lambda_1 + \mu_1)\pi_1 + \mu_2\pi_2 \\
0 &= \lambda_1\pi_1 - (\lambda_2 + \mu_2)\pi_2 + \mu_3\pi_3 \\
&\vdots \\
0 &= \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1}. \\
&\vdots
\end{aligned}$$

We solve these equations in order to find,

$$\begin{aligned}
\pi_1 &= \frac{\lambda_0}{\mu_1}\pi_0, \\
\pi_2 &= \frac{\lambda_1 + \mu_1}{\mu_2}\pi_1 - \frac{\lambda_0}{\mu_2}\pi_0 = \frac{\lambda_1 + \mu_1}{\mu_2}\frac{\lambda_0}{\mu_1}\pi_0 - \frac{\lambda_0}{\mu_2}\pi_0 \\
&= \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0 \\
\pi_3 &= \frac{\lambda_2 + \mu_2}{\mu_3}\pi_2 - \frac{\lambda_1}{\mu_3}\pi_1 = \frac{\lambda_2 + \mu_2}{\mu_3}\frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0 - \frac{\lambda_1}{\mu_3}\frac{\lambda_0}{\mu_1}\pi_0 \\
&= \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}\pi_0 \\
&\vdots \\
\pi_n &= \frac{\lambda_0\lambda_1\lambda_2\dots\lambda_{n-1}}{\mu_1\mu_2\mu_3\dots\mu_n}\pi_0.
\end{aligned}$$

This leads to the following proposition.

Proposition 38.9. *Let $\theta_n := \frac{\lambda_0\lambda_1\lambda_2\dots\lambda_{n-1}}{\mu_1\mu_2\mu_3\dots\mu_n}$ for $n = 1, 2, \dots$ and $\theta_0 := 1$. Then the birth and death process, $\{X(t)\}$ with birth rates $\{\lambda_j\}_{j=0}^\infty$ and death rates $\{\mu_j\}_{j=1}^\infty$ has a stationary distribution, π, iff $\Theta := \sum_{n=0}^\infty \theta_n < \infty$ in which case,*

$$\pi_n = \frac{\theta_n}{\Theta} \text{ for all } n.$$

Lemma 38.10 (Detail balance). *In general, if we can find a distribution, π, satisfying the* ***detail balance equation,***

$$\pi_i Q_{ij} = \pi_j Q_{ji} \text{ for all } i \neq j, \tag{38.8}$$

then π is a stationary distribution, i.e. $\pi Q = 0$.

Proof. First proof. Intuitively, Eq. (38.8) states that sites i and j are always exchanging sand back and forth at equal rates. Hence if all sites are doing this the size of the piles of sand at each site must remain unchanged.

Second Proof. Summing Eq. (38.8) on i making use of the fact that $\sum_i Q_{ji} = 0$ for all j implies, $\sum_i \pi_i Q_{ij} = 0$. ■

We could have used this result on our birth death processes to find the stationary distribution as well. Indeed, looking at the rate diagram,

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2 \underset{\mu_3}{\overset{\lambda_2}{\rightleftarrows}} 3 \dots \underset{\mu_{n-1}}{\overset{\lambda_{n-2}}{\rightleftarrows}} (n-1) \underset{\mu_n}{\overset{\lambda_{n-1}}{\rightleftarrows}} n \underset{\mu_{n+1}}{\overset{\lambda_n}{\rightleftarrows}} (n+1),$$

we see the conditions for detail balance between n and $n = 1$ are,

$$\pi_n\lambda_n = \pi_{n+1}\mu_{n+1}$$

which implies $\frac{\pi_{n+1}}{\pi_n} = \frac{\lambda_n}{\mu_{n+1}}$. Therefore it follows that,

$$\begin{aligned}
\frac{\pi_1}{\pi_0} &= \frac{\lambda_0}{\mu_1}, \\
\frac{\pi_2}{\pi_0} &= \frac{\pi_2}{\pi_1}\frac{\pi_1}{\pi_0} = \frac{\lambda_1}{\mu_2}\frac{\lambda_0}{\mu_1}, \\
&\vdots \\
\frac{\pi_n}{\pi_0} &= \frac{\pi_n}{\pi_{n-1}}\frac{\pi_{n-1}}{\pi_{n-2}}\dots\frac{\pi_1}{\pi_0} = \frac{\lambda_{n-1}}{\mu_n}\dots\frac{\lambda_1}{\mu_2}\frac{\lambda_0}{\mu_1} \\
&= \frac{\lambda_0\lambda_1\dots\lambda_{n-1}}{\mu_1\mu_2\dots\mu_n} = \theta_n
\end{aligned}$$

as before.

Lemma 38.11. *For $|x| < 1$ and $\alpha \in \mathbb{R}$ we have,*

$$(1-x)^{-\alpha} = \sum_{k=0}^\infty \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{k!}x^k, \tag{38.9}$$

where $\frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{k!} := 1$ when $k = 0$.

Proof. This is a consequence of Taylor's theorem with integral remainder. The main point is to observe that

$$\frac{d}{dx}(1-x)^{-\alpha} = \alpha(1-x)^{-(\alpha+1)}$$
$$\left(\frac{d}{dx}\right)^2 (1-x)^{-\alpha} = \alpha(\alpha+1)(1-x)^{-(\alpha+2)}$$
$$\vdots$$
$$\left(\frac{d}{dx}\right)^k (1-x)^{-\alpha} = \alpha(\alpha+1)\dots(\alpha+k-1)(1-x)^{-(\alpha+k)}$$
$$\vdots$$

and hence,

$$\left(\frac{d}{dx}\right)^k (1-x)^{-\alpha}|_{x=0} = \alpha(\alpha+1)\dots(\alpha+k-1). \tag{38.10}$$

Therefore by Taylor's theorem,

$$(1-x)^{-\alpha} = \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{d}{dx}\right)^k (1-x)^{-\alpha}|_{x=0}\cdot x^k$$

which combined with Eq. (38.10) gives Eq. (38.9). ∎

Example 38.12 (Exercise 4.5 on p. 377). Suppose that $\lambda_n = \theta < 1$ and $\mu_n = \frac{n}{n+1}$. In this case,

$$\theta_n = \frac{\theta^n}{\frac{1}{2}\frac{2}{3}\cdots\frac{n}{n+1}} = (n+1)\theta^n$$

and we must have,

$$\pi_n = \frac{(n+1)\theta^n}{\sum_{n=0}^{\infty}(n+1)\theta^n}.$$

We can simplify this answer a bit by noticing that

$$\sum_{n=0}^{\infty}(n+1)\theta^n = \frac{d}{d\theta}\sum_{n=0}^{\infty}\theta^{n+1} = \frac{d}{d\theta}\frac{\theta}{1-\theta} = \frac{(1-\theta)+\theta}{(1-\theta)^2} = \frac{1}{(1-\theta)^2}.$$

(Alternatively, apply Lemma 38.11 with $\alpha = 2$ and $x = \theta$.)Thus we have,

$$\pi_n = (1-\theta)^2(n+1)\theta^n.$$

Example 38.13 (Exercise 4.4 on p. 377). Two machines operate with failure rate μ and there is a repair facility which can repair one machine at a time with rate λ. Let $X(t)$ be the number of operational machines at time t. The state space is thus, $\{0,1,2\}$ with the transition diagram,

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2$$

where $\lambda_0 = \lambda$, $\lambda_1 = \lambda$, $\mu_2 = 2\mu$ and $\mu_1 = \mu$. Thus we find,

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0 = \frac{\lambda}{\mu}\pi_0$$
$$\pi_2 = \frac{\lambda^2}{2\mu^2}\pi_0 = \frac{1}{2}\frac{\lambda^2}{\mu^2}\pi_0.$$

so that

$$1 = \pi_0+\pi_1+\pi_2 = \left(1+\frac{\lambda}{\mu}+\frac{1}{2}\frac{\lambda^2}{\mu^2}\right)\pi_0.$$

So the long run probability that all machines are broken is given by

$$\pi_0 = \left(1+\frac{\lambda}{\mu}+\frac{1}{2}\frac{\lambda^2}{\mu^2}\right)^{-1}.$$

If we now suppose that only one machine can be in operation at a time (perhaps there is only one plug), the new rates become, $\lambda_0 = \lambda$, $\lambda_1 = \lambda$, $\mu_2 = \mu$ and $\mu_1 = \mu$ and working as above we have:

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0 = \frac{\lambda}{\mu}\pi_0$$
$$\pi_2 = \frac{\lambda^2}{\mu^2}\pi_0 = \frac{\lambda^2}{\mu^2}\pi_0.$$

so that

$$1 = \pi_0+\pi_1+\pi_2 = \left(1+\frac{\lambda}{\mu}+\frac{\lambda^2}{\mu^2}\right)\pi_0.$$

So the long run probability that all machines are broken is given by

$$\pi_0 = \left(1+\frac{\lambda}{\mu}+\frac{\lambda^2}{\mu^2}\right)^{-1}.$$

Example 38.14 (Problem VI.4.7, p. 379). A system consists of 3 machines and 2 repairmen. At most 2 machines can operate at any time. The amount of time that an operating machine works before breaking down is exponentially distributed with mean 5 hours. The amount of time that it takes a single repairman to fix a machine is exponentially distributed with mean 4 hours. Only one repairman can work on a failed machine at any given time. Let $X(t)$ be the number of machines in operating condition at time t.

a) Calculate the long run probability distribution of $X(t)$.

b) If an operating machine produces 100 units of output per hour, what is the long run output per hour from the factory.

Solution to Exercise (Problem VI.4.7, p. 379).
The state space of operating machines is $S = \{0, 1, 2, 3\}$ and the system is modeled by a birth death process with rate diagram,

$$0 \underset{1/5}{\overset{2/4}{\rightleftarrows}} 1 \underset{2/5}{\overset{2/4}{\rightleftarrows}} 2 \underset{2/5}{\overset{1/4}{\rightleftarrows}} 3.$$

a) We then have $\theta_0 = 1$,

$$\theta_1 = \frac{1/2}{1/5} = \frac{5}{2}$$
$$\theta_2 = \frac{1/2}{1/5}\frac{1/2}{2/5} = \frac{5^2}{2^3}$$
$$\theta_3 = \frac{1/2}{1/5}\frac{1/2}{2/5}\frac{1/4}{2/5} = \frac{5^3}{2^4}\frac{1}{4} = \frac{5^3}{2^6}$$

and

$$\Theta = \sum_{j=0}^{3} \theta_i = 1 + \frac{5}{2} + \frac{5^2}{2^3} + \frac{5^3}{2^6} = \frac{549}{64}.$$

: $\frac{549}{64}$ Therefore $\pi_i = \theta_i / \Theta$ gives,

$$(\pi_0, \pi_1, \pi_2, \pi_3) = \frac{64}{549}\left(1, \frac{5}{2}, \frac{5^2}{2^3}, \frac{5^3}{2^6}\right)$$
$$= \left(\tfrac{64}{549}\ \tfrac{160}{549}\ \tfrac{200}{549}\ \tfrac{125}{549} \right)$$
$$= \left(0.116\,58\ 0.291\,44\ 0.364\,30\ 0.227\,69 \right).$$

b) If the operating machines can produce 100 units per hour, the long run output per hour is,

$$100 \cdot \pi_1 + 200\,(\pi_2 + \pi_3) = 100 \cdot 0.291\,44 + 200\,(0.364\,30 + 0.227\,69) \cong 147.54 \text{ /hour}.$$

Solution to Exercise (Problem VI.4.7, p. 379 but only one repair person.). Here is the same problems with only one repair person. The state space of operating machines is $S = \{0, 1, 2, 3\}$ and the system is modeled by a birth death process with rate diagram,

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2 \underset{\mu_3}{\overset{\lambda_2}{\rightleftarrows}} 3$$

where, $\lambda_0 = \lambda_1 = \lambda_2 = 1/4$ and $\mu_1 = 1/5$, $\mu_2 = \mu_3 = 2/5$, so the rate diagram is,

$$0 \underset{1/5}{\overset{1/4}{\rightleftarrows}} 1 \underset{2/5}{\overset{1/4}{\rightleftarrows}} 2 \underset{2/5}{\overset{1/4}{\rightleftarrows}} 3.$$

a) We then have $\theta_0 = 1$,

$$\theta_1 = \frac{1/4}{1/5} = \frac{5}{4}$$
$$\theta_2 = \frac{1/4}{1/5}\frac{1/4}{2/5} = \frac{5^2}{2}\frac{1}{4^2}$$
$$\theta_3 = \frac{1/4}{1/5}\frac{1/4}{2/5}\frac{1/4}{2/5} = \frac{5^3}{2^2}\frac{1}{4^3}$$

and

$$\Theta = \sum_{j=0}^{3} \theta_i = 1 + \frac{5}{4} + \frac{5^2}{2}\frac{1}{4^2} + \frac{5^3}{2^2}\frac{1}{4^3} = \frac{901}{256}.$$

Therefore $\pi_i = \theta_i / \Theta$ gives,

$$(\pi_0, \pi_1, \pi_2, \pi_3) = \frac{256}{901}\left(1, \frac{5}{4}, \frac{5^2}{2}\frac{1}{4^2}, \frac{5^3}{2^2}\frac{1}{4^3}\right)$$
$$= \left(\tfrac{256}{901}\ \tfrac{320}{901}\ \tfrac{200}{901}\ \tfrac{125}{901} \right)$$
$$= \left(0.284\ 0.355\ 0.222\ 0.139 \right).$$

b) If the operating machines can produce 100 units per hour, the long run output per hour is,

$$100 \cdot \pi_1 + 200\,(\pi_2 + \pi_3) = 100 \cdot 0.355 + 200\,(0.222 + 0.139) \cong 108 \text{ /hour}.$$

Example 38.15 (Telephone Exchange). Consider as telephone exchange consisting of K out going lines. The mean call time is $1/\mu$ and new call requests arrive at the exchange at rate λ. If all lines are occupied, the call is lost. Let $X(t)$ be the number of outgoing lines which are in service at time t – see Figure 38.3. We model this as a birth death process with state space, $\{0, 1, 2, \dots, K\}$ and birth parameters, $\lambda_k = \lambda$ for $k = 0, 1, 2, \dots, K-1$ and death rates, $\mu_k = k\mu$ for $k = 1, 2, \dots, K$, see Figure 38.4. In this case,

$$\theta = 1,\ \theta_1 = \frac{\lambda}{\mu},\ \theta_2 = \frac{\lambda^2}{2\mu^2},\ \theta_3 = \frac{\lambda^3}{3! \cdot \mu^3}, \dots, \theta_K = \frac{\lambda^K}{K!\mu^K}$$

so that

$$\Theta := \sum_{k=0}^{K} \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \cong e^{\lambda/\mu} \text{ for large } K.$$

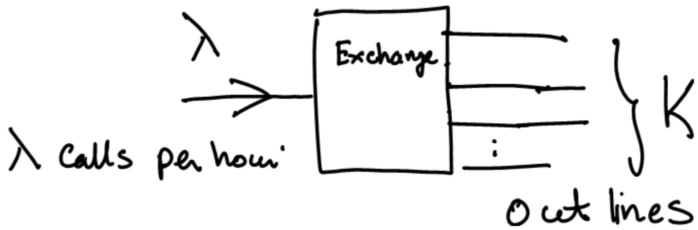


Fig. 38.3. Schematic of a telephone exchange.

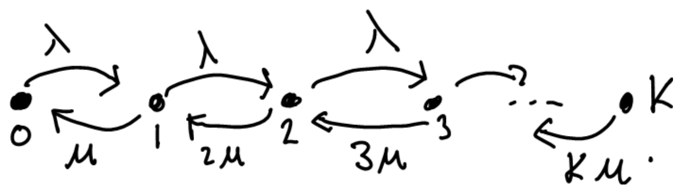


Fig. 38.4. Rate diagram for the telephone exchange.

and hence

$$\pi_k = \Theta^{-1}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \cong \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k e^{-\lambda/\mu}.$$

For example, suppose $\lambda = 100$ calls / hour and average duration of a connected call is $1/4$ of an hour, i.e. $\mu = 4$. Then we have

$$\pi_{25} = \frac{\frac{1}{25!}(25)^{25}}{\sum_{k=0}^{25}\frac{1}{k!}(25)^k} \cong 0.144.$$

so the exchange is busy 14.4% of the time. On the other hand if there are 30 or even 35 lines, then we have,

$$\pi_{30} = \frac{\frac{1}{30!}(25)^{30}}{\sum_{k=0}^{30}\frac{1}{k!}(25)^k} \cong 0.053$$

and

$$\pi_{35} = \frac{\frac{1}{35!}(25)^{35}}{\sum_{k=0}^{35}\frac{1}{k!}(25)^k} \cong .012$$

and hence the exchange is busy 5.3% and 1.2% respectively.

38.4.1 Linear birth and death process with immigration

Suppose now that $\lambda_n = n\lambda + a$ and $\mu_n = n\mu$ for some $\lambda, \mu > 0$ where λ and μ represent the birth rates and deaths of each individual in the population and a represents the rate of migration into the population. In this case,

$$\begin{aligned}\theta_n &= \frac{a(a+\lambda)(a+2\lambda)\dots(a+(n-1)\lambda)}{n!\mu^n} \\ &= \left(\frac{\lambda}{\mu}\right)^n \frac{\frac{a}{\lambda}\left(\frac{a}{\lambda}+1\right)\left(\frac{a}{\lambda}+2\right)\dots\left(\frac{a}{\lambda}+(n-1)\right)}{n!}.\end{aligned}$$

Using Lemma 38.11 with $\alpha = a/\lambda$ and $x = \lambda/\mu$ which we need to assume is less than 1, we find

$$\Theta := \sum_{n=0}^{\infty}\theta_n = \left(1-\frac{\lambda}{\mu}\right)^{-a/\lambda}$$

and therefore,

$$\pi_n = \left(1-\frac{\lambda}{\mu}\right)^{a/\lambda} \frac{\frac{a}{\lambda}\left(\frac{a}{\lambda}+1\right)\left(\frac{a}{\lambda}+2\right)\dots\left(\frac{a}{\lambda}+(n-1)\right)}{n!}\left(\frac{\lambda}{\mu}\right)^n$$

In this case there is an invariant distribution iff $\lambda < \mu$ and $a > 0$. Notice that if $a = 0$, then 0 is an absorbing state so when $\lambda < \mu$, the process actually dies out.

(STOP) Bruce think about this, perhaps we need to start at π_1 instead of π_0. Or we take $\pi_0 = 1$ would be an invariant distribution when $a = 0$. The class $\{1,2,3,\dots$ is now not closed and hence transient!) When $a = 0$ we should compute the hitting probability of site 0!

Now that we have found the stationary distribution in this case, let us try to compute the expected population of this model at time t.

Theorem 38.16. *If*

$$M(t) := \mathbb{E}[X(t)] = \sum_{n=1}^{\infty} nP(X(t)=n) = \sum_{n=1}^{\infty} n\pi_n(t)$$

be the expected population size for our linear birth and death process with immigration, then

$$M(t) = \frac{a}{\lambda-\mu}\left(e^{t(\lambda-\mu)}-1\right) + M(0)e^{t(\lambda-\mu)}$$

which when $\lambda = \mu$ should be interpreted as

$$M(t) = at + M(0).$$

Proof. In this proof we take for granted the fact that it is permissible to interchange the time derivative with the infinite sum. Assuming this fact we find,

$$
\begin{aligned}
\dot{M}(t) &= \sum_{n=1}^{\infty} n\dot{\pi}_n(t) \\
&= \sum_{n=1}^{\infty} n \begin{pmatrix} (a+\lambda(n-1))\,\pi_{n-1}(t) \\ -(a+\lambda n+\mu n)\,\pi_n(t) + \mu(n+1)\,\pi_{n+1}(t) \end{pmatrix} \\
&= \sum_{n=1}^{\infty} n\,(a+\lambda(n-1))\,\pi_{n-1}(t) \\
&\quad - \sum_{n=1}^{\infty} n\,(a+\lambda n+\mu n)\,\pi_n(t) + \sum_{n=1}^{\infty} \mu n\,(n+1)\,\pi_{n+1}(t) \\
&= \sum_{n=0}^{\infty} (n+1)\,(a+\lambda n)\,\pi_n(t) \\
&\quad - \sum_{n=1}^{\infty} n\,(a+\lambda n+\mu n)\,\pi_n(t) + \sum_{n=2}^{\infty} \mu\,(n-1)\,n\pi_n(t) \\
&= a\pi_0(t) + [2(a+\lambda) - (a+\lambda+\mu)]\,\pi_1(t) \\
&\quad + \sum_{n=2}^{\infty} [(n+1)(a+\lambda n) + \mu(n-1)\,n - n\,(a+\lambda n+\mu n)]\,\pi_n(t) \\
&= a\pi_0(t) + [a+\lambda-\mu]\,\pi_1(t) + \sum_{n=2}^{\infty} [(a+\lambda n) - \mu n]\,\pi_n(t) \\
&= a\pi_0(t) + \sum_{n=1}^{\infty} [a+\lambda n-\mu n]\,\pi_n(t) \\
&= \sum_{n=0}^{\infty} [a+\lambda n-\mu n]\,\pi_n(t) = a + (\lambda-\mu)\,M(t)\,.
\end{aligned}
$$

Thus we have shown that

$$
\dot{M}(t) = a + (\lambda-\mu)\,M(t) \;\text{ with }\; M(0) = \sum_{n=1}^{\infty} n\pi_n(0)\,,
$$

where $M(0)$ is the mean size of the initial population. Solving this simple differential equation gives the results. ■

Part VIII

Banach Spaces and Probability

Throughout this part of the book we will assume (unless otherwise noted) that X is a separable real Banach space, $\mathcal{B} = \mathcal{B}(X) = \mathcal{B}_X$ is the Borel σ–algebra, and X^* is the continuous dual of X. We begin with a discussion of the properties of $(X, \mathcal{B}_X)$ and then go on to the Bochner integral and abstract Wiener spaces.

Basic Properties of $(X, \mathcal{B}_X)$

Proposition 39.1. *The Borel σ – algebra of X and the σ – algebra generated by $\varphi \in X^*$ are the same, i.e. $\sigma(X^*) = \mathcal{B}$.*

Proof. If $\varphi \in X^*$, then $\varphi : X \to \mathbb{R}$ is continuous and hence Borel measurable. Therefore $\sigma(X^*) \subset \mathcal{B}$. For the converse. Choose $x_n \in X$ such that $\|x_n\| = 1$ for all n and

$$\overline{\{x_n\}} = S = \{x \in X : \|X\| = 1\}.$$

By the Hahn Banach theorem, there exists $\varphi_n \in X^*$ such that i) $\varphi_n(x_n) = 1$ and ii) $\|\varphi_n\|_{X^*} = 1$ for all n. For $x \in X$, $|\varphi_n(x)| \le \|x\|$ for all n and hence $\sup_n |\varphi_n(x)| \le \|x\|$. Conversely, we may choose an increasing sequence $\{n_k\}_{k=1}^\infty \subset \mathbb{N}$ such that $x = \lim_{k\to\infty} \|x\| x_{n_k}$ in which case

$$|\varphi_{n_k}(x) - \varphi_{n_k}(\|x\| x_{n_k})| \le \|x - \|x\| x_{n_k}\| \to 0 \text{ as } k \to \infty.$$

Hence $\lim_{k\to\infty} |\varphi_{n_k}(x)| = \|x\|$ and we have shown

$$\|x\| = \sup_n |\varphi_n(x)| \text{ for all } x \in X. \tag{39.1}$$

Therefore

$$\|\cdot - x_0\| = \sup_n |\varphi_n(\cdot - x_0)| = \sup_n |\varphi_n(\cdot) - \varphi_n(x_0)|$$

is $\sigma(X^*)$–measurable for each $x_0 \in X$ and hence

$$\{x : \|x - x_0\| < \delta\} \in \sigma(X^*).$$

Hence $\sigma(X^*)$ contains all open balls in X and by separability all open sets of X which implies $\mathcal{B}(X) \subset \sigma(X^*)$. ∎

Corollary 39.2. *Suppose that $(\Omega, \mathcal{F}, \mu)$ is a measure space and $F, G : \Omega \to X$ are $\mathcal{F}/\mathcal{B}(X)$ – measurable functions. Then $F(\omega) = G(\omega)$ for μ – a.e. $\omega \in \Omega$ iff $\varphi \circ F(\omega) = \varphi \circ G(\omega)$ for μ – a.e. $\omega \in \Omega$ and every $\varphi \in X^*$.*

Proof. The direction, " $\Longrightarrow$ ", is clear. For the converse direction let $\{\varphi_n\} \subset X^*$ be as in Proposition 39.1 and for $n \in \mathbb{N}$, let

$$E_n := \{\omega \in \Omega : \varphi_n \circ F(\omega) \ne \varphi_n \circ G(\omega)\}.$$

By assumption $\mu(E_n) = 0$ and therefore $E := \cup_{n=1}^\infty E_n$ is a μ – null set as well. This completes the proof since $\varphi_n(F - G) = 0$ on E^c and therefore, by Eq. (39.1)

$$\|F - G\| = \sup_n |\varphi_n(F - G)| = 0 \text{ on } E^c.$$

∎

Definition 39.3. $\mathcal{F}C_c^\infty(X) = \{f : X \to \mathbb{R} : f = F(\varphi_1, \varphi_2, \ldots, \varphi_n),\ F \in C_c^\infty(\mathbb{R}^n)$ *and* $\varphi_i \in X^*\}$.

Lemma 39.4. *Given a rectangle R in $\mathbb{R}^n$, say $R = [a_1, b_1) \times \cdots \times [a_n, b_n)$, then there exists $f_k \in C_c^\infty(\mathbb{R}^n)$ such that $f_k \to 1_R$ boundedly.*

Proof. It suffices to consider the one dimensional case. Let $\varphi \in C_c^\infty(\mathbb{R})$ such that $\varphi \ge 0$, φ is supported in $(-1, 0)$ and $\int_{\mathbb{R}} \varphi(x)dx = 1$. Set $\varphi_\varepsilon(x) = \frac{1}{\varepsilon}\varphi(\frac{x}{\varepsilon})$. Then

$$\begin{aligned} \varphi_\varepsilon * 1_{[a,b)}(x) &= \int_{\mathbb{R}} \varphi_\varepsilon(y) 1_{[a,b)}(x - y)\, dy = \int_{\mathbb{R}} \varphi(y) 1_{[a,b)}(x - \varepsilon y)\, dy \\ &= \int_{-1}^0 \varphi(y) 1_{[a,b)}(x - \varepsilon y)\, dy \to 1_{[a,b)}(x) \text{ as } \varepsilon \downarrow 0 \end{aligned}$$

for all $x \in \mathbb{R}$. ∎

Lemma 39.5. *The σ – algebra generated by $\mathcal{F}C_c^\infty(X)$ and the functions of the form $\{e^{i\varphi} : \varphi \in X^*\}$ are same as the Borel σ – algebra, i.e. $\sigma(\mathcal{F}C_c^\infty(X)) = \mathcal{B}(X)$ and $\sigma(\{e^{i\varphi} : \varphi \in X^*\}) = \mathcal{B}(X)$.*

Proof. By Lemma 39.4, $1_{\varphi^{-1}([a,b))} = 1_{[a,b)} \circ \varphi$ is a bounded $\sigma(\mathcal{F}C_c^\infty(X))$ – measurable function for all $-\infty < a < b < \infty$ and $\varphi \in X^*$. Therefore it follows $\varphi^{-1}([a,b)) \in \sigma(\mathcal{F}C_c^\infty(X))$ for all $-\infty < a < b < \infty$ and hence φ is all that φ is $\sigma(\mathcal{F}C_c^\infty(X))$ – measurable for all $\varphi \in X^*$. Hence we may conclude that

$$\mathcal{B}(X) = \sigma(X^*) \subset \sigma(\mathcal{F}C_c^\infty(X)).$$

Since every $f \in \mathcal{F}C_c^\infty(X)$ is continuous, $\sigma(\mathcal{F}C_c^\infty(X)) \subset \mathcal{B}(X)$ and the proof that $\sigma(\mathcal{F}C_c^\infty(X)) = \mathcal{B}(X)$ is complete.

Clearly for all $\varphi \in X^*$, $e^{i\varphi}$ is $\sigma(X^*) = \mathcal{B}_X$ – measurable so that $\sigma(\{e^{i\varphi} : \varphi \in X^*\}) \subset \mathcal{B}(X)$. To prove the converse inclusion, suppose that $f \in C_c^\infty(\mathbb{R}, \mathbb{C})$ and for $M >> 1$, let

$$f_M(x) = \sum_{k \in \mathbb{Z}^n} f(x + k2\pi M).$$

Then f_M is a $2\pi M$ – periodic function on $\mathbb{R}$ and therefore by the Weierstrass approximation theorem (see Theorems 4.40 and 4.42) there exists polynomials $p_m(\xi, \bar{\xi})$ for $\xi \in \mathbb{C}$ such that $p_m(e^{i\frac{x}{M}}, e^{-i\frac{x}{M}})$ converges to $f_M(x)$ uniformly in x as $m \to \infty$. In particular this implies for any $\varphi \in X^*$ that $f_M(\varphi)$ is the uniform limit of $p_m(e^{i\varphi/M}, e^{-i\varphi/M})$ and there therefore $f_M(\varphi)$ is a bounded $\sigma(\{e^{i\varphi} : \varphi \in X^*\})$ – measurable function. Because $f_M(\varphi) \to f(\varphi)$ as $M \to \infty$, it now follows that $f(\varphi)$ is also a bounded $\sigma(\{e^{i\varphi} : \varphi \in X^*\})$ – measurable function. Letting f approximate $1_{[a,b)}$ with $-\infty < a < b < \infty$ then shows that $1_{\varphi^{-1}([a,b))} = 1_{[a,b)} \circ \varphi$ is a bounded $\sigma(\{e^{i\varphi} : \varphi \in X^*\})$ – measurable function and therefore

$$\mathcal{B}(X) = \sigma(X^*) \subset \sigma(\{e^{i\varphi} : \varphi \in X^*\}).$$

■

Theorem 39.6 (Uniqueness of the Fourier Transform). *Suppose μ, ν are two probability measures on $(X, \mathcal{B}(X))$ such that $\int_X e^{i\varphi(x)} d\mu(x) = \int_X e^{i\varphi(x)} d\nu(x)$ for all $\varphi \in X^*$ then $\mu = \nu$. Write $\hat{\mu}(\varphi) = \int_X e^{i\varphi(x)} d\mu(x)$*

Proof. (The argument here is basically the same as its finite dimensional cousin in Corollary 8.11.) Let $\mathbb{H}$ be the linear space of bounded $\mathcal{B}(X)$ – measurable functions on X such that $\mu(f) = \nu(f)$ and $\mathbb{M} = \{e^{i\varphi} \in \mathbb{H} : \varphi \in X^*\} \subset \mathbb{H}$. As is easily checked, $\mathbb{H}$ contains 1, $\mathbb{H}$ is closed under bounded convergence and complex conjugation. Hence by the Dynkin's multiplicative system theorem as in Corollary 8.8, $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions which by Lemma 39.5 implies $\mu(f) = \nu(f)$ for all bounded $\mathcal{B}(X)$ – measurable functions. That is to say $\mu = \nu$.

Alternative argument: apply the density Theorem 39.7 below with μ replaced by $\mu + \nu$. In more detail if $f : X \to \mathbb{C}$ is a bounded $\mathcal{B}(X)$ – measurable function, we may find $f_n \in \text{span}\{e^{i\varphi} : \varphi \in X^*\}$ such that $f_n \to f$ in $L^1(\mu + \nu)$. Since this convergence takes place in $L^1(\mu)$ and $L^1(\nu)$ as well and $\mu(f_n) = \nu(f_n)$ for all n, we may conclude,

$$\mu(f) = \lim_{n\to\infty} \mu(f_n) = \lim_{n\to\infty} \nu(f_n) = \nu(f).$$

■

Theorem 39.7 (Density Theorem I). *Let μ be a probability measure on $\mathcal{B}(X)$ and $1 \le p < \infty$. Then*

1. *$\mathcal{F}C_c^\infty(X)$ is dense in $L^p(X, \mathcal{B}, \mu)$ and*
2. *span$\{e^{i\varphi} : \varphi \in X^*\}$ is dense in $L^p(X, \mathcal{B}, \mu)$.*

Proof. 1. Let $\mathbb{M} = \mathcal{F}C_c^\infty(X)$, $\mathbb{H}$ denote the bounded measurable functions in $\bar{\mathbb{M}}^{L^p}$ – the L^p–closure of $\mathbb{M}$. Then $\mathbb{M} \subset \mathbb{H}$, $1 \in \mathbb{H}$ (as the reader should prove) and $\mathbb{H}$ is a linear space which, by the dominated convergence theorem, is closed under bounded convergence. So by the Dynkin's multiplicative system theorem (see Corollary 8.8), it follows that $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions. This result along with Lemma 39.5 implies $\overline{\mathcal{F}C_c^\infty(X)}^{L^p}$ contains all bounded $\mathcal{B}(X)$ – measurable functions and this latter class of functions is dense in L^p.

2. Let $\mathbb{H}$ denote linear subspace consisting of all bounded functions in $\overline{\text{span}\{e^{i\varphi} : \varphi \in X^*\}}^{L^p(\mu)}$. As before $1 \in \mathbb{H}$, $\mathbb{H}$ is closed under bounded convergence and complex conjugation and contains the multiplicative system $\mathbb{M} := \{e^{i\varphi} : \varphi \in X^*\}$. Therefore $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions. This result along with Lemma 39.5 implies $\overline{\text{span}\{e^{i\varphi} : \varphi \in X^*\}}^{L^p(\mu)}$ contains all bounded $\mathcal{B}(X)$ – measurable functions and this latter class of functions is dense in L^p. ■

Proposition 39.8 (Density Theorem II). *Suppose μ is a probability measure such that for all $\varphi \in X^*$ there exists $\varepsilon = \varepsilon(\varphi) > 0$ such that $\mu(e^{\varepsilon|\varphi|}) < \infty$. Then*

$$\mathcal{F} \equiv \{P(\varphi_1, \dots, \varphi_n) : P : \mathbb{R}^n \to \mathbb{R} \text{ a polynomial}\}$$

is dense in $L^p(X, \mathcal{B}, \mu)$ for all $1 \le p < \infty$.

Proof. If $\overline{\mathcal{F}}^{L^p}$ is a proper subspace of L^p then by the Hahn Banach theorem there exists $\lambda \in (L^p)^*$ such that $\lambda \neq 0$ while $\lambda(\mathcal{F}) = \{0\}$. Since $(L^p)^* \cong L^q$ where $\frac{1}{q} + \frac{1}{p} = 1$, there exists $g \in L^q$ such that

$$\lambda(f) = \int_X fg d\mu \text{ for all } f \in L^p.$$

By assumption

$$\int_X fg d\mu = \lambda(f) = 0 \text{ for all } f \in \mathcal{F} \tag{39.2}$$

and $g \neq 0$ because $\lambda \neq 0$. We will arrive at contradiction by showing $g = 0$. Let $\varphi \in X^*$ and choose $\varepsilon > 0$ such that $\int_X e^{\varepsilon|\varphi|} d\mu < \infty$. Then for any $n \in \mathbb{N}_0$ and $\delta \in (0, \varepsilon/p)$,

$$\int_X |\varphi|^n e^{\delta|\varphi|} |g| \, d\mu \le \|g\|_{L^q} \left(\int_X |\varphi|^{pn} e^{p\delta|\varphi|} d\mu \right)^{1/p} < \infty$$

since there exists $C < \infty$ such that

$$|\varphi|^{pn} e^{p\delta|\varphi|} \leq C e^{\varepsilon|\varphi|}.$$

Define $F(z) = \int\limits_X e^{z\varphi} g \cdot d\mu$ for $|\mathrm{Re}\ z| < \delta$. Then F is analytic on $|\mathrm{Re}\ z| < \delta$ and

$$F^{(n)}(z) = \int\limits_X \varphi^n e^{z\varphi} g \ d\mu \text{ for all } n \in \mathbb{N}_0.$$

In particular $F^{(n)}(0) = \int\limits_X \varphi^n g \ d\mu$ and hence $F^{(n)}(0) = 0$ for all $n \in \mathbb{N}_0$ by Eq. (39.2). By analytic continuation it then follows that $F \equiv 0$ on $|\mathrm{Re}\ z| < \delta$ and in particular

$$\int_X e^{i\varphi} g d\mu = F(i) = 0.$$

Since $\varphi \in X^*$ was arbitrary we have shown $\int_X e^{i\varphi} g d\mu = 0$ for all $\varphi \in X^*$. Because $\mathrm{span}\{e^{i\varphi} : \varphi \in X^*\}$ is dense in $L^p(\mu)$ this implies $g = 0 \in L^q$ and we have arrived at the desired contradiction. ■

Bochner Integral

Throughout this chapter we will assume that $(\Omega, \mathcal{F}, \mu)$ be a fixed measure space, X is a separable real Banach space, $\mathcal{B} = \mathcal{B}(X) = \mathcal{B}_X$ is the Borel σ– algebra, and X^* is the continuous dual of X. In this chapter we will define the *Bochner integral*, $\int_\Omega f d\mu \in X$, of a measurable function $f : \Omega \to X$. define integrals of functions taking values in a Banach space. (Shortly we will further assume that μ is a σ – finite on $\mathcal{F}$.) We will also introduce a vector valued conditional expectation and prove some related vector valued martingale results.

40.1 Banach Valued L^p – Spaces

Remark 40.1. Recall that we have already seen in this case that the Borel σ – field $\mathcal{B}$ on X is the same as the σ – field $(\sigma(X^*))$ which is generated by X^* – the continuous linear functionals on X. (This is done in Proposition 39.1 below.) As a consequence $F : \Omega \to X$ is $\mathcal{F}/\mathcal{B}$ measurable iff $\varphi \circ F : \Omega \to \mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$ – measurable for all $\varphi \in X^*$. In particular it follows that if $F, G : \Omega \to X$ are measurable functions then so is $F + G$ and λF for all $\lambda \in \mathbb{F}$ and it follows that $\{F \neq G\} = \{F - G \neq 0\}$ is measurable as well. Also note that $\|\cdot\| : X \to [0, \infty)$ is continuous and hence measurable and hence $\omega \to \|F(\omega)\|_X$ is the composition of two measurable functions and therefore measurable

Definition 40.2. *For $1 \le p < \infty$ let $L^p(\mu; X)$ denote the space of measurable functions $F : \Omega \to X$ such that $\int_\Omega \|F\|^p d\mu < \infty$. For $F \in L^p(\mu; X)$, define*

$$\|F\|_{L^p} = \left(\int_\Omega \|F\|_X^p d\mu \right)^{\frac{1}{p}}.$$

As usual in L^p – spaces we will identify two measurable functions, $F, G : \Omega \to X$, if $F = G$ a.e.

Lemma 40.3. *Suppose $a_n \in X$ and $\|a_{n+1} - a_n\| \le \varepsilon_n$ and $\sum_{n=1}^\infty \varepsilon_n < \infty$. Then $\lim_{n\to\infty} a_n = a \in X$ exists and $\|a - a_n\| \le \delta_n := \sum_{k=n}^\infty \varepsilon_k$.*

Proof. (This is a special case of Exercise 45.18.) Let $m > n$ then

$$\|a_m - a_n\| = \left\| \sum_{k=n}^{m-1} (a_{k+1} - a_k) \right\| \le \sum_{k=n}^{m-1} \|a_{k+1} - a_k\| \le \sum_{k=n}^\infty \varepsilon_k := \delta_n . \tag{40.1}$$

So $\|a_m - a_n\| \le \delta_{\min(m,n)} \to 0$ as $, m, n \to \infty$, i.e. $\{a_n\}$ is Cauchy. Let $m \to \infty$ in (40.1) to find $\|a - a_n\| \le \delta_n$. ∎

Lemma 40.4. *Suppose that $\{F_n\}$ is Cauchy in measure, i.e. $\lim_{m,n\to\infty} \mu(\|F_n - F_m\| \ge \varepsilon) = 0$ for all $\varepsilon > 0$. Then there exists a subsequence $G_j = F_{n_j}$ such that $F := \lim_{j\to\infty} G_j$ exists μ – a.e. and moreover $F_n \xrightarrow{\mu} F$ as $n \to \infty$, i.e. $\lim_{n\to\infty} \mu(\|F_n - F\| \ge \varepsilon) = 0$ for all $\varepsilon > 0$.*

Proof. Let $\varepsilon_n > 0$ such that $\sum_{n=1}^\infty \varepsilon_n < \infty$ ($\varepsilon_n = 2^{-n}$ would do) and set $\delta_n = \sum_{k=n}^\infty \varepsilon_k$. Choose $G_j = F_{n_j}$ where $\{n_j\}$ is a subsequence of $\mathbb{N}$ such that

$$\mu(\{\|G_{j+1} - G_j\| > \varepsilon_j\}) \le \varepsilon_j.$$

Let

$$A_N := \cup_{j\ge N} \{\|G_{j+1} - G_j\| > \varepsilon_j\} \text{ and}$$
$$E := \cap_{N=1}^\infty A_N = \{\|G_{j+1} - G_j\| > \varepsilon_j \text{ i.o.}\}.$$

Since $\mu(A_N) \le \delta_N < \infty$ and $A_N \downarrow E$ it follows[1] that $0 = \mu(E) = \lim_{N\to\infty} \mu(A_N)$. For $\omega \notin E$, $\|G_{j+1}(\omega) - G_j(\omega)\| \le \varepsilon_j$ for a.a. j and hence by Lemma 40.3, $F(\omega) := \lim_{j\to\infty} G_j(\omega)$ exists for $\omega \notin E$. Let us define $F(\omega) = 0$ for all $\omega \in E$.

Next we will show $G_N \xrightarrow{\mu} F$ as $N \to \infty$ where F and G_N are as above. If

$$\omega \in A_N^c = \cap_{j\ge N} \{\|G_{j+1} - G\| \le \varepsilon_j\},$$

then

[1] Alternatively, $\mu(E) = 0$ by the first Borel Cantelli lemma and the fact that $\sum_{j=1}^\infty \mu(\{\|G_{j+1} - G_j\| > \varepsilon_j\}) \le \sum_{j=1}^\infty \varepsilon_j < \infty$.

$$\|G_{j+1}(\omega) - G_j(\omega)\| \le \varepsilon_j \text{ for all } j \ge N.$$

Another application of Lemma 40.3 shows $\|F(\omega) - G_j(\omega)\| \le \delta_j$ for all $j \ge N$, i.e.

$$A_N^c \subset \cap_{j \ge N} \{\|F - G_j\| \le \delta_j\} \subset \{|F - G_N| \le \delta_N\}.$$

Therefore, by taking complements of this equation, $\{\|F - G_N\| > \delta_N\} \subset A_N$ and hence

$$\mu(\|F - G_N\| > \delta_N) \le \mu(A_N) \le \delta_N \to 0 \text{ as } N \to \infty$$

and in particular, $G_N \xrightarrow{\mu} F$ as $N \to \infty$.

With this in hand, it is straightforward to show $F_n \xrightarrow{\mu} F$. Indeed, by the usual trick, for all $j \in \mathbb{N}$,

$$\mu(\{\|F_n - F\| > \varepsilon\}) \le \mu(\{\|F - G_j\| > \varepsilon/2\}) + \mu(\|G_j - F_n\| > \varepsilon/2).$$

Therefore, letting $j \to \infty$ in this inequality gives,

$$\mu(\{\|F_n - F\| > \varepsilon\}) \le \limsup_{j\to\infty} \mu(\|G_j - F_n\| > \varepsilon/2) \to 0 \text{ as } n \to \infty,$$

wherein we have used $\{F_n\}_{n=1}^\infty$ is Cauchy in measure and $G_j \xrightarrow{\mu} F$. ∎

Theorem 40.5. *For each $p \in [0, \infty)$, the space $(L^p(\mu; X), \|\cdot\|_{L^p})$ is a Banach space.*

Proof. It is straightforward to check that $\cdot\|_{L^p}$ is a norm. For example,

$$\begin{aligned}\|F + G\|_{L^p} &= \left(\int_\Omega \|F + G\|_X^p d\mu\right)^{\frac{1}{p}} \le \left(\int_\Omega (\|F\|_X + \|G\|_X)^p d\mu\right)^{\frac{1}{p}} \\ &\le \|F\|_{L^p} + \|G\|_{L^p}.\end{aligned}$$

So the main point is to prove completeness of the norm.

Let $\{F_n\}_{n=1}^\infty \subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev's inequality $\{F_n\}$ is Cauchy in measure and by Lemma 40.4 there exists a subsequence $\{G_j\}$ of $\{F_n\}$ such that $G_j \to F$ a.e. By Fatou's Lemma,

$$\begin{aligned}\|G_j - F\|_p^p &= \int \liminf_{k\to\infty} \|G_j - G_k\|^p d\mu \le \liminf_{k\to\infty} \int \|G_j - G_k\|^p d\mu \\ &= \liminf_{k\to\infty} \|G_j - G_k\|_p^p \to 0 \text{ as } j \to \infty.\end{aligned}$$

In particular, $\|F\|_p \le \|G_j - F\|_p + \|G_j\|_p < \infty$ so the $F \in L^p$ and $G_j \xrightarrow{L^p} F$. The proof is finished because,

$$\|F_n - F\|_p \le \|F_n - G_j\|_p + \|G_j - F\|_p \to 0 \text{ as } j, n \to \infty.$$

∎

We say a function $F : \Omega \to X$ is a simple function if F is measurable and has finite range. If F also satisfies, $\mu(F \ne 0) < \infty$ we say that F is a μ – simple function and let $\mathcal{S}(\mu; X)$ denote the vector space of μ – simple functions.

Proposition 40.6. *For each $1 \le p < \infty$ the μ – simple functions, $\mathcal{S}(\mu; X)$, are dense inside of $L^p(\mu; X)$.*

Proof. Let $\mathbb{D} := \{x_n\}_{n=1}^\infty$ be a countable dense subset of $X \setminus \{0\}$. For each $\varepsilon > 0$ and $n \in \mathbb{N}$ let

$$B_n^\varepsilon := \left\{x \in X : \|x - x_n\| \le \min\left(\varepsilon, \frac{1}{2}\|x_n\|\right)\right\}$$

and then define $A_n^\varepsilon := B_n^\varepsilon \setminus (\cup_{k=1}^n B_k^\varepsilon)$. Thus $\{A_n^\varepsilon\}_{n=1}^\infty$ is a partition of $X \setminus \{0\}$ with the added property that $\|y - x_n\| \le \varepsilon$ and $\frac{1}{2}\|x_n\| \le \|y\| \le \frac{3}{2}\|x_n\|$for all $y \in A_n^\varepsilon$.

Given $F \in L^p(\mu; X)$ let

$$F_\varepsilon := \sum_{n=1}^\infty x_n \cdot 1_{F \in A_n^\varepsilon} = \sum_{n=1}^\infty x_n \cdot 1_{F^{-1}(A_n^\varepsilon)}.$$

For $\omega \in F^{-1}(A_n^\varepsilon)$, i.e. $F(\omega) \in A_n^\varepsilon$, we have

$$\begin{aligned}\|F_\varepsilon(\omega)\| &= \|x_n\| \le 2\|F(\omega)\| \text{ and} \\ \|F_\varepsilon(\omega) - F(\omega)\| &= \|x_n - F(\omega)\| \le \varepsilon.\end{aligned}$$

Putting these two estimates together shows,

$$\|F_\varepsilon - F\| \le \varepsilon \text{ and } \|F_\varepsilon - F\| \le \|F_\varepsilon\| + \|F\| \le 3\|F\|.$$

Hence we may now apply the dominated convergence theorem in order to show

$$\lim_{\varepsilon\downarrow 0} \|F - F_\varepsilon\|_{L^p(\mu;X)} = 0.$$

As the F_ε – have countable range we have not yet completed the proof. To remedy this defect, to each $N \in \mathbb{N}$ let

$$F_\varepsilon^N := \sum_{n=1}^N x_n \cdot 1_{F^{-1}(A_n^\varepsilon)}.$$

Then it is clear that $\lim_{N\to\infty} F_\varepsilon^N = F_\varepsilon$ and that $\left\|F_\varepsilon^N\right\| \le \|F_\varepsilon\| \le 2\|F\|$ for all N. Therefore another application of the dominated convergence theorem implies, $\lim_{N\to\infty}\left\|F_\varepsilon^N - F_\varepsilon\right\|_{L^p(\mu;X)} = 0$. Thus any $F \in L^p(\mu;X)$ may be arbitrarily well approximated by one of the $F_\varepsilon^N \in \mathcal{S}(\mu;X)$ with ε sufficiently small and N sufficiently large. ∎

For later purposes it will be useful to record a result based on the partitions $\{A_n^\varepsilon\}_{n=1}^\infty$ of $X \setminus \{0\}$ introduced in the above proof.

Lemma 40.7. *Suppose that $F : \Omega \to X$ is a measurable function such that $\mu(F \neq 0) > 0$. Then there exists $B \in \mathcal{F}$ and $\varphi \in X^*$ such that $\mu(B) > 0$ and $\inf_{\omega\in B}\varphi\circ F(\omega) > 0$.*

Proof. Let $\varepsilon > 0$ be chosen arbitrarily, for example you might take $\varepsilon = 1$ and let $\{A_n := A_n^\varepsilon\}_{n=1}^\infty$ be the partition of $X \setminus \{0\}$ introduced in the proof of Proposition 40.6 above. Since $\{F \neq 0\} = \sum_{n=1}^\infty \{F \in A_n\}$ and $\mu(F \neq 0) > 0$, it follows that that $\mu(F \in A_n) > 0$ for some $n \in \mathbb{N}$. We now let $B := \{F \in A_n\} = F^{-1}(A_n)$ and choose $\varphi \in X^*$ such that $\varphi(x_n) = \|x_n\|$ and $\|\varphi\|_{X^*} = 1$. For $\omega \in B$ we have $F(\omega) \in A_n$ and therefore $\|F(\omega) - x_n\| \leq \frac{1}{2}\|x_n\|$ and hence,

$$|\varphi(F(\omega)) - \|x_n\|| = |\varphi(F(\omega)) - \varphi(x_n)| \leq \|\varphi\|_{X^*}\|F(\omega) - x_n\| \leq \frac{1}{2}\|x_n\|.$$

From this inequality we see that $\varphi(F(\omega)) \geq \frac{1}{2}\|x_n\| > 0$ for all $\omega \in B$. ∎

Definition 40.8. *To each $F \in \mathcal{S}(\mu;X)$, let*

$$\begin{aligned} I(F) &= \sum_{x\in X} x\mu(F^{-1}(\{x\})) = \sum_{x\in X} x\mu(\{F = x\}) \\ &= \sum_{x\in F(\Omega)} x\mu(F = x) \in X. \end{aligned}$$

The following proposition is straightforward to prove.

Proposition 40.9. *The map $I : \mathcal{S}(\mu;X) \to X$ is linear and satisfies for all $F \in \mathcal{S}(\mu;X)$,*

$$\|I(F)\|_X \leq \int_\Omega \|F\| d\mu \text{ and} \tag{40.2}$$

$$\varphi(I(F)) = \int_X \varphi\circ F \, d\mu \; \forall\, \varphi \in X^*. \tag{40.3}$$

Proof. If $0 \neq c \in \mathbb{R}$ and $F \in \mathcal{S}(\mu;X)$, then

$$\begin{aligned} I(cF) &= \sum_{x\in X} x\mu(cF = x) = \sum_{x\in X} x\mu\left(F = \frac{x}{c}\right) \\ &= \sum_{y\in X} cy\ \mu(F = y) = cI(F) \end{aligned}$$

and if $c = 0$, $I(0F) = 0 = 0I(F)$. If $F, G \in \mathcal{S}(\mu;X)$,

$$\begin{aligned} I(F+G) &= \sum_x x\mu(F+G = x) \\ &= \sum_x x \sum_{y+z=x}\mu(F = y, G = z) \\ &= \sum_{y,z}(y+z)\mu(F = y, G = z) \\ &= \sum_y y\mu(F = y) + \sum_z z\mu(G = z) = I(F) + I(G). \end{aligned}$$

Equation (40.2) is a consequence of the following computation:

$$\|I(F)\|_X = \|\sum_{x\in X} x\mu(F = x)\| \leq \sum_{x\in X}\|x\|\mu(F = x) = \int_\Omega \|F\| d\mu$$

and Eq. (40.3) follows from:

$$\begin{aligned} \varphi(I(F)) &= \varphi(\sum_{x\in X} x\mu(\{F = x\})) \\ &= \sum_{x\in X}\varphi(x)\mu(\{F = x\}) = \int_X \varphi\circ F \, d\mu. \end{aligned}$$

∎

Theorem 40.10 (Bochner Integral). *There is a unique continuous linear map $\bar{I} : L^1(\Omega,\mathcal{F},\mu;X) \to X$ such that $\bar{I}|_{\mathcal{S}(\mu;X)} = I$ where I is defined in Definition 40.8. Moreover, for all $F \in L^1(\Omega,\mathcal{F},\mu;X)$,*

$$\|\bar{I}(F)\|_X \leq \int_\Omega \|F\| d\mu \tag{40.4}$$

and $\bar{I}(F)$ is the unique element in X such that

$$\varphi(\bar{I}(F)) = \int_X \varphi\circ F \, d\mu \; \forall\, \varphi \in X^*. \tag{40.5}$$

The map $\bar{I}(F)$ will be denoted suggestively by $\int_X F d\mu$ or $\mu(F)$ so that Eq. (40.5) may be written as

$$\begin{gathered} \varphi\left(\int_X F d\mu\right) = \int_X \varphi\circ F \, d\mu \; \forall\, \varphi \in X^* \text{ or} \\ \varphi(\mu(F)) = \mu(\varphi\circ F) \;\; \forall\, \varphi \in X^* \end{gathered}$$

Proof. The existence of a continuous linear map $\bar{I} : L^1(\Omega, \mathcal{F}, \mu; X) \to X$ such that $\bar{I}|_{\mathcal{S}(\mu;X)} = I$ and Eq. (40.4) holds follows from Propositions 40.9 and 40.6 and the bounded linear transformation Theorem 13.18. If $\varphi \in X^*$ and $F \in L^1(\Omega, \mathcal{F}, \mu; X)$, choose $F_n \in \mathcal{S}(\mu; X)$ such that $F_n \to F$ in $L^1(\Omega, \mathcal{F}, \mu; X)$ as $n \to \infty$. Then $\bar{I}(F) = \lim_{n\to\infty} I(F_n)$ and hence by Eq. (40.3),

$$\varphi(\bar{I}(F)) = \varphi(\lim_{n\to\infty} I(F_n)) = \lim_{n\to\infty} \varphi(I(F_n)) = \lim_{n\to\infty} \int_X \varphi \circ F_n \, d\mu.$$

This proves Eq. (40.5) since

$$\begin{aligned} \left| \int_\Omega (\varphi \circ F - \varphi \circ F_n) d\mu \right| &\le \int_\Omega |\varphi \circ F - \varphi \circ F_n| \, d\mu \\ &\le \int_\Omega \|\varphi\|_{X^*} \|\varphi \circ F - \varphi \circ F_n\|_X \, d\mu \\ &= \|\varphi\|_{X^*} \|F - F_n\|_{L^1} \to 0 \text{ as } n \to \infty. \end{aligned}$$

The fact that $\bar{I}(F)$ is determined by Eq. (40.5) is a consequence of the Hahn – Banach theorem. ∎

Example 40.11. Suppose that $x \in X$ and $f \in L^1(\mu; \mathbb{R})$, then $F(\omega) := f(\omega)\, x$ defines an element of $L^1(\mu; X)$ and

$$\int_\Omega F d\mu = \left(\int_\Omega f d\mu \right) x. \tag{40.6}$$

To prove this just observe that $\|F\| = |f| \, \|x\| \in L^1(\mu)$ and for $\varphi \in X^*$ we have

$$\begin{aligned} \varphi\left(\left(\int_\Omega f d\mu \right) x \right) &= \left(\int_\Omega f d\mu \right) \cdot \varphi(x) \\ &= \left(\int_\Omega f \varphi(x) \, d\mu \right) = \int_\Omega \varphi \circ F \, d\mu. \end{aligned}$$

Since $\varphi\left(\int_\Omega F d\mu\right) = \int_\Omega \varphi \circ F \; d$ for all $\varphi \in X^*$ it follows that Eq. (40.6) is correct.

Remark 40.12. The separability assumption on X may be relaxed by assuming that $F : \Omega \to X$ has separable essential range. In this case we may still define $\int_X F d\mu$ by applying the above formalism with X replaced by the separable Banach space, $X_0 := \overline{\operatorname{span}(\operatorname{essran}_\mu(F))}$. For example if Ω is a compact topological space and $F : \Omega \to X$ is a continuous map, then $\int_\Omega F d\mu$ is always defined.

40.2 Conditional Expectation

Now that we have defined the Bochner integral it is time to move on to conditional expectations. For now let us now suppose that μ is a probability measure on $(\Omega, \mathcal{F})$ and denote the Bochner integral, $\mu(F) = \int_\Omega F d\mu$ by $\mathbb{E}F$. Given a sub-σ – algebra $\mathcal{G}$ of $\mathcal{F}$ we would like to define $\mathbb{E}[F|\mathcal{G}]$ as an element of $L^1(\Omega, \mathcal{G}, \mu; X)$ for each $F \in L^1(\Omega, \mathcal{F}, \mu; X)$. As with the Bochner integral we will start with simple functions. For any $F \in \mathcal{S}(\mu; X)$ we have the identity,

$$F = \sum_{x \in \operatorname{Ran}(F) \setminus \{0\}} x 1_{\{F=x\}}$$

and hence it is reasonable to require,

$$\mathbb{E}_\mathcal{G} F = \mathbb{E}[F|\mathcal{G}] = \mathbb{E}\left[\sum_{x \in \operatorname{Ran}(F) \setminus \{0\}} x 1_{\{F=x\}} |\mathcal{G} \right] = \sum_{x \in \operatorname{Ran}(F) \setminus \{0\}} \mathbb{E}\left[x 1_{\{F=x\}} |\mathcal{G} \right].$$

Furthermore, since $x \in X$ is constant in the previous formula we should further require,

$$\mathbb{E}_\mathcal{G} F = \sum_{x \in \operatorname{Ran}(F) \setminus \{0\}} x \cdot \mathbb{E}\left[1_{\{F=x\}} |\mathcal{G} \right]. \tag{40.7}$$

Proposition 40.13. *If $\mathbb{E}_\mathcal{G} : \mathcal{S}(\mu; X) \to L^1(\Omega, \mathcal{G}, \mu; X)$ is the map defined by Eq. (40.7), then;*

1. *$\mathbb{E}_\mathcal{G}$ is linear*
2. *$\|\mathbb{E}_\mathcal{G} F\|_{L^1(\mu)} \le \|F\|_{L^1(\mu)}$ for all $F \in \mathcal{S}(\mu; X)$, i.e. $\mathbb{E}_\mathcal{G}$ is a contraction,*
3. *$\mathbb{E}_\mathcal{G} F$ satisfies,*
$$\mathbb{E}[\mathbb{E}_\mathcal{G} F : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}$$
and
$$\varphi \circ \mathbb{E}_\mathcal{G} F = \mathbb{E}_\mathcal{G}[\varphi \circ F] \text{ a.s. for all } \varphi \in X^*.$$

Proof. 1) If $0 \ne c \in \mathbb{R}$ and $F \in \mathcal{S}(\mu; X)$, then

$$\begin{aligned} \mathbb{E}_\mathcal{G}(cF) &= \sum_{x \in X} x \mathbb{E}\left[1_{\{cF=x\}} |\mathcal{G} \right] = \sum_{x \in X} x \mathbb{E}\left[1_{\{F=c^{-1}x\}} |\mathcal{G} \right] \\ &= \sum_{y \in X} cy \; \mathbb{E}\left[1_{\{F=y\}} |\mathcal{G} \right] = c \mathbb{E}_\mathcal{G}(F) \end{aligned}$$

and if $c = 0$, $\mathbb{E}_\mathcal{G}(0F) = 0 = 0\mathbb{E}_\mathcal{G}(F)$.

If $F, G \in \mathcal{S}(\mu; X)$,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}(F+G) &= \sum_x x\mathbb{E}_{\mathcal{G}}\left[1_{F+G=x}\right] \\ &= \sum_x x \sum_{y+z=x} \mathbb{E}_{\mathcal{G}}\left[1_{F=y\ \&\ G=z}\right] \\ &= \sum_{y,z} (y+z)\mathbb{E}_{\mathcal{G}}\left[1_{F=y\ \&\ G=z}\right] \\ &= \sum_y y\mathbb{E}_{\mathcal{G}}\left[1_{F=y}\right] + \sum_z z\mathbb{E}_{\mathcal{G}}\left[1_{G=z}\right] \\ &= \mathbb{E}_{\mathcal{G}}(F) + \mathbb{E}_{\mathcal{G}}(G).\end{aligned}$$

2) Integrating the pointwise inequality,

$$\|\mathbb{E}_{\mathcal{G}}F\| = \sum_{x \in \mathrm{Ran}(F)\setminus\{0\}} \|x\| \cdot \mathbb{E}\left[1_{\{F=x\}}|\mathcal{G}\right]$$

implies,

$$\begin{aligned}\|\mathbb{E}_{\mathcal{G}}F\|_{L^1(\mu)} &\le \sum_{x \in \mathrm{Ran}(F)\setminus\{0\}} \|x\| \cdot \mathbb{E}\left(\mathbb{E}\left[1_{\{F=x\}}|\mathcal{G}\right]\right) \\ &= \sum_{x \in \mathrm{Ran}(F)\setminus\{0\}} \|x\| \cdot \mathbb{E}\left(1_{\{F=x\}}\right) \\ &= \mathbb{E}\left(\sum_{x \in \mathrm{Ran}(F)\setminus\{0\}} \|x\|\, 1_{\{F=x\}}\right) \\ &= \mathbb{E}\,\|F\| = \|F\|_{L^1(\mu)}\,.\end{aligned}$$

3. If $A \in \mathcal{G}$ we have,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}F : A\right] &= \mathbb{E}\left[\sum_{x\neq 0} x \cdot \mathbb{E}\left[1_{\{F=x\}}|\mathcal{G}\right] \cdot 1_A\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0} x \cdot \mathbb{E}\left[1_A 1_{\{F=x\}}|\mathcal{G}\right]\right] \\ &= \sum_{x\neq 0} \mathbb{E}\left[x \cdot \mathbb{E}\left[1_A 1_{\{F=x\}}|\mathcal{G}\right]\right] \\ &= \sum_{x\neq 0} x \cdot \mathbb{E}\left[\mathbb{E}\left[1_A 1_{\{F=x\}}|\mathcal{G}\right]\right] \\ &= \sum_{x\neq 0} x \cdot \mathbb{E}\left[1_A 1_{\{F=x\}}\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0} x \cdot 1_A 1_{\{F=x\}}\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0} x \cdot 1_{\{1_A F=x\}}\right] \\ &= \mathbb{E}\left[F : A\right].\end{aligned}$$

where we have used Example 40.11 two times in the above computation and all of the above sums are really over $x \in \mathrm{Ran}\,(F) \setminus \{0\}$. Finally if $\varphi \in X^*$ we have,

$$\begin{aligned}\varphi \circ \mathbb{E}_{\mathcal{G}}F &= \sum_{x\neq 0} \varphi(x) \cdot \mathbb{E}\left[1_{\{F=x\}}|\mathcal{G}\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0} \varphi(x) \cdot 1_{\{F=x\}}|\mathcal{G}\right] = \mathbb{E}\left[\varphi \circ F|\mathcal{G}\right].\end{aligned}$$

■

We may now apply the bounded linear transformation Theorem 13.18 in order to extend $\mathbb{E}_{\mathcal{G}}$ to all of $L^1(\Omega, \mathcal{F}, \mu; X)$.

Theorem 40.14 (Conditional expectation). *Let $F \in L^1(\Omega, \mathcal{F}, \mu; X)$. There is a linear map, $\mathbb{E}_{\mathcal{G}} : L^1(\Omega, \mathcal{F}, \mu; X) \to L^1(\Omega, \mathcal{G}, \mu; X)$, such that $\mathbb{E}_{\mathcal{G}}F$ is uniquely determined by either;*

1. $\mathbb{E}_{\mathcal{G}}F$ is the unique element in $L^1(\Omega, \mathcal{G}, \mu; X)$ such that

$$\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}F : A\right] = \mathbb{E}\left[F : A\right] \text{ for all } A \in \mathcal{G}$$

or

2. $\mathbb{E}_{\mathcal{G}}F$ *is the unique element in* $L^1(\Omega,\mathcal{G},\mu;X)$ *such that* $\varphi\circ\mathbb{E}_{\mathcal{G}}F=\mathbb{E}_{\mathcal{G}}[\varphi\circ F]$ *a.s. for all* $\varphi\in X^*$.

Moreover, $\mathbb{E}_{\mathcal{G}}:L^1(\Omega,\mathcal{F},\mu;X)\to L^1(\Omega,\mathcal{G},\mu;X)$ *is a contraction.*

Proof. The existence of contraction $\mathbb{E}_{\mathcal{G}}:L^1(\Omega,\mathcal{F},\mu;X)\to L^1(\Omega,\mathcal{G},\mu;X)$ with the desired properties easily follows from Propositions 40.6 and 40.13 along with the bounded linear transformation Theorem 13.18. So it only remains to verify that $\mathbb{E}_{\mathcal{G}}F$ is uniquely determined by either of the two conditions above.

1) If $G\in L^1(\Omega,\mathcal{G},\mu;X)$ satisfies $\mathbb{E}[G:A]=\mathbb{E}[F:A]=\mathbb{E}[\mathbb{E}_{\mathcal{G}}F:A]$ for all $A\in\mathcal{G}$ then $\mathbb{E}[G-\mathbb{E}_{\mathcal{G}}F:A]=0$ for all $A\in\mathcal{G}$. If $G\neq\mathbb{E}_{\mathcal{G}}F$ a.s., we may use Lemma 40.7 in order to find $A\in\mathcal{G}$ with $\mu(A)>0$ and $\varphi\in X^*$ such that $\varphi\circ(G-\mathbb{E}_{\mathcal{G}}F)>0$ on A. We then may conclude,

$$0=\varphi(0)=\varphi(\mathbb{E}[G-\mathbb{E}_{\mathcal{G}}F:A])=\mathbb{E}[\varphi(G-\mathbb{E}_{\mathcal{G}}F):A]>0$$

which is absurd and hence we must have $G\neq\mathbb{E}_{\mathcal{G}}F$ a.s.

2) If $G\in L^1(\Omega,\mathcal{G},\mu;X)$ satisfies $\varphi\circ G=\mathbb{E}_{\mathcal{G}}[\varphi\circ F]=\varphi\circ\mathbb{E}_{\mathcal{G}}F$ a.s. for all $\varphi\in X^*$ then $\varphi(G-\mathbb{E}_{\mathcal{G}}F)=0$ a.s. for all $\varphi\in X^*$ and the result follows from Corollary 39.2. Let us recall the proof here. As in the proof of Proposition 39.1, there exists a countable subset, $\mathbb{D}\subset X^*$, such that $\|x\|=\sup_{\varphi\in\mathbb{D}}\varphi(x)$ for all $x\in X$. Therefore we may conclude,

$$\|G-\mathbb{E}_{\mathcal{G}}F\|=\sup_{\varphi\in\mathbb{D}}\varphi(G-\mathbb{E}_{\mathcal{G}}F)=0\text{ a.s.},$$

i.e. $G=\mathbb{E}_{\mathcal{G}}F$ a.s.

Alternate proof. If $G\in L^1(\Omega,\mathcal{G},\mu;X)$ satisfies $\varphi\circ G=\mathbb{E}_{\mathcal{G}}[\varphi\circ F]=\varphi\circ\mathbb{E}_{\mathcal{G}}F$ a.s. for all $\varphi\in X^*$ then for any $A\in\mathcal{G}$ we have

$$\varphi(\mathbb{E}[G:A])=\mathbb{E}[\varphi\circ G:A]=\mathbb{E}[\varphi\circ\mathbb{E}_{\mathcal{G}}F:A]=\varphi(\mathbb{E}[\mathbb{E}_{\mathcal{G}}F:A]).$$

Therefore by the Hahn - Banach theorem X^* separates points and we may conclude that $\mathbb{E}[G:A]=\mathbb{E}[\mathbb{E}_{\mathcal{G}}F:A]$ and so by item 1., $G=\mathbb{E}_{\mathcal{G}}F$ a.s. ∎

Proposition 40.15. *Let* $\mathcal{G}\subset\mathcal{F}$ *and* $F\in L^1(\Omega,\mathcal{F},\mu;X)$. *The conditional expectation operator* $(\mathbb{E}_{\mathcal{G}})$ *satisfies the following additional properties;*

1. $\|\mathbb{E}_{\mathcal{G}}F\|\le\mathbb{E}_{\mathcal{G}}\|F\|$ *a.s..*
2. $\|\mathbb{E}_{\mathcal{G}}F\|_{L^p(\mu)}\le\|F\|_{L^p(\mu)}$ *for all* $F\in L^p(\Omega,\mathcal{F},\mu;X)$ *where* $1\le p<\infty$.
3. *If* $h\in L^\infty(\Omega,\mathcal{G},\mu)$, *then* $\mathbb{E}_{\mathcal{G}}[hF]=h\cdot\mathbb{E}_{\mathcal{G}}[F]$ *a.s.*
4. *If* $\mathcal{G}_0\subset\mathcal{G}\subset\mathcal{F}$ *then* $\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}}=\mathbb{E}_{\mathcal{G}_0}=\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\mathcal{G}_0}$.

Proof. We prove each item in turn.

1. If $F\in\mathcal{S}(\mu;X)$, then

$$\|\mathbb{E}_{\mathcal{G}}F\|=\left\|\sum_{x\neq0}x\mathbb{E}_{\mathcal{G}}1_{F=x}\right\|\le\sum_{x\neq0}\|x\|\,\mathbb{E}_{\mathcal{G}}1_{F=x}$$
$$=\mathbb{E}_{\mathcal{G}}\left(\sum_{x\neq0}\|x\|\,1_{F=x}\right)=\mathbb{E}_{\mathcal{G}}\|F\|.$$

For general $F\in L^1(\Omega,\mathcal{F},\mu;X)$ we may choose $F_n\in\mathcal{S}(\mu;X)$ such that $F_n\to F$ in L^1 and therefore, $\|F_n\|\to\|F\|$, $\|\mathbb{E}_{\mathcal{G}}F_n\|\to\|\mathbb{E}_{\mathcal{G}}F\|$, and $\mathbb{E}_{\mathcal{G}}\|F_n\|\to\mathbb{E}_{\mathcal{G}}\|F\|$ in $L^1(\mu)$ as $n\to\infty$. Thus we may pass to the limit in the inequality $\|\mathbb{E}_{\mathcal{G}}F_n\|\le\mathbb{E}_{\mathcal{G}}\|F_n\|$ in order to show $\|\mathbb{E}_{\mathcal{G}}F\|\le\mathbb{E}_{\mathcal{G}}\|F\|$ a.s..
Alternative proof. If $\varphi\in X^*$ with $\|\varphi\|_{X^*}=1$, we have

$$|\varphi\circ\mathbb{E}_{\mathcal{G}}F|=|\mathbb{E}_{\mathcal{G}}[\varphi\circ F]|\le\mathbb{E}_{\mathcal{G}}|[\varphi\circ F]|\le\mathbb{E}_{\mathcal{G}}\|F\|\ \text{ a.s.}$$

Therefore if we let $\{\varphi_n\}$ be as defined in the proof of Proposition 39.1, then

$$\|\mathbb{E}_{\mathcal{G}}F\|=\sup_n|\varphi_n\circ\mathbb{E}_{\mathcal{G}}F|\le\mathbb{E}_{\mathcal{G}}\|F\|\ \text{ a.s.}$$

2. For the second item we have,

$$\|\mathbb{E}_{\mathcal{G}}F\|^p_{L^p(\mu)}=\mathbb{E}[\|\mathbb{E}_{\mathcal{G}}F\|^p]\le\mathbb{E}[(\mathbb{E}_{\mathcal{G}}\|F\|)^p]\le\mathbb{E}[\mathbb{E}_{\mathcal{G}}[\|F\|^p]]=\mathbb{E}\|F\|^p.$$

3. If $\varphi\in X^*$ then

$$\begin{aligned}\varphi(h\cdot\mathbb{E}_{\mathcal{G}}[F])&=h\cdot\varphi\circ\mathbb{E}_{\mathcal{G}}[F]=h\cdot\mathbb{E}_{\mathcal{G}}[\varphi\circ F]\\&=\mathbb{E}_{\mathcal{G}}[h\cdot\varphi\circ F]=\mathbb{E}_{\mathcal{G}}[\varphi(h\cdot F)]\\&=\varphi(\mathbb{E}_{\mathcal{G}}[h\cdot F]).\end{aligned}$$

Since $\varphi\in X^*$ was arbitrary it follows that $\mathbb{E}_{\mathcal{G}}[h\cdot F]=h\cdot\mathbb{E}_{\mathcal{G}}[F]$ a.s.

4. If $\varphi\in X^*$ then

$$\begin{aligned}\varphi(\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}}F)&=\mathbb{E}_{\mathcal{G}_0}[\varphi(\mathbb{E}_{\mathcal{G}}F)]=\mathbb{E}_{\mathcal{G}_0}[\mathbb{E}_{\mathcal{G}}(\varphi\circ F)]\\&=\mathbb{E}_{\mathcal{G}_0}(\varphi\circ F)=\varphi\circ\mathbb{E}_{\mathcal{G}_0}F\end{aligned}$$

and similarly,

$$\varphi(\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\mathcal{G}_0}F)=\varphi\circ\mathbb{E}_{\mathcal{G}_0}F.$$

Since $\varphi\in X^*$ was arbitrary the result follows.

∎

40.3 Basic Martingale Results

In this section we will write μ for P. Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, P)$ is a filtered probability space, $1 \leq p < \infty$, and let $\mathcal{F}_\infty := \vee_{n=1}^{\infty} \mathcal{F}_n := \sigma\left(\cup_{n=1}^{\infty} \mathcal{F}_n\right)$. We will say that $M_n : \Omega \to X$ is a **martingale** if $\{M_n\}_{n=1}^{\infty}$ is an adapted integrable process such that $\mathbb{E}_{\mathcal{B}_n} M_{n+1} = M_n$ a.s. Notice that if $\{M_n\}$ is a martingale then $X_n := \|M_n\|$ is a positive submartingale.

Lemma 40.16. *The space $\cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P)$.*

Proof. The spaces $L^p(\Omega, \mathcal{F}_n, P)$ form an increasing sequence of closed subspaces of $L^p(\Omega, \mathcal{F}_\infty, P)$. Further let $\mathbb{A}$ be the algebra of functions consisting of those $f \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P)$ such that f is bounded. As a consequence of the density Theorem 12.27 (from the probability notes), we know that $\mathbb{A}$ and hence $\cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P)$. This completes the proof. However for the readers convenience let us quickly review the proof of Theorem 12.27 (from the probability notes) in this context.

Let $\mathbb{H}$ denote those bounded $\mathcal{F}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{A}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded $\mathcal{F}_\infty$ –measurable $\mathbb{R}$ – valued functions on Ω, $1 \in \mathbb{H}$, $\mathbb{A} \subset \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated (or bounded) convergence theorem, $\lim_{n\to\infty} \|(f - f_n)\|_{L^p(P)} = 0.$[2] We may now choose $\varphi_n \in \mathbb{A}$ such that $\|\varphi_n - f_n\|_{L^p(P)} \leq \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} \leq \limsup_{n\to\infty} \|(f - f_n)\|_{L^p(P)} + \limsup_{n\to\infty} \|f_n - \varphi_n\|_{L^p(P)} = 0,$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem, now shows $\mathbb{H}$ contains all bounded $\sigma(\mathbb{A}) = \mathcal{F}_\infty$ – measurable functions on Ω. Since for any $f \in L^p(\Omega, \mathcal{F}, P)$, $f 1_{|f| \leq n} \in \mathbb{H}$ there exists $\varphi_n \in \mathbb{A}$ such that $\|f_n - \varphi_n\|_p \leq n^{-1}$. Using the DCT we know that $f_n \to f$ in L^p and therefore by Minikowski's inequality it follows that $\varphi_n \to f$ in L^p. ∎

Corollary 40.17. *The space $\cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P; X)$.*

Proof. Since $\mathcal{S}(\Omega, \mathcal{F}_\infty, P; X)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P; X)$ it suffices to show that every element of $\mathcal{S}(\Omega, \mathcal{F}_\infty, P; X)$ is well approximated by some $G \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$ and for this it suffices to show $1_A \cdot x$ is well approximated by by some $G \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$ for all $x \in X$ and $A \in \mathcal{F}_\infty$. But as a consequence Lemma 40.16 we may find $h \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P)$ such that $\|h - 1_A\|_{L^p(P)}$ is as small as we please and therefore

$$\|1_A \cdot x - h \cdot x\|_{L^p(P)} \leq \|x\| \cdot \|h - 1_A\|_{L^p(P)}$$

can be made as small as we please as well. ∎

Theorem 40.18. *For every $F \in L^p(\Omega, \mathcal{F}, P)$, $M_n = \mathbb{E}[F|\mathcal{F}_n]$ is a martingale and $M_n \to M_\infty := \mathbb{E}[F|\mathcal{F}_\infty]$ in $L^p(\Omega, \mathcal{F}_\infty, P; X)$ as $n \to \infty$.*

Proof. The tower property immediately shows $M_n = \mathbb{E}[F|\mathcal{F}_n]$ is a martingale. Since conditional expectation is a contraction on L^p it follows that $\mathbb{E}\|M_n\|^p \leq \mathbb{E}\|F\|^p < \infty$ for all $n \in \mathbb{N} \cup \{\infty\}$. So to finish the proof we need to show $M_n \to M_\infty$ in $L^p(\Omega, \mathcal{F}, P; X)$ as $n \to \infty$.

If $F \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$, then $M_n = F$ for all sufficiently large n and for $n = \infty$ and the result holds. Now suppose that $F \in L^p(\Omega, \mathcal{F}_\infty, P; X)$ and $G \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$. Then

$$\begin{aligned}\|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_n} F\|_p &\leq \|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_\infty} G\|_p + \|\mathbb{E}_{\mathcal{F}_\infty} G - \mathbb{E}_{\mathcal{F}_n} G\|_p + \|\mathbb{E}_{\mathcal{F}_n} G - \mathbb{E}_{\mathcal{F}_n} F\|_p \\ &\leq 2\|F - G\|_p + \|\mathbb{E}_{\mathcal{F}_\infty} G - \mathbb{E}_{\mathcal{F}_n} G\|_p\end{aligned}$$

and hence

$$\limsup_{n\to\infty} \|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_n} F\|_p \leq 2\|F - G\|_p.$$

Using the density Corollary 40.17 we may choose $G \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$ as close to $F \in L^p(\Omega, \mathcal{F}_\infty, P; X)$ as we please and therefore it follows that $\limsup_{n\to\infty} \|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_n} F\|_p = 0$.

For general $F \in L^p(\Omega, \mathcal{F}, P)$ it suffices to observe that $M_\infty := \mathbb{E}[F|\mathcal{F}_\infty] \in L^p(\Omega, \mathcal{F}_\infty, P)$ and by the tower property of conditional expectations,

$$\mathbb{E}[M_\infty|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[F|\mathcal{F}_\infty]|\mathcal{F}_n] = \mathbb{E}[F|\mathcal{F}_n] = M_n.$$

So again $M_n \to M_\infty$ in L^p as desired. ∎

The converse of Theorem 40.18 holds as well but is not really needed for our purposes. It use compactness results from the probability notes which need to be transferred here.

Theorem 40.19 (Probably should skip). *Suppose $1 \leq p < \infty$ and $\{M_n\}_{n=1}^{\infty} \subset L^p(\Omega, \mathcal{F}, P; X)$ is a martingale. Further assume that $\sup_n \|M_n\|_p < \infty$ and that $\{M_n\}_{n=1}^{\infty}$ is uniformly integrable if $p = 1$. Then there exists $M_\infty \in L^p(\Omega, \mathcal{F}_\infty, P; X)$ such that $M_n := \mathbb{E}[M_\infty|\mathcal{F}_\infty]$. Moreover by Theorem 40.18 we know that $M_n \to M_\infty$ in $L^p(\Omega, \mathcal{F}_\infty, P)$ as $n \to \infty$ and hence M_∞ is uniquely determined by $\{M_n\}_{n=1}^{\infty}$.*

[2] It is at this point that the proof would break down if $p = \infty$.

Proof. By Theorems 13.20 and 13.22 exists $M_\infty \in L^p(\Omega, \mathcal{F}_\infty, P)$ and a subsequence, $Y_k = M_{n_k}$ such that

$$\lim_{k\to\infty} \mathbb{E}[Y_k h] = \mathbb{E}[M_\infty h] \text{ for all } h \in L^q(\Omega, \mathcal{F}_\infty, P)$$

where $q := p(p-1)^{-1}$. Using the martingale property, if $h \in (\mathcal{F}_n)_b$ for some n, it follows that $\mathbb{E}[Y_k h] = \mathbb{E}[M_n h]$ for all large k and therefore that

$$\mathbb{E}[M_\infty h] = \mathbb{E}[M_n h] \text{ for all } h \in (\mathcal{F}_n)_b.$$

This implies that $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$ as desired. ∎

Theorem 40.20 (Almost sure convergence). *Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^\infty, P)$ is a filtered probability space, $1 \le p < \infty$, and let $\mathcal{F}_\infty := \vee_{n=1}^\infty \mathcal{F}_n := \sigma(\cup_{n=1}^\infty \mathcal{F}_n)$. Then for every $F \in L^1(\Omega, \mathcal{F}, P; X)$, the martingale, $M_n = \mathbb{E}[F|\mathcal{F}_n]$, converges almost surely to $M_\infty := \mathbb{E}[F|\mathcal{F}_\infty]$.*

Proof. We follow the proof in Stroock [50, Corollary 5.2.7]. Let $\mathbb{H}$ denote those $F \in L^1(\Omega, \mathcal{F}_\infty, P; X)$ such that $M_n := \mathbb{E}[F|\mathcal{F}_n] \to M_\infty = F$ a.s. As we saw above $\mathbb{H}$ contains the dense subspace $\cup_{n=1}^\infty L^1(\Omega, \mathcal{F}_n, P; X)$. It is also easy to see that $\mathbb{H}$ is a linear space. Thus it suffices to show that $\mathbb{H}$ is closed in $L^1(P; X)$. To prove this let $F^{(k)} \in \mathbb{H}$ with $F^{(k)} \to F$ in $L^1(P)$ and let $M_n^{(k)} := \mathbb{E}\left[F^{(k)}|\mathcal{F}_n\right]$. Then by the Doob's maximal inequality applied to the sub-martingale $\left\{\left\|M_n - M_n^{(k)}\right\|\right\}_{n=1}^\infty$ we have

$$P\left(\sup_n \left\|M_n - M_n^{(k)}\right\| \ge a\right) \le \frac{1}{a}\sup_n \mathbb{E}\left\|M_n - M_n^{(k)}\right\| \le \frac{1}{a}\mathbb{E}\left\|F - F^{(k)}\right\|$$

for all $a > 0$ and $k \in \mathbb{N}$. Therefore,

$$\begin{aligned} P&\left(\sup_{n\ge N}\|F - M_n\| \ge 3a\right) \\ &\le P\left(\left\|F - F^{(k)}\right\| \ge a\right) + P\left(\sup_{n\ge N}\left\|F^{(k)} - M_n^{(k)}\right\| \ge a\right) \\ &\quad + P\left(\sup_{n\ge N}\left\|M_n^{(k)} - M_n\right\| \ge a\right) \\ &\le \frac{2}{a}\mathbb{E}\left\|F - F^{(k)}\right\| + P\left(\sup_{n\ge N}\left\|F^{(k)} - M_n^{(k)}\right\| \ge a\right) \end{aligned}$$

and hence

$$\limsup_{N\to\infty} P\left(\sup_{n\ge N}\|F - M_n\| \ge 3a\right) \le \frac{2}{a}\mathbb{E}\left\|F - F^{(k)}\right\| \to 0 \text{ as } k \to \infty.$$

Thus we have shown

$$\limsup_{N\to\infty} P\left(\sup_{n\ge N}\|F - M_n\| \ge 3a\right) = 0 \text{ for all } a > 0.$$

Since

$$\left\{\limsup_{n\to\infty}\|F - M_n\| \ge 3a\right\} \subset \left\{\sup_{n\ge N}\|F - M_n\| \ge 3a\right\} \text{ for all } N,$$

it follows that

$$P\left(\limsup_{n\to\infty}\|F - M_n\| \ge 3a\right) = 0 \text{ for all } a > 0$$

and therefore $\limsup_{n\to\infty}\|F - M_n\| = 0$ (P a.s.) which shows that $F \in \mathbb{H}$. ∎

Abstract Wiener Space Overview

Suppose that X is a real separable Banach space and $\mathcal{B}_X$ is the Borel $\sigma-$ algebra on X.

Definition 41.1. *A measure μ on $(X, \mathcal{B}_X)$ is called a (mean zero, non-degenerate)* ***Gaussian measure*** *provided that its characteristic functional is given by*

$$\hat{\mu}(u) := \int_X e^{iu(x)} d\mu(x) = e^{-\frac{1}{2}q(u,u)} \text{ for all } u \in X^*, \tag{41.1}$$

where $q = q_\mu : X^* \times X^* \to \mathbb{R}$ is a quadratic form such that $q(u,v) = q(v,u)$ and $q(u) = q(u,u) \geqslant 0$ with equality iff $u = 0$, i.e. q is a real inner product on X^*.

Proposition 41.2. *Suppose that $(X, \mathcal{B} = \mathcal{B}_X, \mu)$ is a Gaussian probability space and q is the inner product on X^* defined by*

$$\hat{\mu}(\varphi) = e^{-\frac{1}{2}q(\varphi,\varphi)}.$$

1. *Suppose that $\{\varphi_1, \dots \varphi_n\} \subset X^*$ is a linearly independent set and let A be the $n \times n$ real positive matrix such that $A_{ij}^{-1} = q(\varphi_i, \varphi_j)$.*[1] *Then, for all bounded measurable functions $f : \mathbb{R}^n \to \mathbb{R}$,*
$$\int_X f(\varphi_1, \dots, \varphi_n) d\mu = \int_{\mathbb{R}^n} f(y_1, \dots, y_n) d\mu_A(y) \tag{41.2}$$
where
$$d\mu_A(y) = \frac{1}{Z} e^{-\frac{1}{2}Ay\cdot y}\, dy$$
and $Z = (2\pi)^{n/2} \det^{-1/2}(A)$.

[1] Notice that $(q(\varphi_i, \varphi_j))_{i,j=1}^n$ is an invertible matrix since if $x \in \mathbb{R}^n$ such that

$$0 = \sum_j q(\varphi_i, \varphi_j) x_j = q\left(\varphi_i, \sum_j x_j \varphi_j\right),$$

then we may multiply this equation by x_i and sum to learn that $0 = q\left(\sum_j x_j\varphi_j, \sum_j x_j\varphi_j\right)$ which implies (since q is an inner product) that $\sum_j x_j \varphi_j = 0$. The linear independence of $\{\varphi_j\}_{j=1}^n$ implies that $x = 0$ so that $\mathrm{Nul}\left((q(\varphi_i, \varphi_j))_{i,j=1}^n\right) = \{0\}$.

2. *If $\{\varphi_1, \dots, \varphi_n\} \subset X^*$ is an orthonormal set, i.e. $q(\varphi_i, \varphi_j) = \delta_{ij}$, then for all bounded measurable functions $f : \mathbb{R}^n \to \mathbb{R}$,*
$$\int_X f(\varphi_1, \dots, \varphi_n) d\mu = \int_{\mathbb{R}^n} f(x_1, \dots, x_n) p_1(dx_1) \dots p_1(dx_n), \tag{41.3}$$
where $p_1(dx)$ is the heat kernel on $\mathbb{R}$ defined by
$$p_1(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{41.4}$$
3. *In particular,*
$$q(\varphi) := q(\varphi, \varphi) = \int_X \varphi^2(x) d\mu(x) \text{ and}$$
$$q(\varphi, \psi) = \int_X \varphi(x)\psi(x) d\mu(x).$$
4. *A probability measure μ on $(X, \mathcal{B})$ is a Gaussian measure space iff for all $0 \neq \varphi \in X^*$ there exists $0 \neq K_\varphi \in \mathbb{R}$ such that*
$$\int_X e^{i\lambda\varphi(x)} d\mu(x) = e^{-\frac{1}{2}\lambda^2 K_\varphi}\ \forall\, \lambda \in \mathbb{R}. \tag{41.5}$$
(In this case $q_\mu(\varphi, \varphi) = K_\varphi$.) This may be rephrased as saying the μ is Gaussian iff $\varphi_*\mu$ is Gaussian on $\mathbb{R}$ for all $\varphi \in X^* \setminus \{0\}$.

Proof. For the proof of Item 1. and Item 2., we first observe, for $\lambda \in \mathbb{R}^n$, that

$$\begin{aligned}
&\int_{\mathbb{R}^n} e^{\lambda\cdot y} d\mu_A(y) \\
&\quad = \frac{1}{Z}\int_{\mathbb{R}^n} e^{-\frac{1}{2}Ay\cdot y} e^{\lambda\cdot y} dy \\
&\quad = \frac{1}{Z}\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}A\left(y - A^{-1}\lambda\right)\cdot\left(y - A^{-1}\lambda\right) + \frac{1}{2}AA^{-1}\lambda\cdot A^{-1}\lambda\right) dy \\
&\quad = e^{\frac{1}{2}A^{-1}\lambda\cdot\lambda}\frac{1}{Z}\int_{\mathbb{R}^n} e^{-\frac{1}{2}Ay\cdot y} dy = e^{\frac{1}{2}A^{-1}\lambda\cdot\lambda}.
\end{aligned}$$

A simple analytic continuation argument shows that we may replace λ by $i\lambda$ is this formula to learn

$$\hat{\mu}_A(\lambda) := \int_{\mathbb{R}^n} e^{i\lambda\cdot y} d\mu_A(y) = e^{-\frac{1}{2}A^{-1}\lambda\cdot\lambda}.$$

Secondly, let $\Phi := (\varphi_1, \dots, \varphi_n) : X \to \mathbb{R}^n$ and let $\nu := \Phi_*\mu$ so that ν is a probability measure on $\mathbb{R}^n$. The Fourier transform of ν is given by

$$\begin{aligned}\hat{\nu}(\lambda) &= \int_{\mathbb{R}^n} e^{i\lambda\cdot y} d\nu(y) = \int_X e^{i\lambda\cdot\Phi(x)} d\mu(x) \\ &= \int_X \exp\left(i\sum_{i=1}^n \lambda_i\varphi_i(x)\right) d\mu(x) = \exp\left(-\frac{1}{2}q(\sum_{i=1}^n \lambda_i\varphi_i, \sum_{i=1}^n \lambda_i\varphi_i)\right) \\ &= \exp\left(-\frac{1}{2}\sum_{i,j=1}^n \lambda_i\lambda_j q(\varphi_i,\varphi_j)\right) = \exp\left(-\frac{1}{2}\sum_{i,j=1}^n \lambda_i\lambda_j A_{ij}^{-1}\right) \\ &= \hat{\mu}_A(\lambda).\end{aligned}$$

Therefore by uniqueness of the Fourier transform, $\nu = \mu_A$, or equivalently

$$\frac{1}{Z}\int_{\mathbb{R}^n} f(y_1, \dots, y_n) e^{-\frac{1}{2}Ay\cdot y}\, dy = \int_X f(\varphi_1(x), \dots, \varphi_n(x)) d\mu(x)$$

for all bounded measurable functions $f : \mathbb{R}^n \to \mathbb{R}$.

Item 3. is now a simple exercise. For Item 4. Eq. (41.5) implies

$$\int_X f(\varphi(x)) d\mu(x) = \frac{1}{Z_\varphi}\int_{\mathbb{R}} f(z)\exp\left(-\frac{1}{2K_\varphi}z^2\right)\, dz$$

for all $f \in (\mathcal{B}_{\mathbb{R}})_b$. In particular $\varphi(\cdot) \in L^2(\mu)$ and

$$\int_X \varphi^2(x) d\mu(x) = \frac{1}{Z_\varphi}\int_{\mathbb{R}} z^2\exp\left(-\frac{1}{2K_\varphi}z^2\right)\, dz = K_\varphi.$$

Therefore $K_\varphi = q(\varphi,\varphi)$ where q is defined by

$$q(\varphi,\psi) = \int_X \varphi(x)\psi(x) d\mu(x).$$

■

Proposition 41.3 (A no go theorem). *If H is an infinite dimensional Hilbert space, then there is no probability measure μ on $\mathcal{B} = \mathcal{B}_H$ such that*

$$\hat{\mu}(u) = \exp\left(-\frac{1}{2}q(u,u)\right) \tag{41.6}$$

where

$$q(u,u) := \|u\|_{H^*}^2 = \sum_{k=1}^\infty |u(e_k)|^2$$

where $\{e_k\}_{k=1}^\infty$ is any orthonormal basis for H.

Proof. Suppose such a measure μ were to exists. Let $\{e_i\}_{i=1}^\infty$ be an orthonormal set in H and for $n \in \mathbb{N}$ and $x \in H$ let $T_n(x) = ((x,e_1), \dots, (x,e_n)) \in \mathbb{R}^n$ and $\mu_n := (T_n)_*\mu = \mu \circ T_n^{-1}$. Assuming Eq. (41.6) holds, we learn that

$$\begin{aligned}\int_{\mathbb{R}^n} e^{i\xi\cdot a} d\mu_n(a) &= \int_H e^{i\xi\cdot T_n(x)} d\mu(x) \\ &= \int_H \exp i\left(\sum_{j=1}^n \xi_j e_j, x\right) d\mu(x) = e^{-\frac{1}{2}\xi\cdot\xi}.\end{aligned}$$

Since

$$\int_{\mathbb{R}^n} e^{i\xi\cdot a}\frac{1}{(2\pi)^{n/2}}e^{-a\cdot a/2}da = e^{-\frac{1}{2}\xi\cdot\xi}$$

it follows from Proposition 41.2 that

$$d\mu_n(a) = \frac{1}{(2\pi)^{n/2}}e^{-a\cdot a/2}da.$$

For $M > 0$ and $n \in \mathbb{N}$, let

$$\begin{aligned}W_n^M &= \{x \in H : |(e_i,x)| \le M \text{ for } i = 1,2,\dots,n\} \\ &= \{x \in H : |[T_n(x)]_i| \le M \text{ for } i = 1,2,\dots,n\}.\end{aligned}$$

Then by what we have just proved

$$\begin{aligned}\mu(W_n^M) &= \mu_n(\{a \in \mathbb{R}^n : |(e_i,x)| \le M \text{ for } i = 1,\dots,n\}) \\ &= \left((2\pi)^{-1/2}\int_{-M}^M e^{-\frac{1}{2}a^2}da\right)^n \to 0 \text{ as } n \to \infty.\end{aligned}$$

Since

$$W_n^M \downarrow H_M \equiv \{x \in H : |(e_i,x)| \le M\ \forall\, i = 1,2,\dots\} \text{ as } n \to \infty,$$

this would imply $\mu(H_M) = \lim_{n\to\infty}\mu(W_n^M) = 0$ for all $M > 0$. Since $H_M \uparrow H$ as $M \uparrow \infty$ we learn that $\mu(H) = \lim_{M\to\infty}\mu(H_M) = 0$, i.e. $\mu \equiv 0$. ■

Moral: The measure μ must be defined on a larger space. This is somewhat analogous to trying to define Lebesgue measure on the rational numbers. In each case the measure can only be defined on a certain completion of the naive initial starting space. With this in mind we go back to the general theory.

In what follows we frequently make use of the fact that

$$C_p := \int_X \|x\|_X^p \, d\mu(x) < \infty \text{ for all } 1 \leqslant p < \infty. \tag{41.7}$$

This is a consequence of Skorohod's inequality (see for example [34, Theorem 3.2])

$$\int_X e^{\lambda \|x\|_X} d\mu(x) < \infty \text{ for all } \lambda < \infty; \tag{41.8}$$

or the even stronger Fernique's theorem.

Theorem 41.4 (Ferniques Theorem). *Suppose that μ is a Gaussian measure on $(X, \mathcal{B}_X)$. Then there exists $\varepsilon = \varepsilon(\mu) > 0$ such that*

$$\int_X e^{\varepsilon \|x\|_X^2} d\mu(x) < \infty. \tag{41.9}$$

In particular μ has moments to all orders.

For the proof of this theorem see Theorem 43.3 below (also see [8, Theorem 2.8.5] or [34, Theorem 3.1]).

Lemma 41.5. *If $u, v \in X^*$, then*

$$\int_X u(x)\, v(x)\, d\mu(x) = q(u, v) \tag{41.10}$$

and

$$|q(u,v)| \leqslant C_2 \|u\|_{X^*} \|v\|_{X^*}. \tag{41.11}$$

Proof. Let $u_*\mu := \mu \circ u^{-1}$ denote the law of u under μ. Then by Eq. (41.1),

$$(u_*\mu)(dx) = \frac{1}{\sqrt{2\pi q(u,u)}} e^{-\frac{1}{2q(u,u)} x^2} dx$$

and hence,

$$\int_X u^2(x)\, d\mu(x) = q_\mu(u,u) = q(u,u). \tag{41.12}$$

Polarizing this identity gives Eq. (41.10) which along with Eq. (41.7) implies Eq. (41.11). ∎

Alternate Proof of the Existence of $C_2 < \infty$ such that Eq. (41.11) holds. Note that $q(u,u) = -2\ln[\hat{\mu}(u)]$, and so if $u_n \in X^*$ and $u_n \to 0$ weakly, that is, $u_n(x) \to 0$ as $n \to \infty$ for all $x \in X$, then by the dominated convergence theorem, $q(u_n) \to -2\ln[\hat{\mu}(0)] = 0$. More generally, if $u_n, v_n \to 0$ weakly, then $|q(u_n, v_n)| \leqslant \sqrt{q(u_n)}\sqrt{q(v_n)} \to 0$ as $n \to \infty$. In particular, there exists $C < \infty$ such that $|q(u,v)| \leqslant C\|u\|_{X^*}\|v\|_{X^*}$. Indeed, if no such $C < \infty$ existed, then there would exist u_n, v_n such that $\|u_n\|_{X^*} = \|v_n\|_{X^*} = 1$ while $|q(u_n, v_n)| \geqslant n^2$. We would then have $|q(u_n/n, v_n/n)| \geqslant 1$ while $u_n/n \to 0$ and, $v_n/n \to 0$.

The next theorems summarizes some well known properties of Gaussian measures that we will use freely below. This well known material may be (mostly) found in the books [34] and [8]. In particular, the following theorem is based in part on [8, Lemma 2.4.1 on p. 59] and [8, Theorem 3.9.6 on p. 138].

Theorem 41.6. *Let X be a real separable Banach space and $(X, \mathcal{B}_X, \mu)$ be a Gaussian measure space as in Definition 41.1 For $x \in X$ let*

$$\|x\|_H := \sup_{u \in X^* \setminus \{0\}} \frac{|u(x)|}{\sqrt{q(u,u)}} \tag{41.13}$$

*and define the **Cameron-Martin subspace,** $H \subset X$, by*

$$H = \{h \in X : \|h\|_H < \infty\}. \tag{41.14}$$

Then;

1. *$(H, \|\cdot\|_H)$ is a normed space such that*
$$\|h\|_X \leqslant \sqrt{C_2}\, \|h\|_H \quad \text{for all } h \in H, \tag{41.15}$$
where C_2 is as in (41.7).
2. *For $f \in \operatorname{Re} L^2(\mu)$, let*
$$Jf := h_f := \int_X x\ f(x)\ d\mu(x) \in X, \tag{41.16}$$
where the integral is to be interpreted as a Bochner integral.[2]. Then $Jf \in H$ and $J : \operatorname{Re} L^2(\mu) \to H$ is a contraction. (See Remark 42.3 below for motivation for defining Jf as in Eq. (41.16).)
3. *Now suppose that K is the closure of X^* in $\operatorname{Re} L^2(\mu)$. Then $J_K := J|_K : K \to H$ is an isometry.*
4. *Moreover, $J(K) = H$ and therefore $J_K : K \to H$ is an isometric isomorphism of real Banach spaces. Since K is a real Hilbert space it follows that $\|\cdot\|_H$ is a Hilbertian norm on H.*

[2] Notice that
$$\int_X \|x f(x)\|\, d\mu(x) \leq \sqrt{C_2}\, \|f\|_{L^2(\mu)} < \infty,$$
so the integrand is indeed Bochner integrable.

5. *H is a separable Hilbert space and*

$$(Ju, h)_H = u(h) \text{ for all } u \in X^* \text{ and } h \in H. \tag{41.17}$$

6. *The quadratic form q may be computed as*

$$q(u, v) = \sum_{i=1}^{\infty} u(e_i)v(e_i) \tag{41.18}$$

where $\{e_i\}_{i=1}^{\infty}$ is any orthonormal basis for H.

7. *The Cameron-Martin space, H, is dense in X.*

Notice that by Item 1. $H \overset{i}{\hookrightarrow} X$ is continuous and hence so is $X^* \overset{i^{tr}}{\hookrightarrow} H^* \cong H = (\cdot,\cdot)_{H^*}$. Eq. (41.18) asserts that

$$q = (\cdot,\cdot)_{H^*}\big|_{X^* \times X^*}.$$

Proof. We will prove each item in turn.

1. Using Eq. (41.11) we find

$$\|h\|_X = \sup_{u \in X^* \setminus \{0\}} \frac{|u(h)|}{\|u\|_{X^*}} \leqslant \sup_{u \in X^* \setminus \{0\}} \frac{|u(h)|}{\sqrt{q(u,u)/C_2}} \leqslant \sqrt{C_2}\, \|h\|_H,$$

and hence if $\|h\|_H = 0$ then $\|h\|_X = 0$ and so $h = 0$. If $h, k \in H$, then for all $u \in X^*$, $|u(h)| \leqslant \|h\|_H \sqrt{q(u)}$ and $|u(k)| \leqslant \|k\|_H \sqrt{q(u)}$ so that

$$|u(h+k)| \leqslant |u(h)| + |u(k)| \leqslant (\|h\|_H + \|k\|_H)\sqrt{q(u)}.$$

This shows $h + k \in H$ and $\|h+k\|_H \leqslant \|h\|_H + \|k\|_H$. Similarly, if $\lambda \in \mathbb{R}$ and $h \in H$, then $\lambda h \in H$ and $\|\lambda h\|_H = |\lambda|\, \|h\|_H$. Therefore H is a subspace of W and $(H, \|\cdot\|_H)$ is a normed space.

2. For $f \in \operatorname{Re} L^2(\mu)$ and $u \in X^*$

$$u(Jf) = u\left(\int_X xf(x)d\mu(x)\right) = \int_X u(x)f(x)d\mu(x) = (u, f)_{L^2(\mu)} \tag{41.19}$$

and hence

$$|u(Jf)| \leqslant \|u\|_{L^2(\mu)} \|f\|_{L^2(\mu)} = \sqrt{q(u)}\, \|f\|_{L^2(\mu)}$$

which shows that $Jf \in H$ and $\|Jf\|_H \leqslant \|f\|_{L^2(\mu)}$.

3. Let $f \in K$ and choose $u_n \in X^*$ such that $L^2(\mu) - \lim_{n\to\infty} u_n = f$. Then

$$\lim_{n\to\infty} \frac{|u_n(Jf)|}{\sqrt{q(u_n)}} = \lim_{n\to\infty} \frac{\left|\int_X u_n(x)f(x)d\mu(x)\right|}{\|u_n\|_{L^2(\mu)}} = \frac{\|f\|_{L^2(\mu)}^2}{\|f\|_{L^2(\mu)}} = \|f\|_{L^2(\mu)}$$

from which it follows $\|Jf\|_H = \|f\|_K$. So we have shown that $J : K \to H$ is an isometry.

4. We now wish to show that $J_K := J|_K : K \to H$ is surjective, i.e. given $h \in H$ we are looking for an $f \in K$ such that

$$h = Jf = \int_X xf(x)\, d\mu(x).$$

This will be the case iff

$$\hat{h}(u) := u(h) = u(Jf) = \int_X u(x) f(x)\, d\mu(x) = (u, f)_K \text{ for all } u \in X^*.$$

In order to see that this equation has a solution f, notice that

$$\left|\hat{h}(u)\right| = |u(h)| \leqslant \sqrt{q(u)}\, \|h\|_H = \|u\|_{L^2(\mu)} \|h\|_H = \|u\|_K \|h\|_H$$

for all $u \in X^*$ which is dense in K. Therefore $\hat{h}$ extends continuously to K and so by the Riesz representation theorem for Hilbert spaces, there exists an $f \in K$ such that $\hat{h}(u) = (u, f)_K$ for all $u \in X^* \subset K$.

5. H is a separable since it is unitarily equivalent to $K \subset L^2(X, \mathcal{B}, \mu)$ and $L^2(X, \mathcal{B}, \mu)$ is separable. Suppose that $u \in X^*$, $f \in K$ and $h = Jf \in H$. Then

$$\begin{aligned}(Ju, h)_H &= (Ju, Jf)_H = (u, f)_K \\ &= \int_X u(x)f(x)d\mu(x) = u\left(\int_X xf(x)d\mu(x)\right) \\ &= u(Jf) = u(h).\end{aligned}$$

6. Let $\{e_i\}_{i=1}^{\infty}$ be an orthonormal basis for H, then for $u, v \in X^*$,

$$\begin{aligned}q(u, v) = (u, v)_K = (Ju, Jv)_H &= \sum_{i=1}^{\infty} (Ju, e_i)_H (e_i, Jv)_H \\ &= \sum_{i=1}^{\infty} u(e_i)v(e_i)\end{aligned}$$

wherein the last equality we have again used Eq. (41.17).

7. If $u \in X^*$ such that $u|_H = 0$, then by Eq. (41.18) it follows that $q(u,u) = 0$ and since q is an inner product we must have $u = 0$. Alternatively this last assertion follows from Eq. (41.17);
$$q(u,u) = (Ju, Ju)_H = u(Ju) = 0.$$
It now follows as a consequence of the Hahn–Banach theorem that H must be a dense subspace of X.

■

Remark 41.7. Let $f \in K$ and $\varphi_n \in X^*$ such that $f = L^2(\mu) - \lim_{n\to\infty} \varphi_n$. Then by passing to a subsequence if necessary, we may assume that $\varphi_n \to f$ a.s. as $n \to \infty$. So by the Dominated convergence theorem,
$$\int_X e^{if} d\mu = \lim_{n\to\infty} \int_X e^{i\varphi_n} d\mu = e^{-q(\varphi_n)/2}$$
$$= \lim_{n\to\infty} \exp\left(-\frac{1}{2}\|\varphi_n\|^2_{L^2(\mu)}\right) = \lim_{n\to\infty} \exp\left(-\frac{1}{2}\|f\|^2_{L^2(\mu)}\right).$$
This computation shows that $K \subset L^2(\mu)$ consists of jointly Gaussian random variables.

Lemma 41.8. *Let $(X, \mathcal{B}, \mu)$ be a Gaussian measure space. Then for any $\{u_k\}_{k=1}^\infty \subset X^* \subset K$ which is an orthonormal basis for K we have*
$$\|x\|^2_H = \sum_{k=1}^\infty |u_k(x)|^2 \text{ for all } x \in X. \tag{41.20}$$
In particular,
$$H = \left\{x \in X : \sum_{k=1}^\infty |u_k(x)|^2 < \infty\right\}.$$

Proof. Let $\{u_k\}_{k=1}^\infty \subset X^*$ be an orthonormal basis for K and let $e_k := Ju_k \in H$ so that $\{e_k\}_{k=1}^\infty$ is an orthonormal basis for H. When $h \in H$ we have, using Eq. (41.17), that
$$\|h\|^2_H = \sum_{k=1}^\infty (h, e_k)^2_H = \sum_{k=1}^\infty (h, Ju_k)^2_H = \sum_{k=1}^\infty |u_k(h)|^2.$$

Now suppose that $x \in X$ satisfies $M := \sum_{k=1}^\infty |u_k(x)|^2 < \infty$. Let $K_0 := \mathrm{span}\{u_k : k = 1, 2, \dots\}$ be the **algebraic** span of $\{u_k\}_{k=1}^\infty$. Then for $\varphi := \sum_{k=1}^n \alpha_k u_k \in K_0$ we have

$$|\hat{x}(\varphi)|^2 = |\varphi(x)|^2 = \left|\sum_{k=1}^n \alpha_k u_k(x)\right|^2 \le \sum_{k=1}^n |\alpha_k|^2 \cdot \sum_{k=1}^n |u_k(x)|^2 \le M \cdot \|\varphi\|^2_K.$$
As K_0 is dense in K it follows that $\hat{x}$ extends to a continuous linear functional on K and therefore $\hat{x} = (\cdot, f)_K$ for some $f \in K$. But this then implies, using Eq. (41.19), that
$$\varphi(x) = \hat{x}(\varphi) = (\varphi, f)_K = \varphi(Jf) \text{ for all } \varphi \in X^*.$$
Hence as a consequence of the Hanh Banach theorem, we know that $x = Jf \in H$.

Combining the previous two observations it follows that $\sum_{k=1}^\infty |u_k(x)|^2 = \infty$ implies that $x \notin H$, i.e. that $\|x\|_H = \infty$ and if $\sum_{k=1}^\infty |u_k(x)|^2 < \infty$ then $x \in H$ and again Eq. (41.20) holds. ■

Proposition 41.9. *Let μ be a Gaussian measure on $(X, \mathcal{B})$ as above. If $\dim X = \infty$, then $\mu(H) = 0$ where H is the Cameron-Martin space associated to μ.*

Proof. By Lemma 41.8 we know that
$$H = \left\{x \in X : \sum_{k=1}^\infty |u_k(x)|^2 < \infty\right\}$$
where $\{u_k\}_{k=1}^\infty \subset X^*$ is an orthonormal basis for K. However, relative to μ, $\{u_k\}_{k=1}^\infty$ are i.i.d. standard normal random variables. Hence it follows by the strong law of large numbers that
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n |u_k(x)|^2 = 1 \text{ for } \mu - \text{a.e. } x.$$
On the other hand, for all $x \in H$ we will have
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n |u_k(x)|^2 = 0$$
and therefore $\mu(H) = 0$. ■

Lemma 41.10. *Suppose that $(X, \mathcal{B}, \mu, q = q_\mu)$ is a Gaussian measure space as above. If $\tilde{H} \subset X$ is a subspace of X which is equipped with a Hilbertian inner product, $(\cdot,\cdot)_{\tilde{H}}$, such that*
$$q_\mu(u,u) = \|u|_{\tilde{H}}\|^2_{\tilde{H}^*} \text{ for all } u \in X^*,$$
then $\tilde{H} = H_\mu$.

Proof. Recall that

$$H_\mu = \left\{ h \in X : \|h\|_{H_\mu}^2 := \sup_{u \in X^* \setminus \{0\}} \frac{|u(h)|^2}{q(u,u)} < \infty. \right\}$$

Thus if $\tilde{h} \in \tilde{H}$ and $u \in X^*$, we have

$$\left| u\left(\tilde{h}\right) \right| \le \|u|_{\tilde{H}}\|_{\tilde{H}^*} \left\| \tilde{h} \right\|_{\tilde{H}} = \sqrt{q_\mu(u,u)} \cdot \left\| \tilde{h} \right\|_{\tilde{H}}$$

from which it follows that $\tilde{h} \in H_\mu$. This shows that $\tilde{H} \subset H_\mu$ and that $\left\| \tilde{h} \right\|_{H_\mu} \le \left\| \tilde{h} \right\|_{\tilde{H}}$ for all $\tilde{h} \in \tilde{H}$.

Similarly if $h \in H_\mu$, then

$$|u(h)| \le \sqrt{q_\mu(u,u)} \cdot \|h\|_H = \|u|_{\tilde{H}}\|_{\tilde{H}^*} \|h\|_H . \tag{41.21}$$

To complete the proof we now need to argue that $\{u|_{\tilde{H}} : u \in X^*\}$ is dense in $\tilde{H}^*$. If this were not the case, then there would be a $\varphi = \left(\cdot, \tilde{h}\right)_{\tilde{H}} \in \tilde{H}^* \setminus \{0\}$ such that

$$0 = (u|_{\tilde{H}}, \varphi)_{\tilde{H}^*} = u\left(\tilde{h}\right) \text{ for all } u \in X^*.$$

Hence as a consequence of the Hahn-Banach theorem we find that $\tilde{h}$ must be zero and hence $\varphi = 0$ which is a contradiction.

Let $\hat{h}(u) = u(h)$ for all $u \in X^*$, Eq. (41.21) may be stated as,

$$\left| \hat{h}(u) \right| \le \|h\|_H \cdot \|u|_{\tilde{H}}\|_{\tilde{H}^*} \text{ for all } u \in X^*. \tag{41.22}$$

Notice that if $u|_{\tilde{H}} = 0$ then $\hat{h}(u) = 0$ so $\hat{h}$ may be considered to be a functional on $\{u|_{\tilde{H}} : u \in X^*\}$ and by Eq. (41.22) $\hat{h}$ extends to an element of $\tilde{H}^{**} \cong \tilde{H}$. Thus there is a $\tilde{h} \in \tilde{H}$ such that

$$u(h) = \hat{h}(u) = \left(\tilde{h}\right)^{\hat{}}(u) = u\left(\tilde{h}\right) \text{ for all } u \in X^*.$$

Another application of the Hahn-Banach theorem allows us to conclude that $h = \tilde{h} \in \tilde{H}$, i.e. $H \subset \tilde{H}$. Thus we have shown that $H = \tilde{H}$ and since

$$\|u|_H\|_{H^*}^2 = q_\mu(u,u) = \|u|_{\tilde{H}}\|_{\tilde{H}^*}^2$$

it follows that $(\cdot,\cdot)_H = (\cdot,\cdot)_{\tilde{H}}$. ∎

Remark 41.11. From Eq. (41.15), if $u \in X^*$ then $u|_H \in H^*$. Therefore we have

$$\int_X u^2 d\mu = q(u,u) = \|u\|_{H^*}^2 = \sum_{j=1}^\infty |u(e_j)|^2, \tag{41.23}$$

where $\{e_j\}_{j=1}^\infty$ is an orthonormal basis for H. More generally, if φ is a linear bounded map from X to $\mathbf{C}$, where $\mathbf{C}$ is a real Hilbert space, then

$$\|\varphi\|_{H^* \otimes \mathbf{C}}^2 =: \sum_{j=1}^\infty \|\varphi(e_j)\|_{\mathbf{C}}^2 = \int_X \|\varphi(x)\|_{\mathbf{C}}^2 d\mu(x) < \infty. \tag{41.24}$$

To prove Eq. (41.24), let $\{f_j\}_{j=1}^\infty$ be an orthonormal basis for $\mathbf{C}$. Then

$$\begin{aligned} \int_X \|\varphi(x)\|_{\mathbf{C}}^2 d\mu(x) &= \int_X \sum_{j=1}^\infty \left|\langle \varphi(x), f_j\rangle_{\mathbf{C}}\right|^2 d\mu(x) = \sum_{j=1}^\infty \int_X \left|\langle \varphi(x), f_j\rangle_{\mathbf{C}}\right|^2 d\mu(x) \\ &= \sum_{j=1}^\infty \left\|\langle \varphi(\cdot), f_j\rangle_{\mathbf{C}}\right\|_{H^*}^2 = \sum_{j=1}^\infty \sum_{k=1}^\infty \left|\langle \varphi(e_k), f_j\rangle_{\mathbf{C}}\right|^2 \\ &= \sum_{k=1}^\infty \sum_{j=1}^\infty \left|\langle \varphi(e_k), f_j\rangle_{\mathbf{C}}\right|^2 = \sum_{k=1}^\infty \|\varphi(e_k)\|_{\mathbf{C}}^2 = \|\varphi\|_{H^* \otimes \mathbf{C}}^2 . \end{aligned}$$

A simple consequence of Eq. (41.24) is that

$$\|\varphi\|_{H^* \otimes \mathbf{C}}^2 \leqslant \|\varphi\|_{X^* \otimes \mathbf{C}}^2 \int_X \|x\|_X^2 d\mu(x) = C_2 \|\varphi\|_{X^* \otimes \mathbf{C}}^2 . \tag{41.25}$$

Theorem 41.12 (Cameron-Martin). *Let (H, X, μ) be an abstract Wiener space (i.e. $(X, \mathcal{B}_X, \mu)$ is a Gaussian measure space with Cameron – Martin space being $H \subset X$) and for any $x \in X$ and $A \in \mathcal{B}_X$ let $\mu_x(A) = \mu(A - x)$. If $h \in H$, then $\mu_h \ll \mu$ and*

$$\frac{d\mu_h}{d\mu} = e^{(h,\cdot) - \frac{1}{2}\|h\|_H^2} d\mu$$

where $(h,\cdot)_H \equiv (J_K^{-1} h) \in K \subset L^2$. Conversely if $h \in X \setminus H$, then μ_h and μ are mutually singular ($\mu_h \perp \mu$).

For a proof of this theorem see Theorem 43.6 and Proposition 43.11 below.

42

Example of Gaussian Measure Spaces

We will give three class of examples. Before doing so let me record the following strategy for construction Gaussian measures.

Proposition 42.1. *Suppose that X is a separable Banach space and H is a dense Hilbertian subspace of X such that the inclusion map from H into X is continuous. Further suppose there exists an orthonormal basis $\{e_n\}_{n=1}^\infty$ of H, i.i.d. standard normal random variables $\{Z_n\}_{n=1}^\infty$ on some probability space, and an increasing subsequence, $\{N_k\}_{k=1}^\infty$ of $\mathbb{N}$ such that $S_k := \sum_{n=1}^{N_k} Z_n e_n$ is convergent in probability in the norm-topology on X. Let S denote the limit so that S is an X – valued random variable. Then $\mu := \mathrm{Law}(S)$ is a Gaussian measure on $(X, \mathcal{B}_X)$ with*

$$q_\mu(u,v) = \sum_{n=1}^\infty u(e_n)\, v(e_n) = (u|_H, v|_H)_{H^*}. \tag{42.1}$$

This construction works if $\dim(X) < \infty$ *after some obvious modifications and simplifications, see Example 42.2 below..*

Proof. If $u \in X^*$, then $u \circ S_k = \sum_{n=1}^{N_k} Z_n u(e_n)$ is a Gaussian random variable with

$$\mathrm{Var}(u \circ S_k) = \sum_{n=1}^{N_k} |u(e_n)|^2.$$

Using this result along with the DCT implies,

$$\begin{aligned}
\hat{\mu}(u) &= \int_X e^{iu(x)} d\mu(x) = \mathbb{E}\left[e^{iu\circ S}\right] = \lim_{k\to\infty} \mathbb{E}\left[e^{iu\circ S_k}\right] \\
&= \lim_{k\to\infty} \exp\left(-\frac{1}{2}\sum_{n=1}^{N_k} |u(e_n)|^2\right) = \exp\left(-\frac{1}{2}\sum_{n=1}^{\infty} |u(e_n)|^2\right) \\
&= \exp\left(-\frac{1}{2} q_\mu(u,u)\right)
\end{aligned}$$

This shows that μ is a Gaussian measure and the q_μ is given as in Eq. (42.1). ■

42.1 Finite Dimensional Examples and Results

Example 42.2. Let $\dim(X) < \infty$ and $q : X^* \times X^* \to \mathbb{R}$ be a given inner product. Then there exists a unique measure μ on $(X, \mathcal{B}_X)$ such that $\hat{\mu}(\varphi) = e^{-\frac{1}{2}q(\varphi,\varphi)}$.

Proof. Uniqueness is a consequence of Corollary 8.11 or Theorem 39.6. For existence recall the isomorphisms:

$$\begin{aligned}
X \ni x &\xrightarrow{\cong} \hat{x} \in X^{**} \text{ and} \\
X^* \ni \varphi &\longrightarrow q(\varphi, \cdot) \in X^{**}
\end{aligned}$$

where $\hat{x}(\varphi) = \varphi(x)$. Composing these two isomorphisms gives the following linear isomorphism

$$X^* \ni \varphi \xrightarrow{\cong} x_\varphi \in X \tag{42.2}$$

where x_φ is the unique element in X such that $\hat{x}_\varphi = q(\varphi, \cdot)$, i.e. $\psi(x_\varphi) = q(\varphi, \psi)$ for all $\psi \in X^*$. Using the isomorphism in Eq. (42.2) we may put a inner product $(\cdot,\cdot)$ on X using $(x_\varphi, x_\psi) = q(\varphi, \psi)$ for all $\varphi, \psi \in X^*$. Note with these definitions we have

$$\begin{aligned}
\varphi(x_\psi) &= q(\varphi, \psi) = (x_\varphi, x_\psi) \text{ so} \\
\varphi(x) &= (x_\varphi, x)\ \forall\, x \in X
\end{aligned} \tag{42.3}$$

and q is the dual inner product on X^* to $(\cdot,\cdot)$ on X. Thus if we let $\{e_i\}_{i=1}^n$ be an orthonormal basis for $(X, (\cdot,\cdot))$ and $\{Z_i\}_{i=1}^n$ be i.i.d. normal random variables and $Y := \sum_{i=1}^n Z_i e_i \in X$. It follows for all $u \in X^*$ that $u(Y) = \sum_{i=1}^n u(e_i) Z_i$ is a mean zero Gaussian random variable with

$$\mathrm{Var}[u(Y)] = \sum_{i=1}^n (u(e_i))^2 = q(u,u).$$

Thus if we let $\mu := \mathrm{Law}_P(Y)$ – a measure on $(X, \mathcal{B}_X)$, we will have,

$$\hat{\mu}(u) = \int_X e^{iu(x)} d\mu(x) = \mathbb{E}\left[e^{iu(Y)}\right] = e^{-\frac{1}{2}q(u,u)}$$

as desired. ■

Remark 42.3. Before leaving the finite dimensional example let us give a more constructive method for computing x_φ defined in Eq. (42.2). For $\varphi \in X^*$, we have

$$x_\varphi = \sum_{i=1}^n (x_\varphi, e_i)\, e_i = \sum_{i=1}^n \varphi(e_i)\, e_i$$

while

$$\int_X \varphi(x) x \, d\mu(x) = \mathbb{E}[\varphi(Y) Y] = \sum_{i,j=1}^n \mathbb{E}[Z_i Z_j \varphi(e_i)\, e_j] = \sum_{i=1}^n \varphi(e_i)\, e_i.$$

Thus it follows that $x_\varphi = J\varphi$ and this should help explain the formula for $J\varphi$ used in Theorem 41.6.

42.2 Hilbert Space Completion Examples

Theorem 42.4. *Suppose $(X, (\cdot,\cdot)_X)$ is a Hilbert space and μ is a Gaussian measure on $(X, \mathcal{B}_X)$. Let $H = H_\mu \subset X$ be the Cameron-Martin space associated to μ so that*

$$\hat{\mu}(\varphi) = e^{-\frac{1}{2}(\varphi|_H, \varphi|_H)_{H^*}} \quad \forall \quad \varphi \in X^*. \tag{42.4}$$

Then the inclusion map, $i : H \hookrightarrow X$, is Hilbert Schmidt (H.S.). In fact,

$$\|i\|_{HS}^2 := \sum_k \|e_k\|_X^2 = \int_X \|x\|_X^2 d\mu(x) < \infty$$

where $\{e_k\}_{k=1}^\infty$ is any orthonormal basis for H.

Proof. Let $\{\xi_j\}_{j=1}^\infty \subset X$ be an orthonormal basis for X and $\{e_k\}_{k=1}^\infty$ be an orthonormal basis for H. Then integrating the identity,

$$\|x\|_X^2 = \sum_{j=1}^\infty (x, \xi_j)_X^2 \text{ for all } x \in X,$$

with respect to μ implies,

$$\begin{aligned}
\int_X \|x\|_X^2 d\mu &= \sum_{j=1}^\infty \int_X (x, \xi_j)_X^2 d\mu = \sum_{j=1}^\infty ((\cdot, \xi_j)_X, (\cdot, \xi_j)_X)_{H^*} \\
&= \sum_{j=1}^\infty \sum_{k=1}^\infty (e_k, \xi_j)_X^2 = \sum_{k=1}^\infty \sum_{j=1}^\infty (e_k, \xi_j)_X^2 \\
&= \sum_{k=1}^\infty \|e_k\|_X^2 = \|i\|_{HS}^2 .
\end{aligned}$$

The fact $\int_X \|x\|_X^2 d\mu < \infty$ is a consequence of Fernique's Theorem 43.3. ■

Theorem 42.5. *Let H and X be Hilbert spaces and suppose that $H \overset{i}{\hookrightarrow} X$ is a Hilbert-Schmidt linear injective map. Then there exists μ on $(X, \mathcal{B}_X)$ such that (41.1) holds with*

$$q(u,u) = \|u|_H\|_{H^*}^2 = \sum_{k=1}^\infty |u(e_k)|^2$$

where $\{e_k\}_{k=1}^\infty$ is any O.N. basis for H.

Proof. Let $\{e_k\}_{k=1}^\infty$ be an orthonormal basis for H. The assumption that i is Hilbert-Schmidt means that

$$\sum_{k=1}^\infty \|e_k\|_X^2 = \sum_{k=1}^\infty \|ie_k\|_X^2 = \|i\|_{HS}^2 < \infty.$$

Let $\{Y_k\}_{k=1}^\infty$ be i.i.d. standard normal random variables on some probability space, $(\Omega, \mathcal{B}, P)$ and for $n \in \mathbb{N}$, let

$$\varphi_m := \sum_{k=1}^m Y_k e_k \in H \subset X.$$

We then have for $n \geq m$ that

$$\begin{aligned}
\mathbb{E}\|\varphi_n - \varphi_m\|_X^2 &= \mathbb{E}\left\| \sum_{k=m+1}^n Y_k e_k \right\|_X^2 \\
&= \sum_{k,l=m+1}^n \mathbb{E} Y_k Y_l (e_k, e_l)_X = \sum_{k=m+1}^n \|e_k\|_X^2 \to 0 \text{ as } m,n \to \infty.
\end{aligned}$$

Therefore,

$$L^2(P) - \lim_{m,n\to\infty} \|\varphi_n - \varphi_m\|_X^2 = 0$$

and so by passing to a subsequence if necessary we may assume that $\lim_{m,n\to\infty} \|\varphi_n - \varphi_m\|_X^2 = 0$ a.s. and since X is a Hilbert space, there exists $\varphi : \Omega \to X$ such that $\lim_{n\to\infty} \varphi_n = \varphi$ in X a.s. and in $L^2(P)$. We now let $\mu := \mathrm{Law}_P(\varphi)$ which is a measure on $(X, \mathcal{B}_X)$. We will finish the proof by showing that μ is the desired Gaussian measure on $(X, \mathcal{B}_X)$.

Let $u \in X^*$, then $u \circ \varphi_n = \sum_{k=1}^n Y_k u(e_k)$ is a Gaussian random variable with $\mathrm{Var}(u \circ \varphi_n) = \sum_{k=1}^n |u(e_k)|^2$ and therefore,

$$\mathbb{E}\left[e^{iu\circ\varphi_n}\right] = \exp\left(-\frac{1}{2}\sum_{k=1}^n |u(e_k)|^2\right).$$

Letting $n \to \infty$ in this equation using the DCT on the left side we learn that

$$\hat{\mu}(u) = \mathbb{E}\left[e^{iu\circ\varphi}\right] = \exp\left(-\frac{1}{2}\sum_{k=1}^{\infty}|u(e_k)|^2\right) = e^{-\|u|_H\|_{H^*}^2/2}$$

which shows that μ is the desired Gaussian measure. ∎

Remark 42.6. Suppose that $H \subset X$ and ν is a Gaussian measure on X such that Eq. (41.1) holds and $\{e_j\}_{j=1}^{\infty}$ is any orthonormal basis for H. Then for $x_1, x_2 \in X$, we have

$$\begin{aligned}\int_X (x_1, x)_X (x_2, x)_X \, d\nu(x) &= ((x_1, \cdot)_X (x_2, \cdot)_X)_{H^*} = \sum_{j=1}^{\infty} (x_1, e_j)_X (x_2, e_j)_X \\ &= \sum_{j=1}^{\infty} (x_1, ie_j)_X (x_2, ie_j)_X = \sum_{j=1}^{\infty} (i^* x_1, e_j)_H (i^* x_2, e_j)_H \\ &= (i^* x_1, i^* x_2)_H = (x_1, ii^* x_2)_X = (x_1, i^* x_2)_X .\end{aligned}$$

42.2.1 Deficient Wiener measure

In this sub-section we are going to construct a deficient version of Wiener measure. It will be deficient in the sense that the support of the measure is going to be $L^2([0,1])$ rather than the continuous function on $[0,1]$ which will be dealt with a little later. The goal is to make sense out of the informal expression,

$$d\mu(\omega) = \frac{1}{Z}\exp\left(-\frac{1}{2}\int_0^1 |\omega'(s)|^2 ds\right)\mathcal{D}\omega$$

to be thought of as a measure on path, $\omega : [0,1] \to \mathbb{R}$ such that $\omega(0) = 0$. According to the above expression we expect the Cameron-Martin space for our measure μ to be;

$$H = \{h \in AC([0,1]) : h(0) = 0 \text{ and } \|h\|_H < \infty\} \tag{42.5}$$

where $AC([0,1])$ denotes the space of absolutely continuous functions, $h : [0,1] \to \mathbb{R}$ and

$$\|h\|_H^2 := (h, h) = \int_0^1 |h'(s)|^2 ds.$$

The space H becomes a Hilbert space when equipped with the inner product;

$$(h, k) = \int_0^1 h'(s)k'(s)ds \text{ for all } h, k \in H.$$

For later purposes let

$$W := \{\omega \in C([0,1]) : \omega(0) = 0\}$$

which is a Banach space when equipped with the sup-norm,

$$\|f\| = \max_{x\in[0,1]} |f(x)| .$$

Proposition 42.7. *Let $G(s,t) = \min(s,t)$. Then G is the* ***reproducing kernel*** *for H, i.e. $(G(s,\cdot), h)_H = h(s)$ for all $s \in [0,1]$.*

Proof. The proof follows from the fundamental theorem of calculus for absolutely continuous functions,

$$(G(s,\cdot), h) = \int_0^1 1_{t\le s} h'(t)dt = h(s).$$

∎

Corollary 42.8. *If $\{h_n\}_{n=1}^{\infty}$ is* ***any*** *orthonormal basis for H, then*

$$\sum_{n=1}^{\infty} h_n(s)h_n(t) = G(s,t) = s \wedge t.$$

Proof. The proof is simply Bessel's equality,

$$\begin{aligned}\sum_{n=1}^{\infty} h_n(s)h_n(t) &= \sum_{n=1}^{\infty} (G(s,\cdot), h_n)(G(t,\cdot), h_n) \\ &= (G(s,\cdot), G(t,\cdot)) = G(s,t).\end{aligned}$$

∎

Proposition 42.9 (A Sobolev Theorem). *The inclusion map $i : H \to W$ is continuous and in fact*

$$\|h\|_W \le \|h\|_H \text{ for all } h \in H.$$

Proof. By Proposition 42.7, for $s \in [0,1]$,

$$|h(s)| = |(G(s,\cdot), h)_H| \le \|G(s,\cdot)\|_H \|h\|_H .$$

This proves the Proposition since

$$\|G(s,\cdot)\|_H^2 = \int_0^1 (1_{t\le s})^2 \, dt = s \le 1.$$

■

Let $W = \{\omega \in C([0,1] \to \mathbb{R}) : \omega(0) = 0\}$ and let H denote the set of functions $h \in W$ which are absolutely continuous and satisfy $(h,h) = \int_0^1 |h'(s)|^2 ds < \infty$. The space H is called the Cameron-Martin space and is a Hilbert space when equipped with the inner product

$$(h,k)_H = \int_0^1 h'(s)k'(s)ds \text{ for all } h,k \in H.$$

The space W is a Banach space when equipped with the sup-norm,

$$\|f\| = \max_{x\in[0,1]} |f(x)| .$$

Example 42.10 (Wiener measure). Let H denote the set of functions $h : [0,T] \to \mathbb{R}^d$ which are absolutely continuous and satisfy $h(0) = 0$ and $(h,h)_H = \int_0^1 |h'(s)|^2 ds < \infty$. The informal expression for Wiener measure is then given by

$$d\mu(\omega) = \frac{1}{Z}\exp\left(-\frac{1}{2}\int_0^T |\dot{\omega}(t)|^2 dt\right)\mathcal{D}\omega \text{ for } \omega \in H.$$

We are going to show that the measure μ may be constructed on $L^2\left([0,T],\mathbb{R}^d\right)$. Before doing this let us compute i^* where $i : H \to L^2$ is the inclusion operator. To this end, let $h \in H$ and $f \in L^2$, then

$$\begin{aligned}(ih,f)_{L^2} &= \int_0^T h(t)\cdot f(t)\,dt = \int_0^T \left(\int_0^t h'(s)\,ds\right)\cdot f(t)\,dt \\ &= \int_{[0,T]^2} 1_{0\le s\le t} h'(s)\cdot f(t)\,dtds = \int_0^T h'(s)\cdot\left(\int_s^T f(t)\,dt\right)ds \\ &= (h, i^* f)_H ,\end{aligned}$$

where

$$\begin{aligned}(i^* f)(\tau) &= \int_0^\tau \left(\int_s^T f(t)\,dt\right)ds = \int_{[0,T]^2} 1_{0\le s\le\tau}\cdot 1_{s\le t\le T} f(t)\,dtds \\ &= \int_0^T \min(t,\tau)\,f(\tau)\,d\tau.\end{aligned}$$

From this we expect that

$$\operatorname{tr}(i\,i^*) = d\cdot\int_0^T \min(t,t)\,dt = d\cdot T^2/2$$

which is finite. This is indeed correct as is seen using Corollary 42.8 above to conclude

$$\sum_{n=1}^\infty \int_0^T |h_n(t)|^2\,dt = d\int_0^T t dt = d\cdot T^2/2.$$

It now follows from Remark 42.6 that

$$\int_{L^2} (f,\omega)_{L^2}(g,\omega)_{L^2}\,d\mu(\omega) = (f, i^* g)_{L^2} = \int_{[0,T]^2} f(s)\,g(t)\min(s,t)\,dsdt.$$

In particular "taking" $f = \delta_s$ and $g = \delta_t$ leads us to expect that

$$\int_{L^2} \omega(s)\otimes\omega(t)\,d\mu(\omega) = \min(s,t)\sum_{i=1}^d e_i\otimes e_i.$$

From this "computation" it is reasonable to expect that we may construct μ on W and this is in fact true as Wiener [53, 52] proved in 1923.

42.2.2 Hints at Quantum Field Theoretic Complications

In this section we are going to explore what happens when we try to make sense out of Gaussian measures which are given informally as;

$$d\mu(\varphi) = \frac{1}{Z}\exp\left(-\frac{1}{2}\int_{\mathbb{R}^d}\left[|\nabla\varphi|^2 + m^2\varphi^2\right]dm\right)\mathcal{D}\varphi. \tag{42.6}$$

This is a prototypical expression which occurs in the physics literature on quantum field theory.

Let us recall that the trace of a positive operator, $B : H \to H$ is defined in the usual way as;

$$\operatorname{tr}(B) := \sum_{k=1}^\infty (Be_k, e_k)_H .$$

The sum is independent of basis since, by the general spectral theorem we may write $B = A^2$ with $A \ge 0$ and then

$$\sum_{k=1}^\infty (Be_k,e_k)_H = \sum_{k=1}^\infty \left(A^2 e_k, e_k\right)_H = \sum_{k=1}^\infty (Ae_k, Ae_k)_H = \|A\|_{HS}^2$$

so that $\operatorname{tr}(B) = \|A\|_{HS}^2$ which we know is basis independent. We say that $B \ge 0$ is trace class if $\operatorname{tr}(B) < \infty$. Recall that we have seen that Hilbert-Schmidt operators are compact and therefore if $B \ge 0$ and is trace class then $B = A^2$ where A is Hilbert-Schmidt and hence compact and therefore B is also compact. As an aside here are some equivalent conditions for a linear operator, $A : H \to X$ to be Hilbert-Schmidt.

Proposition 42.11. *Suppose that H and X are Hilbert spaces and $A : H \to X$ is a bounded linear operator. Then;*

$$\operatorname{tr}(A^*A) = \|A\|_{HS}^2 = \|A^*\|_{HS}^2 = \operatorname{tr}(AA^*).$$

Proof. If $\{e_k\}$ is an O.N. basis for H, then

$$\|A\|_{HS}^2 = \sum_k \|Ae_k\|_X^2 = \sum_k (Ae_k, Ae_k)_X = \sum_k (A^*Ae_k, e_k)_H = \operatorname{tr}(A^*A).$$

Replacing A by A^* above then shows $\|A^*\|_{HS}^2 = \operatorname{tr}(AA^*)$ and we have already seen that $\|A^*\|_{HS}^2 = \|A\|_{HS}^2$. ∎

Lemma 42.12. *If H is a Hilbert space and $A : H \to H$ is a positive trace class operator. We may define $(x,y)_A := (Ax, y)$ for all $x, y \in H$. Then let X denote the completion of H in the norm, $\|\cdot\|_A := \sqrt{(\cdot,\cdot)_A}$. Then the inclusion map, $i : H \to X$ will be Hilbert Schmidt and hence X will support a Gaussian measure with variance determined by H.*

Example 42.13. Now consider the "measure" $d\mu(\varphi)$ of Eq. (42.6). In this setting one may show that μ can not be constructed on $L^2(\mathbb{R}^d)$ no matter the dimension d.[1] To see this let H be the Sobolev space of one derivative in L^2. In this case we have $i^* = (-\Delta + m^2)^{-1}$. Indeed, if $u := i^*g$, then

$$(f,g)_{L^2} = (if, g)_{L^2} = (f,u)_H = \int_{\mathbb{R}^d} (\nabla f \cdot \nabla u + m^2 f \cdot u)\, dx \;\forall\; f \in H,$$

which, by Elliptic regularity or by the Fourier transform, implies the distribution, Δu, is an L^2 – function and

$$(f,g)_{L^2} = \left(f, (-\Delta + m^2)\, u\right).$$

Hence we have $(-\Delta + m^2)\, u = g$ or $i^*g = u = (-\Delta + m^2)^{-1} g$ as claimed.

Informally, now,

$$\operatorname{tr}(i\, i^*) = \int_{\mathbb{R}^d} (-\Delta + m^2)^{-1}(x,x)\, dx = \infty.$$

because when $d \geq 2$, $(-\Delta + m^2)^{-1}(x,x) = \infty$ and or $d = 1$ $(-\Delta + m^2)^{-1}(x,x) = c > 0$ for all $x \in \mathbb{R}$. To give a rigorous proof, notice that $i\, i^*$ is unitarily equivalent to the the multiplication operator, $(k^2 + m^2)^{-1}$ which has continuous spectrum, it follows that $(-\Delta + m^2)^{-1}$ is not compact let alone trace class. In general on non-atomic spaces with no infinite atoms any non-zero multiplication operator is not trace class. Indeed, suppose M_f is the multiplication operator on $L^2(X,m)$ and observe that we may assume $f \geq 0$ since $|M_f| = M_{|f|}$. Then for some $\varepsilon > 0$ we will have $\mu(f \geq \varepsilon) > 0$. Since m is non-atomic, we may write $\{f \geq \varepsilon\}$ as a disjoint union of $\{A_j\}$ where $\infty > m(A_j) > 0$. Then the functions, $\left\{\frac{1}{\sqrt{m(A_j)}} 1_{A_j}\right\}_{j=1}^\infty$ forms an orthonormal subset of $L^2(m)$ and therefore,

$$\begin{aligned}\operatorname{tr}(M_f) &\geq \sum_{j=1}^\infty \left(\frac{1}{\sqrt{m(A_j)}} 1_{A_j}, f\frac{1}{\sqrt{m(A_j)}} 1_{A_j}\right) = \sum_{j=1}^\infty \frac{1}{m(A_j)} \int_{A_j} f dm \\ &\geq \sum_{j=1}^\infty \frac{1}{m(A_j)} \int_{A_j} \varepsilon dm = \varepsilon \cdot \infty = \infty.\end{aligned}$$

Example 42.14. Part of the problem above was the non-compactness of $\mathbb{R}^d$. To avoid this issue, let us replace $\mathbb{R}^d$ be a $\mathbb{T}^d$ – the d – dimensional torus which we identify with $[0,2\pi]^d / \backsim$ where $\backsim$ is the usual identification of the endpoints. We will denote points in $[0,2\pi]^d$ by θ and let $d\theta$ denote normalized Haar measure on $\mathbb{T}^d$. In this case $\left\{\frac{e^{in\cdot\theta}}{\sqrt{\|n\|^2 + m^2}}\right\}_{n \in \mathbb{Z}^d}$ is an orthonormal basis for $H(\mathbb{T}^d)$ equipped with the inner product;

$$(f,g)_{H(\mathbb{T}^d)} := \int_{\mathbb{T}^d} \left[\nabla f(\theta) \cdot \nabla g(\theta) + m^2 f(\theta)\, g(\theta)\right] d\theta.$$

Hence if $i : H(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ is the inclusion map, then

$$\|i\|_{H.S.}^2 = \sum_{n \in \mathbb{Z}^d} \left\| \frac{e^{in\cdot\theta}}{\sqrt{\|n\|^2 + m^2}} \right\|_{L^2(\mathbb{T}^d)}^2 = \sum_{n \in \mathbb{Z}^d} \frac{1}{\|n\|^2 + m^2}$$

which is finite iff $d = 1$. When $d = 2$ the sum is logarithmically divergent and it is worse when $d = 3$. This can also be understood by noting that $(-\Delta + m^2)^{-1}(\theta - \alpha)$ has the same singularity structure as Example 42.13 above.

Hence we need to take X even bigger than $L^2(\mathbb{T}^d)$ – however only just barely when $d = 2$. For example for any $s \in \mathbb{R}$ and $f \in C^\infty(\mathbb{T}^2)$,

$$\|f\|_s^2 := \sum_{n \in \mathbb{Z}^2} \left(\|n\|^2 + m^2\right)^s \left|\hat{f}(n)\right|^2$$

where $\hat{f}(n) := \int_{\mathbb{T}^2} f(\theta)\, e^{-in\cdot\theta} d\theta$. So $\|f\|_0^2 = \|f\|_{L^2(\mathbb{T}^2)}^2$, $\|f\|_1^2 = \|f\|_H$ and for any $s < 0$, $\|f\|_s^2 \leq \|f\|_0^2$. Let X_s denote the completion of $C^\infty(\mathbb{T}^2)$ in the s

[1] When $d = 1$ it comes close to working but there are problems due to the fact that $\mathbb{R}$ is not compact.

– norm. This is the Sobolev space of s – derivatives in L^2. For any $s < 0$, the inclusion map, $i : H = X_1 \to X_s$ is Hilbert Schmidt. Indeed, we now have

$$\|i\|_{H.S.}^2 = \sum_{n\in\mathbb{Z}^2} \left\| \frac{e^{in\cdot\theta}}{\sqrt{\|n\|^2+m^2}} \right\|_{X_s}^2 = \sum_{n\in\mathbb{Z}^2} \frac{1}{\|n\|^2+m^2} \left(\|n\|^2+m^2\right)^s$$
$$= \sum_{n\in\mathbb{Z}^2} \frac{1}{\left(\|n\|^2+m^2\right)^{1+|s|}} < \infty.$$

42.3 Classical Wiener Space

We now want to consider the case of classical Wiener measure or Brownian motion if you like.

Theorem 42.15. *Let $W := \{\omega \in C([0,1],\mathbb{R}) : \omega(0) = 0\}$ be Wiener space equipped with the uniform norm, $\mathcal{B}$ be the Borel σ – algebra on W, and μ be Wiener measure on $(W,\mathcal{B})$, i.e. μ is the law of a Brownian motion. Further let*

$$H := \left\{ h \in W : h \text{ is A.C. and } (h,h)_H := \int_0^1 |h'(s)|^2 ds < \infty \right\}.$$

Then;

1. *H is a dense subspace of W and the inclusion map, $i : H \to W$ is continuous.*
2. *μ is a Gaussian measure on W and*
$$\hat{\mu}(\ell) = \exp\left(-\frac{1}{2} q(\ell,\ell)\right) \text{ for all } \ell \in W^*,$$
where
$$q(\ell,\ell) := \|\ell|_H\|_{H^*}^2 = \sum_{n=1}^{\infty} |\ell(h_n)|^2$$
for any orthonormal basis $\{h_n\}_{n=1}^{\infty}$ of H.
3. *The space $H = H_\mu$ is the Cameron-Martin space for $(W,\mathcal{B},\mu)$.*

Proof. We prove each item in turn.

1. For $h \in H$ we have have
$$h(t) = \int_0^t h'(\tau)\, d\tau$$
and therefore,
$$|h(t) - h(s)| = \left| \int_s^t h'(\tau)\, d\tau \right| \le \int_s^t |h'(\tau)|\, d\tau$$
$$\le \sqrt{t-s} \cdot \sqrt{\int_s^t |h'(\tau)|^2 d\tau} \le \|h\|_H \cdot \sqrt{t-s}$$
from which it follows that h is $1/2$ – Hölder continuous and in particular, $h \in W$. Moreover taking $s = 0$ in this inequality then implies,
$$|h(t)| \le \|h\|_H \cdot \sqrt{t} \le \|h\|_H$$
which shows that $\|h\|_W \le \|h\|_H$ and hence $\|i\|_{op} \le 1 < \infty$. The fact that H is dense in W follows by any number of means including the classical Weierstrass approximation theorem. We will give a direct proof here as we will need the approximations described in this proof in order to prove item 2.
Let $D := \{0 = t_0 < t_1 < \cdots < t_n = 1\}$ be a partition of $[0,1]$ and $|D| := \max_i |t_i - t_{i-1}|$ be the mesh size of D. For such a partition let $\pi_D : W \to H$ be define by $h = \pi_D(\omega)$ provided $h = \omega$ on D and $h''(t) = 0$ for $t \notin D$. We let
$$H_D := \mathrm{Ran}(\pi_D) = \{h \in H : h''(t) = 0 \text{ for all } t \notin D\}.$$
If $\{D_n\}_{n=1}^{\infty}$ is an increasing sequence of partitions such that $|D_n| \downarrow 0$ as $n \to \infty$, then by the uniform continuity of ω it is easily seen that
$$\|\omega - \pi_{D_n}(\omega)\|_W \to 0 \text{ as } n \to \infty.$$
This clearly shows that H is dense in W
We also claim that $\pi_D|_H$ is orthogonal projection onto H_D. To see this it suffices to show $k := h - \pi_D(h)$ is orthogonal to H_D for all $h \in H$. In order to verify this we need to compute $H_D^\perp$. If $k \perp H_D$, then for all $h \in H_D$ we have,
$$0 = (k,h) = \int_0^1 k'(s)\, h'(s)\, ds = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} k'(s)\, m_i ds$$
$$= \sum_{i=1}^n \left[k(t_i) - k(t_{i-1})\right] m_i,$$
where $m_i := \frac{h(t_i)-h(t_{i-1})}{t_i-t_{i-1}} = h'(s)$ for any $s \in (s_{i-1}, s_i)$. As the m_i may be chosen arbitrarily it follows that $k(t_i) - k(t_{i-1}) = 0$ for all i and since $k(t_0) = k(0) = 0$ it follows that $k(t_i) = 0$ for all i, i.e. $k|_D = 0$. Conversely if $k|_D = 0$ then the above computation shows $k \perp H_D$ and we have shown $H_D = \{k \in H : k|_D = 0\}$. Since $h - \pi_D(h) = 0$ on D it follows that $h - \pi_D(h) \in H_D^\perp$ and therefore $\pi_D : H \to H_D$ is orthogonal projection.

2. Now suppose that $h \in H$. Then

$$\begin{aligned}(h, \pi_D(\omega)) &= \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} h'(s)\, \pi_D(\omega)'(s)\, ds \\ &= \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} h'(s) \cdot \frac{\Delta_i \omega}{\Delta_i} ds = \sum_{i=1}^{n} \frac{\Delta_i h}{\Delta_i} \cdot \frac{\Delta_i \omega}{\Delta_i} \Delta_i \\ &= \sum_{i=1}^{n} \frac{\Delta_i h}{\Delta_i} \cdot \Delta_i \omega\end{aligned}$$

and therefore by the basic properties of Brownian motion, $\omega \to (h, \pi_D(\omega))$ is a mean zero Gaussian random variable with

$$\mathrm{Var}(h, \pi_D(\cdot)) = \sum_{i=1}^{n} \left| \frac{\Delta_i h}{\Delta_i} \right|^2 \Delta_i = (h, h)_H .$$

Therefore it follows that

$$\int_W e^{i(h, \pi_D(\omega))_H} d\mu(\omega) = \exp\left(-\frac{1}{2}(h,h)_H\right)$$

for all $h \in H_D$ and hence we may conclude that

$$\mathrm{Law}_\mu\, \pi_D(\cdot) = \frac{1}{Z_D} \exp\left(-\frac{1}{2}\|h\|_H^2\right) dm_D(h) = d\mu_D(h)$$

where m_D is a Lebesgue measure on H_D and Z_D is a normalization constant. Thus we may conclude that

$$\int_W f(\pi_D(\omega))\, d\mu(\omega) = \int_{H_D} f(h)\, d\mu_D(h).$$

So if $f : W \to \mathbb{R}$ is a bounded continuous function or non-negative continuous function on W it follows by either DCT or MCT that

$$\int_W f(\omega)\, d\mu(\omega) = \lim_{n\to\infty} \int_W f(\pi_{D_n}(\omega))\, d\mu(\omega) = \lim_{n\to\infty} \int_{H_{D_n}} f(h)\, d\mu_{D_n}(h). \tag{42.7}$$

In particular if $\ell \in W^*$ we may conclude that

$$\begin{aligned}\int_W e^{i\ell(\omega)} d\mu(\omega) &= \lim_{n\to\infty} \int_{H_{D_n}} e^{i\ell(h)} d\mu_{D_n}(h) \\ &= \lim_{n\to\infty} \exp\left(-\frac{1}{2}\left\|\ell|_{H_{D_n}}\right\|_{H^*_{D_n}}\right) \\ &= \lim_{n\to\infty} \exp\left(-\frac{1}{2}\left\|\ell \circ \pi_{D_n}\right\|_{H^*}\right) \\ &= \exp\left(-\frac{1}{2}\left\|\ell|_H\right\|_{H^*}\right)\end{aligned}$$

as $\pi_{D_n} \xrightarrow{s} I_H$ as $n \to \infty$. The point is that if $k \in H$ with $k \perp \cup_n H_{D_n}$ then $k|_{D_n} = 0$ for all n, i.e. $k = 0$ on the dense set, $\cup_n D_n$ and as k is continuous it follows that $k \equiv 0$.

3. The proof of item 3. is the consequence of Lemma 41.10 above.

■

Remark 42.16 (Heuristics). From either item 2. of the Theorem or from Eq. (42.7) we have informally that,

$$d\mu(\omega) = \frac{1}{Z} \exp\left(-\frac{1}{2}\|\omega\|_H^2\right) \mathcal{D}\omega.$$

42.4 Construction of Classical Wiener Measure

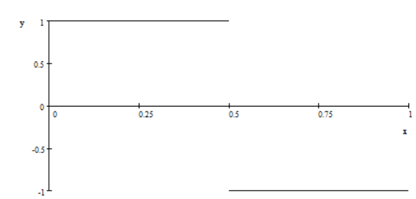
Exercise 42.1 (Haar Basis). In this problem, let L^2 denote $L^2([0,1], m)$ with the standard inner product,

$$\psi(x) = 1_{[0,1/2)}(x) - 1_{[1/2,1)}(x)$$

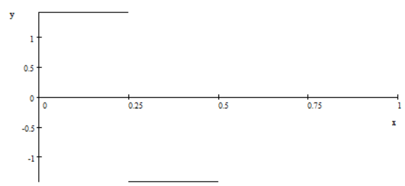
and for $k, j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ with $0 \le j < 2^k$ let

$$\begin{aligned}\psi_{kj}(x) &= 2^{k/2}\psi(2^k x - j) \\ &= 2^{k/2}\left(1_{2^{-k}[j,j+1/2)}(x) - 1_{2^{-k}[j+1/2,j+1)}(x)\right).\end{aligned}$$

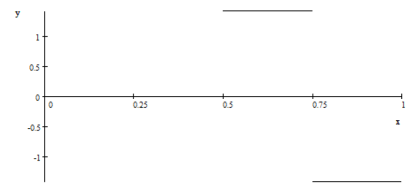
The following pictures shows the graphs of $\psi_{0,0}$, $\psi_{1,0}, \psi_{1,1}, \psi_{2,1}, \psi_{2,2}$ and $\psi_{2,3}$ respectively.



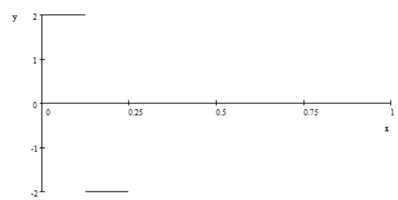
Plot of $\psi_0, 0$.



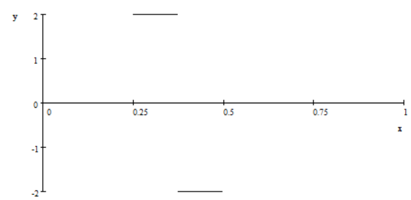
Plot of $\psi_1 0$.



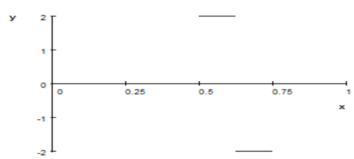
Plot of $\psi_1 1$.



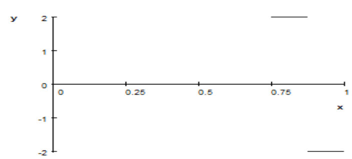
Plot of $\psi_2 0$.



Plot of $\psi_2 1$.



Plot of $\psi_2 2$.



Plot of $\psi_2 3$.

1. Let $M_0 = \operatorname{span}(\{\mathbf{1}\})$ and for $n \in \mathbb{N}$ let
$$M_n := \operatorname{span}\left(\{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k < n \text{ and } 0 \le j < 2^k\right\}\right),$$
where $\mathbf{1}$ denotes the constant function 1. Show
$$M_n = \operatorname{span}\left(\{1_{[j2^{-n},(j+1)2^{-n})} : \text{and } 0 \le j < 2^n\right).$$
2. Show $\beta := \{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k \text{ and } 0 \le j < 2^k\right\}$ is an orthonormal set. **Hint:** show $\psi_{k+1,j} \in M_k^{\perp}$ for all $0 \le j < 2^{k+1}$ and show $\left\{\psi_{kj} : 0 \le j < 2^k\right\}$ is an orthonormal set for fixed k.
3. Show $\cup_{n=1}^{\infty} M_n$ is a dense subspace of L^2 and therefore β is an orthonormal basis for L^2. **Hint:** see Theorem 12.30.
4. For $f \in L^2$, let
$$H_n f := \langle f|\mathbf{1}\rangle \mathbf{1} + \sum_{k=0}^{n-1} \sum_{j=0}^{2^k-1} \langle f|\psi_{kj}\rangle \psi_{kj}.$$
Show
$$H_n f = \sum_{j=0}^{2^n-1} \left(2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(x)dx\right) 1_{[j2^{-n},(j+1)2^{-n})}$$
and use this to show $\|f - H_n f\|_\infty \to 0$ as $n \to \infty$ for all $f \in C([0,1])$. **Hint**: Compute orthogonal projection onto M_n using a judiciously chosen basis for M_n.

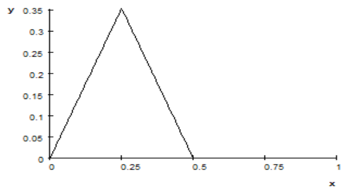
As above let $W = \{\omega \in C([0,1] \to \mathbb{R}) : \omega(0) = 0\}$ and H be the Cameron-Martin space defined in Eq. (42.5). The goal of this section is to give Paul Levy's proof of Wiener's Theorem [53, 52]. Because of Proposition 42.9, the transpose $i^{tr} : W^* \to H^*$ is continuous as well. Hence by the Riesz theorem, for each $\varphi \in W^*$ there exists a unique $h_\varphi \in H$ such that $(h_\varphi, \cdot)_H = \varphi|_H$. Hence we may define an inner product q on W^* by $q(\varphi, \psi) = (h_\varphi, h_\psi)$ for all $\varphi, \psi \in W^*$.

Theorem 42.17. *There exists a unique (Gaussian) measure μ on W such that*
$$\int_W e^{i\varphi(x)} d\mu(x) = e^{-q(\varphi)/2}, \tag{42.8}$$
where $q(\varphi, \psi) = (h_\varphi, h_\psi)$ as described above.

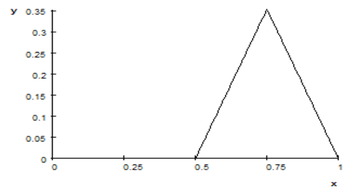
Proof. Let $\beta := \{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k \text{ and } 0 \le j < 2^k\right\}$ be the Haar basis for $L^2([0,1], d\lambda)$ introduced in Exercise 42.1 above and let $\Psi_0(t) := \int_0^t 1 ds = t$ and $\Psi_{kj}(t) := \int_0^t \psi_{kj}(s)ds$ for $0 \le k$ and $0 \le j < 2^k$. Then $\tilde{\beta} = \{\Psi_0\} \cup \left\{\Psi_{kj} : 0 \le k \text{ and } 0 \le j < 2^k\right\}$ is an orthonormal basis for H which satisfies,
$$\|\Psi_0\|_\infty = 1 \text{ and } \|\Psi_{kj}\|_\infty = 2^{k/2} 2^{-(k+1)} = \frac{1}{2} 2^{-k/2}.$$

The following pictures shows the graphs of Ψ_{00}, $\Psi_{1,0}, \Psi_{1,1}, \Psi_{2,1}, \Psi_{2,2}$ and $\Psi_{2,3}$ respectively.

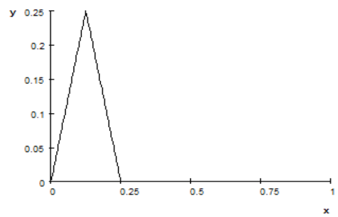
Plot of $\Psi_0, 0$.



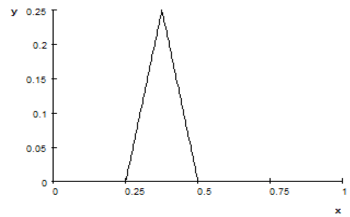
Plot of $\Psi_1 0$.



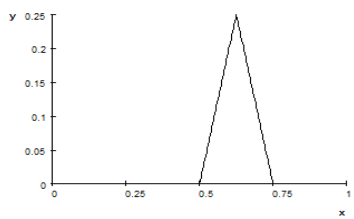
Plot of $\Psi_1 1$.



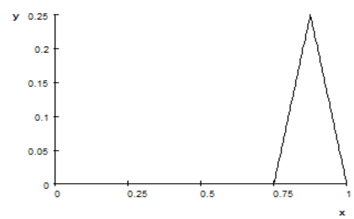
Plot of $\Psi_2 0$.



Plot of $\Psi_2 1$.



Plot of $\Psi_2 2$.



Plot of $\Psi_2 3$.

Because $\left\{\Psi_{kj} : 0 \le k \text{ and } 0 \le j < 2^k\right\}$ have disjoint supports, if $\{x_j\}_{j<2^k} \subset \mathbb{R}$, we have,

$$\max_t \left| \sum_{j<2^k} x_j \Psi_{kj}(t) \right| \le \frac{1}{2} 2^{-\frac{k}{2}} \max\left\{|x_j| : j < 2^k\right\} \le \frac{1}{2} 2^{-\frac{k}{2}} \left(\sum_{j<2^k} |x_j|^p \right)^{\frac{1}{p}} \tag{42.9}$$

for any $p \in [1, \infty)$.

Let $\{Z_0\} \cup \left[\cup_{k=0}^{\infty} \left\{Z_{k,j} : 0 \le j < 2^k\right\}\right]$ be a collection of i.i.d. standard normal random variables. Making use of Eq. (42.9) along with Hölder's inequality we find,

$$\begin{aligned}\mathbb{E}\left\| \sum_{j<2^k} Z_{kj} \Psi_{kj} \right\|_\infty &\le \frac{1}{2} 2^{-\frac{k}{2}} \mathbb{E}\left(\sum_{j<2^k} |Z_{kj}|^p \right)^{\frac{1}{p}} \le \frac{1}{2} 2^{-\frac{k}{2}} \left(\sum_{j<2^k} \mathbb{E}|Z_{kj}|^p \right)^{\frac{1}{p}} \\ &= \frac{1}{2} 2^{-\frac{k}{2}} 2^{\frac{k}{p}} C_p\end{aligned}$$

where

$$C_p = \left(\int_{\mathbb{R}} |x|^p \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \, dx \right)^{1/p} < \infty.$$

Therefore

$$\mathbb{E} \sum_k \left\| \sum_{j<2^k} Z_{kj} \Psi_{kj} \right\|_\infty \le \frac{1}{2} C_p \sum_{k=1}^{\infty} 2^{-\left(\frac{1}{2} - \frac{1}{p}\right)k} < \infty$$

providing $p > 2$. In particular it follows that $\left\| \sum_{j<2^k} Z_{kj} \Psi_{kj} \right\|_\infty < \infty$ a.s. and therefore if we define,

$$S_N := Z_0 \Psi_0 + \sum_{k \le N} \left(\sum_{j<2^k} Z_{kj} \Psi_{kj}(\cdot) \right),$$

then $S := \|\cdot\|_\infty\text{–}\lim_{N\to\infty} S_N$ exists a.s. Combining this convergence result with Proposition 42.1 completes the proof of the theorem. ■

Remark 42.18. We can use the above argument to prove more. Namely that Wiener measure lives on the space of paths which are α − Hölder continuous provided $\alpha < 1/2$.

Proof. To prove this refinement let $\alpha \in [0, 1]$ and consider

$$\Gamma_k(t) := \sum_{j<2^k} x_j \Psi_{kj}(t)$$

where $\{x_j\}_{j<2^k}$ are a fixed sequence of real numbers. By the mean value theorem and using the disjoint supports of $\{\Psi_{kj}\}$ and the fact that $\left\|\dot{\Psi}_{kj}\right\|_\infty = 2^{k/2}$, we find

$$|\Gamma_k(t) - \Gamma_k(s)| \le M 2^{k/2} |t - s|$$

where $M := \max\left\{x_j : j < 2^k\right\}$. Hence

$$\frac{|\Gamma_k(t) - \Gamma_k(s)|}{|t - s|^\alpha} \le 2^{k/2} M |t - s|^{1-\alpha}.$$

If $|t-s| \le 2^{-k}$ this gives

$$\frac{|\Gamma_k(t) - \Gamma_k(s)|}{|t-s|^{\alpha}} \le 2^{k/2} 2^{-k(1-\alpha)} M = 2^{-k(\frac{1}{2}-\alpha)} M$$

If $|t-s| \ge 2^{-k}$, recalling $\|\Gamma_k\|_\infty \le \frac{1}{2} 2^{-\frac{k}{2}} M$,we have

$$\frac{|\Gamma_k(t) - \Gamma_k(s)|}{|t-s|^{\alpha}} \le 2 2^{k\alpha} \frac{1}{2} 2^{-\frac{k}{2}} M = 2^{-k(\frac{1}{2}-\alpha)} M.$$

That is to say, $\|\Gamma_k\|_\alpha \le 2^{-k(\frac{1}{2}-\alpha)} M$ and working as above this implies

$$\begin{aligned} \sum_k \mathbb{E} \left\| \sum_{j<2^k} Z_{kj} \Psi_{kj}(t) \right\|_\alpha &\le \sum_k \frac{1}{2} 2^{-k(\frac{1}{2}-\alpha)} \left(\sum_{j<2^k} \mathbb{E} |Z_{kj}|^p \right)^{\frac{1}{p}} \\ &= \sum_k \frac{1}{2} C_p 2^{-k(\frac{1}{2}-\alpha)} 2^{\frac{k}{p}} < \infty \end{aligned}$$

provided that

$$\frac{1}{2} - \alpha - \frac{1}{p} > 0$$

or $\alpha < 1/2 - 1/p$. Since we may take p as large as we like it follows that

$$\lim_{N\to\infty} \|S - S_N\|_\alpha = 0 \text{ a.s.}$$

and therefore S takes values in $C^{0,\alpha}([0,1],\mathbb{R})$ a.s. and therefore $\mu = \mathrm{Law}(S)$ is a probability measure on $C^{0,\alpha}([0,1],\mathbb{R})$. ■

Lemma 42.19. *Let $\{b_s\}_{s\ge 0}$ be a one dimensional Brownian motion. Then for any $T > 0$ and $\lambda \in [0, \pi^2/(4T^2))$;*

$$\mathbb{E}\left[\exp\left(\frac{\lambda}{2}\int_0^T b_s^2 ds\right)\right] = \cos^{-1/2}\left(\sqrt{\lambda} T\right) < \infty. \tag{42.10}$$

(All we really need for the qualitative result is to observe that, using Fernique's theorem,

$$\begin{aligned} \mathbb{E}\left[\exp\left(\frac{\lambda}{2}\int_0^T b_s^2 ds\right)\right] &= \mathbb{E}\left[\exp\left(\frac{\lambda T^2}{2}\int_0^1 b_s^2 ds\right)\right] = 1 + \frac{\lambda T^2}{2}\mathbb{E}\left[\int_0^1 b_s^2 ds\right] + O\left(\lambda^2 T^4\right) \\ &= 1 + \frac{\lambda T^2}{2}\int_0^1 s ds + O\left(\lambda^2 T^4\right) = 1 + \frac{\lambda T^2}{4} + O\left(\lambda^2 T^4\right) \end{aligned}$$

for sufficiently small λT^2.)

Proof. When $T = 1$, simply follow the proof of [25, Eq. (6.9) on p. 472] with λ replaced by $-\lambda$. For general $T > 0$ observe that

$$\int_0^T b_s^2 ds = T \int_0^1 b_{tT}^2 dt \overset{d}{=} T^2 \int_0^1 b_t^2 dt$$

and therefore,

$$\mathbb{E}\left[\exp\left(\frac{\lambda}{2}\int_0^T b_s^2 ds\right)\right] = \mathbb{E}\left[\exp\left(\frac{\lambda T^2}{2}\int_0^1 b_s^2 ds\right)\right] = \cos^{-1/2}\left(\sqrt{\lambda T^2}\right)$$

provided that $\lambda \in [0, \pi^2/(4T^2))$.

Heuristic Proof.

$$\begin{aligned} \mathbb{E}\left[\exp\left(\frac{\lambda}{2}\int_0^1 b_s^2 ds\right)\right] &= \frac{1}{Z}\int_{H([0,1],\mathbb{R})} \exp\left(-\frac{1}{2}\int_0^1 \left[\dot{x}^2(t) - \lambda x^2(t)\right] dt\right) \mathcal{D}x \\ &= \frac{1}{Z}\int_{H([0,1],\mathbb{R})} \exp\left(-\frac{1}{2}\int_0^1 \left[-D^2 x(t) - \lambda x(t)\right] x(t)\, dt\right) \mathcal{D}x \end{aligned}$$

where $D^2 = \frac{d^2}{dt^2}$ with Dirichlet boundary conditions at 0 and Neumann boundary conditions at 1. The eigenfunctions for D^2 are proportional to $\{\sin(l\pi x/2)\}_{l\in 2\mathbb{N}-1}$ with eigenvalues being $-l^2\pi^2/4$. Therefore,

$$\begin{aligned} \int_{H([0,1],\mathbb{R})} \exp\left(-\frac{1}{2}\int_0^1 \left[\dot{x}^2(t) - \lambda x^2(t)\right] dt\right) \mathcal{D}x &= \frac{(2\pi)^{\dim H/2}}{\sqrt{\det(-D^2 - \lambda)}} \\ &= \sqrt{\prod_{l \text{ odd}} \frac{2\pi}{l^2\pi^2/4 - \lambda}} \end{aligned}$$

and Z is this expression with $\lambda = 0$. Hence we find,

$$\mathbb{E}\left[\exp\left(\frac{\lambda}{2}\int_0^1 b_s^2 ds\right)\right] = \sqrt{\prod_{l \text{ odd}} \frac{l^2\pi^2/4}{l^2\pi^2/4 - \lambda}} = \left(\prod_{l \text{ odd}}\left(1 - \frac{4\lambda}{l^2\pi^2}\right)\right)^{-1/2}$$

provided that $4\lambda < \pi^2$.

Rigorous Proof. Let $\{N_k\}_{k=1}^\infty$ be a sequence of i.i.d. normal random variables and $\{h_k\}_{k=1}^\infty$ be an orthonormal basis for $H = H([0,1],\mathbb{R})$. Then (see Theorem 44.3 below)

$$b_s := \sum_{k=1}^\infty N_k h_k(s)$$

is a standard one dimensional Brownian motion. Letting $a_k := \int_0^1 h_k^2(s)\, ds$, we then have

$$\int_0^1 b_s^2 ds = \sum_{k=1}^{\infty} N_k^2 a_k$$

and therefore that

$$\mathbb{E}\left[\exp\left(\frac{\lambda}{2}\int_0^1 b_s^2 ds\right)\right] = \mathbb{E}\left[\exp\left(\frac{\lambda}{2}\sum_{k=1}^{\infty} N_k^2 a_k\right)\right]$$
$$= \prod_{k=1}^{\infty} \mathbb{E}\left[\exp\left(\frac{\lambda}{2} a_k N^2\right)\right] = \prod_{k=1}^{\infty} \frac{1}{\sqrt{1-\lambda a_k}}. \qquad (42.11)$$

provided that $\lambda < 1/a_k$ for all k. Assuming this restriction on λ, the latter product is finite iff $\sum_{k=1}^{\infty} a_k < \infty$. But

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} \int_0^1 h_k^2(s)\, ds = \int_0^1 \sum_{k=1}^{\infty} h_k^2(s)\, ds = \int_0^1 s ds = 1/2 < \infty.$$

A calculus of variation exercises show

$$\sup_{h \neq 0} \frac{\int_0^1 h^2(s)\, ds}{\int_0^1 \dot{h}^2(s)\, ds} = \frac{4}{\pi^2}$$

and therefore $a_k \leq 4/\pi^2$ for all k and the condition on λ becomes $\lambda < \pi^2/4$.

To get the more explicit expression for $\mathbb{E}\left[\exp\left(\frac{\lambda}{2}\int_0^1 b_s^2 ds\right)\right]$ let us observe that

$$\left\{h_l(t) = \frac{2\sqrt{2}}{\pi l} \sin\left(l\frac{\pi}{2}t\right)\right\}_{l \text{ odd}}$$

is an orthonormal basis for H. Indeed, $\left\{\dot{h}_l(t) = \sqrt{2}\cos\left(l\frac{\pi}{2}t\right)\right\}_{l \text{ odd}}$ is the orthonormal basis of eigenfunctions associated to d^2/dt^2 with Neumann boundary conditions at 0 and Dirichlet boundary conditions at 1. With this choice of basis we find

$$a_l = \frac{8}{\pi^2 l^2}\int_0^1 \sin^2\left(l\frac{\pi}{2}t\right) dt = \frac{4}{\pi^2 l^2}$$

which combined with Eq. (42.11) gives Eq. (42.10). ■

43

Fernique and Cameron-Martin Theorems

In this section we take care of some details which have been postponed. The first is the important exponential integrability result of Fernique and second is the quasi-invariance results of Cameron and Martin.

43.1 Fernique's Theorem

Lemma 43.1. *Let $(X,\mathcal{B}_X,\mu)$ and $(Y,\mathcal{B}_Y,\nu)$ be Gaussian measure spaces then so is $(X\times Y,\mathcal{B}_{(X\times Y)},\mu\times\nu)$.*

Proof. Let $\psi\in(X\times Y)^*$, $\alpha(x)=\psi(x,0)$ and $\beta(y)=\psi(0,y)$. Then $\alpha\in X^*$, $\beta\in Y^*$ and $\psi(x,y)=\alpha(x)+\beta(y)$. Therefore

$$\begin{aligned}(\mu\times\nu)\hat{}\,(\psi)&=\int_{X\times Y}e^{i\psi(x,y)}d\mu(x)d\nu(y)=\int_{X\times Y}e^{i\alpha(x)}e^{i\beta(y)}d\mu(x)d\nu(y)\\&=\exp\left(-\frac{1}{2}(q_\mu(\alpha,\alpha)+q_\nu(\beta,\beta))\right)\\&=\exp\left(-\frac{1}{2}q(\psi,\psi)\right)\end{aligned}$$

where $q=q_{\mu\times\nu}$ is the quadratic form on $X\times Y$ given by

$$q(\psi,\psi)=q_\mu(\alpha,\alpha)+q_\nu(\beta,\beta). \tag{43.1}$$

■

Lemma 43.2. *Suppose that μ is a Gaussian measure on $(X,\mathcal{B})$ and for $\theta\in\mathbb{R}$, let $R_\theta: X\times X\to X\times X$ be the "rotation" map given by*

$$R_\theta(x,y)=(x\cos\theta-y\sin\theta,y\cos\theta+x\sin\theta).$$

Then for all $f\in(\mathcal{B}_{(X\times X)})_b=(\mathcal{B}_X\otimes\mathcal{B}_X)_b$ and $\theta\in\mathbb{R}$,

$$\int_{X\times X}f(x,y)d\mu(x)d\mu(y)=\int_{X\times X}f(R_\theta(x,y))d\mu(x)d\mu(y), \tag{43.2}$$

i.e. $\mu\times\mu$ is invariant under the rotations R_θ for all $\theta\in\mathbb{R}$. (See Corollary 43.4 for the converse of this Lemma.)

Proof. Let q_μ be the quadratic form on X^* such that $\hat{\mu}=e^{-q_\mu/2}$, then by Lemma 43.1 the quadratic form associated to $\mu\times\mu$ is

$$q(\psi)=q_\mu(\psi(\cdot,0))+q_\mu(\psi(0,\cdot)),$$

where $q(\psi)$ is short hand for $q(\psi,\psi)$ with similar notation for $q_\mu(\psi(\cdot,0))$ and $q_\mu(\psi(0,\cdot))$. Using the theory of the Fourier transform, the Lemma will be proved provided that we show, $q(\psi\circ R_\theta)=q(\psi)$ for all $\psi\in(X\times X)^*$. As in the proof of Lemma 43.1 let $\alpha(x)=\psi(x,0)$ and $\beta(x)=\psi(0,x)$, and notice that

$$(\psi\circ R_\theta)(x,y)=\psi(x\cos\theta-y\sin\theta,y\cos\theta+\sin\theta x).$$

So if we define

$$\begin{aligned}\alpha_\theta(x)&=\psi(x\cos\theta,x\sin\theta)=\cos\theta\cdot\alpha(x)+\sin\theta\cdot\beta(x)\text{ and}\\\beta_\theta(x)&=\psi(-y\sin\theta,y\cos\theta)=-\sin\theta\cdot\alpha(y)+\cos\theta\cdot\beta(y)\end{aligned}$$

then

$$\begin{aligned}q(\psi\circ R_\theta)&=q_\mu(\alpha_\theta)+q_\mu(\beta_\theta)\\&=q_\mu(\cos\theta\cdot\alpha+\sin\theta\cdot\beta)+q_\mu(-\sin\theta\cdot\alpha+\cos\theta\cdot\beta).\end{aligned}$$

However,

$$q_\mu(\cos\theta\cdot\alpha+\sin\theta\cdot\beta)=\cos^2\theta\cdot q_\mu(\alpha)+\sin^2\theta\cdot q_\mu(\beta)+2\cos\theta\sin\theta\cdot q_\mu(\alpha,\beta)$$

and

$$q_\mu(-\sin\theta\cdot\alpha+\cos\theta\cdot\beta)=\sin^2\theta\cdot q_\mu(\alpha)+\cos^2\theta\cdot q_\mu(\beta)-2\cos\theta\sin\theta\cdot q_\mu(\alpha,\beta)$$

which shows that

$$q(\psi\circ R_\theta)=q_\mu(\alpha)+q_\mu(\beta)=q(\psi).$$

■

Theorem 43.3 (Ferniques Theorem). *Suppose that μ is a probability measure on $(X,\mathcal{B})$ such that $\mu\times\mu$ is invariant under $R_{\pi/4}$. (For example this is true for any Gaussian measure.) Then there exists $\varepsilon=\varepsilon(\mu)>0$ such that*

$$\int_X e^{\varepsilon\|x\|^2}d\mu(x)<\infty \tag{43.3}$$

and in particular μ has moments to all orders.

Proof. Since

$$e^z = 1 + \int_0^z e^y dy = 1 + \int_{\mathbb{R}} 1_{0\le y\le z} e^y dy,$$

$$\int_X e^{\varepsilon\|x\|^2} d\mu(x) = \int_X \left(1 + \int_{\mathbb{R}} 1_{0\le y\le \varepsilon\|x\|^2} e^y dy\right) d\mu(x)$$
$$= 1 + \int_0^\infty dy\ e^y \mu(\varepsilon\|x\|^2 \ge y). \tag{43.4}$$

Because of this formula it suffices to show that there are constants, $C \in (0,\infty)$ and $\beta \in (1,\infty)$ such that

$$\mu(\varepsilon\|x\|^2 \ge y) \le Ce^{-\beta y} \text{ for all } y \ge 0. \tag{43.5}$$

For if we use Eq. (43.5) in Eq. (43.4), we will have

$$\int_X e^{\varepsilon\|x\|^2} d\mu(x) \le 1 + C\int_0^\infty dy\ e^y e^{-\beta y} = 1 + \frac{C}{\beta - 1} < \infty. \tag{43.6}$$

We will now prove Eq. (43.5). By replacing y with εt^2, Eq. (43.5) is equivalent to showing

$$\mu(\|x\| \ge t) \le Ce^{-\beta\varepsilon t^2} = Ce^{-\gamma t^2} \text{ for all } t \ge 0, \tag{43.7}$$

where $\gamma := \beta\varepsilon$. Because we are free to choose $\varepsilon > 0$ as small as we like, it suffices to prove that Eq. (43.7) for some $\gamma > 0$. Let $P = \mu \times \mu$ on $X \times X$ and let $t \ge s \ge 0$. Then by the $R_{\pi/4}$ invariance of P,

$$\mu(\|x\| \le s)\mu(\|x\| \ge t) = P(\|x\| \le s \text{ and } \|y\| \ge t)$$
$$= P\left(\left\|\frac{x+y}{\sqrt{2}}\right\| \le s \text{ and } \left\|\frac{x-y}{\sqrt{2}}\right\| \ge t\right)$$
$$\le P(|\|x\| - \|y\|| \le \sqrt{2}s \text{ and } \|x\| + \|y\| \ge \sqrt{2}t).$$

Let $a = \|x\|$ and $b = \|y\|$ and notice that if $|a-b| \le \sqrt{2}s$ and $a+b \ge \sqrt{2}t$ then

$$\sqrt{2}t \le a + b \le b + \sqrt{2}s + b = 2b + \sqrt{2}s \text{ and}$$
$$\sqrt{2}t \le a + b \le a + a + \sqrt{2}s = 2a + \sqrt{2}s$$

from which it follows that

$$a \ge \frac{t-s}{\sqrt{2}} \text{ and } b \ge \frac{t-s}{\sqrt{2}},$$

as seen in Figure 43.1 below. Combining these expressions shows

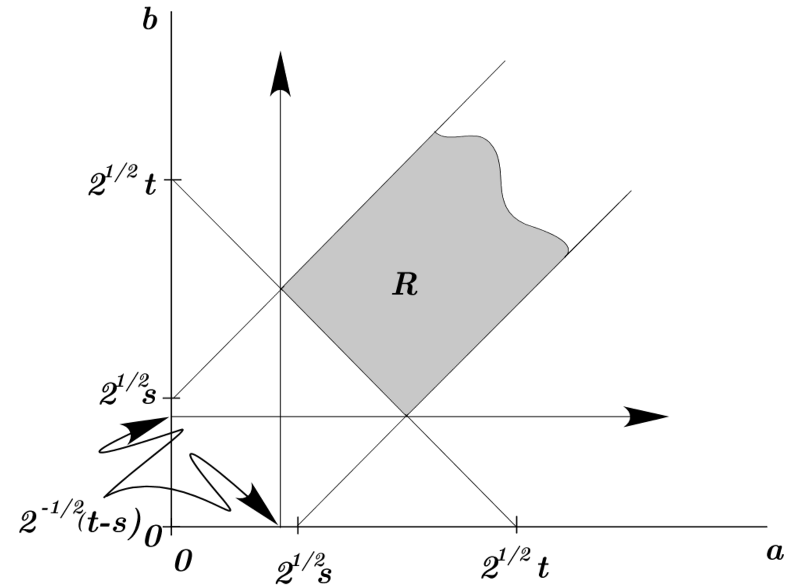


Fig. 43.1. The region R is contained in the region $a \ge \frac{t-s}{\sqrt{2}}$ and $b \ge \frac{t-s}{\sqrt{2}}$.

$$\mu(\|x\| \le s)\mu(\|x\| \ge t) \le P\left(\|x\| \ge \frac{t-s}{\sqrt{2}} \text{ and } \|y\| \ge \frac{t-s}{\sqrt{2}}\right)$$
$$= \left[\mu(\|x\| \ge \frac{t-s}{\sqrt{2}})\right]^2,$$

which is to say for $t \ge s \ge 0$,

$$\mu(\|x\| \ge t) \le \frac{1}{\mu(\|x\| \le s)}\left[\mu(\|x\| \ge \frac{t-s}{\sqrt{2}})\right]^2. \tag{43.8}$$

We will now complete the proof by iterating Eq. (43.8). Define $t_0 = s$ and then define $\{t_n\}_{n=1}^\infty$ inductively so that

$$\frac{t_{n+1} - s}{\sqrt{2}} = t_n \text{ for all } n \tag{43.9}$$

i.e.

$$t_{n+1} = s + \sqrt{2}t_n \tag{43.10}$$

and (by a simple induction argument)

$$t_n = s\sum_{i=0}^n 2^{i/2} = s\frac{2^{\frac{n+1}{2}} - 1}{2^{1/2} - 1} \le s\frac{2^{\frac{n+1}{2}}}{2^{1/2} - 1}. \tag{43.11}$$

Then by Eq. (43.8)

$$\mu(\|x\| \geq t_{n+1}) \leq \frac{1}{\mu(\|x\| \leq s)} \left[\mu(\|x\| \geq \frac{t_{n+1}-s}{\sqrt{2}})\right]^2 = \frac{[\mu(\|x\| \geq t_n)]^2}{\mu(\|x\| \leq s)}$$

or equivalently

$$\alpha_{n+1}(s) := \frac{\mu(\|x\| \geq t_{n+1})}{\mu(\|x\| \leq s)} \leq \left(\frac{\mu(\|x\| \geq t_n)}{\mu(\|x\| \leq s)}\right)^2 =: \alpha_n^2(s).$$

Iterating this inequality implies

$$\alpha_n(s) \leq \alpha_0^{2^n}(s) \text{ with } \alpha_0(s) = \frac{\mu(\|x\| \geq s)}{\mu(\|x\| \leq s)}, \tag{43.12}$$

i.e.

$$\mu(\|x\| \geq t_n) \leq \mu(\|x\| \leq s) \left(\alpha_0(s)\right)^{2^n} \text{ for all } n. \tag{43.13}$$

We now **fix** an $s > 0$ sufficiently large so that $\alpha_0(s) < 1$ and suppose $t \geq 2s$ is given. Choose n so that $t_n \leq t \leq t_{n+1} = s + \sqrt{2} t_n$ in which case (as $0 \leq t - s \leq \sqrt{2} t_n$)

$$\left(\frac{t-s}{\sqrt{2}}\right)^2 \leq t_n^2 \leq \frac{2s^2}{\left(2^{1/2}-1\right)^2} 2^n,$$

i.e.

$$2^n \geq \frac{\left(2^{1/2}-1\right)^2}{4s^2} (t-s)^2.$$

Combining this with Eq. (43.13), using $\alpha_0(s) < 1$, shows

$$\begin{aligned} \mu(\|x\| \geq t) &\leq \mu(\|x\| \geq t_n) \\ &\leq \mu(\|x\| \leq s) \left(\alpha_0(s)\right)^{2^n} \leq \mu(\|x\| \leq s) \rho^{(t-s)^2} \end{aligned}$$

where

$$\rho := \left(\alpha_0(s)\right)^{\frac{\left(2^{1/2}-1\right)^2}{4s^2}} \in (0,1).$$

Since $t \geq 2s$ is equivalent to $(t-s) \geq t/2$, we have $(t-s)^2 \geq (t/2)^2$ and therefore

$$\mu(\|x\| \geq t) \leq \mu(\|x\| \leq s) e^{-\frac{|\ln \rho|}{4} t^2}$$

which is sufficient to prove Eq. (43.7) with $\gamma = \frac{|\ln \rho|}{4}$ and C chosen to be a sufficiently large constant. ∎

Corollary 43.4. *The converse of Lemma 43.2 holds, i.e. if μ is a probability measure such that $\mu \times \mu$ is invariant under R_θ for all $\theta \in \mathbb{R}$, then μ is a (possibly degenerate) Gaussian measure.*

Proof. By Fernique's Theorem 43.3, μ has moments to all orders and the bound in Eq. (43.3) holds from which it follows that $\psi(\alpha) := \hat{\mu}(\alpha)$ is infinitely differentiable in α. The invariance of $\mu \times \mu$ under rotations and the formula, $\widehat{\mu \times \mu}(\alpha, \beta) = \psi(\alpha)\psi(\beta)$, implies

$$\psi(\alpha)\psi(\beta) = \psi\left(\alpha \cos\theta + \beta \sin\theta\right) \psi\left(-\alpha \sin\theta + \beta \cos\theta\right) \tag{43.14}$$

for all $\theta \in \mathbb{R}$ and $\alpha, \beta \in X^*$. I now claim that ψ is never 0. Indeed, if $\psi(\alpha) = 0$ for some $\alpha \in X^*$, let $t_0 := \inf\{t > 0 : \psi(t\alpha) = 0\} > 0$. By replacing α by $t_0\alpha$ if necessary, we may assume $t_0 = 1$. Taking $\beta = 0$ and $\theta = -\pi/4$ in Eq. (43.14) then gives rise to the contradiction:

$$0 = \psi(\alpha) = \psi\left(\frac{1}{\sqrt{2}}\alpha\right) \psi\left(\frac{1}{\sqrt{2}}\alpha\right) \neq 0.$$

Hence ψ is never zero and we may define $q(\alpha, \beta)$ for $\alpha, \beta \in X^*$ by

$$q(\alpha, \beta) := -\frac{\partial_\beta \psi(\alpha)}{\psi(\alpha)}. \tag{43.15}$$

Differentiating Eq. (43.14) at $\theta = 0$ implies

$$0 = \partial_\beta \psi(\alpha) \psi(\beta) - \psi(\alpha) \partial_\alpha \psi(\beta) \;\; \forall \alpha, \beta \in X^*$$

or equivalently $q(\alpha, \beta) = q(\beta, \alpha)$ for all $\alpha, \beta \in X^*$. From the symmetry of $q(\alpha, \beta)$ in α and β and from Eq. (43.15) it follows that q is a symmetric bilinear form on $X \times X$. Since $\psi(t\alpha)|_{t=0} = 1$ and

$$\begin{aligned} \frac{d}{dt}\psi(t\alpha) &= \frac{(\partial_\alpha \psi)(t\alpha)}{\psi(t\alpha)} \psi(t\alpha) \\ &= -q(t\alpha, \alpha)\psi(t\alpha) = -tq(\alpha, \alpha)\psi(t\alpha) \end{aligned}$$

we learn

$$\psi(\alpha) = \psi(t\alpha)|_{t=1} = \exp\left(-\frac{t^2}{2} q(\alpha, \alpha)\right)|_{t=1} = e^{-q(\alpha,\alpha)/2}.$$

Moreover,

$$\begin{aligned} -\int_X \alpha^2(x) d\mu(x) &= \frac{d^2}{dt^2}|_0 \int_X e^{it\alpha(x)} d\mu(x) \\ &= \frac{d^2}{dt^2}|_0 \exp\left(-\frac{t^2}{2} q(\alpha, \alpha)\right) = -q(\alpha, \alpha), \end{aligned}$$

i.e.

$$q(\alpha,\alpha) = \int_X \alpha^2(x) d\mu(x) \geq 0 \text{ for all } \alpha \in X^*.$$

Hence we have shown $\hat{\mu} = e^{-q/2}$ and thus μ is a (possibly degenerate) Gaussian measure. ■

Actually we can do significantly better than this corollary as we show in the next theorem.

Theorem 43.5. *If μ is a probability measure such that $\mu \times \mu$ is invariant under $R_{\pi/4}$, then μ is a (possibly degenerate) Gaussian measure with $\hat{\mu} = e^{-\frac{1}{2}q}$ where $q(\alpha) := q(\alpha,\alpha)$ and*

$$q(\alpha,\beta) := \int_X \alpha(x)\beta(x)\, d\mu(x) \text{ for all } \alpha,\beta \in X^*.$$

Proof. By Fernique's Theorem 43.3, μ has moments to all orders and the bound in Eq. (43.3) holds from which it follows that $\psi(\alpha) := \hat{\mu}(\alpha)$ is infinitely differentiable in α and in fact for any $\alpha \in X^*$, the function,

$$\mathbb{C} \ni z \to \hat{\mu}(z\alpha) \text{ is holomorphic.}$$

The invariance of $\mu \times \mu$ under rotation by $\pi/4$ and the formula, $\widehat{\mu \times \mu}(\alpha,\beta) = \psi(\alpha)\psi(\beta)$, implies

$$\psi(\alpha)\psi(\beta) = \psi\left(\frac{1}{\sqrt{2}}(\alpha+\beta)\right)\psi\left(\frac{1}{\sqrt{2}}(-\alpha+\beta)\right) \; \forall\, \alpha,\beta \in X^*. \tag{43.16}$$

Taking $\alpha = 0$ and then $\beta = 0$ in this equation implies,

$$\psi(\beta) = \psi\left(\frac{1}{\sqrt{2}}\beta\right)\psi\left(\frac{1}{\sqrt{2}}\beta\right) \; \forall\, \beta \in X^* \tag{43.17}$$

$$\psi(\alpha) = \psi\left(\frac{1}{\sqrt{2}}\alpha\right)\psi\left(-\frac{1}{\sqrt{2}}\alpha\right) \; \forall\, \alpha \in X^* \tag{43.18}$$

and in particular we learn

$$\psi\left(\frac{1}{\sqrt{2}}\alpha\right)\psi\left(-\frac{1}{\sqrt{2}}\alpha\right) = \psi\left(\frac{1}{\sqrt{2}}\alpha\right)\psi\left(\frac{1}{\sqrt{2}}\alpha\right) \; \forall\, \alpha \in X^*.$$

For $t \in \mathbb{R}$ near zero we know that $\psi\left(\frac{t}{\sqrt{2}}\alpha\right) \neq 0$ and hence we may conclude that

$$\psi\left(-\frac{t}{\sqrt{2}}\alpha\right) = \psi\left(\frac{t}{\sqrt{2}}\alpha\right)$$

for t near zero and hence by the principle of analytic continuation this equation in fact holds for all $t \in \mathbb{C}$ and in particular for $t = \sqrt{2}$ from which we learn that $\psi(-\alpha) = \psi(\alpha)$ for all $\alpha \in X^*$. Since $\overline{\psi(\alpha)} = \psi(-\alpha) = \psi(\alpha)$ we may now conclude that ψ is real.

Iterating Eq. (43.17) implies,

$$\psi(\beta) = \psi\left(\left(\frac{1}{\sqrt{2}}\right)^n \beta\right)^{2^n} \text{ for all } n \in \mathbb{N}_0. \tag{43.19}$$

Let $f(t) := \psi(t\beta)$ so that $f(0) = 1$, $\dot{f}(0) = 0$ (since $f(t)$ is odd), and

$$\ddot{f}(0) = \left(\frac{d}{dt}\right)^2_{t=0} \int_X e^{it\beta(x)} d\mu(x) = -\int_X \beta^2(x)\, d\mu(x) := -q(\beta).$$

So by Taylor's theorem we have,

$$f(t) = 1 - \frac{1}{2}q(\beta)t^2 + O\left(t^3\right)$$

while by Eq. (43.19) we have

$$f(1) = \left[f\left(2^{-n/2}\right)\right]^{2^n} = \left[1 - \frac{1}{2}q(\beta)2^{-n} + O\left(2^{-3n/2}\right)\right]^{2^n}.$$

So by elementary calculus we have,

$$\ln f(1) = 2^n \cdot \ln\left(1 - \frac{1}{2}q(\beta)2^{-n} + O\left(2^{-3n/2}\right)\right) \to -\frac{1}{2}q(\beta) \text{ as } n \to \infty.$$

Thus we have shown

$$\hat{\mu}(\beta) = \psi(\beta) = f(1) = e^{-\frac{1}{2}q(\beta)} = e^{-\frac{1}{2}\operatorname{Var}_\mu(\beta)},$$

i.e. $\hat{\mu} = e^{-q/2}$ and thus μ is a (possibly degenerate) Gaussian measure. ■

43.2 The Cameron-Martin Theorem

Theorem 43.6 (Cameron-Martin). *Let (H, X, μ) be a Gaussian measure space as above and for $h \in X$ let $\mu_h(A) = \mu(A - h)$ for all $A \in \mathcal{B}_X$. Then $\mu_h \ll \mu$ for all $h \in H$ and*

$$\frac{d\mu_h}{d\mu} = \exp\left((h,\cdot) - \frac{1}{2}\|h\|_H^2\right)$$

where $(h,\cdot)_H := (J^{-1}h) \in K \subset L^2$.

Proof. We must show
$$\int_X f(x+h)d\mu(x) = \int_X e^{(h,x)-\frac{1}{2}\|h\|^2} f(x)d\mu(x)$$
for all $f \in \mathcal{B}(X)_b$. It suffices to show that
$$\int_X e^{i\varphi(x+h)}d\mu(x) = \int_X e^{i\varphi(x)}e^{(h,x)-\frac{1}{2}\|h\|_H^2}d\mu(x)$$
for all $\varphi \in X^*$.The left hand side of the above equation is $e^{i\varphi(h)}e^{-\frac{1}{2}q(\varphi,\varphi)}$. To compute the right hand side, first let $h = h_\psi$, with $\psi \in X^*$. Then
$$(h_\psi, h_\varphi) = q(\psi,\varphi) = \psi(h_\varphi)$$
for all $\varphi \in X^*$ and by density $(h_\psi, h)_H = \psi(h)$ for all $h \in H$. So $(h_\psi, x) \equiv \psi(x)$ and
$$\begin{aligned}\int_X & e^{i\varphi(x)}e^{\psi(x)-\frac{1}{2}q(\psi,\psi)}d\mu(x)\\ &= e^{+\frac{1}{2}q(i\varphi+\psi,i\varphi+\psi)-\frac{1}{2}q(\psi,\psi)}\\ &= e^{-\frac{1}{2}q(\varphi,\varphi)+iq(\varphi,\psi)+\frac{1}{2}q(\psi,\psi)-\frac{1}{2}q(\psi,\psi)}\\ &= e^{-\frac{1}{2}q(\varphi,\psi)+i\varphi(h_\psi)}.\end{aligned}$$
Therefore
$$\int_X e^{i\varphi(x+h)}d\mu(x) = \int e^{i\varphi(x)}e^{J^{-1}h-\frac{1}{2}\|h\|_K^2}d\mu(x) \tag{43.20}$$
for all $h \in JX^*$. If $h \in H$, $f := J^{-1}h \in K \subset X$, and $\varphi_n \in X^*$ such that $J\varphi_n \to h$ in H, then by a simple computation (see Lemma 43.7 below) it follows that $\lim_{n\to\infty}\mathbb{E}_\mu\left|e^f - e^{f_n}\right|^2 = 0$. Using this remark it is now easy to see that Eq. (43.20) holds for all $h \in H$. ∎

Lemma 43.7. *Let $\{f_n, f\}$ be Random variables on a probability space $(\Omega,\mathcal{B},\mathcal{P})$ such that (f_n, f) is jointly Gaussian for all n. If $f_n \to f \in L^2$ then $e^{f_n} \to e^f$ in L^2.*

Proof. By assumption, $\lambda f_n + \alpha f$ is Gaussian for all λ,α and therefore,
$$\mathbb{E}e^{(\lambda f_n+\alpha f)} = e^{\frac{1}{2}\mathbb{E}(\lambda f_n+\alpha f)^2}$$
and
$$\begin{aligned}\mathbb{E}\left(\left|e^f - e^{f_n}\right|^2\right) =& \mathbb{E}\left(e^{2f}+e^{2f_n}-2e^{f_n+f}\right)\\ =& e^{\frac{1}{2}4\|f\|^2}+e^{\frac{1}{2}4\|f_n\|^2}-2e^{\frac{1}{2}\|f_n+f\|^2}\\ &\to e^{2\|f\|^2}+e^{2\|f\|^2}-2e^{\frac{1}{2}\|2f\|^2} = 0\end{aligned}$$
∎

Lemma 43.8. *If μ and ν are two probability measures on $(\Omega,\mathcal{B})$. Then $\mu \perp \nu$ iff $\|\mu-\nu\|_{\mathrm{var}} = 2$.*

Proof. Recall that if φ is a signed measure on $(\Omega,\mathcal{B})$ then Ω may be partitioned as $\Omega = A\cup B$ with $A, B \in \mathcal{B}$ such that $A\cap B = \emptyset$, $\varphi \geq 0$ on A and $\varphi \leq 0$ on B and for any such partition we have
$$\|\varphi\|_{\mathrm{var}} = \varphi(A) - \varphi(B).$$
If $\mu \perp \nu$ there exists a partition $\{A,B\} \subset \mathcal{B}$ of Ω such that $\mu(B) = 0$ and $\nu(A) = 0$. Since $\mu-\nu \geq 0$ on A and $\mu-\nu \leq 0$ on B it follows that
$$\|\mu-\nu\|_{\mathrm{var}} = (\mu-\nu)(A) - (\mu-\nu)(B) = 2.$$
Conversely if $\|\mu-\nu\|_{\mathrm{var}} = 2$ and $\{A,B\} \subset \mathcal{B}$ is a partition of Ω such that $\mu-\nu \geq 0$ on A and $\mu-\nu \leq 0$, then
$$2 = (\mu-\nu)(A) - (\mu-\nu)(B).$$
As $0 \leq (\mu-\nu)(A) \leq 1$ and $0 \leq -(\mu-\nu)(B) \leq 1$ must in fact have $(\mu-\nu)(A) = 1$ and $(\nu-\mu)(B) = 1$. But this can only happen if $\mu(A) = 1$, $\nu(A) = 0$, $\nu(B) = 1$ and $\mu(B) = 0$. This then clearly implies that μ and ν live on the disjoint sets, A and B, respectively and therefore that $\mu \perp \nu$. ∎

Lemma 43.9. *If f, g are unit vectors in $L^2(\nu)$ with $f, g \geq 0$, then*
$$2(1-(f,g)) \leq \int\left|f^2-g^2\right|d\nu \leq 2\sqrt{1-(f,g)^2}. \tag{43.21}$$
In particular, $\int\left|f^2-g^2\right|d\nu = 2$ iff $(f,g) = 0$ which happens iff $f\cdot g = 0$, ν – a.e. as $f, g \geq 0$.

Proof. First observe that $|a-b|^2 \leq \left|a^2-b^2\right|$ for all $a, b \in \mathbb{R}_+$. To prove this suppose with out loss of generality that $a > b$ in which case
$$|a-b|^2 = (a-b)(a-b) \leq (a-b)(a+b) = a^2-b^2 \leq \left|a^2-b^2\right|.$$
From this simple inequality it follows that
$$\int|f^2-g^2|d\nu \geq \int|f-g|^2d\nu = \|f\|_2^2+\|g\|_2^2-2(f,g) = 2-2(f,g)$$
which is the lower bound in Eq. (43.21). For the upper bound we have;
$$\begin{aligned}\int|f^2-g^2|d\nu &= \int|f-g|\,|f+g|d\nu\\ &\leq \sqrt{\int|f-g|^2d\nu\cdot\int|f+g|^2\,d\nu}\\ &= \sqrt{(2-2(f,g))(2+2(f,g))}\\ &= 2\sqrt{1-(f,g)^2}.\end{aligned}$$

■

Lemma 43.10. *The following limit holds;*

$$\lim_{a\to\infty}\int_{\mathbb{R}}\left|\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}-\frac{e^{-\frac{1}{2}(x-a)^2}}{\sqrt{2\pi}}\right|dx=2. \tag{43.22}$$

More precisely if we let $n(dx)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx$ *and* $n_a(dx)=\frac{e^{-\frac{1}{2}(x-a)^2}}{\sqrt{2\pi}}\,dx$, *then*

$$\|n-n_a\|_{\mathrm{var}}\geq 2(1-e^{-\frac{1}{8}a^2})\text{ for all } a\in\mathbb{R}. \tag{43.23}$$

Proof. The computation of the limit in Eq. (43.22) is essentially an obvious consequence of Figure 43.2. In order to derive the analytic estimate in Eq.

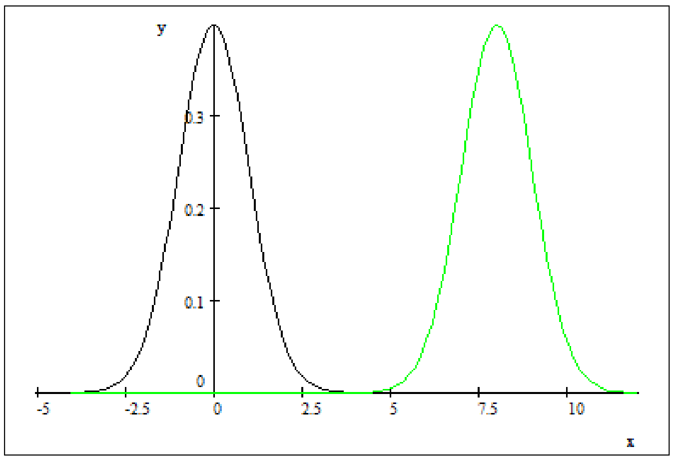


Fig. 43.2. Plot of two Gaussians centered at $x=0$ and $x=8$.

(43.23) we will make use of the lower bound in Eq. (43.21) with

$$f^2=\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}},\quad g^2=\frac{e^{-\frac{1}{2}(x-a)^2}}{\sqrt{2\pi}},\text{ and } d\nu(x)=dx.$$

Since

$$\begin{aligned}(f,g)&=\int_{\mathbb{R}}fg\,dx=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}e^{-\frac{1}{4}x^2}e^{-\frac{1}{4}(x-a)^2}dx\\&=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{4}a^2}\int_{\mathbb{R}}e^{-\frac{1}{2}x^2}e^{\frac{1}{2}x\cdot a}dx\\&=e^{-\frac{1}{4}a^2}e^{\frac{1}{2}\left(\frac{1}{2}a\right)^2}=e^{-\frac{1}{8}a^2},\end{aligned}$$

it follows from Eq. (43.21) that

$$\|n-n_a\|_{\mathrm{var}}=\int_{\mathbb{R}}|f^2-g^2|dx\geq 2(1-e^{-\frac{1}{8}a^2}).$$

■

Proposition 43.11. *For all* $h\in X$ *we have,*

$$\|\mu-\mu_h\|_{\mathrm{var}}\geq 2\cdot\left(1-e^{-\frac{1}{8}\|h\|_H^2}\right). \tag{43.24}$$

In particular if $h\in X\setminus H$, *so that* $\|h\|_H=\infty$, *then* $\mu_h\perp\mu$.

Proof. By Lemma 43.8 it suffices to prove Eq. (43.24) since if $h\in X\setminus H$ then $\|\mu_h-\mu\|_{\mathrm{var}}=2$ and therefore $\mu_h\perp\mu$. To prove Eq. (43.24), let $\varphi\in X^*$ such that $q(\varphi,\varphi)=1$. Then for $f\in(\mathcal{B}_{\mathbb{R}})_b$ we have

$$\mu(f(\varphi(\cdot)))=\int_X f(x)\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}dx$$

and

$$\begin{aligned}\mu_h(f(\varphi(\cdot)))&=\mu(f(\varphi(\cdot)+\varphi(h)))\\&=\int_X f(x)\frac{e^{-\frac{1}{2}(x-\varphi(h))^2}}{\sqrt{2\pi}}\,dx.\end{aligned}$$

Therefore we have,

$$\begin{aligned}\|\mu-\mu_h\|_{\mathrm{var}}&=\sup_{|g|\leq1}|\mu(g)-\mu_h(g)|\\&\geq\sup_{|f|\leq1}|\mu_h(f\circ\varphi)-\mu(f\circ\varphi)|\\&=\|n-n_{\varphi(h)}\|_{\mathrm{var}}\geq 2\left(1-e^{-\frac{1}{8}\varphi(h)^2}\right),\end{aligned}$$

wherein the last inequality we have made use of Lemma 43.10. As $\varphi\in X^*$ with $q(\varphi,\varphi)=1$ is arbitrary it follows that

$$\|\mu-\mu_h\|_{\mathrm{var}}\geq 2\sup_{\varphi:q(\varphi,\varphi)=1}\left(1-e^{-\frac{1}{8}\varphi(h)^2}\right)=2\cdot\left(1-e^{-\frac{1}{8}\|h\|_H^2}\right).$$

■

43.3 Exercises

For the exercises to follows let

$$p_t\left(dx\right) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2} dx$$

be the normal distribution on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ with variance t and $\mu = p_1^{\otimes\mathbb{N}}$ be the infinite product of p_1 as a measure on $\mathbb{R}^\infty$ equipped with the product $\sigma -$ algebra, $\mathcal{B}_{\mathbb{R}^\infty}$. We also let

$$H = \ell^2 := \left\{ x = (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}} : (x, x) = \sum_{j=1}^{\infty} x_j^2 < \infty \right\}.$$

Formally we have

$$d\mu\left(x\right) = \frac{1}{(\sqrt{2\pi})^\infty} \exp\left(-\frac{1}{2}\sum\nolimits_{i=1}^{\infty} x_i^2\right) \prod_{i=1}^{\infty} dx_i.$$

Exercise 43.1. $\mu(\left\{x \in \mathbb{R}^\infty : \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N x_i^2 = 1\right\}) = 1$, in particular $\mu_\lambda \perp \mu$ for all $\lambda \neq 1$. Use strong law of large numbers.

Exercise 43.2. For $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n, \dots) \in X$ with $\alpha_i > 0$, let

$$\ell^2(\alpha) = \left\{ a \in X : (a, a)_\alpha := \sum_{i=1}^{\infty} \alpha_i a_i^2 < \infty \right\}.$$

Show

$$\mu(\ell^2(\alpha)) = \begin{cases} 1 & \text{if} \quad \alpha \in \ell^1 \\ 0 \text{ ortherwise} \end{cases}$$

Gaussian Measure exercises. Support properties of $\mu(\ell^2(\alpha)) = 1$ iff α satisfies ..

Exercise 43.3. $d\mu(\cdot + h) = \rho_h d\mu$ if $h \in \ell^2$.

Gross' Abstract Wiener Spaces

44.1 Conditional expectation results

Let $(\Omega, \mathcal{B}, P)$ be a probability space and X be a separable Banach space. Further let

$$L^p(P; X) = \{f : \Omega \to X : \text{ measurable } \ni \mathbb{E}[\|f\|^p] < \infty\}.$$

In this section we assume that $1 \leq p < \infty$. Recall that conditional expectation relative to a sub-σ – algebra, $\mathcal{G} \subset \mathcal{B}$, is a contraction on $L^p(P; X)$. Let us also recall the following result (which may be found in Stroock): namely if $\mathcal{B}_n \uparrow \mathcal{B}_\infty$ then for $f \in L^p(P, X)$ we have $\mathbb{E}_{\mathcal{B}_n} f \to \mathbb{E}_{\mathcal{B}_\infty} f$ a.s. and in $L^p(P, X)$. See Chapter 40 for the necessary results about Bochner integration and associated conditional expectations.

Now suppose that $f : \Omega \to X$ is a Gaussian mean zero random variable such that $\mu := \mathrm{Law}_P(f)$ is supported on X. Let $H = H_\mu$ be the Cameron – Martin space associated to μ and suppose that $\{e_n\}_{n=1}^\infty$ is any orthonormal subset of H. Let $\varphi_n := (e_n, \cdot)_H^\sim$ be the stochastic extension of $(e_n, \cdot)$ to X which is μ – a.e. well defined. Further let $\mathcal{B}_n := \sigma(\tilde{e}_k : 1 \leq k \leq n)$ and $\mathcal{B}_\infty := \vee_n \mathcal{B}_n = \sigma(\tilde{e}_k : 1 \leq k < \infty)$.

Lemma 44.1. *Let $f_n := \sum_{k=1}^n (\varphi_k \circ f) \cdot e_k : \Omega \to H$ for all $n \in \mathbb{N}$. Then $f_n = \mathbb{E}\left[f | f^{-1}\mathcal{B}_n\right]$ and $f_\infty := \sum_{k=1}^\infty (\varphi_k \circ f) \cdot e_k$ is convergent a.s. and L^p and moreover is a representative of $\mathbb{E}\left[f | f^{-1}\mathcal{B}_\infty\right]$.*

Proof. It is clear that f_n is $\mathcal{B}_n$ – measurable so it suffices to show that

$$\mathbb{E}\left[[f - f_n] g(\varphi_1 \circ f, \ldots, \varphi_n \circ f)\right] = 0$$

for all bounded and measurable functions $g : \mathbb{R}^n \to \mathbb{R}$. In order to check if this is true it suffices to show for all $\psi \in X^*$ that

$$0 = \psi\left(\mathbb{E}\left[[f - f_n] g(\varphi_1 \circ f, \ldots, \varphi_n \circ f)\right]\right) = \mathbb{E}\left[\psi(f - f_n) \cdot g(\varphi_1 \circ f, \ldots, \varphi_n \circ f)\right].$$

However now things are quite easy since $(\psi(f) - \psi(f_n), \varphi_1 \circ f, \ldots, \varphi_n \circ f)$ are jointly Gaussian it suffices to show that

$$\mathbb{E}\left[\psi(f - f_n) \cdot \varphi_i \circ f\right] = 0 \text{ for } i = 1, 2, \ldots, n.$$

This is easily seen to be the case since $\mathbb{E}\left[\psi(f) \cdot \varphi_i \circ f\right] = \psi(e_i)$ and

$$\mathbb{E}\left[\psi(f_n) \varphi_i \circ f\right] = \sum_{k=1}^n \mathbb{E}\left[(\varphi_k \circ f) \varphi_i \circ f\right] \cdot \psi(e_k) = \sum_{k=1}^n \delta_{ik} \cdot \psi(e_k) = \psi(e_i).$$

The second assertion about f_∞ follows from the facts about conditional expectation stated before the lemma. ■

Corollary 44.2 (Kuo [34, Corollary 4.4 on p. 85]). *Suppose that (X, μ) is a Gaussian measure space with $H = H_\mu$ – being the associated Cameron-Martin space. If K is a Hilbert space and $T : X \to K$ is a bounded operator, then $T|_H : H \to K$ is Hilbert-Schmidt,*

$$\|T|_H\|_{HS(H,K)}^2 = \int_X \|Tx\|_K^2 \, d\mu(x) \leq \left(\int_X \|x\|_X^2 \, d\mu(x)\right) \cdot \|T\|_{\mathrm{Hom}(X,K)}. \tag{44.1}$$

Proof. Let $\Omega = X$, $P = \mu$, and $f : \Omega \to X$ be the identity map. Further let $\{f_n\}_{n=1}^\infty$ be as in Lemma 44.1 so that $f_n \to f$ in $L^2(\mu)$ as $n \to \infty$. Since

$$\left|\|Tf\|_K - \|Tf_n\|_K\right| \leq \|T(f - f_n)\|_K \leq \|T\|_{\mathrm{Hom}(X,K)} \|f - f_n\|_X$$

it easily follows that $\|Tf_n\|_K \to \|Tf\|_K$ in $L^2(\mu)$ and therefore,

$$\int_X \|Tx\|_K^2 \, d\mu(x) = \mathbb{E}\|Tf\|_K^2 = \lim_{n\to\infty} \mathbb{E}\|Tf_n\|_K^2.$$

On the other hand,

$$\begin{aligned}
\mathbb{E}\|Tf_n\|_K^2 &= \mathbb{E}\left\|\sum_{k=1}^n (\varphi_k \circ f) \cdot Te_k\right\|_K^2 \\
&= \mathbb{E}\left[\sum_{k,l=1}^n (\varphi_k \circ f)(\varphi_l \circ f) \cdot (Te_k, Te_l)_K\right] \\
&= \sum_{k,l=1}^n \delta_{kl} \cdot (Te_k, Te_l)_K = \sum_{k=1}^n \|Te_k\|_K^2.
\end{aligned}$$

Combining these last two equations proves the equality in Eq. (44.1). By Fernique's theorem $\int_X \|x\|_X^2 \, d\mu(x) < \infty$ and therefore $\|T|_H\|_{HS(H,K)}^2 < \infty$. ■

Exercise 44.1. Suppose that $q : X \times X \to \mathbb{R}$ is a continuous quadratic form on X, i.e. there exists $C < \infty$ such that

$$|q(x,y)| \le C \|x\|_X \|y\|_X \text{ for all } x, y \in X.$$

Given any orthonormal basis, $\{e_k\}_{k=1}^\infty$, of H, show

$$\sum_{k=1}^\infty q(e_k, e_k) = \int_X q(x,x)\, d\mu(x) \in \mathbb{R}. \tag{44.2}$$

Solution to Exercise (44.1). First observe by Fernique's theorem that

$$\int_X |q(x,x)|\, d\mu(x) \le C \int_X \|x\|_X^2\, d\mu(x) < \infty$$

so the integral in Eq. (44.2) is well defined. To finish the proof we will use the notation and ideas in the proof of Corollary 44.2.

Let $\delta_n := f_n - f$. Then

$$\begin{aligned} q(f_n, f_n) &= q(f + \delta_n, f + \delta_n) \\ &= q(f,f) + q(f, \delta_n) + q(\delta_n, f) + q(\delta_n, \delta_n) \end{aligned}$$

so that

$$|q(f_n, f_n) - q(f,f)| \le C \left[2 \|f\|_X \|\delta_n\|_X + \|\delta_n\|_X^2\right]$$

from which it follows that $q(f_n, f_n) \to q(f,f)$ in $L^1(\mu)$ as $n \to \infty$. Therefore

$$\int_X q(x,x)\, d\mu(x) = \mathbb{E}[q(f,f)] = \lim_{n\to\infty} \mathbb{E}[q(f_n, f_n)].$$

On the other hand,

$$\begin{aligned} \mathbb{E}[q(f_n, f_n)] &= \mathbb{E}\left[q\left(\sum_{k=1}^n (\varphi_k \circ f) \cdot e_k, \sum_{l=1}^n (\varphi_l \circ f) \cdot e_l\right)\right] \\ &= \sum_{k,l=1}^n q(e_k, e_l)\, \mathbb{E}[(\varphi_k \circ f)(\varphi_l \circ f)] = \sum_{k,l=1}^n q(e_k, e_l)\, \delta_{kl} \\ &= \sum_{k=1}^n q(e_k, e_k). \end{aligned}$$

The last two equations prove Eq. (44.2).

Given an orthogonal projection Q on H we may define $\tilde{Q} : X \to X$ by $\tilde{Q}x = \sum_{k=1}^\infty \varphi_k(x)\, e_k$ (μ – a.e. convergent) where $\{e_k\}_{k=1}^\infty$ is any orthonormal basis for $\operatorname{Ran}(Q)$. With this definition we have

$$\tilde{Q} = \mathbb{E}[Z | \sigma(\tilde{e}_k : 1 \le k < \infty)] \;\; \mu \text{ a.e.}$$

where $Z : X \to X$ is the identity map. Let us observe that if $u \in \operatorname{Ran}(Q)$, then $u = \sum_{k=1}^\infty (u, e_k)\, e_k$ and we have, since the map, $H \ni u \to \tilde{u} \in L^2(X, \mu)$ is an isometry, that $\tilde{u} = \sum_{k=1}^\infty (u, e_k)\, \tilde{e}_k$ as $L^2(\mu)$ – functions and hence μ – a.e. In this way we see that $\tilde{u}$ is $\overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu$ – measurable for all $u \in \operatorname{Ran}(Q)$. Thus if $\{u_l\}_{l=1}^\infty$ is another orthonormal basis for $\operatorname{Ran}(Q)$, then

$$\sigma(\tilde{u}_l : 1 \le l < \infty) \subset \overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu$$

and therefore,

$$\overline{\sigma(\tilde{u}_l : 1 \le l < \infty)}^\mu \subset \overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu.$$

By symmetry we may also conclude that

$$\overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu \subset \overline{\sigma(\tilde{u}_l : 1 \le l < \infty)}^\mu$$

so that

$$\overline{\sigma(\tilde{u}_l : 1 \le l < \infty)}^\mu = \overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu.$$

Since

$$\begin{aligned} \mathbb{E}[Z | \sigma(\tilde{e}_k : 1 \le k < \infty)] &= \mathbb{E}\left[Z \middle| \overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu\right] \text{ a.e.} \\ &= \mathbb{E}\left[Z \middle| \overline{\sigma(\tilde{u}_l : 1 \le l < \infty)}^\mu\right] \\ &= \mathbb{E}[Z | \sigma(\tilde{u}_l : 1 \le l < \infty)] \text{ a.e.} \end{aligned}$$

it follows that $\tilde{Q}$ is in fact (modulo μ – null sets) well defined independent of the choice of orthonormal basis for $\operatorname{Ran}(Q)$. We now let $\sigma(Q) := \overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu$ so that

$$\tilde{Q}x = \mathbb{E}[x | \sigma(Q)].$$

Now suppose $\alpha \in X^*$. Let us observe that

$$\alpha\left(\tilde{Q}x\right) = \sum_{k=1}^\infty \varphi_k(x)\, \alpha(e_k) \text{ a.s.}$$

so that

$$\left\|\alpha \circ \tilde{Q}\right\|_{L^2} = \sum_{k=1}^\infty |\alpha(e_k)|^2 \le (\alpha, \alpha)_{H^*}.$$

Also observe that $\alpha\left(\tilde{Q}x\right)$ is $\mathcal{B}_\infty$ – measurable, $\left(\alpha \circ \tilde{Q}, \{\varphi_k\}\right)$ is jointly Gaussian and

$$\mathbb{E}\left[\alpha \circ \tilde{Q} \cdot \varphi_k\right] = \alpha(e_k) \text{ for all } k$$

while

$$\mathbb{E}\left[\alpha \circ Z \cdot \varphi_k\right] = \alpha(e_k) \text{ for all } k.$$

From these observations it follows that $\mathbb{E}[\alpha \circ Z|\sigma(Q)] = \alpha \circ \tilde{Q}$ for all $\alpha \in X^*$.

Let us further observe that $Q = Q_1 + Q_2$ where Q_1 and Q_2 are commuting projections. Notice that if Q_i are orthogonal projections then the following are equivalent, $[Q_1, Q_2] = 0$, $Q_1 Q_2 = 0 = Q_2 Q_1$, and $\operatorname{Ran}(Q_1) \perp \operatorname{Ran}(Q_2)$. I now claim that $\tilde{Q}x = \tilde{Q}_1 x + \tilde{Q}_2 x$ a.s. To prove this let $\{e_n\}_{n=1}^\infty$ be a basis for $\operatorname{Ran}(Q_1)$ and $\{u_n\}_{n=1}^\infty$ be a basis for $\operatorname{Ran}(Q_2)$. Then $\{e_1, u_1, e_2, u_2, \dots\}$ is an ON basis for $\operatorname{Ran}(Q) = \operatorname{Ran}(Q_1) \oplus \operatorname{Ran}(Q_2)$ and we have

$$\begin{aligned}
\tilde{Q}x &= \lim_{N\to\infty} \sum_{k=1}^N \left(\tilde{e}_k(x) \cdot e_k + \tilde{u}_k(x) \cdot u_k\right) \\
&= \lim_{N\to\infty} \sum_{k=1}^N \tilde{e}_k(x) \cdot e_k + \lim_{N\to\infty} \sum_{k=1}^N \tilde{u}_k(x) \cdot u_k \\
&= \tilde{Q}_1 x + \tilde{Q}_2 x \text{ for } \mu - \text{a.e. } x.
\end{aligned}$$

If $\{e_n\}_{n=1}^\infty$ is a basis for H then as we saw above $\tilde{u}$ is $\overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu$ – measurable for all $u \in H$. Thus if $\varphi \in X^*$ and $u \in H$ is chosen so that $\varphi|_H = (\cdot, u)_H$, then $\tilde{u} = \varphi$ μ – a.e. on X and therefore φ is $\overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu$ – measurable for all $\varphi \in X^*$. This shows that

$$\mathcal{B}_X = \sigma(X^*) \subset \overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu \subset \mathcal{B}_X^\mu.$$

Thus is follows that

$$\sigma(I) = \overline{\sigma(\tilde{e}_k : 1 \le k < \infty)}^\mu = \mathcal{B}_X^\mu$$

and hence we may conclude in this case that $\tilde{I}x = x$ for μ – a.e. x.

Now suppose that $u \in H$ and $u_n \in H_*$ such that $u_n \to u$ in H and $(u_n, \cdot)_H = \alpha_n|_H$ with $\alpha_n \in X^*$ for all n. Then by above,

$$\mathbb{E}[\alpha_n|\sigma(Q)] = \mathbb{E}[\alpha_n \circ Z|\sigma(Q)] = \alpha_n \circ \tilde{Q}$$

and moreover we have by the contractive properties of conditional expectation that

$$\mathbb{E}[\tilde{u}|\sigma(Q)] = L^p(\mu)\text{-}\lim_{n\to\infty} \alpha_n \circ \tilde{Q}$$

where $\tilde{u} = L^p(\mu)\text{-}\lim_{n\to\infty} \alpha_n$.

Putting this all together we have, among other things, proved the following theorem.

Theorem 44.3. *Suppose that $(X, \mathcal{B}_X, \mu)$ is a Gaussian measure space with Cameron-Martin space, $H = H_\mu$. Then for any orthonormal subset $\{e_k\}_{k=1}^\infty$ of H and any i.i.d. standard normal random variables $\{Z_k\}_{k=1}^\infty$ the sum $\sum_{k=1}^\infty Z_k e_k$ is convergent in X a.s. and $L^p(\mu)$ for all $1 \le p < \infty$. Moreover if $\nu := \operatorname{Law}\left(\sum_{k=1}^\infty Z_k e_k\right)$, then*

$$\hat{\nu}(u) = \exp\left(-\frac{1}{2}\sum_{k=1}^\infty [u(e_k)]^2\right).$$

44.2 Measurable Norms

Definition 44.4. *A norm $\|\cdot\|$ on H is* ***measurable*** *if for every $\varepsilon > 0$ there exists a finite rank projection P_0 such that for all other finite rank projections with $P \perp P_0$ we have*

$$\mu_V(\{x \in V : \|Px\| > \varepsilon\}) < \varepsilon$$

where $V = \operatorname{Ran}(P) \oplus \operatorname{Ran}(P_0)$.

The results to follow may be found in [22, 23, 34, 10, 8]. In what follows we will denote the Hilbert norm on H via $|h| := \sqrt{(h,h)_H}$.

Lemma 44.5 ([34, Lemma 4.2 on p.61.]). *There is a constant $c < \infty$ such that*

$$\|x\| \le c\,|x| \text{ for all } x \in H.$$

Proof. Choose P_0 such that for all $P \perp P_0$,

$$\mu_V\left(\left\{x \in V : \|Px\| > \frac{1}{2}\right\}\right) < \frac{1}{2}$$

where $V = \operatorname{Ran}(P) \oplus \operatorname{Ran}(P_0)$. For $z \in \operatorname{Ran}(P_0)^\perp$, let $Px := |z|^{-2}(z,x)z$ then

$$\|Px\| = |z|^{-2}\,|(z,x)|\,\|z\|$$

and therefore, with $\hat{z} := z/|z|$

$$\frac{1}{2} > \mu_V\left(\left\{x \in V : |(\hat{z},x)| > \frac{|z|}{2\|z\|}\right\}\right) = \frac{2}{\sqrt{2\pi}} \int_{\frac{|z|}{2\|z\|}}^\infty e^{-y^2/2} dy$$

which implies $\frac{|z|}{2\|z\|} \ge M$ for some M, i.e.

$$\|z\| \le \frac{1}{2M}|z|.$$

For general $x \in H$, we thus have

$$\|x\| = \|P_0 x + (1-P_0)\,x\| \le \|P_0 x\| + \|(1-P_0)\,x\|$$
$$\le \|P_0\|\,|x| + \frac{1}{2M}\,|(1-P_0)\,x| \le \left[\|P_0\| + \frac{1}{2M}\right]|x|\,.$$

■

Lemma 44.6. *Let $\|\cdot\|$ be a norm on H such that $\|\cdot\| \le C\,|\cdot|$ and $X := \bar{H}^{\|\cdot\|}$ – the completion of H in with respect to $\|\cdot\|$. Then the transpose $i^* : X^* \to H^*$ of the inclusion map $i : H \to X$ satisfies the following properties:*

1. *i^* is injective.*
2. *The range of i^* is dense in H^*.*
3. *If we let*

$$H_* := \{h \in H : (h,\cdot) : H \to \mathbb{R} \text{ extends continuously to } X\}$$

then H_* is a dense subspace of H and $\{(h,\cdot) : h \in H_*\} = X^*$, wherein we are abusing notation and writing $(h,\cdot)$ for the unique continuous extension of $(h,\cdot)$ to X.

Proof. 1. If $u \in \mathrm{Nul}(i^*)$, then $u \circ i \equiv 0$, i.e. $u|_H \equiv 0$ and therefore $u \equiv 0$ since H is dense in X.

2. Suppose $h \in H$ and $\varphi = (h,\cdot) \in H^*$ is orthogonal to $\mathrm{Ran}\,(i^*)$. Then for all $u \in X^*$ we will have $0 = (\varphi, i^*(u))_{H^*} = u\,(h)$. It now follows by the Hahn Banach theorem that $h = 0$, i.e. $\varphi = 0$. This proves item 2. as every element of H^* is of the form $\varphi = (h,\cdot)$ for some $h \in H$.

3. Let $J : H \to H^*$ be the unitary map of Hilbert spaces defined by $Jh = (h,\cdot)$. Then $h \in H_*$ iff $Jh = \varphi|_H = \varphi \circ i = i^*\varphi$ for some $\varphi \in X^*$. That is to say $H_* = J^{-1} i^*(X^*) = J^{-1}\,\mathrm{Ran}\,(i^*)$ which is dense in H since $J^{-1} : H^* \to H$ is unitary and $i^*(X^*)$ is dense in H^*. ■

Theorem 44.7. *Let H be a real Hilbert space, $\|\cdot\|$ be a measurable norm on H, and let $X := \bar{H}^{\|\cdot\|}$ be the completion of H in this norm. Then there exists a unique Gaussian measure μ on X (i.e. on the Borel σ – field of X) such that*

$$\int_X e^{i\varphi(x)} d\mu(x) = e^{-\frac{1}{2}|\varphi|^2_{H^*}} = e^{-\frac{1}{2}\sum_{h\in S}\varphi^2(h)}\ \forall\ \varphi \in X^*.$$

(See Kuo [34, Theorem 4.1 on p.63], Bogachev [8, Theorem 3.8.6 of on p. 126], Daprato and Zabczyk [10, Theorem 2.12], and section 4.2 and Theorem 5.3.32 of Stroock [50].)

Proof. We follow Kallianpur's proof here. We begin by choosing a sequence of finite rank projections P_n such that $P_n \uparrow I_H$ strongly as $n \to \infty$ and

$$\nu_H\left(\left\{x \in H : \|Px\| > 2^{-n}\right\}\right) < 2^{-n} \text{ for all } P \perp P_n.$$

We now choose an orthonormal basis $\{e_k\}_{k=1}^\infty$ for H such that $P_n h = \sum_{k=1}^{N_n}(h,e_k)e_k$ for some sequence $N_n \uparrow \infty$. Let $\{\xi_k\}_{k=1}^\infty$ be an i.i.d. sequence of standard normal random variables and let

$$S_n := \sum_{k=1}^{N_n} \xi_k e_k.$$

Then by assumption we have

$$P\left(\|S_{n+1} - S_n\| > 2^{-n}\right) = \nu_H\left(\left\{x \in H : \|(P_{n+1} - P_n)\,x\| > 2^{-n}\right\}\right) \le 2^{-n}$$

from which it follows that

$$\sum_{n=1}^\infty P\left(\|S_{n+1} - S_n\| > 2^{-n}\right) < \infty.$$

So by the Borel - Cantelli lemma, P – a.e. we know $\|S_{n+1} - S_n\| \le 2^{-n}$ for almost all n. In particular, with $S_0 = 0$,

$$S = \lim_{N\to\infty} S_N = \sum_{n=0}^\infty (S_{n+1} - S_n)$$

exists in $(X, \|\cdot\|)$, P – a.e. We now define $\mu := S_* P = \mathrm{Law}(S)$. Then for $\varphi \in X^*$ we have using the dominated convergence theorem that

$$\mu\left(e^{i\varphi}\right) = P\left(e^{i\varphi\circ S}\right) = \lim_{N\to\infty} P\left(e^{i\varphi\circ S_N}\right) = \lim_{N\to\infty} e^{-\frac{1}{2}\sum_{k=1}^{N_n}\varphi^2(e_k)} = e^{-\frac{1}{2}|\varphi|^2_{H^*}}.$$

■

Our next goal is to show the converse of Theorem 44.7 is true as well. The proof I will give here I learned from Dan Stroock's talk at the Len Gross conference at Cornell, April 2010 which greatly simplifies L. Gross' original proof.

Theorem 44.8. *Suppose that μ is a non-degenerate mean zero Gaussian measure on a separable Banach space, $(X, \|\cdot\|)$. Let $H = H_\mu$ be the associated reproducing kernel Hilbert space for μ. Then $\|\cdot\|$ is a measurable norm on H. More precisely, if $\{h_n\}_{n=1}^\infty$ is any orthonormal basis for H and $L_n :=$* span $\{h_1,\dots,h_n\}$ *for all $n \in \mathbb{N}$, then for every $\varepsilon > 0$ there $n = n_\varepsilon \in \mathbb{N}$ such that for all $L \subset H$ with $\dim L < \infty$ and $L \perp L_n$ we have*

$$\mu\left(\|P_L(\cdot)\|^2\right) = \int_X \|P_L(x)\|^2\, d\mu(x) \le \varepsilon, \tag{44.3}$$

where $P_L x = \sum_{h\in ONB(L)}(h,x)\,h$.

Proof. Let $\delta > 0$. Assuming Eq. (44.3) holds, we have by Chebyshev's inequality that

$$\mu\left(\{x : \|P_L(x)\| > \delta\}\right) \le \frac{1}{\delta^2}\mu\left(\|P_L(\cdot)\|^2\right) \le \frac{\varepsilon}{\delta^2}.$$

Taking $\varepsilon = \delta^3$ shows that $\mu(\{x : \|P_L(x)\| > \delta\}) < \delta$ for all $L \perp L_n$ with $\dim L < \infty$, i.e. $\|\cdot\|$ is a measurable norm on H. So to finish the proof it suffices to verify Eq. (44.3).

Our proof will be by contradiction. Suppose there exists $\varepsilon > 0$ such that for all $n \in \mathbb{N}$ there exists $L \subset_{f.d.} H$ such that $L \perp L_n$ and

$$\mu\left(\|P_L(\cdot)\|^2\right) > \varepsilon.$$

Under this assumption we are going to produce an orthonormal basis $\{f_k\}_{k=1}^\infty$ for H such that $S_n := \sum_{k=1}^n (\cdot, f_k)_H f_k$ is **not** an $L^2(\mu; X)$ – convergent sequence in violation of Theorem 44.3. To be more specific we will construct inductively a strictly increasing sequence, $\{n_m\}_{m=1}^\infty$, of $\mathbb{N}$ and an an orthonormal basis $\{f_k\}_{k=1}^\infty$ of H such that for each $m \in \mathbb{N}$; $\{f_k\}_{k=1}^{n_m}$ is an orthonormal basis for L_{n_m} and

$$\max_{n_m < k \le n_{m+1}} \mathbb{E}_\mu\left[\|S_k - S_{n_m}\|^2\right] = \max_{n_m < k \le n_{m+1}} \mathbb{E}_\mu\left[\left\|\sum_{i=n_m+1}^k (f_i, \cdot) f_i\right\|_X^2\right] > \varepsilon.$$

The last equation clearly shows that $\{S_n\}_{n=1}^\infty$ is not $L^2(\mu; X)$ Cauchy and hence not $L^2(\mu; X)$ convergent.

We now give the inductive construction. We being by taking $f_1 = h_1$ and $n_1 = 1$. Now assume that we have constructed $n_1 < n_2 < \cdots < n_m$ and $\{f_i\}_{i=1}^{n_m}$ for some level m with the desired properties. By our hypothesis we may find a finite dimensional subspace, $L \subset H$ such that $L \perp L_{n_m}$ and $\mu\left(\|P_L(\cdot)\|^2\right) > \varepsilon$. Let $\{g_i\}_{i=1}^k$ be an orthonormal basis for L and recall that

$$\mathbb{E}_\mu\left[\left\|\sum_{i=1}^k (g_i, \cdot) g_i\right\|_X^2\right] = \mu\left(\|P_L(\cdot)\|_X^2\right) > \varepsilon.$$

For any $N > n_m$, let $P_N : H \ominus L_{n_m} \to L_N \ominus L_{n_m}$ be orthogonal projection. Since $L \subset H \ominus L_{n_m}$ and $P_N \xrightarrow{s} I$ on $H \ominus L_{n_m}$, it follows that $P_N g_i \to g_i$ as $N \to \infty$ for $1 \le i \le k$. We are going to choose N very large so that $|P_N g_i - g_i| \le \delta_k$ where δ_k is a small number depending only on k to be describe shortly. With δ_k sufficiently small $\{P_N g_i\}_{i=1}^k$ will be linearly independent and we may then apply the Gram-Schmidt procedure to the $\{P_N g_i\}_{i=1}^k$ to product an orthonormal subset $\{\bar{g}_i\}_{i=1}^k$ of $L_N \ominus L_{n_m}$. By taking δ_k sufficiently small we may assume that $\max_i |g_i - \bar{g}_i|$ is as small as we please. Integrating the identity,

$$\left\|\sum_{i=1}^k (g_i, \cdot) g_i - \sum_{i=1}^k (\bar{g}_i, \cdot) \bar{g}_i\right\|_H^2 = \sum_{i=1}^k (g_i, \cdot)_H^2 + \sum_{i=1}^k (\bar{g}_i, \cdot)_H^2 - 2\sum_{i,l=1}^k (g_i, \cdot)(\bar{g}_l, \cdot)(g_i, \bar{g}_l)$$

then implies

$$\mathbb{E}_\mu\left\|\sum_{i=1}^k (g_i, \cdot) g_i - \sum_{i=1}^k (\bar{g}_i, \cdot) \bar{g}_i\right\|_H^2 = 2k - \sum_{i,l=1}^k (g_i, \bar{g}_l)^2$$

which again can be made as close to zero as we please by taking δ_k sufficiently small. Therefore using $\|\cdot\|_X \le c\|\cdot\|_H$ for some $c < \infty$, it is possible to choose N so large that

$$\begin{aligned}
\left\|\sum_{i=1}^k (\bar{g}_i, \cdot) \bar{g}_i\right\|_{L^2(\mu;X)} &\ge \left\|\sum_{i=1}^k (g_i, \cdot) g_i\right\|_{L^2(\mu;X)} - \left\|\sum_{i=1}^k (g_i, \cdot) g_i - \sum_{i=1}^k (\bar{g}_i, \cdot) \bar{g}_i\right\|_{L^2(\mu;X)} \\
&\ge \left\|\sum_{i=1}^k (g_i, \cdot) g_i\right\|_{L^2(\mu;X)} - c\left\|\sum_{i=1}^k (g_i, \cdot) g_i - \sum_{i=1}^k (\bar{g}_i, \cdot) \bar{g}_i\right\|_{L^2(\mu;H)} \\
&> \sqrt{\varepsilon}.
\end{aligned}$$

For this large N, we take $n_{m+1} = N$ and let $\{f_{n_m+1}, \ldots, f_N\}$ be an orthonormal basis for $L_N \ominus L_{n_m}$ such that $f_{n_m+l} = \bar{g}_l$ for $1 \le l \le k$ which completes the construction and hence the proof. ∎

Part IX

Appendices

45

Basic Metric Space Facts

Definition 45.1. *A function $d : X \times X \to [0, \infty)$ is called a metric if*

1. *(Symmetry) $d(x, y) = d(y, x)$ for all $x, y \in X$*
2. *(Non-degenerate) $d(x, y) = 0$ if and only if $x = y \in X$*
3. *(Triangle inequality) $d(x, z) \le d(x, y) + d(y, z)$ for all $x, y, z \in X$.*

As primary examples, any normed space $(X, \|\cdot\|)$ (see Definition **??**) is a metric space with $d(x, y) := \|x - y\|$. Thus the space $\ell^p(\mu)$ (as in Theorem **??**) is a metric space for all $p \in [1, \infty]$. Also any subset of a metric space is a metric space. For example a surface Σ in $\mathbb{R}^3$ is a metric space with the distance between two points on Σ being the usual distance in $\mathbb{R}^3$.

Definition 45.2. *Let (X, d) be a metric space. The **open ball** $B(x, \delta) \subset X$ centered at $x \in X$ with radius $\delta > 0$ is the set*

$$B(x, \delta) := \{y \in X : d(x, y) < \delta\}.$$

*We will often also write $B(x, \delta)$ as $B_x(\delta)$. We also define the **closed ball** centered at $x \in X$ with radius $\delta > 0$ as the set $C_x(\delta) := \{y \in X : d(x, y) \le \delta\}$.*

Definition 45.3. *A sequence $\{x_n\}_{n=1}^\infty$ in a metric space (X, d) is said to be **convergent** if there exists a point $x \in X$ such that $\lim_{n\to\infty} d(x, x_n) = 0$. In this case we write $\lim_{n\to\infty} x_n = x$ or $x_n \to x$ as $n \to \infty$.*

Exercise 45.1. Show that x in Definition 45.3 is necessarily unique.

Definition 45.4. *A set $E \subset X$ is **bounded** if $E \subset B(x, R)$ for some $x \in X$ and $R < \infty$. A set $F \subset X$ is **closed** iff every convergent sequence $\{x_n\}_{n=1}^\infty$ which is contained in F has its limit back in F. A set $V \subset X$ is **open** iff V^c is closed. We will write $F \sqsubset X$ to indicate F is a closed subset of X and $V \subset_o X$ to indicate the V is an open subset of X. We also let τ_d denote the collection of open subsets of X relative to the metric d.*

Definition 45.5. *A subset $A \subset X$ is a **neighborhood** of x if there exists an open set $V \subset_o X$ such that $x \in V \subset A$. We will say that $A \subset X$ is an **open neighborhood** of x if A is open and $x \in A$.*

Exercise 45.2. Let $\mathcal{F}$ be a collection of closed subsets of X, show $\cap\mathcal{F} := \cap_{F\in\mathcal{F}} F$ is closed. Also show that finite unions of closed sets are closed, i.e. if $\{F_k\}_{k=1}^n$ are closed sets then $\cup_{k=1}^n F_k$ is closed. (By taking complements, this shows that the collection of open sets, τ_d, is closed under finite intersections and arbitrary unions.)

The following "continuity" facts of the metric d will be used frequently in the remainder of this book.

Lemma 45.6. *For any non empty subset $A \subset X$, let $d_A(x) := \inf\{d(x, a) | a \in A\}$, then*

$$|d_A(x) - d_A(y)| \le d(x, y) \;\forall\, x, y \in X \tag{45.1}$$

and in particular if $x_n \to x$ in X then $d_A(x_n) \to d_A(x)$ as $n \to \infty$. Moreover the set $F_\varepsilon := \{x \in X | d_A(x) \ge \varepsilon\}$ is closed in X.

Proof. Let $a \in A$ and $x, y \in X$, then

$$d_A(x) \le d(x, a) \le d(x, y) + d(y, a).$$

Take the infimum over a in the above equation shows that

$$d_A(x) \le d(x, y) + d_A(y) \quad \forall x, y \in X.$$

Therefore, $d_A(x) - d_A(y) \le d(x, y)$ and by interchanging x and y we also have that $d_A(y) - d_A(x) \le d(x, y)$ which implies Eq. (45.1). If $x_n \to x \in X$, then by Eq. (45.1),

$$|d_A(x) - d_A(x_n)| \le d(x, x_n) \to 0 \text{ as } n \to \infty$$

so that $\lim_{n\to\infty} d_A(x_n) = d_A(x)$. Now suppose that $\{x_n\}_{n=1}^\infty \subset F_\varepsilon$ and $x_n \to x$ in X, then

$$d_A(x) = \lim_{n\to\infty} d_A(x_n) \ge \varepsilon$$

since $d_A(x_n) \ge \varepsilon$ for all n. This shows that $x \in F_\varepsilon$ and hence F_ε is closed. ∎

Corollary 45.7. *The function d satisfies,*

$$|d(x, y) - d(x', y')| \le d(y, y') + d(x, x').$$

In particular $d : X \times X \to [0, \infty)$ is "continuous" in the sense that $d(x, y)$ is close to $d(x', y')$ if x is close to x' and y is close to y'. (The notion of continuity will be developed shortly.)

Proof. By Lemma 45.6 for single point sets and the triangle inequality for the absolute value of real numbers,

$$\begin{aligned}|d(x,y)-d(x',y')| &\le |d(x,y)-d(x,y')|+|d(x,y')-d(x',y')| \\ &\le d(y,y')+d(x,x').\end{aligned}$$

∎

Example 45.8. Let $x \in X$ and $\delta > 0$, then $C_x(\delta)$ and $B_x(\delta)^c$ are closed subsets of X. For example if $\{y_n\}_{n=1}^\infty \subset C_x(\delta)$ and $y_n \to y \in X$, then $d(y_n, x) \le \delta$ for all n and using Corollary 45.7 it follows $d(y,x) \le \delta$, i.e. $y \in C_x(\delta)$. A similar proof shows $B_x(\delta)^c$ is open, see Exercise 45.3.

Exercise 45.3. Show that $V \subset X$ is open iff for every $x \in V$ there is a $\delta > 0$ such that $B_x(\delta) \subset V$. In particular show $B_x(\delta)$ is open for all $x \in X$ and $\delta > 0$. **Hint:** by definition V is not open iff V^c is not closed.

Lemma 45.9 (Approximating open sets from the inside by closed sets). *Let A be a closed subset of X and $F_\varepsilon := \{x \in X | d_A(x) \ge \varepsilon\} \sqsubset X$ be as in Lemma 45.6. Then $F_\varepsilon \uparrow A^c$ as $\varepsilon \downarrow 0$.*

Proof. It is clear that $d_A(x) = 0$ for $x \in A$ so that $F_\varepsilon \subset A^c$ for each $\varepsilon > 0$ and hence $\cup_{\varepsilon>0} F_\varepsilon \subset A^c$. Now suppose that $x \in A^c \subset_o X$. By Exercise 45.3 there exists an $\varepsilon > 0$ such that $B_x(\varepsilon) \subset A^c$, i.e. $d(x,y) \ge \varepsilon$ for all $y \in A$. Hence $x \in F_\varepsilon$ and we have shown that $A^c \subset \cup_{\varepsilon>0} F_\varepsilon$. Finally it is clear that $F_\varepsilon \subset F_{\varepsilon'}$ whenever $\varepsilon' \le \varepsilon$. ∎

Definition 45.10. *Given a set A contained in a metric space X, let $\bar{A} \subset X$ be the* ***closure of*** *A defined by*

$$\bar{A} := \{x \in X : \exists \{x_n\} \subset A \ni x = \lim_{n\to\infty} x_n\}.$$

That is to say $\bar{A}$ contains all ***limit points*** *of A. We say A is* ***dense in*** *X if $\bar{A} = X$, i.e. every element $x \in X$ is a limit of a sequence of elements from A.*

Exercise 45.4. Given $A \subset X$, show $\bar{A}$ is a closed set and in fact

$$\bar{A} = \cap\{F : A \subset F \subset X \text{ with } F \text{ closed}\}. \tag{45.2}$$

That is to say $\bar{A}$ is the smallest closed set containing A.

Definition 45.11. *A subset D of X is* ***dense*** *if $\bar{D} = X$ and we say that X is* ***separable*** *if it contains a countable dense subset.*

45.1 Continuity

Suppose that (X,ρ) and (Y,d) are two metric spaces and $f : X \to Y$ is a function.

Definition 45.12. *A function $f : X \to Y$ is* ***continuous at*** *$x \in X$ if for all $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$d(f(x), f(x')) < \varepsilon \text{ provided that } \rho(x,x') < \delta. \tag{45.3}$$

The function f is said to be ***continuous*** *if f is continuous at all points $x \in X$.*

The following lemma gives two other characterizations of continuity of a function at a point.

Lemma 45.13 (Local Continuity Lemma). *Suppose that (X,ρ) and (Y,d) are two metric spaces and $f : X \to Y$ is a function defined in a neighborhood of a point $x \in X$. Then the following are equivalent:*

1. *f is continuous at $x \in X$.*
2. *For all neighborhoods $A \subset Y$ of $f(x)$, $f^{-1}(A)$ is a neighborhood of $x \in X$.*
3. *For all sequences $\{x_n\}_{n=1}^\infty \subset X$ such that $x = \lim_{n\to\infty} x_n$, $\{f(x_n)\}$ is convergent in Y and*
$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

Proof. 1 $\implies$ 2. If $A \subset Y$ is a neighborhood of $f(x)$, there exists $\varepsilon > 0$ such that $B_{f(x)}(\varepsilon) \subset A$ and because f is continuous there exists a $\delta > 0$ such that Eq. (45.3) holds. Therefore

$$B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right) \subset f^{-1}(A)$$

showing $f^{-1}(A)$ is a neighborhood of x.

2 $\implies$ 3. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ and $x = \lim_{n\to\infty} x_n$. Then for any $\varepsilon > 0$, $B_{f(x)}(\varepsilon)$ is a neighborhood of $f(x)$ and so $f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ is a neighborhood of x which must contain $B_x(\delta)$ for some $\delta > 0$. Because $x_n \to x$, it follows that $x_n \in B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ for a.a. n and this implies $f(x_n) \in B_{f(x)}(\varepsilon)$ for a.a. n, i.e. $d(f(x), f(x_n)) < \varepsilon$ for a.a. n. Since $\varepsilon > 0$ is arbitrary it follows that $\lim_{n\to\infty} f(x_n) = f(x)$.

3. $\implies$ 1. We will show not 1. $\implies$ not 3. If f is not continuous at x, there exists an $\varepsilon > 0$ such that for all $n \in \mathbb{N}$ there exists a point $x_n \in X$ with $\rho(x_n, x) < \frac{1}{n}$ yet $d(f(x_n), f(x)) \ge \varepsilon$. Hence $x_n \to x$ as $n \to \infty$ yet $f(x_n)$ does not converge to $f(x)$. ∎

Here is a global version of the previous lemma.

Lemma 45.14 (Global Continuity Lemma). *Suppose that (X,ρ) and (Y,d) are two metric spaces and $f : X \to Y$ is a function defined on all of X. Then the following are equivalent:*

1. *f is continuous.*
2. *$f^{-1}(V) \in \tau_\rho$ for all $V \in \tau_d$, i.e. $f^{-1}(V)$ is open in X if V is open in Y.*
3. *$f^{-1}(C)$ is closed in X if C is closed in Y.*
4. *For all convergent sequences $\{x_n\} \subset X$, $\{f(x_n)\}$ is convergent in Y and*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

Proof. Since $f^{-1}(A^c) = \left[f^{-1}(A)\right]^c$, it is easily seen that 2. and 3. are equivalent. So because of Lemma 45.13 it only remains to show 1. and 2. are equivalent. If f is continuous and $V \subset Y$ is open, then for every $x \in f^{-1}(V)$, V is a neighborhood of $f(x)$ and so $f^{-1}(V)$ is a neighborhood of x. Hence $f^{-1}(V)$ is a neighborhood of all of its points and from this and Exercise 45.3 it follows that $f^{-1}(V)$ is open. Conversely, if $x \in X$ and $A \subset Y$ is a neighborhood of $f(x)$ then there exists $V \subset_o X$ such that $f(x) \in V \subset A$. Hence $x \in f^{-1}(V) \subset f^{-1}(A)$ and by assumption $f^{-1}(V)$ is open showing $f^{-1}(A)$ is a neighborhood of x. Therefore f is continuous at x and since $x \in X$ was arbitrary, f is continuous. ■

Example 45.15. The function d_A defined in Lemma 45.6 is continuous for each $A \subset X$. In particular, if $A = \{x\}$, it follows that $y \in X \to d(y,x)$ is continuous for each $x \in X$.

Exercise 45.5. Use Example 45.15 and Lemma 45.14 to recover the results of Example 45.8.

The next result shows that there are lots of continuous functions on a metric space (X,d).

Lemma 45.16 (Urysohn's Lemma for Metric Spaces). *Let (X,d) be a metric space and suppose that A and B are two disjoint closed subsets of X. Then*

$$f(x) = \frac{d_B(x)}{d_A(x) + d_B(x)} \text{ for } x \in X \tag{45.4}$$

defines a continuous function, $f : X \to [0,1]$, such that $f(x) = 1$ for $x \in A$ and $f(x) = 0$ if $x \in B$.

Proof. By Lemma 45.6, d_A and d_B are continuous functions on X. Since A and B are closed, $d_A(x) > 0$ if $x \notin A$ and $d_B(x) > 0$ if $x \notin B$. Since $A \cap B = \emptyset$, $d_A(x) + d_B(x) > 0$ for all x and $(d_A + d_B)^{-1}$ is continuous as well. The remaining assertions about f are all easy to verify. ■

Sometimes Urysohn's lemma will be use in the following form. Suppose $F \subset V \subset X$ with F being closed and V being open, then there exists $f \in C(X,[0,1]))$ such that $f = 1$ on F while $f = 0$ on V^c. This of course follows from Lemma 45.16 by taking $A = F$ and $B = V^c$.

45.2 Completeness in Metric Spaces

Definition 45.17 (Cauchy sequences). *A sequence $\{x_n\}_{n=1}^\infty$ in a metric space (X,d) is* ***Cauchy*** *provided that*

$$\lim_{m,n\to\infty} d(x_n, x_m) = 0.$$

Exercise 45.6. Show that convergent sequences are always Cauchy sequences. The converse is not always true. For example, let $X = \mathbb{Q}$ be the set of rational numbers and $d(x,y) = |x-y|$. Choose a sequence $\{x_n\}_{n=1}^\infty \subset \mathbb{Q}$ which converges to $\sqrt{2} \in \mathbb{R}$, then $\{x_n\}_{n=1}^\infty$ is $(\mathbb{Q},d)$ – Cauchy but not $(\mathbb{Q},d)$ – convergent. The sequence does converge in $\mathbb{R}$ however.

Definition 45.18. *A metric space (X,d) is* ***complete*** *if all Cauchy sequences are convergent sequences.*

Exercise 45.7. Let (X,d) be a complete metric space. Let $A \subset X$ be a subset of X viewed as a metric space using $d|_{A\times A}$. Show that $(A, d|_{A\times A})$ is complete iff A is a closed subset of X.

Example 45.19. Examples 2. – 4. of complete metric spaces will be verified in Chapter **??** below.

1. $X = \mathbb{R}$ and $d(x,y) = |x-y|$, see Theorem **??** above.
2. $X = \mathbb{R}^n$ and $d(x,y) = \|x-y\|_2 = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$.
3. $X = \ell^p(\mu)$ for $p \in [1,\infty]$ and any weight function $\mu : X \to (0,\infty)$.
4. $X = C([0,1],\mathbb{R})$ – the space of continuous functions from $[0,1]$ to $\mathbb{R}$ and

$$d(f,g) := \max_{t\in[0,1]} |f(t) - g(t)|.$$

   This is a special case of Lemma **??** below.

5. Let $X = C([0,1],\mathbb{R})$ and

$$d(f,g) := \int_0^1 |f(t) - g(t)|\, dt.$$

   You are asked in Exercise **??** to verify that (X,d) is a metric space which is **not** complete.

Exercise 45.8 (Completions of Metric Spaces). Suppose that (X,d) is a (not necessarily complete) metric space. Using the following outline show there exists a complete metric space $\left(\bar{X},\bar{d}\right)$ and an isometric map $i: X \to \bar{X}$ such that $i(X)$ is dense in $\bar{X}$, see Definition 45.10.

1. Let $\mathcal{C}$ denote the collection of Cauchy sequences $a = \{a_n\}_{n=1}^\infty \subset X$. Given two element $a,b \in \mathcal{C}$ show $d_{\mathcal{C}}(a,b) := \lim_{n\to\infty} d(a_n,b_n)$ exists, $d_{\mathcal{C}}(a,b) \geq 0$ for all $a,b \in \mathcal{C}$ and $d_{\mathcal{C}}$ satisfies the triangle inequality,
$$d_{\mathcal{C}}(a,c) \leq d_{\mathcal{C}}(a,b) + d_{\mathcal{C}}(b,c) \text{ for all } a,b,c \in \mathcal{C}.$$
Thus $(\mathcal{C}, d_{\mathcal{C}})$ would be a metric space if it were true that $d_{\mathcal{C}}(a,b) = 0$ iff $a = b$. This however is false, for example if $a_n = b_n$ for all $n \geq 100$, then $d_{\mathcal{C}}(a,b) = 0$ while a need not equal b.
2. Define two elements $a,b \in \mathcal{C}$ to be equivalent (write $a \sim b$) whenever $d_{\mathcal{C}}(a,b) = 0$. Show " $\sim$ " is an equivalence relation on $\mathcal{C}$ and that $d_{\mathcal{C}}(a',b') = d_{\mathcal{C}}(a,b)$ if $a \sim a'$ and $b \sim b'$. (**Hint:** see Corollary 45.7.)
3. Given $a \in \mathcal{C}$ let $\bar{a} := \{b \in \mathcal{C} : b \sim a\}$ denote the equivalence class containing a and let $\bar{X} := \{\bar{a} : a \in \mathcal{C}\}$ denote the collection of such equivalence classes. Show that $\bar{d}\left(\bar{a},\bar{b}\right) := d_{\mathcal{C}}(a,b)$ is well defined on $\bar{X} \times \bar{X}$ and verify $\left(\bar{X},\bar{d}\right)$ is a metric space.
4. For $x \in X$ let $i(x) = \bar{a}$ where a is the constant sequence, $a_n = x$ for all n. Verify that $i: X \to \bar{X}$ is an isometric map and that $i(X)$ is dense in $\bar{X}$.
5. Verify $\left(\bar{X},\bar{d}\right)$ is complete. **Hint:** if $\{\bar{a}(m)\}_{m=1}^\infty$ is a Cauchy sequence in $\bar{X}$ choose $b_m \in X$ such that $\bar{d}(i(b_m), \bar{a}(m)) \leq 1/m$. Then show $\bar{a}(m) \to \bar{b}$ where $b = \{b_m\}_{m=1}^\infty$.

45.3 Compactness

Definition 45.20. *The subset A of a topological space $(X\ \tau)$ is said to be* ***compact*** *if every open cover (Definition **??**) of A has finite a sub-cover, i.e. if $\mathcal{U}$ is an open cover of A there exists $\mathcal{U}_0 \subset\subset \mathcal{U}$ such that $\mathcal{U}_0$ is a cover of A. (We will write $A \sqsubset\sqsubset X$ to denote that $A \subset X$ and A is compact.) A subset $A \subset X$ is* ***precompact*** *if $\bar{A}$ is compact.*

Proposition 45.21. *Suppose that $K \subset X$ is a compact set and $F \subset K$ is a closed subset. Then F is compact. If $\{K_i\}_{i=1}^n$ is a finite collections of compact subsets of X then $K = \cup_{i=1}^n K_i$ is also a compact subset of X.*

Proof. Let $\mathcal{U} \subset \tau$ be an open cover of F, then $\mathcal{U} \cup \{F^c\}$ is an open cover of K. The cover $\mathcal{U} \cup \{F^c\}$ of K has a finite subcover which we denote by $\mathcal{U}_0 \cup \{F^c\}$ where $\mathcal{U}_0 \subset\subset \mathcal{U}$. Since $F \cap F^c = \emptyset$, it follows that $\mathcal{U}_0$ is the desired subcover of F. For the second assertion suppose $\mathcal{U} \subset \tau$ is an open cover of K. Then $\mathcal{U}$ covers each compact set K_i and therefore there exists a finite subset $\mathcal{U}_i \subset\subset \mathcal{U}$ for each i such that $K_i \subset \cup \mathcal{U}_i$. Then $\mathcal{U}_0 := \cup_{i=1}^n \mathcal{U}_i$ is a finite cover of K. ■

Exercise 45.9 (Suggested by Michael Gurvich). Show by example that the intersection of two compact sets need not be compact. (This pathology disappears if one assumes the topology is Hausdorff, see Definition **??** below.)

Exercise 45.10. Suppose $f: X \to Y$ is continuous and $K \subset X$ is compact, then $f(K)$ is a compact subset of Y. Give an example of continuous map, $f: X \to Y$, and a compact subset K of Y such that $f^{-1}(K)$ is not compact.

Exercise 45.11 (Dini's Theorem). Let X be a compact topological space and $f_n: X \to [0,\infty)$ be a sequence of continuous functions such that $f_n(x) \downarrow 0$ as $n \to \infty$ for each $x \in X$. Show that in fact $f_n \downarrow 0$ uniformly in x, i.e. $\sup_{x\in X} f_n(x) \downarrow 0$ as $n \to \infty$. **Hint:** Given $\varepsilon > 0$, consider the open sets $V_n := \{x \in X : f_n(x) < \varepsilon\}$.

Definition 45.22. *A collection $\mathcal{F}$ of closed subsets of a topological space (X,τ) has the* ***finite intersection property*** *if $\cap \mathcal{F}_0 \neq \emptyset$ for all $\mathcal{F}_0 \subset\subset \mathcal{F}$.*

The notion of compactness may be expressed in terms of closed sets as follows.

Proposition 45.23. *A topological space X is compact iff every family of closed sets $\mathcal{F} \subset 2^X$ having the* ***finite intersection property*** *satisfies $\bigcap \mathcal{F} \neq \emptyset$.*

Proof. $(\Rightarrow)$ Suppose that X is compact and $\mathcal{F} \subset 2^X$ is a collection of closed sets such that $\bigcap \mathcal{F} = \emptyset$. Let
$$\mathcal{U} = \mathcal{F}^c := \{C^c : C \in \mathcal{F}\} \subset \tau,$$
then $\mathcal{U}$ is a cover of X and hence has a finite subcover, $\mathcal{U}_0$. Let $\mathcal{F}_0 = \mathcal{U}_0^c \subset\subset \mathcal{F}$, then $\cap \mathcal{F}_0 = \emptyset$ so that $\mathcal{F}$ does not have the finite intersection property. $(\Leftarrow)$ If X is not compact, there exists an open cover $\mathcal{U}$ of X with no finite subcover. Let
$$\mathcal{F} = \mathcal{U}^c := \{U^c : U \in \mathcal{U}\},$$
then $\mathcal{F}$ is a collection of closed sets with the finite intersection property while $\bigcap \mathcal{F} = \emptyset$. ■

Exercise 45.12. Let (X,τ) be a topological space. Show that $A \subset X$ is compact iff (A, τ_A) is a compact topological space.

Metric Space Compactness Criteria

Let (X,d) be a metric space and for $x \in X$ and $\varepsilon > 0$ let
$$B'_x(\varepsilon) := B_x(\varepsilon) \setminus \{x\}$$
be the ball centered at x of radius $\varepsilon > 0$ with x deleted. Recall from Definition **??** that a point $x \in X$ is an accumulation point of a subset $E \subset X$ if $\emptyset \neq E \cap V \setminus \{x\}$ for all open neighborhoods, V, of x. The proof of the following elementary lemma is left to the reader.

Lemma 45.24. *Let $E \subset X$ be a subset of a metric space (X, d). Then the following are equivalent:*

1. *$x \in X$ is an accumulation point of E.*
2. *$B'_x(\varepsilon) \cap E \neq \emptyset$ for all $\varepsilon > 0$.*
3. *$B_x(\varepsilon) \cap E$ is an infinite set for all $\varepsilon > 0$.*
4. *There exists $\{x_n\}_{n=1}^{\infty} \subset E \setminus \{x\}$ with $\lim_{n\to\infty} x_n = x$.*

Definition 45.25. *A metric space (X, d) is ε – **bounded** $(\varepsilon > 0)$ if there exists a finite cover of X by balls of radius ε and it is **totally bounded** if it is ε – bounded for all $\varepsilon > 0$.*

Theorem 45.26. *Let (X, d) be a metric space. The following are equivalent.*

(a) X is compact.
(b) Every infinite subset of X has an accumulation point.
(c) Every sequence $\{x_n\}_{n=1}^{\infty} \subset X$ has a convergent subsequence.
(d) X is totally bounded and complete.

Proof. The proof will consist of showing that $a \Rightarrow b \Rightarrow c \Rightarrow d \Rightarrow a$.

$(a \Rightarrow b)$ We will show that **not** $b \Rightarrow$ **not** a. Suppose there exists an infinite subset $E \subset X$ which has no accumulation points. Then for all $x \in X$ there exists $\delta_x > 0$ such that $V_x := B_x(\delta_x)$ satisfies $(V_x \setminus \{x\}) \cap E = \emptyset$. Clearly $\mathcal{V} = \{V_x\}_{x\in X}$ is a cover of X, yet $\mathcal{V}$ has no finite sub cover. Indeed, for each $x \in X$, $V_x \cap E \subset \{x\}$ and hence if $\Lambda \subset\subset X$, $\cup_{x\in\Lambda} V_x$ can only contain a finite number of points from E (namely $\Lambda \cap E$). Thus for any $\Lambda \subset\subset X$, $E \nsubseteq \cup_{x\in\Lambda} V_x$ and in particular $X \neq \cup_{x\in\Lambda} V_x$. (See Figure 45.1.)

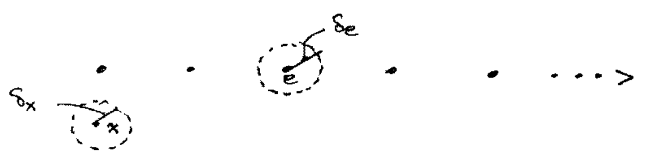


Fig. 45.1. The construction of an open cover with no finite sub-cover.

$(b \Rightarrow c)$ Let $\{x_n\}_{n=1}^{\infty} \subset X$ be a sequence and $E := \{x_n : n \in \mathbb{N}\}$. If $\#(E) < \infty$, then $\{x_n\}_{n=1}^{\infty}$ has a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ which is constant and hence convergent. On the other hand if $\#(E) = \infty$ then by assumption E has an accumulation point and hence by Lemma 45.24, $\{x_n\}_{n=1}^{\infty}$ has a convergent subsequence.

$(c \Rightarrow d)$ Suppose $\{x_n\}_{n=1}^{\infty} \subset X$ is a Cauchy sequence. By assumption there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ which is convergent to some point $x \in X$. Since $\{x_n\}_{n=1}^{\infty}$ is Cauchy it follows that $x_n \to x$ as $n \to \infty$ showing X is complete. We now show that X is totally bounded. Let $\varepsilon > 0$ be given and choose an arbitrary point $x_1 \in X$. If possible choose $x_2 \in X$ such that $d(x_2, x_1) \geq \varepsilon$, then if possible choose $x_3 \in X$ such that $d_{\{x_1,x_2\}}(x_3) \geq \varepsilon$ and continue inductively choosing points $\{x_j\}_{j=1}^{n} \subset X$ such that $d_{\{x_1,\dots,x_{n-1}\}}(x_n) \geq \varepsilon$. This process must terminate, for otherwise we would produce a sequence $\{x_n\}_{n=1}^{\infty} \subset X$ which can have no convergent subsequences. Indeed, the x_n have been chosen so that $d(x_n, x_m) \geq \varepsilon > 0$ for every $m \neq n$ and hence no subsequence of $\{x_n\}_{n=1}^{\infty}$ can be Cauchy.

$(d \Rightarrow a)$ For sake of contradiction, assume there exists an open cover $\mathcal{V} = \{V_\alpha\}_{\alpha\in A}$ of X with no finite subcover. Since X is totally bounded for each $n \in \mathbb{N}$ there exists $\Lambda_n \subset\subset X$ such that

$$X = \bigcup_{x\in\Lambda_n} B_x(1/n) \subset \bigcup_{x\in\Lambda_n} C_x(1/n).$$

Choose $x_1 \in \Lambda_1$ such that no finite subset of $\mathcal{V}$ covers $K_1 := C_{x_1}(1)$. Since $K_1 = \cup_{x\in\Lambda_2} K_1 \cap C_x(1/2)$, there exists $x_2 \in \Lambda_2$ such that $K_2 := K_1 \cap C_{x_2}(1/2)$ can not be covered by a finite subset of $\mathcal{V}$, see Figure 45.2. Continuing this way inductively, we construct sets $K_n = K_{n-1} \cap C_{x_n}(1/n)$ with $x_n \in \Lambda_n$ such that no K_n can be covered by a finite subset of $\mathcal{V}$. Now choose $y_n \in K_n$ for each n. Since $\{K_n\}_{n=1}^{\infty}$ is a decreasing sequence of closed sets such that $\operatorname{diam}(K_n) \leq 2/n$, it follows that $\{y_n\}$ is a Cauchy and hence convergent with

$$y = \lim_{n\to\infty} y_n \in \cap_{m=1}^{\infty} K_m.$$

Since $\mathcal{V}$ is a cover of X, there exists $V \in \mathcal{V}$ such that $y \in V$. Since $K_n \downarrow \{y\}$ and $\operatorname{diam}(K_n) \to 0$, it now follows that $K_n \subset V$ for some n large. But this violates the assertion that K_n can not be covered by a finite subset of $\mathcal{V}$. ■

Corollary 45.27. *Any compact metric space (X, d) is second countable and hence also separable by Exercise **??**. (See Example **??** below for an example of a compact topological space which is not separable.)*

Proof. To each integer n, there exists $\Lambda_n \subset\subset X$ such that $X = \cup_{x\in\Lambda_n} B(x, 1/n)$. The collection of open balls,

$$\mathcal{V} := \cup_{n\in\mathbb{N}} \cup_{x\in\Lambda_n} \{B(x, 1/n)\}$$

forms a countable basis for the metric topology on X. To check this, suppose that $x_0 \in X$ and $\varepsilon > 0$ are given and choose $n \in \mathbb{N}$ such that $1/n < \varepsilon/2$ and $x \in \Lambda_n$ such that $d(x_0, x) < 1/n$. Then $B(x, 1/n) \subset B(x_0, \varepsilon)$ because for $y \in B(x, 1/n)$,

$$d(y, x_0) \leq d(y, x) + d(x, x_0) < 2/n < \varepsilon.$$

■

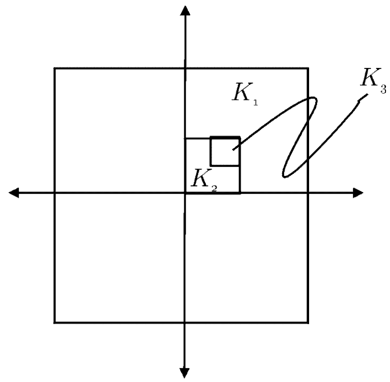


Fig. 45.2. Nested Sequence of cubes.

Corollary 45.28. *The compact subsets of $\mathbb{R}^n$ are the closed and bounded sets.*

Proof. This is a consequence of Theorem **??** and Theorem 45.26. Here is another proof. If K is closed and bounded then K is complete (being the closed subset of a complete space) and K is contained in $[-M,M]^n$ for some positive integer M. For $\delta > 0$, let

$$\Lambda_\delta = \delta\mathbb{Z}^n \cap [-M,M]^n := \{\delta x : x \in \mathbb{Z}^n \text{ and } \delta|x_i| \le M \text{ for } i = 1,2,\dots,n\}.$$

We will show, by choosing $\delta > 0$ sufficiently small, that

$$K \subset [-M,M]^n \subset \cup_{x\in\Lambda_\delta} B(x,\varepsilon) \tag{45.5}$$

which shows that K is totally bounded. Hence by Theorem 45.26, K is compact. Suppose that $y \in [-M,M]^n$, then there exists $x \in \Lambda_\delta$ such that $|y_i - x_i| \le \delta$ for $i = 1,2,\dots,n$. Hence

$$d^2(x,y) = \sum_{i=1}^n (y_i - x_i)^2 \le n\delta^2$$

which shows that $d(x,y) \le \sqrt{n}\delta$. Hence if choose $\delta < \varepsilon/\sqrt{n}$ we have shows that $d(x,y) < \varepsilon$, i.e. Eq. (45.5) holds. ∎

Example 45.29. Let $X = \ell^p(\mathbb{N})$ with $p \in [1,\infty)$ and $\mu \in \ell^p(\mathbb{N})$ such that $\mu(k) \ge 0$ for all $k \in \mathbb{N}$. The set

$$K := \{x \in X : |x(k)| \le \mu(k) \text{ for all } k \in \mathbb{N}\}$$

is compact. To prove this, let $\{x_n\}_{n=1}^\infty \subset K$ be a sequence. By compactness of closed bounded sets in $\mathbb{C}$, for each $k \in \mathbb{N}$ there is a subsequence of $\{x_n(k)\}_{n=1}^\infty \subset \mathbb{C}$ which is convergent. By Cantor's diagonalization trick, we may choose a subsequence $\{y_n\}_{n=1}^\infty$ of $\{x_n\}_{n=1}^\infty$ such that $y(k) := \lim_{n\to\infty} y_n(k)$ exists for all $k \in \mathbb{N}$.[1] Since $|y_n(k)| \le \mu(k)$ for all n it follows that $|y(k)| \le \mu(k)$, i.e. $y \in K$. Finally

$$\lim_{n\to\infty} \|y - y_n\|_p^p = \lim_{n\to\infty} \sum_{k=1}^\infty |y(k) - y_n(k)|^p = \sum_{k=1}^\infty \lim_{n\to\infty} |y(k) - y_n(k)|^p = 0$$

wherein we have used the Dominated convergence theorem. (Note

$$|y(k) - y_n(k)|^p \le 2^p \mu^p(k)$$

and μ^p is summable.) Therefore $y_n \to y$ and we are done.

Alternatively, we can prove K is compact by showing that K is closed and totally bounded. It is simple to show K is closed, for if $\{x_n\}_{n=1}^\infty \subset K$ is a convergent sequence in X, $x := \lim_{n\to\infty} x_n$, then

$$|x(k)| \le \lim_{n\to\infty} |x_n(k)| \le \mu(k) \;\forall\; k \in \mathbb{N}.$$

This shows that $x \in K$ and hence K is closed. To see that K is totally bounded, let $\varepsilon > 0$ and choose N such that $\left(\sum_{k=N+1}^\infty |\mu(k)|^p\right)^{1/p} < \varepsilon$. Since $\prod_{k=1}^N C_{\mu(k)}(0) \subset \mathbb{C}^N$ is closed and bounded, it is compact. Therefore there exists a finite subset $\Lambda \subset \prod_{k=1}^N C_{\mu(k)}(0)$ such that

$$\prod_{k=1}^N C_{\mu(k)}(0) \subset \cup_{z\in\Lambda} B_z^N(\varepsilon)$$

where $B_z^N(\varepsilon)$ is the open ball centered at $z \in \mathbb{C}^N$ relative to the $\ell^p(\{1,2,3,\dots,N\})$ – norm. For each $z \in \Lambda$, let $\tilde{z} \in X$ be defined by $\tilde{z}(k) = z(k)$ if $k \le N$ and $\tilde{z}(k) = 0$ for $k \ge N+1$. I now claim that

$$K \subset \cup_{z\in\Lambda} B_{\tilde{z}}(2\varepsilon) \tag{45.6}$$

[1] The argument is as follows. Let $\{n_j^1\}_{j=1}^\infty$ be a subsequence of $\mathbb{N} = \{n\}_{n=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^1}(1)$ exists. Now choose a subsequence $\{n_j^2\}_{j=1}^\infty$ of $\{n_j^1\}_{j=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^2}(2)$ exists and similarly $\{n_j^3\}_{j=1}^\infty$ of $\{n_j^2\}_{j=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^3}(3)$ exists. Continue on this way inductively to get

$$\{n\}_{n=1}^\infty \supset \{n_j^1\}_{j=1}^\infty \supset \{n_j^2\}_{j=1}^\infty \supset \{n_j^3\}_{j=1}^\infty \supset \dots$$

such that $\lim_{j\to\infty} x_{n_j^k}(k)$ exists for all $k \in \mathbb{N}$. Let $m_j := n_j^j$ so that eventually $\{m_j\}_{j=1}^\infty$ is a subsequence of $\{n_j^k\}_{j=1}^\infty$ for all k. Therefore, we may take $y_j := x_{m_j}$.

which, when verified, shows K is totally bounded. To verify Eq. (45.6), let $x \in K$ and write $x = u + v$ where $u(k) = x(k)$ for $k \le N$ and $u(k) = 0$ for $k < N$. Then by construction $u \in B_{\tilde{z}}(\varepsilon)$ for some $\tilde{z} \in \Lambda$ and

$$\|v\|_p \le \left(\sum_{k=N+1}^{\infty} |\mu(k)|^p \right)^{1/p} < \varepsilon.$$

So we have

$$\|x - \tilde{z}\|_p = \|u + v - \tilde{z}\|_p \le \|u - \tilde{z}\|_p + \|v\|_p < 2\varepsilon.$$

Exercise 45.13 (Extreme value theorem). Let (X, τ) be a compact topological space and $f : X \to \mathbb{R}$ be a continuous function. Show $-\infty < \inf f \le \sup f < \infty$ and there exists $a, b \in X$ such that $f(a) = \inf f$ and $f(b) = \sup f$[2]. **Hint:** use Exercise 45.10 and Corollary 45.28.

Exercise 45.14 (Uniform Continuity). Let (X, d) be a compact metric space, (Y, ρ) be a metric space and $f : X \to Y$ be a continuous function. Show that f is uniformly continuous, i.e. if $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(y), f(x)) < \varepsilon$ if $x, y \in X$ with $d(x, y) < \delta$. **Hint:** you could follow the argument in the proof of Theorem **??**.

Definition 45.30. *Let L be a vector space. We say that two norms, $|\cdot|$ and $\|\cdot\|$, on L are equivalent if there exists constants $\alpha, \beta \in (0, \infty)$ such that*

$$\|f\| \le \alpha |f| \text{ and } |f| \le \beta \|f\| \text{ for all } f \in L.$$

Theorem 45.31. *Let L be a finite dimensional vector space. Then any two norms $|\cdot|$ and $\|\cdot\|$ on L are equivalent. (This is typically not true for norms on infinite dimensional spaces, see for example Exercise* **??***.)*

Proof. Let $\{f_i\}_{i=1}^n$ be a basis for L and define a new norm on L by

$$\left\| \sum_{i=1}^n a_i f_i \right\|_2 := \sqrt{\sum_{i=1}^n |a_i|^2} \text{ for } a_i \in \mathbb{F}.$$

By the triangle inequality for the norm $|\cdot|$, we find

$$\left| \sum_{i=1}^n a_i f_i \right| \le \sum_{i=1}^n |a_i| |f_i| \le \sqrt{\sum_{i=1}^n |f_i|^2} \sqrt{\sum_{i=1}^n |a_i|^2} \le M \left\| \sum_{i=1}^n a_i f_i \right\|_2$$

where $M = \sqrt{\sum_{i=1}^n |f_i|^2}$. Thus we have

$$|f| \le M \|f\|_2$$

for all $f \in L$ and this inequality shows that $|\cdot|$ is continuous relative to $\|\cdot\|_2$. Since the normed space $(L, \|\cdot\|_2)$ is homeomorphic and isomorphic to $\mathbb{F}^n$ with the standard euclidean norm, the closed bounded set, $S := \{f \in L : \|f\|_2 = 1\} \subset L$, is a compact subset of L relative to $\|\cdot\|_2$. Therefore by Exercise 45.13 there exists $f_0 \in S$ such that

$$m = \inf \{|f| : f \in S\} = |f_0| > 0.$$

Hence given $0 \ne f \in L$, then $\frac{f}{\|f\|_2} \in S$ so that

$$m \le \left| \frac{f}{\|f\|_2} \right| = |f| \frac{1}{\|f\|_2}$$

or equivalently

$$\|f\|_2 \le \frac{1}{m} |f|.$$

This shows that $|\cdot|$ and $\|\cdot\|_2$ are equivalent norms. Similarly one shows that $\|\cdot\|$ and $\|\cdot\|_2$ are equivalent and hence so are $|\cdot|$ and $\|\cdot\|$. ■

Corollary 45.32. *If $(L, \|\cdot\|)$ is a finite dimensional normed space, then $A \subset L$ is compact iff A is closed and bounded relative to the given norm, $\|\cdot\|$.*

Corollary 45.33. *Every finite dimensional normed vector space $(L, \|\cdot\|)$ is complete. In particular any finite dimensional subspace of a normed vector space is automatically closed.*

Proof. If $\{f_n\}_{n=1}^\infty \subset L$ is a Cauchy sequence, then $\{f_n\}_{n=1}^\infty$ is bounded and hence has a convergent subsequence, $g_k = f_{n_k}$, by Corollary 45.32. It is now routine to show $\lim_{n\to\infty} f_n = f := \lim_{k\to\infty} g_k$. ■

Theorem 45.34. *Suppose that $(X, \|\cdot\|)$ is a normed vector in which the unit ball, $V := B_0(1)$, is precompact. Then $\dim X < \infty$.*

Proof. Since $\bar{V}$ is compact, we may choose $\Lambda \subset\subset X$ such that

$$\bar{V} \subset \cup_{x \in \Lambda} \left(x + \frac{1}{2} V \right) \tag{45.7}$$

where, for any $\delta > 0$,

$$\delta V := \{\delta x : x \in V\} = B_0(\delta).$$

[2] Here is a proof if X is a metric space. Let $\{x_n\}_{n=1}^\infty \subset X$ be a sequence such that $f(x_n) \uparrow \sup f$. By compactness of X we may assume, by passing to a subsequence if necessary that $x_n \to b \in X$ as $n \to \infty$. By continuity of f, $f(b) = \sup f$.

Let $Y := \text{span}(\Lambda)$, then Eq. (45.7) implies,

$$V \subset \bar{V} \subset Y + \frac{1}{2}V.$$

Multiplying this equation by $\frac{1}{2}$ then shows

$$\frac{1}{2}V \subset \frac{1}{2}Y + \frac{1}{4}V = Y + \frac{1}{4}V$$

and hence

$$V \subset Y + \frac{1}{2}V \subset Y + Y + \frac{1}{4}V = Y + \frac{1}{4}V.$$

Continuing this way inductively then shows that

$$V \subset Y + \frac{1}{2^n}V \text{ for all } n \in \mathbb{N}. \tag{45.8}$$

Indeed, if Eq. (45.8) holds, then

$$V \subset Y + \frac{1}{2}V \subset Y + \frac{1}{2}\left(Y + \frac{1}{2^n}V\right) = Y + \frac{1}{2^{n+1}}V.$$

Hence if $x \in V$, there exists $y_n \in Y$ and $z_n \in B_0(2^{-n})$ such that $y_n + z_n \to x$. Since $\lim_{n\to\infty} z_n = 0$, it follows that $x = \lim_{n\to\infty} y_n \in \bar{Y}$. Since $\dim Y \leq \#(\Lambda) < \infty$, Corollary 45.33 implies $Y = \bar{Y}$ and so we have shown that $V \subset Y$. Since for any $x \in X$, $\frac{1}{2\|x\|}x \in V \subset Y$, we have $x \in Y$ for all $x \in X$, i.e. $X = Y$. ■

Exercise 45.15. Suppose $(Y, \|\cdot\|_Y)$ is a normed space and $(X, \|\cdot\|_X)$ is a finite dimensional normed space. Show **every** linear transformation $T : X \to Y$ is necessarily bounded.

45.4 Function Space Compactness Criteria

In this section, let (X, τ) be a topological space.

Definition 45.35. *Let $\mathcal{F} \subset C(X)$.*

1. $\mathcal{F}$ is **equicontinuous at** $x \in X$ iff for all $\varepsilon > 0$ there exists $U \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon$ for all $y \in U$ and $f \in \mathcal{F}$.
2. $\mathcal{F}$ is **equicontinuous** if $\mathcal{F}$ is equicontinuous at all points $x \in X$.
3. $\mathcal{F}$ is **pointwise bounded** if $\sup\{|f(x)| : f \in \mathcal{F}\} < \infty$ for all $x \in X$.

Theorem 45.36 (Ascoli-Arzela Theorem). *Let (X, τ) be a compact topological space and $\mathcal{F} \subset C(X)$. Then $\mathcal{F}$ is precompact in $C(X)$ iff $\mathcal{F}$ is equicontinuous and point-wise bounded.*

Proof. ($\Leftarrow$) Since $C(X) \subset \ell^\infty(X)$ is a complete metric space, we must show $\mathcal{F}$ is totally bounded. Let $\varepsilon > 0$ be given. By equicontinuity, for all $x \in X$, there exists $V_x \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon/2$ if $y \in V_x$ and $f \in \mathcal{F}$. Since X is compact we may choose $\Lambda \subset\subset X$ such that $X = \cup_{x\in\Lambda} V_x$. We have now decomposed X into "blocks" $\{V_x\}_{x\in\Lambda}$ such that each $f \in \mathcal{F}$ is constant to within ε on V_x. Since $\sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} < \infty$, it is now evident that

$$\begin{aligned} M &= \sup\{|f(x)| : x \in X \text{ and } f \in \mathcal{F}\} \\ &\leq \sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} + \varepsilon < \infty. \end{aligned}$$

Let $\mathbb{D} := \{k\varepsilon/2 : k \in \mathbb{Z}\} \cap [-M, M]$. If $f \in \mathcal{F}$ and $\varphi \in \mathbb{D}^\Lambda$ (i.e. $\varphi : \Lambda \to \mathbb{D}$ is a function) is chosen so that $|\varphi(x) - f(x)| \leq \varepsilon/2$ for all $x \in \Lambda$, then

$$|f(y) - \varphi(x)| \leq |f(y) - f(x)| + |f(x) - \varphi(x)| < \varepsilon \;\forall\; x \in \Lambda \text{ and } y \in V_x.$$

From this it follows that $\mathcal{F} = \bigcup\left\{\mathcal{F}_\varphi : \varphi \in \mathbb{D}^\Lambda\right\}$ where, for $\varphi \in \mathbb{D}^\Lambda$,

$$\mathcal{F}_\varphi := \{f \in \mathcal{F} : |f(y) - \varphi(x)| < \varepsilon \text{ for } y \in V_x \text{ and } x \in \Lambda\}.$$

Let $\Gamma := \left\{\varphi \in \mathbb{D}^\Lambda : \mathcal{F}_\varphi \neq \emptyset\right\}$ and for each $\varphi \in \Gamma$ choose $f_\varphi \in \mathcal{F}_\varphi \cap \mathcal{F}$. For $f \in \mathcal{F}_\varphi$, $x \in \Lambda$ and $y \in V_x$ we have

$$|f(y) - f_\varphi(y)| \leq |f(y) - \varphi(x))| + |\varphi(x) - f_\varphi(y)| < 2\varepsilon.$$

So $\|f - f_\varphi\|_\infty < 2\varepsilon$ for all $f \in \mathcal{F}_\varphi$ showing that $\mathcal{F}_\varphi \subset B_{f_\varphi}(2\varepsilon)$. Therefore,

$$\mathcal{F} = \cup_{\varphi\in\Gamma}\mathcal{F}_\varphi \subset \cup_{\varphi\in\Gamma} B_{f_\varphi}(2\varepsilon)$$

and because $\varepsilon > 0$ was arbitrary we have shown that $\mathcal{F}$ is totally bounded.

($\Rightarrow$) (*The rest of this proof may safely be skipped.) Since $\|\cdot\|_\infty : C(X) \to [0, \infty)$ is a continuous function on $C(X)$ it is bounded on any compact subset $\mathcal{F} \subset C(X)$. This shows that $\sup\{\|f\|_\infty : f \in \mathcal{F}\} < \infty$ which clearly implies that $\mathcal{F}$ is pointwise bounded.[3] Suppose $\mathcal{F}$ were **not** equicontinuous at some point $x \in X$ that is to say there exists $\varepsilon > 0$ such that for all $V \in \tau_x$, $\sup_{y\in V}\sup_{f\in\mathcal{F}} |f(y) - f(x)| > \varepsilon$.[4] Equivalently said, to each $V \in \tau_x$ we may choose

$$f_V \in \mathcal{F} \text{ and } x_V \in V \;\ni\; |f_V(x) - f_V(x_V)| \geq \varepsilon. \tag{45.9}$$

[3] One could also prove that $\mathcal{F}$ is pointwise bounded by considering the continuous evaluation maps $e_x : C(X) \to \mathbb{R}$ given by $e_x(f) = f(x)$ for all $x \in X$.

[4] If X is first countable we could finish the proof with the following argument. Let $\{V_n\}_{n=1}^\infty$ be a neighborhood base at x such that $V_1 \supset V_2 \supset V_3 \supset \dots$. By the assumption that $\mathcal{F}$ is not equicontinuous at x, there exist $f_n \in \mathcal{F}$ and $x_n \in V_n$ such that $|f_n(x) - f_n(x_n)| \geq \epsilon \;\forall\; n$. Since $\mathcal{F}$ is a compact metric space by passing to a subsequence if necessary we may assume that f_n converges uniformly to some

Set $\mathcal{C}_V = \overline{\{f_W : W \in \tau_x \text{ and } W \subset V\}}^{\|\cdot\|_\infty} \subset \mathcal{F}$ and notice for any $\mathcal{V} \subset\subset \tau_x$ that

$$\cap_{V\in\mathcal{V}}\mathcal{C}_V \supseteq \mathcal{C}_{\cap\mathcal{V}} \neq \emptyset,$$

so that $\{\mathcal{C}_V\}_V \in \tau_x \subset \mathcal{F}$ has the finite intersection property.[5] Since $\mathcal{F}$ is compact, it follows that there exists some

$$f \in \bigcap_{V\in\tau_x} \mathcal{C}_V \neq \emptyset.$$

Since f is continuous, there exists $V \in \tau_x$ such that $|f(x) - f(y)| < \varepsilon/3$ for all $y \in V$. Because $f \in \mathcal{C}_V$, there exists $W \subset V$ such that $\|f - f_W\| < \varepsilon/3$. We now arrive at a contradiction;

$$\begin{aligned}\varepsilon &\le |f_W(x) - f_W(x_W)| \\ &\le |f_W(x) - f(x)| + |f(x) - f(x_W)| + |f(x_W) - f_W(x_W)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\end{aligned}$$

Alternate proof. For $\varepsilon > 0$ let $\Lambda_\varepsilon \subset_f X$ and $\{V_x^\varepsilon\}_{x\in\Lambda_\varepsilon}$ be a finite open cover of X with the property; for all $x \in X$ we have

$$|f(y) - f(x)| < \varepsilon \;\forall\; y \in V_x \text{ and } f \in \mathcal{F}.$$

Let $D := \cup_{m=1}^\infty \Lambda_{1/m}$ – countable set and suppose that $\{f_n\} \subset \mathcal{F}$ is a given sequence. Since $\{f_n(x)\}_{n=1}^\infty$ is bounded in $\mathbb{R}$ for all $x \in D$, by Cantor's diagonalization argument, we may choose a subsequence, $g_k := f_{n_k}$ such that $g_0(x) := \lim_{k\to\infty} g_k(x)$ exists for all $x \in D$. To finish the proof we need only show $\{g_k\}$ is uniformly Cauchy. To this end, observe that for $y \in X$ and $m \in \mathbb{N}$ we may choose an $x \in \Lambda_{1/m}$ such that $y \in V_x^{1/m}$ and therefore,

$$\begin{aligned}|g_k(y) - g_l(y)| &\le |g_k(y) - g_k(x)| + |g_k(x) - g_l(x)| + |g_l(x) - g_l(y)| \\ &\le 2/m + |g_k(x) - g_l(x)|\end{aligned}$$

$f \in \mathcal{F}$. Because $x_n \to x$ as $n \to \infty$ we learn that

$$\begin{aligned}\epsilon &\le |f_n(x) - f_n(x_n)| \le |f_n(x) - f(x)| + |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \\ &\le 2\|f_n - f\| + |f(x) - f(x_n)| \to 0 \text{ as } n \to \infty\end{aligned}$$

which is a contradiction.

[5] If we are willing to use Net's described in Appendix **??** below we could finish the proof as follows. Since $\mathcal{F}$ is compact, the net $\{f_V\}_{V\in\tau_x} \subset \mathcal{F}$ has a cluster point $f \in \mathcal{F} \subset C(X)$. Choose a subnet $\{g_\alpha\}_{\alpha\in A}$ of $\{f_V\}_{V\in\tau_X}$ such that $g_\alpha \to f$ uniformly. Then, since $x_V \to x$ implies $x_{V_\alpha} \to x$, we may conclude from Eq. (45.9) that

$$\epsilon \le |g_\alpha(x) - g_\alpha(x_{V_\alpha})| \to |g(x) - g(x)| = 0$$

which is a contradiction.

and therefore,

$$\|g_k - g_l\|_u \le 2/m + \max_{x\in\Lambda_{1/m}} |g_k(x) - g_l(x)|.$$

Passing to the limit as $k, l \to \infty$ then shows

$$\limsup_{k,l\to\infty} \|g_k - g_l\|_u \le 2/m \to 0 \text{ as } m \to \infty.$$

■

Remark 45.37. The above theorem may be easily generalized to cover $C(X, S)$ where (S, ρ) is a complete metric space with the Heine–Borel property, namely closed and bounded sets are compact. For example if $\mathcal{F} \subset C(X, S)$ is pointwise bounded and equicontinuous then it is sequentially compact. This is proved by first showing $C(X, S)$ is a complete metric space relative to the metric,

$$\rho_u(f, g) := \sup_{x\in X} \rho(f(x), g(x)).$$

Now follow the alternative proof in the previous theorem replacing $\mathbb{R}$ by S, absolute values by the metric ρ, and $\|\cdot\|_u$ by ρ_u everywhere. For example the proof should start as; for $\varepsilon > 0$ let $\Lambda_\varepsilon \subset_f X$ and $\{V_x^\varepsilon\}_{x\in\Lambda_\varepsilon}$ be a finite open cover of X with the property; for all $x \in X$ we have

$$\rho(f(y), f(x)) < \varepsilon \;\forall\; y \in V_x \text{ and } f \in \mathcal{F}.$$

Exercise 45.16. Give an alternative proof of the implication, $(\Leftarrow)$, in Theorem 45.36 by showing every subsequence $\{f_n : n \in \mathbb{N}\} \subset \mathcal{F}$ has a convergence subsequence.

Exercise 45.17. Suppose $k \in C\left([0,1]^2, \mathbb{R}\right)$ and for $f \in C([0,1], \mathbb{R})$, let

$$Kf(x) := \int_0^1 k(x,y) f(y)\, dy \text{ for all } x \in [0,1].$$

Show K is a compact operator on $(C([0,1], \mathbb{R}), \|\cdot\|_\infty)$.

The following result is a corollary of Lemma **??** and Theorem 45.36.

Corollary 45.38 (Locally Compact Ascoli-Arzela Theorem). *Let (X, τ) be a locally compact and σ – compact topological space and $\{f_m\} \subset C(X)$ be a pointwise bounded sequence of functions such that $\{f_m|_K\}$ is equicontinuous for any compact subset $K \subset X$. Then there exists a subsequence $\{m_n\} \subset \{m\}$ such that $\{g_n := f_{m_n}\}_{n=1}^\infty \subset C(X)$ is a sequence which is uniformly convergent on compact subsets of X.*

Proof. Let $\{K_n\}_{n=1}^\infty$ be the compact subsets of X constructed in Lemma **??**. We may now apply Theorem 45.36 repeatedly to find a nested family of subsequences

$$\{f_m\} \supset \{g_m^1\} \supset \{g_m^2\} \supset \{g_m^3\} \supset \dots$$

such that the sequence $\{g_m^n\}_{m=1}^\infty \subset C(X)$ is uniformly convergent on K_n. Using Cantor's trick, define the subsequence $\{h_n\}$ of $\{f_m\}$ by $h_n := g_n^n$. Then $\{h_n\}$ is uniformly convergent on K_l for each $l \in \mathbb{N}$. Now if $K \subset X$ is an arbitrary compact set, there exists $l < \infty$ such that $K \subset K_l^o \subset K_l$ and therefore $\{h_n\}$ is uniformly convergent on K as well. ∎

Proposition 45.39. *Let $\Omega \subset_o \mathbb{R}^d$ such that $\bar{\Omega}$ is compact and $0 \le \alpha < \beta \le 1$. Then the inclusion map $i : C^\beta(\overline{\Omega}) \hookrightarrow C^\alpha(\overline{\Omega})$ is a compact operator. See Chapter* **??** *and Lemma* **??** *for the notation being used here.*

Let $\{u_n\}_{n=1}^\infty \subset C^\beta(\overline{\Omega})$ such that $\|u_n\|_{C^\beta} \le 1$, i.e. $\|u_n\|_\infty \le 1$ and

$$|u_n(x) - u_n(y)| \le |x-y|^\beta \text{ for all } x, y \in \overline{\Omega}.$$

By the Arzela-Ascoli Theorem 45.36, there exists a subsequence of $\{\tilde{u}_n\}_{n=1}^\infty$ of $\{u_n\}_{n=1}^\infty$ and $u \in C^o(\bar{\Omega})$ such that $\tilde{u}_n \to u$ in C^0. Since

$$|u(x) - u(y)| = \lim_{n\to\infty} |\tilde{u}_n(x) - \tilde{u}_n(y)| \le |x-y|^\beta,$$

$u \in C^\beta$ as well. Define $g_n := u - \tilde{u}_n \in C^\beta$, then

$$[g_n]_\beta + \|g_n\|_{C^0} = \|g_n\|_{C^\beta} \le 2$$

and $g_n \to 0$ in C^0. To finish the proof we must show that $g_n \to 0$ in C^α. Given $\delta > 0$,

$$[g_n]_\alpha = \sup_{x \ne y} \frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} \le A_n + B_n$$

where

$$\begin{aligned} A_n &= \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} : x \ne y \text{ and } |x-y| \le \delta\right\} \\ &= \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\beta} \cdot |x-y|^{\beta-\alpha} : x \ne y \text{ and } |x-y| \le \delta\right\} \\ &\le \delta^{\beta-\alpha} \cdot [g_n]_\beta \le 2\delta^{\beta-\alpha} \end{aligned}$$

and

$$B_n = \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} : |x-y| > \delta\right\} \le 2\delta^{-\alpha} \|g_n\|_{C^0} \to 0 \text{ as } n \to \infty.$$

Therefore,

$$\limsup_{n\to\infty} [g_n]_\alpha \le \limsup_{n\to\infty} A_n + \limsup_{n\to\infty} B_n \le 2\delta^{\beta-\alpha} + 0 \to 0 \text{ as } \delta \downarrow 0.$$

This proposition generalizes to the following theorem which the reader is asked to prove in Exercise **??** below.

Theorem 45.40. *Let Ω be a precompact open subset of $\mathbb{R}^d$, $\alpha, \beta \in [0,1]$ and $k, j \in \mathbb{N}_0$. If $j + \beta > k + \alpha$, then $C^{j,\beta}(\bar{\Omega})$ is compactly contained in $C^{k,\alpha}(\bar{\Omega})$.*

45.5 Supplementary Remarks

45.5.1 Word of Caution

Example 45.41. Let (X, d) be a metric space. It is always true that $\overline{B_x(\varepsilon)} \subset C_x(\varepsilon)$ since $C_x(\varepsilon)$ is a closed set containing $B_x(\varepsilon)$. However, it is not always true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$. For example let $X = \{1, 2\}$ and $d(1,2) = 1$, then $B_1(1) = \{1\}$, $\overline{B_1(1)} = \{1\}$ while $C_1(1) = X$. For another counterexample, take

$$X = \left\{(x, y) \in \mathbb{R}^2 : x = 0 \text{ or } x = 1\right\}$$

with the usually Euclidean metric coming from the plane. Then

$$\begin{aligned} B_{(0,0)}(1) &= \left\{(0, y) \in \mathbb{R}^2 : |y| < 1\right\}, \\ \overline{B_{(0,0)}(1)} &= \left\{(0, y) \in \mathbb{R}^2 : |y| \le 1\right\}, \text{ while} \\ C_{(0,0)}(1) &= \overline{B_{(0,0)}(1)} \cup \{(0,1)\}. \end{aligned}$$

In spite of the above examples, Lemmas 45.42 and 45.43 below shows that for certain metric spaces of interest it is true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$.

Lemma 45.42. *Suppose that $(X, |\cdot|)$ is a normed vector space and d is the metric on X defined by $d(x, y) = |x - y|$. Then*

$$\overline{B_x(\varepsilon)} = C_x(\varepsilon) \text{ and}$$
$$\mathrm{bd}(B_x(\varepsilon)) = \{y \in X : d(x, y) = \varepsilon\}.$$

where the boundary operation, $\mathrm{bd}(\cdot)$ *is defined in Definition* **??** *(BRUCE: Forward Reference.) below.*

Proof. We must show that $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. For $y \in C$, let $v = y - x$, then

$$|v| = |y - x| = d(x, y) \le \varepsilon.$$

Let $\alpha_n = 1 - 1/n$ so that $\alpha_n \uparrow 1$ as $n \to \infty$. Let $y_n = x + \alpha_n v$, then $d(x, y_n) = \alpha_n d(x, y) < \varepsilon$, so that $y_n \in B_x(\varepsilon)$ and $d(y, y_n) = (1 - \alpha_n)|v| \to 0$ as $n \to \infty$. This shows that $y_n \to y$ as $n \to \infty$ and hence that $y \in \bar{B}$. ∎

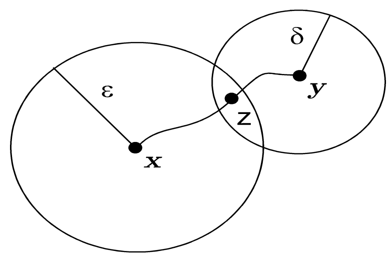


Fig. 45.3. An almost length minimizing curve joining x to y.

45.5.2 Riemannian Metrics

This subsection is not completely self contained and may safely be skipped.

Lemma 45.43. *Suppose that X is a Riemannian (or sub-Riemannian) manifold and d is the metric on X defined by*

$$d(x,y) = \inf\{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}$$

where $\ell(\sigma)$ is the length of the curve σ. We define $\ell(\sigma) = \infty$ if σ is not piecewise smooth.

Then

$$\overline{B_x(\varepsilon)} = C_x(\varepsilon) \text{ and}$$
$$\mathrm{bd}(B_x(\varepsilon)) = \{y \in X : d(x,y) = \varepsilon\}$$

where the boundary operation, $\mathrm{bd}(\cdot)$ *is defined in Definition* **??** *below.*

Proof. Let $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. We will show that $C \subset \bar{B}$ by showing $\bar{B}^c \subset C^c$. Suppose that $y \in \bar{B}^c$ and choose $\delta > 0$ such that $B_y(\delta) \cap \bar{B} = \emptyset$. In particular this implies that

$$B_y(\delta) \cap B_x(\varepsilon) = \emptyset.$$

We will finish the proof by showing that $d(x,y) \geq \varepsilon + \delta > \varepsilon$ and hence that $y \in C^c$. This will be accomplished by showing: if $d(x,y) < \varepsilon + \delta$ then $B_y(\delta) \cap B_x(\varepsilon) \neq \emptyset$. If $d(x,y) < \max(\varepsilon,\delta)$ then either $x \in B_y(\delta)$ or $y \in B_x(\varepsilon)$. In either case $B_y(\delta) \cap B_x(\varepsilon) \neq \emptyset$. Hence we may assume that $\max(\varepsilon,\delta) \leq d(x,y) < \varepsilon + \delta$. Let $\alpha > 0$ be a number such that

$$\max(\varepsilon,\delta) \leq d(x,y) < \alpha < \varepsilon + \delta$$

and choose a curve σ from x to y such that $\ell(\sigma) < \alpha$. Also choose $0 < \delta' < \delta$ such that $0 < \alpha - \delta' < \varepsilon$ which can be done since $\alpha - \delta < \varepsilon$. Let $k(t) = d(y, \sigma(t))$ a

continuous function on $[0,1]$ and therefore $k([0,1]) \subset \mathbb{R}$ is a connected set which contains 0 and $d(x,y)$. Therefore there exists $t_0 \in [0,1]$ such that $d(y, \sigma(t_0)) = k(t_0) = \delta'$. Let $z = \sigma(t_0) \in B_y(\delta)$ then

$$d(x,z) \leq \ell(\sigma|_{[0,t_0]}) = \ell(\sigma) - \ell(\sigma|_{[t_0,1]}) < \alpha - d(z,y) = \alpha - \delta' < \varepsilon$$

and therefore $z \in B_x(\varepsilon) \cap B_x(\delta) \neq \emptyset$. ∎

Remark 45.44. Suppose again that X is a Riemannian (or sub-Riemannian) manifold and

$$d(x,y) = \inf\{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}.$$

Let σ be a curve from x to y and let $\varepsilon = \ell(\sigma) - d(x,y)$. Then for all $0 \leq u < v \leq 1$,

$$\begin{aligned} d(x,y) + \varepsilon = \ell(\sigma) &= \ell(\sigma|_{[0,u]}) + \ell(\sigma|_{[u,v]}) + \ell(\sigma|_{[v,1]}) \\ &\geq d(x,\sigma(u)) + \ell(\sigma|_{[u,v]}) + d(\sigma(v),y) \end{aligned}$$

and therefore, using the triangle inequality,

$$\begin{aligned} \ell(\sigma|_{[u,v]}) &\leq d(x,y) + \varepsilon - d(x,\sigma(u)) - d(\sigma(v),y) \\ &\leq d(\sigma(u),\sigma(v)) + \varepsilon. \end{aligned}$$

This leads to the following conclusions. If σ is within ε of a length minimizing curve from x to y then $\sigma|_{[u,v]}$ is within ε of a length minimizing curve from $\sigma(u)$ to $\sigma(v)$. In particular if σ is a length minimizing curve from x to y then $\sigma|_{[u,v]}$ is a length minimizing curve from $\sigma(u)$ to $\sigma(v)$.

45.6 Exercises

Exercise 45.18. Let (X,d) be a metric space. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ is a sequence and set $\varepsilon_n := d(x_n, x_{n+1})$. Show that for $m > n$ that

$$d(x_n, x_m) \leq \sum_{k=n}^{m-1} \varepsilon_k \leq \sum_{k=n}^{\infty} \varepsilon_k.$$

Conclude from this that if

$$\sum_{k=1}^\infty \varepsilon_k = \sum_{n=1}^\infty d(x_n, x_{n+1}) < \infty$$

then $\{x_n\}_{n=1}^\infty$ is Cauchy. Moreover, show that if $\{x_n\}_{n=1}^\infty$ is a convergent sequence and $x = \lim_{n\to\infty} x_n$ then

$$d(x, x_n) \leq \sum_{k=n}^\infty \varepsilon_k.$$

Exercise 45.19. Show that (X, d) is a complete metric space iff every sequence $\{x_n\}_{n=1}^\infty \subset X$ such that $\sum_{n=1}^\infty d(x_n, x_{n+1}) < \infty$ is a convergent sequence in X. You may find it useful to prove the following statements in the course of the proof.

1. If $\{x_n\}$ is Cauchy sequence, then there is a subsequence $y_j := x_{n_j}$ such that $\sum_{j=1}^\infty d(y_{j+1}, y_j) < \infty$.
2. If $\{x_n\}_{n=1}^\infty$ is Cauchy and there exists a subsequence $y_j := x_{n_j}$ of $\{x_n\}$ such that $x = \lim_{j\to\infty} y_j$ exists, then $\lim_{n\to\infty} x_n$ also exists and is equal to x.

Exercise 45.20. Suppose that $f : [0, \infty) \to [0, \infty)$ is a C^2 – function such that $f(0) = 0$, $f' > 0$ and $f'' \le 0$ and (X, ρ) is a metric space. Show that $d(x, y) = f(\rho(x, y))$ is a metric on X. In particular show that

$$d(x, y) := \frac{\rho(x, y)}{1 + \rho(x, y)}$$

is a metric on X. (Hint: use calculus to verify that $f(a + b) \le f(a) + f(b)$ for all $a, b \in [0, \infty)$.)

Exercise 45.21. Let $\{(X_n, d_n)\}_{n=1}^\infty$ be a sequence of metric spaces, $X := \prod_{n=1}^\infty X_n$, and for $x = (x(n))_{n=1}^\infty$ and $y = (y(n))_{n=1}^\infty$ in X let

$$d(x, y) = \sum_{n=1}^\infty 2^{-n} \frac{d_n(x(n), y(n))}{1 + d_n(x(n), y(n))}.$$

Show:

1. (X, d) is a metric space,
2. a sequence $\{x_k\}_{k=1}^\infty \subset X$ converges to $x \in X$ iff $x_k(n) \to x(n) \in X_n$ as $k \to \infty$ for each $n \in \mathbb{N}$ and
3. X is complete if X_n is complete for all n.

Exercise 45.22. Suppose (X, ρ) and (Y, d) are metric spaces and A is a dense subset of X.

1. Show that if $F : X \to Y$ and $G : X \to Y$ are two continuous functions such that $F = G$ on A then $F = G$ on X. **Hint:** consider the set $C := \{x \in X : F(x) = G(x)\}$.
2. Suppose $f : A \to Y$ is a function which is uniformly continuous, i.e. for every $\varepsilon > 0$ there exists a $\delta > 0$ such that
$$d(f(a), f(b)) < \varepsilon \text{ for all } a, b \in A \text{ with } \rho(a, b) < \delta.$$
Show there is a unique continuous function $F : X \to Y$ such that $F = f$ on A. **Hint:** each point $x \in X$ is a limit of a sequence consisting of elements from A.
3. Let $X = \mathbb{R} = Y$ and $A = \mathbb{Q} \subset X$, find a function $f : \mathbb{Q} \to \mathbb{R}$ which is continuous on $\mathbb{Q}$ but does **not** extend to a continuous function on $\mathbb{R}$.

References

1. David Applebaum, *Lévy processes and stochastic calculus*, second ed., Cambridge Studies in Advanced Mathematics, vol. 116, Cambridge University Press, Cambridge, 2009. MR MR2512800
2. Richard F. Bass, *Probabilistic techniques in analysis*, Probability and its Applications (New York), Springer-Verlag, New York, 1995. MR MR1329542 (96e:60001)
3. ———, *Diffusions and elliptic operators*, Probability and its Applications (New York), Springer-Verlag, New York, 1998. MR MR1483890 (99h:60136)
4. Patrick Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication. MR MR1324786 (95k:60001)
5. ———, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication. MR MR1700749 (2000e:60008)
6. Ottó J. Björnsson, *A note on the characterization of standard Borel spaces*, Math. Scand. **47** (1980), no. 1, 135–136. MR MR600083 (82a:54070)
7. R. M. Blumenthal and R. K. Getoor, *Markov processes and potential theory*, Pure and Applied Mathematics, Vol. 29, Academic Press, New York, 1968. MR MR0264757 (41 #9348)
8. Vladimir I. Bogachev, *Gaussian measures*, Mathematical Surveys and Monographs, vol. 62, American Mathematical Society, Providence, RI, 1998. MR 2000a:60004
9. Leo Breiman, *Probability*, Addison-Wesley Publishing Company, Reading, Mass., 1968. MR MR0229267 (37 #4841)
10. Giuseppe Da Prato and Jerzy Zabczyk, *Stochastic equations in infinite dimensions*, Cambridge University Press, Cambridge, 1992. MR 95g:60073
11. Claude Dellacherie, *Capacités et processus stochastiques*, Springer-Verlag, Berlin, 1972, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 67. MR MR0448504 (56 #6810)
12. J. L. Doob, *Stochastic processes*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1990, Reprint of the 1953 original, A Wiley-Interscience Publication. MR MR1038526 (91d:60002)
13. Nelson Dunford and Jacob T. Schwartz, *Linear operators. Part I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1988, General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. MR MR1009162 (90g:47001a)
14. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001)
15. ———, *Stochastic calculus*, Probability and Stochastics Series, CRC Press, Boca Raton, FL, 1996, A practical introduction. MR MR1398879 (97k:60148)
16. Evgenii B. Dynkin and Aleksandr A. Yushkevich, *Markov processes: Theorems and problems*, Translated from the Russian by James S. Wood, Plenum Press, New York, 1969. MR MR0242252 (39 #3585a)
17. Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986, Characterization and convergence. MR MR838085 (88a:60130)
18. William Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971. MR MR0270403 (42 #5292)

19. Peter Friz and Nicolas Victoir, *Multidimensional stochastic processes as rough paths theory and applications*, first ed., Cambridge Studies in Advanced Mathematics, vol. 120, Cambridge University Press, Cambridge, 2010.
20. Masatoshi Fukushima, Yōichi Ōshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, de Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 1994. MR MR1303354 (96f:60126)
21. Robert D. Gordon, *Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument*, Ann. Math. Statistics **12** (1941), 364–366. MR MR0005558 (3,171e)
22. Leonard Gross, *Integration and nonlinear transformations in Hilbert space*, Trans. Amer. Math. Soc. **94** (1960), 404–440. MR 22 #2883
23. ———, *Abstract Wiener spaces*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1, Univ. California Press, Berkeley, Calif., 1967, pp. 31–42. MR 35 #3027
24. Paul R. Halmos, *Lectures on ergodic theory*, (1960), vii+101. MR MR0111817 (22 #2677)
25. Nobuyuki Ikeda and Shinzo Watanabe, *Stochastic differential equations and diffusion processes*, second ed., North-Holland Mathematical Library, vol. 24, North-Holland Publishing Co., Amsterdam, 1989. MR MR1011252 (90m:60069)
26. Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 1987. MR MR959133 (89k:60044)
27. ———, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003. MR MR1943877 (2003j:60001)
28. Svante Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997. MR MR1474726 (99f:60082)
29. Shizuo Kakutani, *On equivalence of infinite product measures*, Ann. of Math. (2) **49** (1948), 214–224. MR MR0023331 (9,340e)
30. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
31. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR MR1121940 (92h:60127)
32. Oliver Knill, *Probability and stochastic processes with applications*, Havard Web-Based, 1994.
33. Leonid B. Koralov and Yakov G. Sinai, *Theory of probability and random processes*, second ed., Universitext, Springer, Berlin, 2007. MR MR2343262 (2008k:60002)
34. Hui Hsiung Kuo, *Gaussian measures in Banach spaces*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 463. MR 57 #1628
35. Zhi Ming Ma and Michael Röckner, *Introduction to the theory of (nonsymmetric) Dirichlet forms*, Universitext, Springer-Verlag, Berlin, 1992. MR MR1214375 (94d:60119)
36. Edward Nelson, *An existence theorem for second order parabolic equations*, Trans. Amer. Math. Soc. **88** (1958), 414–429. MR MR0095341 (20 #1844)
37. ———, *Feynman integrals and the Schrödinger equation*, J. Mathematical Phys. **5** (1964), 332–343. MR MR0161189 (28 #4397)
38. J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR MR1600720 (99c:60144)
39. ———, *Probability and measure*, Tech. report, Mathematics Department, University of Cambridge, 2009.
40. K. R. Parthasarathy, *Probability measures on metric spaces*, AMS Chelsea Publishing, Providence, RI, 2005, Reprint of the 1967 original. MR MR2169627 (2006d:60004)
41. Yuval Peres, *An invitation to sample paths of brownian motion*, stat-www.berkeley.edu/ peres/bmall.pdf (2001), 1–68.
42. Philip Protter, *Stochastic integration and differential equations*, Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 1990, A new approach. MR MR1037262 (91i:60148)
43. Philip E. Protter, *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer-Verlag, Berlin, 2005, Second edition. Version 2.1, Corrected third printing. MR MR2273672
44. L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. MR 2001g:60188
45. Sheldon M. Ross, *Stochastic processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1983, Lectures in Mathematics, 14. MR MR683455 (84m:60001)
46. H. L. Royden, *Real analysis*, third ed., Macmillan Publishing Company, New York, 1988. MR MR1013117 (90g:00004)
47. Gennady Samorodnitsky and Murad S. Taqqu, *Stable non-Gaussian random processes*, Stochastic Modeling, Chapman & Hall, New York, 1994, Stochastic models with infinite variance. MR MR1280932 (95f:60024)
48. Michael Sharpe, *General theory of Markov processes*, Pure and Applied Mathematics, vol. 133, Academic Press Inc., Boston, MA, 1988. MR MR958914 (89m:60169)
49. Barry Simon, *Functional integration and quantum physics*, second ed., AMS Chelsea Publishing, Providence, RI, 2005. MR MR2105995 (2005f:81003)
50. Daniel W. Stroock, *Probability theory, an analytic view*, Cambridge University Press, Cambridge, 1993. MR MR1267569 (95f:60003)
51. S. R. S. Varadhan, *Probability theory*, Courant Lecture Notes in Mathematics, vol. 7, New York University Courant Institute of Mathematical Sciences, New York, 2001. MR MR1852999 (2003a:60001)
52. Norbert Wiener, *Differential space*, J. of Math. and Phys. **2** (1923), 131–174.
53. ———, *The Homogeneous Chaos*, Amer. J. Math. **60** (1938), no. 4, 897–936. MR MR1507356