Lecture Notes On

An Analytical Introduction to Probability Theory

Wee Peng Tay
School of Electrical and Electronic Engineering
Nanyang Technological University

An Analytical Introduction to Probability Theory

1. Basics of Real Analysis - I

SIGNAL, NTU `https://personal.ntu.edu.sg/wptay/`

1.1 Introduction

This basic real analysis introduction covers topics in sequences, compact spaces and continuity. Recommended textbook is "Principles of Mathematical Analysis" by Walter Rudin.

1.2 Sequences

Inner product space $\subset$ normed space $\subset$ metric space $\subset$ topological space. Working with metric spaces suffices for our purposes.

Definition 1.1. *A metric space is an ordered pair $(\mathcal{X}, d)$ where $\mathcal{X}$ is a set and d is a metric on $\mathcal{X}$, such that $\forall\, x, y, z \in \mathcal{X}$, the following holds:*

1) $d(x, y) \geq 0$, where equality holds iff $x = y$.

2) $d(x, y) = d(y, x)$.

3) Triangle inequality: $d(x, y) \leq d(x, z) + d(y, z)$.

Example 1.1. *Examples of metric space:*

- $(\mathbb{R}, |x - y|)$ *(normed space)*
- $\left(\mathbb{R}^k, \|x - y\|_p\right)$, *where $p \geq 1$ (normed space)*
- $\ell^p(\mathbb{R}) = \left\{ x = (x_n)_{n\geq 1} \,:\, \|x\|_p = \left(\sum_{n\geq 1} |x_n|^p\right)^{1/p} < \infty \right\}$ *(normed space)*
- $(\mathcal{X}, d)$, *where $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ otherwise.*

We write (x_n) or $(x_n)_{n\geq 1}$ as shorthand for the sequence $x_1, x_2, \ldots$.

Definition 1.2. *For $(x_n) \subset \mathcal{X}$, we say that $x_n \to x$ as $n \to \infty$ or $\lim_{n\to\infty} x_n = x$ if*

$$\forall\, \epsilon > 0,\ \exists\, N \ \text{s.t.}\ d(x_n, x) < \epsilon,\ \forall\, n \geq N.$$

Definition 1.3. *We say x_n diverges to ∞ if there exists $x_0 \in \mathcal{X}$ and*

$$\forall\, K \geq 0,\ \exists\, N \ \text{s.t.}\ d(x_n, x_0) \geq K,\ \forall\, n \geq N.$$

We define extended real numbers $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. For $x \in \mathbb{R}$, we have:

1) $x + \infty = \infty$
2) $x - \infty = -\infty$
3) $x \cdot \pm\infty = \pm\infty, \forall\, x > 0$
4) $0 \cdot \pm\infty = 0$ (by convention)

The following lemmas are based on $(\mathbb{R}, |\cdot|)$ but most can be generalized to $(\mathcal{X}, d)$.

Lemma 1.1. *If (a_n) converges, then its limit is unique.*

Proof. Supposing $a_n \to a$ and $a_n \to a'$, we have

$$\forall\, \epsilon > 0,\ \exists\, N \ \text{ s.t. } |a_n - a| < \epsilon \text{ and } |a_n - a'| < \epsilon,\ \forall\, n \geq N.$$

By the triangle inequality, we obtain:

$$|a - a'| \leq |a_n - a| + |a_n - a'| < 2\epsilon.$$

Since ϵ is arbitrary, $a = a'$. □

Lemma 1.2. *If (a_n) converges, $|a_n| < \infty$.*

Proof. Supposing $a_n \to a$, $\exists\, N$, s.t. $|a_n - a| < 1$, $\forall\, n \geq N$. Then we obtain

$$|a_n| \leq |a| + 1, \forall\ n \geq N,$$

and

$$|a_n| \leq \max\{a_1, \ldots, a_{N-1}, |a| + 1\} < \infty.$$

□

Lemma 1.3. *If (a_n) is increasing (i.e., $a_{n+1} \geq a_n$ for all $n \geq 1$) and is bounded above, then (a_n) converges.*

Proof. To prove the lemma, we need some definitions. We define $\sup A$ = least upper bound (LUB) of A if $A \neq \emptyset$ and $A \subset \mathbb{R}$, i.e.,

1) $\forall\, a \in A$, $a \leq \sup A$.
2) if $\gamma < \sup A$, $\exists\, a \in A$, s. t. $\gamma < a$.

For the definition to be well-defined, we need the LUB/Completeness axiom: $\sup A$ exists in $\mathbb{R}$ for all $A \neq \emptyset$ bounded above. (Note that this axiom does not hold for $\mathbb{Q}$.) If A is not bounded above, we set $\sup A = \infty$. From Item 2), we obtain a useful property:

$$\forall\, \epsilon > 0,\ \exists\, a \in A,\ \text{s. t. } \sup A - \epsilon < a.$$

The greatest lower bound of A is written as $\inf A$.

Returning to the proof of the lemma, let $a = \sup a_n < \infty$ since (a_n) is bounded. We have

$$\begin{aligned} &\forall\, \epsilon > 0,\ \exists\, N,\ \text{s. t. } a - \epsilon \leq a_N \leq a_n \leq a,\ \forall\, n \geq N. \\ &\implies a_n \to a. \end{aligned}$$

□

Applying Lemma 1.3 to the negative of a sequence, we also conclude that any decreasing sequence that is bounded below converges.

Remark 1.1.

1. *If* $a_n \in \overline{\mathbb{R}}_+$ *and is increasing, then* $\lim_{n\to\infty} a_n$ *exists (maybe* ∞*).*
2. *If* $a_n \in \mathbb{R}_+$*, then from Lemma 1.3,* $\sum_{k=1}^{\infty} a_k \triangleq \lim_{n\to\infty} \sum_{k=1}^{n} a_k$ *exists.*

Definition 1.4. *A subsequence of* (a_n) *is a sequence* (a_{n_i})*, where* $n_1 < n_2 < \dots$ *is an increasing sequence of indices.*

Lemma 1.4. *If* $a_n \to a \iff a_{n_i} \to a$*, for all subsequences* (a_{n_i})*.*

Proof. The direction "$\Leftarrow$" is obvious. To prove the other direction, we have $\forall\, \epsilon > 0$, $\exists\, N$ s. t. $|a_n - a| < \epsilon$, $\forall\, n \geq N$. Therefore, $\forall\, n_i \geq N$, $|a_{n_i} - a| < \epsilon$. ☐

Lemma 1.5. *Every sequence* (a_n) *in* $\mathbb{R}$ *has a monotone subsequence.*

Proof. Suppose there are no increasing subsequences. Then, $\exists\, n_1$, s. t. $a_{n_1} \geq a_n$, $\forall\, n \geq n_1$. Similarly, $\exists\, n_2 > n_1$, s. t. $a_{n_2} \geq a_n$, $\forall\, n \geq n_2$, and so on. This constructs a decreasing subsequence $a_{n_1} \geq a_{n_2} \geq \dots$. ☐

Lemma 1.6. *Every bounded sequence in* $\mathbb{R}$ *contains a convergent subsequence.*

Proof. From Lemma 1.5, a bounded sequence has a monotone subsequence, which converges from Lemma 1.3. ☐

Let S = set of subsequence limits of $(a_n) \subset \overline{\mathbb{R}}$. We define $\limsup_{n\to\infty} a_n = \sup S$, and $\liminf_{n\to\infty} a_n = \inf S$. We have

$$\forall\, \epsilon > 0,\ \exists\, s \in S,\ \text{s. t. } \sup S - \epsilon/2 \leq s \leq \sup S.$$

Let $a_{n_i} \to s$ (can you see why such a subsequence must exist?) so that

$$\forall\, \epsilon > 0,\ \exists\, N,\ \text{s. t. } |a_{n_i} - s| < \epsilon/2,\ \forall\, n_i > N.$$

Therefore,

$$|a_{n_i} - \sup S| \leq |a_{n_i} - s| + |s - \sup S| < \epsilon,$$

which implies that $\sup S \in S$ and $\sup S = \max S$, i.e., $\limsup_{n\to\infty} a_n$ exists. Similarly, we have $\inf S = \min S$.

Lemma 1.7. *Suppose* $s = \limsup_{n\to\infty} a_n > -\infty$*. Then* $\forall\, \epsilon > 0$*,* $\exists\, N$ s. t. $a_n \leq s + \epsilon$*,* $\forall\, n \geq N$*.*

Proof. If $s = \infty$, the lemma obviously holds. Suppose $s < \infty$. We prove by contradiction. Suppose $\exists\, \epsilon > 0$ and subsequence (a_{n_i}) s. t. $a_{n_i} > s + \epsilon$. Since $s < \infty$, (a_{n_i}) is bounded. Then from Lemma 1.6, $\exists$ subsubsequence $(a_{n_i'})$, s. t. $a_{n_i'} \to a' \geq s + \epsilon > s$, a contradiction to the definition of s. ☐

Lemma 1.8. $\limsup\limits_{n\to\infty} a_n = \inf\limits_{n\geq 1} \sup\limits_{m\geq n} a_m$ *and* $\liminf\limits_{n\to\infty} a_n = \sup\limits_{n\geq 1} \inf\limits_{m\geq n} a_m$.

Proof. If $\limsup_{n\to\infty} a_n = \pm\infty$, then the result clearly holds. Suppose $-\infty < \limsup_{n\to\infty} a_n = \sup S < \infty$, where S is the set of subsequential limits.

For any $\epsilon > 0$, there exists a subsequence (a_{n_i}) s. t. $\sup S - \epsilon \le a_{n_i}$. Then,

$$\sup S - \epsilon \le \sup_{m\ge n} a_m, \ \forall\, n \ge n_1.$$

Letting $n \to \infty$, we have

$$\sup S - \epsilon \le \lim_{n\to\infty} \sup_{m\ge n} a_m = \inf_{n\ge 1} \sup_{m\ge n} a_m,$$

since $\sup_{m\ge n} a_m$ is a decreasing sequence in n. As ϵ is arbitrary, we have

$$\sup S \le \inf_{n\ge 1} \sup_{m\ge n} a_m.$$

On the other hand, for each $n \ge 1$, $\exists\, a_{m_n} \ge \sup_{m\ge n} a_m - 1/n$. Then we have

$$\sup S \ge \lim_{n\to\infty} a_{m_n} \ge \inf_{n\ge 1} \sup_{m\ge n} a_m.$$

The other claim is proved similarly and the proof is now complete. □

1.3 Cauchy Sequences

Definition 1.5. *A sequence (a_n) is a Cauchy sequence if $\forall\, \epsilon > 0$, $\exists\, N$ s. t. $d(a_n, a_m) < \epsilon$, $\forall\, n, m \ge N$.*

Lemma 1.9. *If (a_n) converges, then it is Cauchy.*

Proof. Suppose $a_n \to a$. Then $\forall\, \epsilon > 0$, $\exists\, N$ s. t. $d(a_n, a) < \epsilon/2$ $\forall\, n \ge N$. Then $\forall\, n, m \ge N$, we have

$$d(a_n, a_m) \le d(a_n, a) + d(a_m, a) < \epsilon.$$

□

Definition 1.6. *A metric space $(\mathcal{X}, d)$ in which every Cauchy sequence converges in $\mathcal{X}$ is complete.*

Example of an incomplete space is $\mathbb{Q}$, the set of rational numbers: for every irrational number, one can construct a sequence of rational numbers that converges to it.

Theorem 1.1. *$\mathbb{R}^k$ is complete.*

Proof. WLOG, we prove for $\mathbb{R}$. We first show 2 lemmas.

Lemma 1.10. *A Cauchy sequence is bounded.*

Proof. Similar to Lemma 1.2. □

Lemma 1.11. *If (a_n) is Cauchy and there is a subsequence $a_{n_i} \to a$, then $a_n \to a$.*

Proof. For any $\epsilon > 0$, $\exists\, N$ such that $\forall\, n, n_i \ge N$,

$$d(a_n, a) \le d(a_n, a_{n_i}) + d(a_{n_i}, a) \le \epsilon.$$

□

We now return to the proof of Theorem 1.1. Let (a_n) be a Cauchy sequence in $\mathbb{R}$. Then (a_n) is bounded from Lemma 1.10. From Lemma 1.6, there exists a convergent subsequence. Lemma 1.11 then implies (a_n) converges and the proof is complete. $\square$

2. Basics of Real Analysis - II

SIGNAL, NTU https://personal.ntu.edu.sg/wptay/

Let $(\mathcal{X}, d)$ be a metric space.

2.1 Open and Closed Sets

Definition 2.1. *The open ball of radius $\epsilon > 0$ is defined by*

$$B(x, \epsilon) = \{y \in \mathcal{X} \,:\, d(x, y) < \epsilon\}.$$

Definition 2.2. *A set U is open if $\forall\, x \subset U$, $\exists\, \epsilon > 0$ such that $B(x, \epsilon) \subset U$. A set F is closed if $F^c = \mathcal{X} \backslash F$ is open.*

A set can be both open and closed, e.g., $\mathcal{X}, \emptyset$.

Example 2.1. *Let $\mathcal{X} = \{x_1, x_2, \ldots\}$ be a discrete space. Consider the discrete metric*

$$d(x, y) = \begin{cases} 0, & \text{if } x = y, \\ 1, & \text{if } x \neq y. \end{cases}$$

For any $A \subset \mathcal{X}$ and $\epsilon \in (0, 1)$, $B(a, \epsilon) = \{a\} \subset A$. Therefore A is open. This also proves that A is closed as A^c is open.

If U_i, $i \geq 1$ are open, then $\bigcup_{i=1}^{\infty} U_i$ is open, $\bigcap_{i=1}^{n} U_i$ is open but $\bigcap_{i=1}^{\infty} U_i$ may not be open.

Example 2.2. *Suppose that $U_i = (\frac{1}{2} - \frac{1}{i}, \frac{1}{2} + \frac{1}{i})$, then U_i is open while $\bigcap_{i=1}^{\infty} U_i = \{1/2\}$ is closed.*

If F_i is closed, then $\bigcup_{i=1}^{\infty} F_i$ is closed and $\bigcap_{i=1}^{\infty} F_i$ is closed.

Definition 2.3. *x is a limit point of A if $\forall\, \epsilon > 0$, $\exists\, y \in B(x, \epsilon) \cap A$, and $y \neq x$.*

Therefore, x is a limit point of A if $\exists\, y_1, y_2, \ldots \in A$, $y_i \neq x$ for all $i \geq 1$, s.t. $y_i \to x$.

Lemma 2.1. *A is closed if and only if all limit points of A are in A.*

Proof. Suppose that A is closed, so that A^c is open. Suppose there exists a limit point x of A s.t. $x \in A^c$. Then $\exists\, \epsilon > 0$, s.t. $B(x, \epsilon) \subset A^c$, which is a contradiction to x being a limit point of A.

Suppose that all limit points of A belongs to A. Consider $x \in A^c$. There exists $\epsilon > 0$ s.t. $B(x, \epsilon) \subset A^c$ because otherwise there exists $(y_i) \subset A$ s.t. $y_i \to x$, which means that x is a limit point of A leading to a contradiction. This shows that A^c is open and hence A is closed. $\square$

2.2 Compact Spaces

Definition 2.4. *$\mathcal{X}$ is sequentially compact if every sequence in $\mathcal{X}$ has a convergent subsequence in $\mathcal{X}$.*

Definition 2.5. *$\mathcal{X}$ is totally bounded if $\forall\, \epsilon > 0$, there exists a finite collection $\{B(x_i, \epsilon) : i = 1, 2, \ldots, N_\epsilon\}$ such that*

$$\mathcal{X} \subset \bigcup_{i=1}^{N_\epsilon} B(x_i, \epsilon).$$

Note that if a set $\mathcal{X}$ is totally bounded, then it is bounded. The converse is not true: consider the discrete space in Example 2.1, it is bounded but not totally bounded if it is infinite.

Theorem 2.1. *For a metric space $(\mathcal{X}, d)$, the following are equivalent:*

(i) $\mathcal{X}$ is sequentially compact.

(ii) $\mathcal{X}$ is complete and totally bounded.

(iii) Every open cover of $\mathcal{X}$ has a finite subcover. We say that $\mathcal{X}$ is compact.

Proof.

1) (i) $\Leftrightarrow$ (ii):
We first show that (i) $\Rightarrow$ (ii). Since $\mathcal{X}$ is sequentially compact, every Cauchy sequence in $\mathcal{X}$ has a convergent subsequence. From Lemma 1.11, the Cauchy sequence also converges in $\mathcal{X}$, so $\mathcal{X}$ is complete. Suppose that $\mathcal{X}$ is not totally bounded. Then $\exists\, \epsilon > 0$ so that $\mathcal{X}$ cannot be covered by a finite collection of open balls. Choose any $x_1 \in \mathcal{X}$. Then $\exists\, x_2 \notin B(x_1, \epsilon)$. Similarly, $\exists\, x_3 \notin B(x_1, \epsilon) \cup B(x_2, \epsilon)$, and so on. The sequence (x_n) does not contain any convergent subsequence since $d(x_i, x_j) \geq \epsilon$ for any $i \neq j$. This is a contraction to (i).

   We next show that (ii) $\Rightarrow$ (i). Since $\mathcal{X}$ is totally bounded, for each $m \geq 1$, there exists a finite cover $\{B(x_{m,k}, 1/m) : k = 1, \ldots, M_m\}$ of $\mathcal{X}$. Consider any infinite sequence (y_n) in $\mathcal{X}$. We assume that y_n are distinct because if there are infinitely many y_n that are the same, then there is a trivial convergent subsequence. Then there is a $B(x_{1,k_1}, 1)$ that contains a subsequence $(y_{1,n})$ of (y_n). Similarly, there is a $B(x_{2,k_2}, 1/2)$ that contains a further subsequence $(y_{2,n})$ of $(y_{1,n})$, and so on. Consider the "diagonal" subsequence $(y_{m,m})_{m=1}^{\infty}$. This sequence is Cauchy and since $\mathcal{X}$ is complete, it converges.

2) (i) $\Leftrightarrow$ (iii):
We show (iii) $\Rightarrow$ (i). To do that, we first prove the following facts:

   (a) Any compact $A \subset \mathcal{X}$ is closed. In particular, if $\mathcal{X}$ is compact, then it is closed.
   Let $x \in A^c$ and $U_n = \{y : d(y, x) > 1/n\}$ for $n \geq 1$. Every $y \in \mathcal{X}$ with $y \neq x$ has $d(y, x) > 0$ so y belongs to some U_n. Therefore, $\{U_n : n \geq 1\}$ covers A and there must be a finite subcover. Let N to be the largest index in the subcover, i.e., every $y \in A$ lies in some U_n where $n \leq N$. Then $B(x, 1/N) \subset A^c$ and A is closed.

   (b) If $\mathcal{X}$ is compact and $A \subset \mathcal{X}$ is closed, then A is compact.
   Let $\{U_n\}$ be an open cover of A. Then $\{U_n\} \cup \{A^c\}$ is an open cover of $\mathcal{X}$. There is a finite subcover, say, $\{U_1, \ldots, U_N, A^c\}$ of $\mathcal{X}$. Then $\{U_1, \ldots, U_N\}$ is a finite open cover of A.

   Suppose $\mathcal{X}$ is compact. Assume there is a sequence (x_n) that has no convergent subsequences. In particular, this sequence has infinitely distinct points $y_1, y_2, \ldots$. Since there is no convergent subsequence, there is some open ball B_k containing each y_k and no other y_i. The set $A = \{y_1, y_2, \ldots\}$ is closed as it has no limit points, so it is compact. But $\{B_k\}$ is an open cover of A and has no finite subcover, a contradiction. Therefore (x_n) has a convergent subsequence whose limit lies in $\mathcal{X}$ as $\mathcal{X}$ is closed.

We now show (i) $\Rightarrow$ (iii). Suppose that $\mathcal{X}$ is sequentially compact. Let $\{G_\alpha \subset \mathcal{X} : \alpha \in I\}$ be an open cover of $\mathcal{X}$. We claim that there exists $\epsilon > 0$ such that every ball $B(x, \epsilon)$ is contained in some G_α. Suppose not. Then for each positive integer n, $\exists\, y_n \in \mathcal{X}$ such that $B(y_n, 1/n)$ is not contained in any G_α. By hypothesis, there exists a subsequence $y_{n_i} \to y \in \mathcal{X}$. Since $\{G_\alpha \subset \mathcal{X} : \alpha \in I\}$ is an open cover of $\mathcal{X}$, there exists $G_{\alpha_0} \ni y$ and $\epsilon > 0$ such that $B(y, \epsilon) \subset G_{\alpha_0}$. Choose n_i sufficiently large so that $d(y_{n_i}, y) < \epsilon/2$ and $1/n_i < \epsilon/2$. Then $B(y_{n_i}, 1/n_i) \subset G_{\alpha_0}$, a contradiction.

Since $\mathcal{X}$ is sequentially compact, it is totally bounded, so there exists a finite collection of balls of radius ϵ s.t. $\{B(x_i, \epsilon) : i = 1, 2, \ldots, N\}$ covers $\mathcal{X}$. Choose $\alpha_i \in I$ s.t. $B(x_i, \epsilon) \subset G_{\alpha_i}$. Then $\{G_{\alpha_i} : i = 1, 2, \ldots, N\}$ is a finite subcover of $\mathcal{X}$, which means $\mathcal{X}$ is compact.

$\square$

Theorem 2.2 (Heine-Borel)**.** *$A \subset \mathbb{R}^k$ is compact iff A is closed and bounded.*

Proof. Since the following proof can be repeated in every dimension, it suffices to prove only for $A \subset \mathbb{R}$.

We first prove sufficiency. Suppose A is closed and bounded and $(x_n) \subset A$. Then x_n is bounded since A is bounded. By Lemma 1.6, there exists convergent (x_{n_i}) such that $x_{n_i} \to x \in \mathbb{R}$. Since A is closed, $x \in A$. Therefore, A is compact.

Next, suppose that A is compact. Then from Theorem 2.1, A is complete and totally bounded, hence closed and bounded in $\mathbb{R}$.

$\square$

2.3 Continuity

Definition 2.6. *A function $f : (\mathcal{X}, d_{\mathcal{X}}) \mapsto (\mathcal{Y}, d_{\mathcal{Y}})$ is (pointwise) continuous at $x \in \mathcal{X}$ if $\forall\, \epsilon > 0$, $\exists\, \delta_x > 0$ s.t. $d_{\mathcal{Y}}(f(x), f(z)) < \epsilon$, whenever $d_{\mathcal{X}}(x, z) < \delta_x$.*

Note that the definition is equivalent to saying that $\forall\, \epsilon > 0$, $\exists\, \delta_x > 0$ s.t. $B(x, \delta_x) \subset f^{-1}(B(f(x), \epsilon))$.

Lemma 2.2. *$f : (\mathcal{X}, d_X) \mapsto (\mathcal{Y}, d_Y)$ is continuous iff $f^{-1}(U)$ is open in $\mathcal{X}$ for every open $U \subset \mathcal{Y}$.*

Proof. Suppose f is continuous and U is open in $\mathcal{Y}$. For each $y \in U$, there exists an open ball $B(y, r) \subset U$. Since f is continuous, for each $x \in f^{-1}(\{y\})$, $\exists\, \delta_x > 0$ s.t. $B(x, \delta_x) \subset f^{-1}(B(y, r)) \subset f^{-1}(U)$. Therefore, $f^{-1}(U) = \bigcup_{x \in f^{-1}(U)} B(x, \delta_x)$ is open.

To prove the other direction, we have for every $x \in \mathcal{X}$ and $\epsilon > 0$, $f^{-1}(B(f(x), \epsilon))$ is open and contains x, so it contains an open ball around x. Thus, f is continuous at x. $\square$

Lemma 2.3. *If f is continuous and $B \in \mathcal{X}$ is compact, then $f(B) \triangleq \{f(x) : x \in B\}$ is compact.*

Proof. Consider $y_n = f(x_n)$. $\exists\, (x_{n_i})$ s.t. $x_{ni} \to x \in B$. From the continuity of f, $y_{n_i} = f(x_{n_i}) \to f(x) \in f(B)$. $\square$

From Theorem 2.2, $f(B)$ is closed and bounded. Therefore, $\sup_{x \in B} f(x)$ and $\inf_{x \in B} f(x)$ are both achieved on B.

Definition 2.7. *f is uniformly continuous if* $\forall\, \epsilon > 0$, $\exists\, \delta > 0$ *s.t.* $d_{\mathcal{Y}}(f(x), f(z)) < \epsilon$, $\forall\, d_{\mathcal{X}}(x, z) < \delta$.

Example 2.3.

(a) *If* $|f(x) - f(y)| < Ld(x, y)$ *for some positive constant L and all $x, y \in \mathcal{X}$, we say that f is Lipschitz continuous. It is clear that f is uniformly continuous on $\mathcal{X}$.*

(b) *If $f : \mathbb{R} \mapsto \mathbb{R}$ is differentiable with* $\sup |f'| < \infty$*, where f' is the derivative of f, then f is uniformly continuous from the mean value theorem.*

Lemma 2.4. *Suppose that $f\colon K \mapsto \mathcal{Y}$, where $K \subset \mathcal{X}$ is compact. If f is continuous, then f is uniformly continuous.*

Proof. Fix $\epsilon > 0$. For each $x \in \mathcal{X}$, $\exists\, \delta_x > 0$, s.t. $d_{\mathcal{Y}}(f(x), f(y)) < \epsilon/2$ whenever $d_{\mathcal{X}}(x, y) < \delta_x$. We have $\{B(x, \delta_x/2) : x \in K\}$ is an open cover of K. Since K is compact, exists a finite subcover $\{B(x_i, \delta_{x_i}/2) : 1 \leq i \leq n\}$. Let $\delta = \min\{\delta_{x_i}, \cdots, \delta_{x_n}\} > 0$.

For each $x \in K$, $\exists\, x_i$, s.t. $d_{\mathcal{X}}(x, x_i) < \delta_{x_i}/2$. Then for all y s.t. $d_{\mathcal{X}}(x, y) < \delta/2$, we have

$$d_{\mathcal{X}}(y, x_i) \leq d_{\mathcal{X}}(x, y) + d_{\mathcal{X}}(x, x_i) < \delta_{x_i}.$$

Therefore, we obtain

$$d_{\mathcal{Y}}(f(x), f(y)) \leq d_{\mathcal{Y}}(f(x), f(x_i)) + d_{\mathcal{Y}}(f(x_i), f(y)) \leq \epsilon,$$

which shows that f is uniformly continuous. □

Note that in the above proof, y need not be in K. We obtain a slightly stronger result here.

Corollary 2.1. *Suppose $f : \mathcal{X} \mapsto \mathcal{Y}$ is continuous, and $K \subset \mathcal{X}$ is compact. Then* $\forall\, \epsilon > 0$, $\exists\, \delta > 0$ *such that* $d_{\mathcal{Y}}(f(x), f(y)) < \epsilon$ *whenever $x \in K$, $y \in \mathcal{X}$ and* $d_{\mathcal{X}}(x, y) < \delta$.

2.4 Riemann Integral

We consider a function $f : [a, b] \mapsto \mathbb{R}$ in this section. A partition $P = (x_0, \ldots, x_N)$ is defined by $x_0 = a < x_1 < x_2 < \cdots < x_N = b$. We say that Q is a refinement of P if $P \subset Q$. Let $\mathcal{P}$ be the collection of all partitions. For $P \in \mathcal{P}$, define

$$\begin{aligned} U(f, P) &= \sum_{i=1}^{N} \sup_{[x_{i-1}, x_i]} f \cdot (x_i - x_{i-1}), \\ L(f, P) &= \sum_{i=1}^{N} \inf_{[x_{i-1}, x_i]} f \cdot (x_i - x_{i-1}). \end{aligned}$$

Lemma 2.5. $L(f, P) \leq U(f, Q)$, $\forall\, P, Q \in \mathcal{P}$

Proof.

$$\begin{aligned} L(f, P) &\leq L(f, P \cup Q) \quad \text{since } \inf_{I_1} f \leq \inf_{I_2} f \text{ if } I_1 \supset I_2 \\ &\leq U(f, P \cup Q) \\ &\leq U(f, Q) \quad \text{since } \sup_{I_1} f \geq \sup_{I_2} f \text{ if } I_1 \supset I_2. \end{aligned}$$

□

From the above lemma, we have

$$\sup_{P\in\mathcal{P}} L(f,P) \le \inf_{P\in\mathcal{P}} U(f,P) \tag{1}$$

Definition 2.8. *$f:[a,b]\mapsto\mathbb{R}$ is Riemann integrable if equality in (1) holds, i.e.,*

$$\forall\,\epsilon>0,\ \exists\,P\in\mathcal{P}\ \text{ s.t. }\ \ U(f,P)-L(f,P)<\epsilon.$$

Example 2.4. *The following function is not Riemann integrable as $L(f,P)=0$ and $U(f,P)=1$ for all $P\in\mathcal{P}$:*

$$f(x)=\begin{cases} 1, & \text{if } x\in\mathbb{Q}\cap[0,1],\\ 0, & \text{otherwise.}\end{cases}$$

Definition 2.9. *A set $A\subset\mathbb{R}$ has Lebesgue measure zero if $\forall\,\epsilon>0$, there exists open intervals $(\alpha_1,\beta_1),(\alpha_2,\beta_2),\dots$ s.t.*

$$A\subset\bigcup_{i=1}^{\infty}(\alpha_i,\beta_i)\ \text{ and }\ \sum_{i=1}^{\infty}(\alpha_i-\beta_i)<\epsilon.$$

If a countable sequence of sets $A_1,A_2,\dots$ each of which has Lebesgue measure zero, then the union $\bigcup_{i=1}^{\infty}A_i$ has Lebesgue measure zero. To see this, let $\epsilon>0$ and A_j be covered by $\bigcup_i(\alpha_{ij},\beta_{ij})$ with $\sum_i(\alpha_{ij}-\beta_{ij})<\epsilon/2^j$.

Theorem 2.3 (Henri Lebesgue)**.** *Suppose $f:[a,b]\mapsto\mathbb{R}$ is bounded. Then f is Riemann integrable iff $\exists\,A\subset[a,b]$ of Lebesgue measure zero s.t. f is continuous on $[a,b]\backslash A$.*[1]

Proof. We first show that if f is Riemann integrable, then its set of discontinuities has Lebesgue measure zero. Observe that $y\in(a,b)$ is a point of discontinuity of f iff $\exists\,j\in\mathbb{Z}_+$ s.t. $\sup_I f-\inf_I f\ge 1/j$ for all open intervals $I\subset(a,b)$ containing y. Let

$$S_j=\left\{y\in(a,b)\ :\ \sup_I f-\inf_I f\ge\frac{1}{j}\ \forall\ \text{open intervals } I\subset(a,b) \text{ with } y\in I\right\}.$$

Then, the set of discontinuities of f in (a,b) is $\bigcup_{j=1}^{\infty}S_j$. For $\epsilon>0$, since f is Riemann integrable, there exists some partition $P=(x_0,\dots,x_N)$ s.t.

$$U(f,P)-L(f,P)=\sum_{i=1}^{N}\Big(\sup_{[x_{i-1},x_i]}f-\inf_{[x_{i-1},x_i]}f\Big)\cdot(x_i-x_{i-1})<\frac{\epsilon}{j}. \tag{2}$$

Let $B=\{i:(x_{i-1},x_i)\cap S_j\neq\emptyset\}$. Then from (2), we have

$$\sum_B\Big(\sup_{[x_{i-1},x_i]}f-\inf_{[x_{i-1},x_i]}f\Big)\cdot(x_i-x_{i-1})+\sum_{B^c}\Big(\sup_{[x_{i-1},x_i]}f-\inf_{[x_{i-1},x_i]}f\Big)\cdot(x_i-x_{i-1})<\frac{\epsilon}{j} \tag{3}$$

Since

$$\sum_{B^c}\Big(\sup_{[x_{i-1},x_i]}f-\inf_{[x_{i-1},x_i]}f\Big)\cdot(x_i-x_{i-1})\ge 0$$

and

$$\sum_B\Big(\sup_{[x_{i-1},x_i]}f-\inf_{[x_{i-1},x_i]}f\Big)\cdot(x_i-x_{i-1})\ge\frac{1}{j}\sum_{i\in B}(x_i-x_{i-1}),$$

[1] f is also said to be continuous almost everywhere on $[a,b]$.

we obtain from (3),

$$\sum_{i \in B} (x_i - x_{i-1}) < \epsilon.$$

We have

$$S_j \subset \bigcup_{i \in B} (x_{i-1}, x_i) \bigcup \{x_0, x_1, \cdots, x_N\},$$

therefore S_j has Lebesgue measure zero.

We next prove the converse. Fix an $\epsilon > 0$. Assume that we have A with Lebesgue measure zero, which means that there is a cover $\bigcup_{j=1}^{\infty} (\alpha_j, \beta_j) \supset A$ s.t. $\sum_{j=1}^{\infty} (\beta_j - \alpha_j) < \epsilon$. Let $K = [a, b] \setminus \bigcup_{j \geq 1} (\alpha_j, \beta_j)$, which is closed and bounded and therefore compact by Theorem 2.2. Since f is continuous, from Corollary 2.1, $\exists\, \delta > 0$, s.t. $|f(x) - f(y)| < \epsilon$ whenever $x \in K$, $y \in [a, b]$ and $|x - y| < \delta$.

We choose a partition P with $a = x_0 < x_1 < x_2 < \cdots < x_N = b$ s.t. $\max_{1 \leq i \leq N}(x_i - x_{i-1}) < \delta$. If $[x_{i-1}, x_i] \cap K = \emptyset$, then $[x_{i-1}, x_i] \subset \bigcup_j (\alpha_j, \beta_j)$ and

$$\sum_{i:[x_{i-1},x_i] \cap K = \emptyset} \Big(\sup_{[x_{i-1},x_i]} f - \inf_{[x_{i-1},x_i]} f \Big) \cdot (x_i - x_{i-1}) \leq \big(\sup_{[a,b]} f - \inf_{[a,b]} f\big) \cdot \sum_j (\beta_j - \alpha_j) < M\epsilon,$$

where $M = \sup_{[a,b]} f - \inf_{[a,b]} f < \infty$. Suppose $[x_{i-1}, x_i] \cap K \neq \emptyset$. Then for any $y, z \in [x_{i-1}, x_i]$ and $y_i \in [x_{i-1}, x_i] \cap K$, we have

$$\begin{aligned} |f(y) - f(z)| &\leq |f(y) - f(y_i)| + |f(y_i) - f(z)| \\ &< \epsilon + \epsilon \\ &= 2\epsilon, \end{aligned}$$

where the last inequality follows because $|y - y_i|,\ |z - y_i| < \delta$. Therefore,

$$\sum_{i:[x_{i-1},x_i] \cap K \neq \emptyset} \Big(\sup_{[x_{i-1},x_i]} f - \inf_{[x_{i-1},x_i]} f \Big) \cdot (x_i - x_{i-1}) \leq 2\epsilon(b - a).$$

We finally obtain

$$U(f, P) - L(f, P) \leq M\epsilon + 2\epsilon(b - a) = (M + 2(b - a))\epsilon,$$

and the proof is complete. □

From Theorem 2.3, we see that only a very limited class of functions f is Riemann integrable. This is not sufficient to model many practical applications. Therefore, Henri Lebesgue, a French mathematician in the 17th century, embarked on a program to introduce a much more versatile integral known as the Lebesgue integral. We will introduce this in the coming week as part of the theory of probability.

An Analytical Introduction to Probability Theory

3. Probability Spaces

SIGNAL, NTU `https://personal.ntu.edu.sg/wptay/`

3.1 Introduction

Recommended reference: "Probability: Theory and Examples" by Rick Durrett.

Let Ω be a sample space. An event is a subset of Ω. We are interested to define a "likelihood" or "chance" for each event to happen in the future. We call this the probability of the event.

Example 3.1. *Let $\Omega = [0,1]$, the probability of the event $(a,b]$, where $0 \leq a \leq b < 1$ can be defined by*

$$\mathbb{P}((a,b]) = F(b) - F(a),$$

where F is a non-decreasing and right-continuous (we will see later why this is needed) function with

$$\lim_{x\to 0} F(x) = 0,$$
$$\lim_{x\to 1} F(x) = 1.$$

However, there are many other events like $\bigcup_{i=1}^{\infty}(a_i,b_i]$ whose probabilities we are interested in. In particular, $\mathbb{P}$ should have the following properties:

(i) $\mathbb{P}(\Omega) = 1$.

(ii) If $A_1, A_2, \ldots$ are disjoint sets, then

$$\mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} \mathbb{P}(A_i).$$

(iii) If A is congruent to B (i.e., A is B transformed by translation, rotation or reflection), then $\mathbb{P}(A) = \mathbb{P}(B)$.

Unfortunately, for these conditions to hold for all *events would lead to inconsistency. To see why, define an equivalence $x \sim y$ iff $x - y$ is rational. Then Ω can be partitioned into equivalence classes. Let $N \subset \Omega$ be a subset that contains exactly one member of each equivalence class (we need the axiom of choice here). For each rational number $r \in \mathbb{Q} \cap [0,1)$, let*

$$N_r = \{x + r \,:\, x \in N \cap [0, 1-r)\} \cup \{x + r - 1 \,:\, x \in N \cap [1-r, 1]\},$$

i.e., N_r is N translated to the right by r with the part after $[0,1)$ shifted to the front (wrapped around) so that $N_r \subset \Omega = [0,1]$. From properties (ii) and (iii), we have for any rational $r \in \mathbb{Q} \cap [0,1)$,

$$\mathbb{P}(N) = \mathbb{P}(N \cap [0, 1-r)) + \mathbb{P}(N \cap [1-r, 1)) = \mathbb{P}(N_r). \tag{4}$$

We also have the following:

1. *Every $x \in \Omega$ belongs to a N_r because if $y \in N$ is an element of the equivalence class of x, then $x \in N_r$ where $r = x - y$ if $x \geq y$ or $r = x - y + 1$ if $x < y$.*
2. *Every $x \in \Omega$ belongs to exactly one N_r because if $x \in N_r \cap N_s$ for $r \neq s$, then $x - r$ or $x - r + 1$ and $x - s$ or $x - s + 1$ would be distinct elements of N belonging to the same equivalence class, contradicting how we chose N.*

Therefore, Ω is the disjoint union of N_r over all rational $r \in \mathbb{Q} \cap [0, 1)$. From properties (i) and (ii), we also have $1 = \mathbb{P}(\Omega) = \sum_r \mathbb{P}(N_r)$. But $\mathbb{P}(N_r) = \mathbb{P}(N)$ from (4), so the sum is either 0 if $\mathbb{P}(N) = 0$ or ∞ if $\mathbb{P}(N) > 0$, a contradiction.

This example shows that it is impossible to define a suitable $\mathbb{P}$ for all possible events, some of which are very weird objects (Banach and Tarski (1924) showed that in $\mathbb{R}^n$ where $n \geq 3$, even stranger subsets can be constructed!). The solution that mathematicians have come up with is to restrict to a collection of subsets and *a $\mathbb{P}$ with "nice" properties, i.e., a σ-algebra and measure, respectively.*

3.2 σ-algebras and Measures

Let $\mathcal{A}$ be a collection of events (collection of subsets of Ω).

Definition 3.1. *$\mathcal{A}$ is an algebra if*

(i) $\Omega \in \mathcal{A}$.

(ii) $A \in \mathcal{A} \implies A^c = \Omega \setminus A \in \mathcal{A}$.

(iii) $A_1, A_2 \in \mathcal{A} \implies A_1 \cup A_2 \in \mathcal{A}$. *By induction,* $A_i \in \mathcal{A}, \forall\, i = 1, \ldots, n, \implies \bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$.

Definition 3.2. *$\mathcal{A}$ is a σ-algebra or σ-field if*

(i) $\Omega \in \mathcal{A}$.

(ii) $A \in \mathcal{A} \implies A^c = \Omega \setminus A \in \mathcal{A}$.

(iii) $A_i \in \mathcal{A}, \forall\, i = 1, 2, \ldots \implies \bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{A}$.

$(\Omega, \mathcal{A})$ is called a measurable space if $\mathcal{A}$ is a σ-algebra. A set $A \in \mathcal{A}$ is said to be measurable.

Definition 3.3. *For a measurable space $(\Omega, \mathcal{A})$, a function $\mathbb{P} : \mathcal{A} \mapsto [0, 1]$ is a probability measure if*

(i) $\mathbb{P}(\Omega) = 1$.

(ii) $A_1, A_2, \ldots \in \mathcal{A}$ *with* $A_i \cap A_j = \emptyset$, $\forall\, i \neq j \implies \mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} \mathbb{P}(A_i)$ *(countably additive).*

Let $\mathcal{A}$ be a σ-algebra.

Lemma 3.1. *Suppose $B_i \in \mathcal{A}, B_i \subset B_{i+1}, \forall\, i \geq 1$, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{i \to \infty} \mathbb{P}(B_i)$$

Proof. Let $C_1 = B_1, C_i = B_i \cap B_{i-1}^c, \forall\, i \geq 2$, then the C_i's are disjoint, and we have

$$B_n = \bigcup_{i=1}^{n} C_i,$$
$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} C_i.$$

Then we obtain

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(C_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(C_i) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} C_i\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

□

Corollary 3.1. *For $A_i \in \mathcal{A}, \forall\ i = 1, 2, \ldots,$ we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right).$$

Proof. Let $B_n = \bigcup\limits_{i=1}^{n} A_i$, which is an increasing sequence. We have $\mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} B_n\right)$. From Lemma 3.1, we obtain $\mathbb{P}\left(\bigcup\limits_{n=1}^{\infty} B_n\right) = \lim\limits_{n\to\infty} \mathbb{P}(B_n) = \lim\limits_{n\to\infty} \mathbb{P}\left(\bigcup\limits_{i=1}^{n} A_i\right)$. □

Corollary 3.2. *For a decreasing sequence $B_i \supset B_{i+1}, \forall\ i \geq 1,$ we have*

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{i\to\infty} \mathbb{P}(B_i).$$

Proof. Similar to the proof of Lemma 3.1. □

Lemma 3.2. *For $A, B \in \mathcal{A}, A \subset B,$ we have $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

Proof. $B = A \cup (B \setminus A) \implies \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$. □

Lemma 3.3 (Union bound)**.** *For $A_1, A_2, \ldots \in \mathcal{A},$ we have*

$$\mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) \leq \sum_{i\geq 1} \mathbb{P}(A_i).$$

Proof. Let $B_i = A_i \setminus \bigcup_{j<i} A_j$ for $i \geq 1$. Then the B_i's are disjoint, $B_i \subset A_i$, $\bigcup_{i\geq 1} A_i = \bigcup_{i\geq 1} B_i$ and

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) &= \mathbb{P}\left(\bigcup_{i\geq 1} B_i\right) \\ &= \sum_{i\geq 1} \mathbb{P}(B_i) \\ &\leq \sum_{i\geq 1} \mathbb{P}(A_i). \end{aligned}$$

□

In Example 3.1, let

$$\mathcal{A}' = \left\{ \bigcup_{i=1}^{n} (a_i, b_i] \; : \; n \geq 1, (a_i, b_i] \subset (0,1], (a_i, b_i] \cap (a_j, b_j] = \emptyset, \forall \;\; i \neq j \right\}.$$

Check that $\mathcal{A}'$ is an algebra. For each element of $\mathcal{A}'$, we define

$$\mathbb{P}\left(\bigcup_{i=1}^{n} (a_i, b_i] \right) = \sum_{i=1}^{n} \mathbb{P}((a_i, b_i])$$

for each $n \geq 1$. One can show that with this definition, $\mathbb{P}$ is countably additive on $\mathcal{A}'$, i.e., whenever $A_i \in \mathcal{A}'$, $i \geq 1$ and $\bigcup_{i \geq 1} A_i \in \mathcal{A}$ are finite unions of disjoint intervals, we have

$$\mathbb{P}\left(\bigcup_{i \geq 1} (a_i, b_i] \right) = \sum_{i \geq 1} \mathbb{P}((a_i, b_i]).$$

We also have

$$\begin{aligned}
\mathbb{P}((a,b]) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} (a, b + 1/n] \right) \\
&= \lim_{n \to \infty} \mathbb{P}((a, b + 1/n]) \\
&= \lim_{n \to \infty} (F(b + 1/n) - F(a)) \\
&= F(b) - F(a),
\end{aligned}$$

where the second equality follows from Corollary 3.2 and the last equality requires the right-continuity of F.

Let $\mathcal{A} = \sigma(\mathcal{A}')$ be the σ-algebra generated by $\mathcal{A}'$, i.e., the intersection of all σ-algebras that contain $\mathcal{A}'$. Note that this is well-defined as a trivial σ-algebra containing $\mathcal{A}'$ is the power set of Ω. It is easy to show that $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}'$.

Theorem 3.1. *Carathéodory's Extension Theorem. If $\mathcal{A}'$ is an algebra, $\mathbb{P} : \mathcal{A}' \mapsto [0,1]$ is countably additive on $\mathcal{A}'$ and $\mathbb{P}(\emptyset) = 0$, then $\mathbb{P}$ has a unique extension to $\mathcal{A} = \sigma(\mathcal{A}')$.*

The proof of the existence of such an extension $\mathbb{P} : \mathcal{A} \mapsto [0,1]$ can be found in the book by Durrett. We focus on the proof of uniqueness here. We make use of the very useful Dynkin's π-λ Theorem.

Definition 3.4. *$\mathcal{P}$ is a π-system if $A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$.*

Definition 3.5. *$\mathcal{L}$ is a λ-system if*

- *(i)* $\Omega \in \mathcal{L}$.
- *(ii)* $A \in \mathcal{L} \implies A^c = \Omega \setminus A \in \mathcal{L}$.
- *(iii)* $A_i \in \mathcal{L}, \forall \;\; i \geq 1, \;\; A_i \cap A_j = \emptyset, \forall \;\; i \neq j \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$.

Remark 3.1. *If $\mathcal{A}$ is both a π-system and λ-system, then $\mathcal{A}$ is σ-algebra.*

Theorem 3.2 (Dynkin's π-λ Theorem)**.** *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system with $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. Let $\ell(\mathcal{P})$ be the smallest λ-system that contains $\mathcal{P}$. If $\ell(\mathcal{P})$ is a π-system, $\ell(\mathcal{P})$ is a σ-algebra. Then we have

$$\sigma(\mathcal{P}) \subset \ell(\mathcal{P}) \subset \mathcal{L}.$$

Therefore, it suffices to prove that $\ell(\mathcal{P})$ is a π-system. We prove that $\ell(\mathcal{P})$ is a π-system in the following three steps. For any $A \subset \Omega$, let $\mathcal{G}_A = \{B \subset \Omega : B \cap A \in \ell(\mathcal{P})\}$.

Step 1: We show that if $A \in \ell(\mathcal{P})$, then $\mathcal{G}_A$ is λ-system. This is done by checking the following conditions:

(i) $\Omega \cap A = A \in \ell(\mathcal{P}) \implies \Omega \in \mathcal{G}_A$.

(ii) Suppose $B \in \mathcal{G}_A$. We have $B^c \cap A = ((B \cap A) \cup A^c)^c$. Then we have

$$B \cap A,\ A^c \in \ell(\mathcal{P}) \implies (B \cap A) \cup A^c \in \ell(\mathcal{P}) \implies ((B \cap A) \cup A^c)^c \in \ell(\mathcal{P}),$$

which implies that $B^c \in \mathcal{G}_A$.

(iii) Let $B_i \in \mathcal{G}_A, \forall\ i = 1, 2, \ldots$ with $B_i \cap B_j = \emptyset, \forall\ i \neq j$. Therefore, $B_i \cap A \in \ell(\mathcal{P}), \forall\ i = 1, 2, \ldots$ are also disjoint. Then we have

$$\left(\bigcup_{i=1}^{\infty} B_i\right) \cap A = \bigcup_{i=1}^{\infty} (B_i \cap A) \in \ell(\mathcal{P}) \implies \bigcup_{i=1}^{\infty} B_i \in \mathcal{G}_A.$$

Step 2: We show that if $B \in \mathcal{P} \subset \ell(\mathcal{P})$, then $\ell(\mathcal{P}) \subset \mathcal{G}_B$. Since $\mathcal{P}$ is a π-system, we have

$$\forall\ C \in \mathcal{P}, C \cap B \in \mathcal{P} \subset \ell(\mathcal{P}),$$

which means that

$$\mathcal{P} \subset \mathcal{G}_B,$$

where $\mathcal{G}_B$ is a λ-system from Step 1 since $B \in \ell(\mathcal{P})$. Therefore,

$$\ell(\mathcal{P}) \subset \mathcal{G}_B. \tag{5}$$

Step 3: Consider any $B \in \mathcal{P}$ and an $A \in \ell(\mathcal{P})$. From Step 2, we have $A \in \mathcal{G}_B$. Therefore, $A \cap B \in \ell(\mathcal{P})$ and

$$\mathcal{P} \subset \mathcal{G}_A,$$

where $\mathcal{G}_A$ is a λ-system from Step 1. Thus $\ell(\mathcal{P}) \subset \mathcal{G}_A$. This means that for any $C \in \ell(\mathcal{P})$, we have

$$C \in \mathcal{G}_A \implies C \cap A \in \ell(\mathcal{P}).$$

Since $A, C \in \ell(\mathcal{P})$, the above result immediately shows that $\ell(\mathcal{P})$ is a π-system. $\square$

We now return to the uniqueness proof of Theorem 3.1. Suppose $\mathbb{P}_1$ and $\mathbb{P}_2$ are extensions of $\mathbb{P}$ with $\mathbb{P}_1(A) = \mathbb{P}_2(A), \forall\ A \in \mathcal{A}'$. Let

$$\mathcal{L} = \{A \in \mathcal{A} : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}.$$

It is easy to see that $\mathcal{L}$ is a λ-system due to the properties of probability measures. $\mathcal{A}'$ is a π-system since it is an algebra, and $\mathcal{A}' \subset \mathcal{L}$ by definition. According to Theorem 3.2, $\mathcal{A} = \sigma(\mathcal{A}') \subset \mathcal{L}$. Thus, $\mathbb{P}_1 = \mathbb{P}_2$ on $\mathcal{A}$, meaning the extension is unique on $\mathcal{A}$.

3.3 Regularity

Definition 3.6. *Let (Ω, d) be a metric space. The Borel σ-algebra is a σ-algebra generated by the open sets of Ω (or equivalently, by the closed sets).*

Lemma 3.4. *Let $\mathcal{B}$ be the Borel σ-algebra. Then, $\forall\, A \in \mathcal{B}$,*

$$\mathbb{P}(A) = \sup\{\mathbb{P}(F)\,:\, F \subset A,\ F \text{ is closed}\}. \tag{6}$$

Proof. Let

$$\mathcal{L} = \{A \in \mathcal{B}\,:\, \text{both } A \text{ and } A^c \text{ satisfy (6)}\}.$$

It can be checked that $\mathcal{L}$ is a λ-system (exercise). Let F be closed. It is obvious that F satisfies (6). Let $U = F^c$. We show that U also satisfies (6). To do this, since $\sup \mathbb{P}(C) \le \mathbb{P}(U)$ for all closed $C \subset U$, it suffices to show that there is a sequence of closed subsets $F_n \subset U$ such that $\mathbb{P}(U) = \sup_n \mathbb{P}(F_n)$. To this end, for $n \ge 1$, let

$$F_n = \left\{\omega \in \Omega : \min_{x \in F} d(\omega, x) \ge 1/n\right\}.$$

Then we have $F_n \subset F_{n+1}$ and $U = \bigcup\limits_{n=1}^{\infty} F_n$. From Lemma 3.1, we obtain

$$\mathbb{P}(U) = \lim_{n\to\infty} \mathbb{P}(F_n) = \sup_{n \ge 1} \mathbb{P}(F_n).$$

Therefore, U satisfies (6). As a consequence, $F \in \mathcal{L}$ for any closed F. Since $\mathcal{B}$ is generated by the closed sets, we have $\mathcal{B} \in \mathcal{L}$ and the proof is complete. $\square$

A metric space is separable if it has a countable dense subset, i.e., $\exists\,\{x_n\}_{n=1}^{\infty}$ such that $\forall$ open $U \subset \Omega$, $x_i \in U$ for some x_i. Exercise: show that totally bounded implies separable. The converse is not true (e.g., consider the discrete metric space in Example 2.1).

Definition 3.7. *We say that a probability measure $\mathbb{P}$ for $(\Omega, \mathcal{B})$ is regular if*

$$\mathbb{P}(A) = \sup\{\mathbb{P}(K) : K \subset A, K \text{ is compact}\}$$

for all $A \in \mathcal{B}$.

Theorem 3.3 (Ulam)**.** *If (Ω, d) is a complete separable space with Borel σ-algebra $\mathcal{B}$ and probability measure $\mathbb{P}$, then $\mathbb{P}$ is regular.*

Proof. Fix $\epsilon > 0$ and let $\{x_i\}_{i\ge 1}$ be a dense subset of Ω. Then for any $m \ge 1$, we have

$$\Omega = \bigcup_{i \ge 1} \overline{B}(x_i, 1/m),$$

where $\overline{B}(x_i, 1/m)$ is the closed ball of radius $1/m$ centered at x_i. Since $\mathbb{P}(\Omega) = 1$, there exists $n(m)$ sufficiently large so that

$$\mathbb{P}\left(\Omega \backslash \bigcup_{i=1}^{n(m)} \overline{B}(x_i, \frac{1}{m})\right) \le \frac{\epsilon}{2^m}.$$

Let $K = \bigcap_{m\geq 1} \bigcup_{i=1}^{n(m)} \overline{B}(x_i, 1/m)$, which is closed and totally bounded. Since Ω is complete, K is also complete and hence compact. We then have

$$\begin{aligned}\mathbb{P}(\Omega\backslash K) &= \mathbb{P}\left(\bigcap_{m\geq 1}\left(\Omega\backslash \bigcup_{i=1}^{n(m)} \overline{B}(x_i, 1/m)\right)\right)\\ &\leq \sum_{m\geq 1} \frac{\epsilon}{2^m}\\ &\leq \epsilon.\end{aligned}$$

From Lemma 3.4, for any $A \in \mathcal{A}$, there exists a closed $F \subset A$ such that $\mathbb{P}(A\backslash F) \leq \epsilon$. Therefore,

$$\mathbb{P}(A\backslash(F \cap K)) \leq 2\epsilon,$$

and $F \cap K \subset A$ is a compact set. The theorem is now proved. □

3.4 Notes

(a) $([0,1], \mathcal{B}([0,1]), \lambda)$ is a probability space, where $\lambda((a,b)) = b - a$ is called the Lebesgue measure. This is the uniform distribution on $[0,1]$.

(b) Let $f : \mathbb{R} \mapsto \mathbb{R}_+$ be a function whose set of discontinuities has Lebesgue measure zero and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. Then $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ where $\mathbb{P}(A) = \int_A f(x)\,\mathrm{d}x$, is a probability space. f is an example of a *probability density function* (pdf).

(c) Let $\mathcal{X}$ be a discrete set. Then $(\mathcal{X}, 2^{\mathcal{X}}, \mathbb{P})$ where $\mathbb{P}(\{x\}) = p(x)$ with $\sum_{x\in\mathcal{X}} p(x) = 1$, is a probability space. Here, $2^{\mathcal{X}}$ denotes the power set of $\mathcal{X}$, or the collection of all subsets of $\mathcal{X}$.

(d) There exist non-measureable sets, i.e., one cannot assign a measure to these sets without running into logical consistency issues (see Example 3.1 or Durrett). This is why the existence of probability spaces is non-trivial as we cannot simply define a measure over the power set 2^{Ω}.

(e) Exercise: If A is an algebra, then for any $B \in \sigma(A)$, $\exists\, B_n \in A$ such that

$$\lim_{n\to\infty} \mathbb{P}(B \triangle B_n) = 0,$$

where $B \triangle B_n = (B \cup B_n) \setminus (B \cap B_n)$ is the symmetric difference of two sets.

An Analytical Introduction to Probability Theory

4. Random Variables

SIGNAL, NTU `https://personal.ntu.edu.sg/wptay/`

Throughout this note, we consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

4.1 Measurable Functions

Definition 4.1. *A function $X : (\Omega, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable if $\forall\, B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) \triangleq \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$. In probability theory, X is called a random variable or random element.*

Lemma 4.1. *X is a random variable iff $\forall\, t \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \le t\} \in \mathcal{A}$.*

Proof. '$\Rightarrow$': Suppose X is a random variable. Since $(-\infty, t] \in \mathcal{B}$, we have $\{\omega \in \Omega : X(\omega) \le t\} \in \mathcal{A}$.

'$\Leftarrow$': Let $\mathcal{D} = \{B \in \mathcal{B} : X^{-1}(B) \in \mathcal{A}\}$. Since

$$\begin{aligned} X^{-1}(B^c) &= (X^{-1}(B))^c, \\ X^{-1}\left(\bigcup_{i\ge1} B_i\right) &= \bigcup_{i\ge1} X^{-1}(B_i), \end{aligned}$$

$\mathcal{D}$ is a σ-algebra. Since $(-\infty, t] \in \mathcal{D}$ and $\{(-\infty, t]\}$ generates $\mathcal{B}(\mathbb{R})$, we have $\mathcal{B}(\mathbb{R}) \subset \mathcal{D}$ and X is a random variable. □

Lemma 4.2. *Suppose X is a random variable, then $\inf_{n\ge1} X_n$, $\sup_{n\ge1} X_n$, $\limsup_{n\to\infty} X_n$, $\liminf_{n\to\infty} X_n$ are random variables.*

Proof. We show that $\inf_{n\ge1} X_n$ is a random variable. We have for any $t \in \mathbb{R}$,

$$\left\{\inf_{n\ge1} X_n \le t\right\} = \bigcup_{n\ge1} \{X_n \le t\} \in \mathcal{A},$$

since each X_n is a random variable. By Lemma 4.1, the result follows. The proofs for the other claims are similar. □

The law or distribution of a random variable X is given by

$$\begin{aligned} \mathbb{P}_X(B) \triangleq \mathbb{P}(X \in B) &= \mathbb{P}(\{\omega : X(\omega) \in B\}) \\ &= \mathbb{P}\big(X^{-1}(B)\big) \\ &= \mathbb{P} \circ X^{-1}(B), \end{aligned}$$

and we write

$$\begin{aligned} \sigma(X) &= \sigma\text{-algebra generated by } X \\ &= \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}. \end{aligned}$$

Note that the RHS of the last equality is a σ-algebra.

Suppose $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable, i.e., $g^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for all $B \in \mathcal{B}(\mathbb{R})$, then $g(X)$ is a random variable since

$$(g \circ X)^{-1}(B) = X^{-1}(g^{-1}(B)) \in \mathcal{A}.$$

4.2 Expectation

For $A \in \mathcal{A}$, the indicator function for A is given by

$$\mathbf{1}_A(\omega) = \begin{cases} 1, \text{ if } \omega \in A, \\ 0, \text{ otherwise.} \end{cases} \tag{7}$$

Definition 4.2. *$X(\omega)$ is a simple random variable if there is a discrete set of numbers $x_1 < x_2 < \ldots < x_n$, such that $X(\omega) = \sum_{i=1}^n x_i \mathbf{1}_{A_i}(\omega)$, where $A_i \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ for $i \neq j$, and $\bigcup_{i=1}^n A_i = \Omega$.*

Definition 4.3. *The expectation (Lebesgue integral) of a simple random variable (measurable function) $X : \Omega \to \mathbb{R}$ is given by*

$$\int_\Omega X(\omega)\, \mathrm{d}\mathbb{P} = \int_\Omega X(\omega)\mathbb{P}(\,\mathrm{d}\omega) = \mathbb{E}[X] \triangleq \sum_{i=1}^n x_i \mathbb{P}(A_i).$$

Lemma 4.3. *For any measurable partition $B_1, B_2, \ldots, B_n$ of Ω (i.e., for $i \neq j$, $B_i \cap B_j = \emptyset$ and $\bigcup_{i=1}^n B_i = \Omega$), with $X(\omega) = \sum_{j=1}^m b_j \mathbf{1}_{B_j}(\omega)$, then $\mathbb{E}X = \sum_{j=1}^m b_j \mathbb{P}(B_j)$.*

Proof. For $\omega \in A_i \cap B_j$, we have $x_i = X(\omega) = b_j$. Therefore, we obtain

$$\begin{aligned} \sum_{i=1}^n x_i \mathbb{P}(A_i) &= \sum_{i=1}^n x_i \sum_{j=1}^m \mathbb{P}(A_i \cap B_j) \\ &= \sum_i \sum_j x_i \mathbb{P}(A_i \cap B_j) \\ &= \sum_i \sum_j b_j \mathbb{P}(A_i \cap B_j) \\ &= \sum_j b_j \mathbb{P}(B_j), \end{aligned}$$

where the interchange of summations is allowed because the number of terms in the summand is finite. $\square$

The following result follows from Definition 4.3 using similar computation as in Lemma 4.3.

Lemma 4.4. *Suppose X and Y are simple random variables. Then we have:*

(i) $X(\omega) \leq Y(\omega) \implies \mathbb{E}X \leq \mathbb{E}Y$.

(ii) For any $a \in \mathbb{R}$, aX is a simple random variable and $\mathbb{E}[aX] = a\mathbb{E}X$.

(iii) $X + Y$ is a simple random variable and $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$.

Definition 4.4. *If $X \geq 0$, we define*

$$\mathbb{E}X = \sup\{\mathbb{E}Z : Z \leq X,\ Z \text{ is simple}\}.$$

Let $X^+ = \max(X,0) \geq 0$ and $X^- = -\min(X,0) \geq 0$. We have $X = X^+ - X^-$. We can then define $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$, if not both $\mathbb{E}X^+$ and $\mathbb{E}X^-$ are infinite. If $\mathbb{E}X^+$ and $\mathbb{E}X^-$ are both infinite, then the expectation of X is *undefined*.

If $\mathbb{E}|X| = \mathbb{E}X^+ + \mathbb{E}X^- < \infty$, we say that it is integrable. This implies that both $\mathbb{E}X^+$ and $\mathbb{E}X^-$ are finite, hence $\mathbb{E}X$ exists and is finite.

Lemma 4.5. *If* $0 \leq X \leq Y$*, then* $\mathbb{E}X \leq \mathbb{E}Y$*.*

Proof. For any simple r.v. Z, since $\{Z \leq X\} \subset \{Z \leq Y\}$, $\mathbb{E}X \leq \mathbb{E}Y$ follows from Definition 4.4. ☐

Our goal is to prove the following theorem, which shows that expectations are additive.

Theorem 4.1. *Suppose* $X, Y \geq 0$*, then* $\mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y$*.*

We first show the following.

Lemma 4.6. *Suppose* $X, Y \geq 0$*, then* $\mathbb{E}[X+Y] \geq \mathbb{E}X + \mathbb{E}Y$*.*

Proof. $\forall\, \epsilon > 0$, $\exists$ simple random variables $Z_1 \leq X$ and $Z_2 \leq Y$ such that

$$\begin{aligned}\mathbb{E}X &\leq \mathbb{E}Z_1 + \frac{\epsilon}{2},\\ \mathbb{E}Y &\leq \mathbb{E}Z_2 + \frac{\epsilon}{2},\end{aligned}$$

so that

$$\begin{aligned}\mathbb{E}X + \mathbb{E}Y &\leq \mathbb{E}Z_1 + \mathbb{E}Z_2 + \epsilon\\ &= \mathbb{E}[Z_1 + Z_2] + \epsilon\\ &\leq \mathbb{E}[X+Y] + \epsilon,\end{aligned}$$

where the first equality follows from Lemma 4.4 and the last inequality from $Z_1 + Z_2 \leq X + Y$ and Definition 4.4. Since ϵ can be arbitrarily small, the proof is complete. ☐

To complete the proof of Theorem 4.1, we need to show subadditivity: $\mathbb{E}[X+Y] \leq \mathbb{E}X + \mathbb{E}Y$. It turns out that this is highly non-trivial and in the process we learn several nice proof techniques and the very useful theorem below.

Theorem 4.2 (Monotone Convergence Theorem (MCT))**.** *Suppose* $0 \leq X_n \leq X_{n+1}$*, for* $n \geq 1$ *and* $X_n(\omega) \to X(\omega)$ *for all* $\omega \in \Omega$*. Then*

$$\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}\Big[\lim_{n\to\infty} X_n\Big] = \mathbb{E}X.$$

We first give an example why we cannot always interchange the order of integration and limit.

Example 4.1. *Consider the probability space* $((0,1), \mathcal{B}((0,1)), \lambda)$*, where* λ *is the Lebesgue measure, and let*

$$X_n(\omega) = \begin{cases} n, & \text{if } 0 < \omega \leq \dfrac{1}{n},\\ 0, & \text{if } \dfrac{1}{n} < \omega < 1.\end{cases} \tag{8}$$

Then we have $\mathbb{E}X_n = 1$ *while* $\lim_{n\to\infty} X_n(\omega) = 0$ *for all* ω*.*

To prove the MCT, we first show the following.

Lemma 4.7. *Suppose that $X \geq 0$ is a simple random variable and $B_n \subset B_{n+1}$ for $n \geq 1$. Let $B = \bigcup_{n\geq 1} B_n$. We have*

$$\lim_{n\to\infty} \mathbb{E}[X\mathbf{1}_{B_n}] = \mathbb{E}[X\mathbf{1}_B].$$

Proof. Suppose

$$X = \sum_{i=1}^{m} x_i \mathbf{1}_{A_i},$$

then

$$X\mathbf{1}_{B_n} = \sum_{i=1}^{m} x_i \mathbf{1}_{A_i \cap B_n},$$

and from Lemma 4.3, we obtain

$$\mathbb{E}[X\mathbf{1}_{B_n}] = \sum_{i=1}^{m} x_i \mathbb{P}(A_i \cap B_n).$$

Since $\mathbb{E}[X\mathbf{1}_{B_n}] \leq \mathbb{E}\big[X\mathbf{1}_{\{B_{n+1}\}}\big]$ is an increasing sequence in $\mathbb{R}$, its limit exists (can be ∞). Taking limit as $n \to \infty$ on both sides, we obtain

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[X\mathbf{1}_{B_n}] &= \sum_{i=1}^{m} x_i \lim_{n\to\infty} \mathbb{P}(A_i \cap B_n) \\
&= \sum_{i=1}^{m} x_i \mathbb{P}\left(\bigcup_{n\geq 1} \{A_i \cap B_n\}\right) \\
&= \sum_{i=1}^{m} x_i \mathbb{P}(A_i \cap B) \\
&= \mathbb{E}[X\mathbf{1}_B].
\end{aligned}$$

□

Proof of MCT. Since $X_n \leq X$ for all $n \geq 1$, we have $\mathbb{E}X_n \leq \mathbb{E}X$ from Lemma 4.5. As $\mathbb{E}X_n \leq \mathbb{E}X_{n+1}$ is an increasing bounded sequence, its limit exists and

$$\lim_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}X.$$

Let Z be a simple random variable such that $0 \leq Z \leq X$. Let $0 < \rho < 1$, and define

$$B_n = \{\omega : X_n(\omega) \geq \rho Z(\omega)\}.$$

We have $B_n \subset B_{n+1}$. Furthermore, for n sufficiently large, we have $X_n(\omega) \geq \rho X(\omega) \geq \rho Z(\omega)\ \forall\, \omega \in \Omega$ (i.e., $B_n = \Omega$), so that $\bigcup_{j\geq 1} B_j = \Omega$. For such n, we then have

$$\begin{aligned}
&\rho Z\mathbf{1}_{B_n} \leq X_n\mathbf{1}_{B_n} \leq X_m,\ \forall\, m \geq n, \\
&\rho\mathbb{E}[Z\mathbf{1}_{B_n}] \leq \mathbb{E}X_m,
\end{aligned}$$

from Definition 4.4 since $Z\mathbf{1}_{B_n}$ is simple. Letting $m \to \infty$,

$$\rho\mathbb{E}[Z\mathbf{1}_{B_n}] \leq \lim_{m\to\infty} \mathbb{E}X_m,$$

and $n \to \infty$, we have from Lemma 4.7,

$$\begin{aligned} \rho\mathbb{E}\left[Z\mathbf{1}_{\bigcup_{n\geq 1} B_n}\right] &\leq \lim_{m\to\infty} \mathbb{E}X_m, \\ \rho\mathbb{E}[Z] &\leq \lim_{m\to\infty} \mathbb{E}X_m. \end{aligned}$$

Taking sup over all simple $Z \leq X$, we have

$$\rho\mathbb{E}X \leq \lim_{n\to\infty} \mathbb{E}X_n,$$

and $\rho \to 1$ completes the proof. $\square$

The following procedure gives an explicit construction of a sequence of increasing simple random variables that approximate $X \geq 0$. For each $k \geq 1$ and $0 \leq j < 2^{2k}$, let

$$\begin{aligned} B_{k,j} &= \left\{\omega \,:\, \frac{j}{2^k} < X(\omega) \leq \frac{j+1}{2^k}\right\}, \\ B_{k,2^{2k}} &= \left\{\omega \,:\, X(\omega) > 2^k\right\}, \end{aligned}$$

and

$$Z_k = \sum_{j=0}^{2^{2k}} \frac{j}{2^k}\mathbf{1}_{B_{k,j}}.$$

We have $Z_k(\omega) \leq Z_{k+1}(\omega)$ and

$$\begin{aligned} &0 \leq X(\omega) - Z_k(\omega) \leq 2^{-k},\ \forall\omega\ \text{ s.t. }\ X(\omega) \leq 2^k, \\ &Z_k(\omega) = 2^k,\ X(\omega) > 2^k. \end{aligned}$$

Therefore $Z_k(\omega) \uparrow X(\omega)$ as $k \to \infty$. We finally have all the tools to prove Theorem 4.1.

Proof of Theorem 4.1. For any $X, Y \geq 0$, $\exists$ simple $X_n \uparrow X$ and simple $Y_n \uparrow Y$. We then have $X_n + Y_n \uparrow X + Y$ and $\mathbb{E}[X_n + Y_n] = \mathbb{E}X_n + \mathbb{E}Y_n$. From the MCT, taking $n \to \infty$, we obtain $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$. $\square$

Note that additivity of expectations for general random variables follows from Theorem 4.1 since $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$.

The steps in the proof of Theorem 4.1 are quite standard in measure theory:

(i) Prove for simple random variables.

(ii) Extend to non-negative random variables X by using simple random variables $\uparrow X$, then apply MCT.

(iii) Extend to general $X = X^+ - X^-$.

4.3 Fatou's Lemma and the DCT

Lemma 4.8 (Fatou's Lemma)**.** *Suppose that $X_n \geq 0$ for $n \geq 1$. We have*

$$\mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big] \leq \liminf_{n\to\infty} \mathbb{E}X_n.$$

Proof. Let $Y_k = \inf_{n\geq k} X_n \geq 0$. We have $Y_k \leq Y_{k+1}$ and $Y_k \leq X_n$, $\forall\, n \geq k$ so that

$$\begin{aligned}\mathbb{E}Y_k &\leq \inf_{n\geq k} \mathbb{E}X_n \\ \lim_{k\to\infty} \mathbb{E}Y_k &\leq \lim_{k\to\infty} \inf_{n\geq k} \mathbb{E}X_n = \liminf_{n\to\infty} \mathbb{E}X_n.\end{aligned}$$

From MCT and Lemma 1.8, we have

$$\begin{aligned}\lim_{k\to\infty} \mathbb{E}Y_k &= \mathbb{E}\Big[\lim_{k\to\infty} Y_k\Big] \\ &= \mathbb{E}\Big[\lim_{k\to\infty} \inf_{n\geq k} X_n\Big] \\ &= \mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big],\end{aligned}$$

and we obtain

$$\mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big] \leq \liminf_{n\to\infty} \mathbb{E}X_n.$$

□

A similar proof as that for Lemma 4.8 shows that if $X_n \leq Y$ where $\mathbb{E}|Y| < \infty$, then

$$\mathbb{E}\Big[\limsup_{n\to\infty} X_n\Big] \geq \limsup_{n\to\infty} \mathbb{E}X_n.$$

Theorem 4.3 (Dominated Convergence Theorem (DCT))**.** *Suppose $|X_n(\omega)| \leq Y(\omega)\ \forall\, \omega \in \Omega$, and $\mathbb{E}|Y| < \infty$. If $\lim_{n\to\infty} X_n = X$, then $\mathbb{E}X$ exists, $\mathbb{E}|X| \leq \mathbb{E}Y$ and*

$$\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X.$$

Proof. From Lemma 4.5, we have $\mathbb{E}|X_n| \leq \mathbb{E}Y$, and Fatou's Lemma (Lemma 4.8) yields $\mathbb{E}|X| \leq \mathbb{E}Y$. Therefore, $\mathbb{E}X$ exists. Since $Y + X_n \geq 0$, from Fatou's Lemma, we have

$$\mathbb{E}\Big[\liminf_{n\to\infty} (Y + X_n)\Big] \leq \liminf_{n\to\infty} \mathbb{E}[Y + X_n]. \tag{9}$$

We also have

$$\mathbb{E}\Big[\liminf_{n\to\infty} (Y + X_n)\Big] = \mathbb{E}[Y + X] = \mathbb{E}Y + \mathbb{E}X,$$

and

$$\liminf_{n\to\infty} \mathbb{E}[Y + X_n] = \mathbb{E}Y + \liminf_{n\to\infty} \mathbb{E}[X_n],$$

so that (9) yields

$$\mathbb{E}X \leq \liminf_{n\to\infty} \mathbb{E}X_n.$$

Similarly, we have $Y - X_n \geq 0$ and Fatou's Lemma implies that

$$\begin{aligned}
&\mathbb{E}\Big[\liminf_{n\to\infty} (Y - X_n)\Big] \leq \liminf_{n\to\infty} \mathbb{E}[Y - X_n], \\
&\mathbb{E}X \geq \limsup_{n\to\infty} \mathbb{E}X_n.
\end{aligned}$$

Therefore, $\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_n$. □

Lemma 4.9 (Scheffe)**.** *Suppose that $X_n \geq 0$ for $n \geq 1$, $\lim_{n\to\infty} X_n = X$, both X_n and X are integrable, and $\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$. Then, $\lim_{n\to\infty} \mathbb{E}|X_n - X| = 0$.*

Proof. Since $X_n \geq 0$ and $X \geq 0$, we have

$$0 \leq (X_n - X)^- \leq X.$$

As $\mathbb{E}X < \infty$ and $(X_n - X)^- \to 0$ since $X_n \to X$, DCT yields $\mathbb{E}(X_n - X)^- \to 0$. From

$$\mathbb{E}(X_n - X) = \mathbb{E}(X_n - X)^+ - \mathbb{E}(X_n - X)^-,$$

and $\mathbb{E}(X_n - X) \to 0$, we obtain $\mathbb{E}(X_n - X)^+ \to 0$ and

$$\mathbb{E}|X_n - X| = \mathbb{E}(X_n - X)^+ + \mathbb{E}(X_n - X)^- \to 0.$$

□

Hence, any sequence of r.v.s satisfying the conditions of the DCT also converges in L^1.

4.4 Notes

Suppose $f : (\Omega, \mathcal{A}, \mu) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a non-negative measurable function, where μ is a general measure (i.e., does not have the restriction that $\mu(\Omega) = 1$), then the Lebesgue integral of f over a measurable set $A \in \mathcal{A}$ denoted by

$$\int_A f \,\mathrm{d}\mu$$

is defined in exactly the same way as in Definition 4.4. In fact, the proofs of MCT, Fatou's Lemma and DCT do not require that the underlying measure is a probability measure! Therefore, all the results we have seen so far are true for the general Lebesgue integral. If f is a.e. continuous, then its Lebesgue integral and Riemann integral give the same value, but their constructions are different.

For $p \geq 1$, we define

$$L^p(\Omega, \mathcal{A}, \mu) = \left\{ f : (\Omega, \mathcal{A}, \mu) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R})) \,\middle|\, \int_\Omega |f|^p \,\mathrm{d}\mu < \infty \right\}.$$

If obvious from the context, we usually shorten the notation to $L^p(\mu)$ or $L^p(\Omega)$. An integrable function means it is in $L^1(\mu)$. Using Minkowski's inequality, it can be shown that

$$\|f\|_p = \left(\int |f|^p \,\mathrm{d}\mu \right)^{1/p}$$

is a norm, hence $L^p(\mu)$ is a normed space.

If g is a measurable function, then from Definition 4.3, we have

$$\mathbb{E}g(X) = \int g(X(\omega))\,\mathrm{d}\mathbb{P}(\omega).$$

Letting $x = X(\omega)$, we obtain

$$\begin{aligned}\mathbb{E}g(X) &= \int g(x)\,\mathrm{d}\mathbb{P}\circ X^{-1}(x) \\ &= \int g(x)\,\mathrm{d}\mathbb{P}_X(x). \end{aligned} \tag{10}$$

Note that $X^{-1}(B) = \{\omega : X(\omega) \in B\}$ is defined as the inverse map for Borel sets, and is different from a traditional inverse function, which requires bijectiveness. In particular, $X^{-1}(\cdot)$ is always well-defined.

For discrete random variables $X \in \{x_1, x_2, \ldots\}$, suppose $\mathbb{P}_X(\{x_i\}) = p_i$, a discrete measure. Then the expectation (10) reduces to a sum $\sum_i g(x_i)p_i$. For "continuous" random variables you are familiar with in undergrad probability courses, we need a definition.

Definition 4.5. *If $\mu(A) = 0 \implies \nu(A) = 0$ for all measurable sets A, then we say that ν is absolutely continuous w.r.t. μ and write $\nu \ll \mu$.*

A measure μ is σ-finite if $\Omega = \bigcup_{i\geq 1} \Omega_i$, where the Ω_i are disjoint and $\mu(\Omega_i) < \infty$.

Theorem 4.4 (Radon-Nikodym)**.** *Suppose the ν is finite and μ is σ-finite. If $\nu \ll \mu$, then there exists measurable $f \geq 0$ with $\int |f|\,\mathrm{d}\mu < \infty$ (i.e., $f \in L^1(\mu)$) such that*

$$\nu(A) = \int_A f\,\mathrm{d}\mu.$$

Notation-wise, we usually write $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, which is called the Radon-Nikodym derivative.

Definition 4.6. *X is a continuous random variable if $\exists\, f \in L^1(\mathbb{R}, \lambda)$, where λ is the Lebesgue measure, i.e., $\lambda([a, b]) = b - a$, s.t.*

$$\mathbb{P}_X(A) = \int_A f\,\mathrm{d}\lambda = \int_A f(x)\,\mathrm{d}x$$

for all measurable A. The function f is called the probability density function (pdf) of X.

We also have $\mathbb{E}g(X) = \int g(x)\,\mathrm{d}\mathbb{P}_X(x) = \int g(x)f(x)\,\mathrm{d}x$.

5.1 Convergence

Definition 5.1 (Almost sure convergence)**.** *We say $X_n \to X$ almost surely (a.s.) or with probability 1 if*

$$\mathbb{P}\Big(\Big\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}\Big) = 1.$$

Recall that in the original MCT in Theorem 4.2, we suppose that $0 \le X_n(\omega) \le X_{n+1}(\omega)$ and $X_n(\omega) \to X(\omega)$ for *all* $\omega \in \Omega$. Here, we show that we can replace the condition "for all $\omega \in \Omega$" with "almost surely". Let

$$A = \{\omega \,:\, X_n(\omega) \to X(\omega),\, 0 \le X_n(\omega) \le X_{n+1}(\omega),\, n \ge 1\}$$

and suppose $\mathbb{P}(A) = 1$. We have $0 \le X_n \mathbf{1}_A(\omega) \le X_{n+1}\mathbf{1}_A(\omega)$, and $X_n\mathbf{1}_A(\omega) \to X\mathbf{1}_A(\omega)\ \forall\, \omega \in \Omega$. Then from the MCT in Theorem 4.2, we have

$$\mathbb{E}[X_n\mathbf{1}_A] \to \mathbb{E}[X\mathbf{1}_A].$$

For any $Y \ge 0$, let $Y \wedge n = \min(Y, n)$ for $n \ge 1$. Observe that

$$\begin{aligned}\mathbb{E}[Y\mathbf{1}_{A^c}] &= \mathbb{E}\Big[\lim_{n\to\infty} (Y \wedge n)\mathbf{1}_{A^c}\Big]\\ &= \lim_{n\to\infty} \mathbb{E}[(Y\wedge n)\mathbf{1}_{A^c}] \qquad (11)\\ &\le \lim_{n\to\infty} n\mathbb{P}(A^c)\\ &= 0,\end{aligned}$$

where we have use the MCT in (11). Therefore, $\mathbb{E}[X_n\mathbf{1}_A] = \mathbb{E}X_n$ and $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}X$, and the MCT holds if "$\forall\,\omega \in \Omega$" is replaced with "almost surely" in its theorem statement. The same thing applies for Fatou's Lemma and the DCT.

Definition 5.2 (Convergence in probability)**.** *If $\forall\,\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0,$$

we say that X_n converges to X in probability and write $X_n \xrightarrow{\text{p}} X$.

Lemma 5.1. *If $X_n \to X$ a.s., then $X_n \xrightarrow{\text{p}} X$.*

Proof. Suppose $X_n \to X$ a.s., then $\mathbf{1}_{\{|X_n - X| > \epsilon\}} \to 0$ a.s. From DCT (since probability measure is finite), we have $\mathbb{P}(|X_n - X| > \epsilon) \to 0$. □

The converse is not true and here is an example.

Example 5.1. *Suppose $\Omega = [0, 1]$. Consider the random variables X_n shown in Fig. 1, where X_n follows a similar pattern for $n \ge 5$. We have $X_n \xrightarrow{\text{p}} 0$, but clearly $X_n \nrightarrow 0$ a.s. as $X_n(\omega) = 1$ for infinitely many values of n for every $\omega \in \Omega$.*

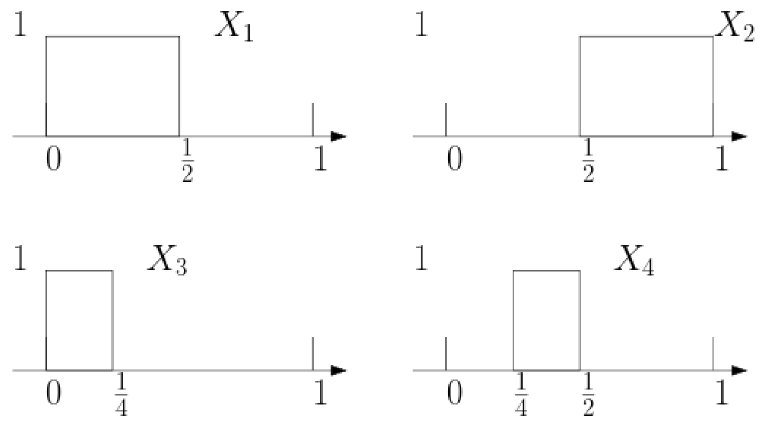

Figure 1: X_n converges in probability but not almost surely.

5.2 Weak Law of Large Numbers

Markov's inequality: Suppose $a > 0$ and $X \geq 0$. Then,

$$a\mathbf{1}_{\{X \geq a\}}(\omega) \leq X\mathbf{1}_{\{X \geq a\}}(\omega) \leq X.$$

Taking expectations, we have

$$\begin{aligned} a\mathbb{P}(X \geq a) &\leq \mathbb{E}X, \\ \mathbb{P}(X \geq a) &\leq \frac{1}{a}\mathbb{E}X. \end{aligned}$$

Chebyshev's inequality follows from Markov's inequality by replacing X with $|X - \mathbb{E}X|^2$ and setting $a = \epsilon^2$: for $\epsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq \epsilon) \leq \frac{1}{\epsilon^2}\,\mathrm{var}(X).$$

Theorem 5.1 (Weak Law of Large Numbers (WLLN))**.** *Suppose $X_1, X_2, \ldots$ are such that $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = \sigma^2 < \infty$ and $\mathbb{E}[X_iX_j] \leq 0$ for $i \neq j$. Then,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{p}} 0 \text{ as } n \to \infty.$$

Proof. We have

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] &= \frac{1}{n^2}\mathbb{E}\left[\sum_{i}^{n} X_i^2 + 2\sum_{i<j} X_iX_j\right] \\ &\leq \frac{1}{n^2}\sum_{i}^{n}\mathbb{E}X_i^2 = \frac{\sigma^2}{n}. \end{aligned}$$

From Chebyshev's inequality, we then have for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2}\frac{\sigma^2}{n} \to 0,$$

as $n \to \infty$. □

5.3 Product Measures

Given two measure spaces $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$, where both $\mu_1$and μ_2 are σ-finite, we can define a new measurable space $(\Omega, \mathcal{F})$, where $\Omega = \Omega_1 \times \Omega_2$ and

$$\mathcal{F} = \sigma\left\{A \times B : A \in \mathcal{A}_1, B \in \mathcal{A}_2\right\}.$$

We call $A \times B$ a rectangle. A natural measure μ for this measurable space satisfies

$$\mu(A \times B) = \mu_1(A)\mu_2(B) \tag{12}$$

for all rectanges $A \times B$. It can be checked that the collection of finite disjoint unions of rectangles is an algebra (exercise). To extend this measure μ to $\mathcal{F}$, we make use of Caratheodory's Extension Theorem (Theorem 3.1): we show that for $A \times B = \bigcup_{i\geq 1} A_i \times B_i$ a disjoint union of rectangles, we have $\mu(A \times B) = \sum_{i\geq 1} \mu_1(A_i)\mu_2(B_i)$.

For $x \in A$, let $I(x) = \{i : x \in A_i\}$ and $B = \bigcup_{i\in I(x)} B_i$ a disjoint union. We have

$$\begin{aligned}
\mathbf{1}_A(x)\mu_2(B) &= \mathbf{1}_A(x)\mu_2\left(\bigcup_{i\in I(x)} B_i\right) \\
&= \sum_{i\in I(x)} \mathbf{1}_A(x)\mu_2(B_i) \\
&= \sum_{i\geq 1} \mathbf{1}_{A_i}(x)\mu_2(B_i).
\end{aligned}$$

Integrating w.r.t. μ_1, we have

$$\mu(A \times B) = \int \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{1}_{A_i}(x)\mu_2(B_i)\,\mathrm{d}\mu_1$$

$$stackrel\text{MCT}= \lim_{n\to\infty} \int \sum_{i=1}^{n} \mathbf{1}_{A_i}(x)\mu_2(B_i)\,\mathrm{d}\mu_1$$

$$\begin{aligned}
&= \lim_{n\to\infty} \sum_{i=1}^{n} \int \mathbf{1}_{A_i}(x)\mu_2(B_i)\,\mathrm{d}\mu_1 \\
&= \sum_{i\geq 1} \mu_1(A_i)\mu_2(B_i),
\end{aligned}$$

where the interchange of the sum and integral in the penultimate equality holds because the number of terms is finite. Therefore, Caratheodory's Extension Theorem shows that there is a unique extension of μ defined by (12) to $\mathcal{F}$. This is called a *product measure*. Notationally, we write $\mu = \mu_1 \times \mu_2$.

The next theorem tells us when we can interchange integrals in general.

Theorem 5.2 (Fubini or Fubini-Tonelli)**.** *Suppose* $(\Omega_1, \mathcal{A}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{A}_2, \mu_2)$ *are* σ*-finite and* $\mu = \mu_1 \times \mu_2$ *is the product measure. Consider a measurable function* $f : \Omega = \Omega_1 \times \Omega_2 \mapsto \mathbb{R}$*. If* $f \geq 0$ *or* $\int |f|\,\mathrm{d}\mu < \infty$*, then*

$$\int_{\Omega_1}\int_{\Omega_2} f(x, y)\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \int_{\Omega} f\,\mathrm{d}\mu = \int_{\Omega_2}\int_{\Omega_1} f(x, y)\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2.$$

Proof. Note that implicit in the theorem statement are the following that we need to prove:

(i) For each $x \in \Omega_1$, $y \mapsto f(x, y)$ is $\mathcal{A}_2$-measurable.

(ii) $x \mapsto \int_{\Omega_2} f(x,y)\,\mathrm{d}\mu_2$ is $\mathcal{A}_1$-measurable.

(iii) $\int_{\Omega_1}\int_{\Omega_2} f(x,y)\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \int_\Omega f\,\mathrm{d}\mu$.

Without loss of generality, we may assume $\mu_1, \mu_2 < \infty$ as the same proof is valid on each partition of Ω and then we can apply the MCT.

The proof follows the steps discussed at the end of Section 4.2. We first prove the theorem for the simplest case where $f = \mathbf{1}_E$, where $E \in \mathcal{F}$, the product σ-algebra. Let $E_x = \{y : (x,y) \in E\}$.

(i): Fix x, then $y \mapsto f(x,y) = \mathbf{1}_{E_x}(y)$. We need to show $E_x \in \mathcal{A}_2$. Let $\mathcal{E} = \{E \in \mathcal{F} : E_x \in \mathcal{A}_2\}$. We have $(E^c)_x = (E_x)^c$ since $y \in (E^c)_x \Leftrightarrow (x,y) \in E^c \Leftrightarrow y \in (E_x)^c$, and $\left(\bigcup_{i\geq 1} E_i\right)_x = \bigcup_{i\geq 1}(E_i)_x$. Therefore, $\mathcal{E}$ is a σ-algebra and it contains all rectangles $A \times B$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$, which implies that $\mathcal{F} \subset \mathcal{E}$.

(ii) & (iii): Let $\mathcal{L} = \{E \in \mathcal{F} \,:\, f = \mathbf{1}_E \text{ satisfies (ii) \&(iii)}\}$.

- $\Omega \in \mathcal{L}$.
- If $E \in \mathcal{L}$, $\mu_2((E^c)_x) = \mu_2((E_x)^c) = \mu_2(\Omega_2) - \mu_2(E_x)$. Since $\mu_2(\Omega_2) < \infty$ and $\mu_2(E_x)$ is $\mathcal{A}_1$-measurable, $\mu_2((E^c)_x)$ is $\mathcal{A}_1$-measurable. We also have

$$\begin{aligned}\int \mu_2((E^c)_x)\,\mathrm{d}\mu_1 &= \mu_2(\Omega_2)\mu_1(\Omega_1) - \int \mu_2(E_x)\,\mathrm{d}\mu_1\\ &= \mu(\Omega) - \mu(E)\\ &= \mu(E^c).\end{aligned}$$

  Therefore, $E^c \in \mathcal{L}$.
- If $E_i \in \mathcal{L}$, $i \geq 1$, are disjoint, then

$$\begin{aligned}\mu_2\left((\bigcup_{i\geq 1} E_i)_x\right) &= \mu_2\left(\bigcup_{i\geq 1}(E_i)_x\right)\\ &= \sum_{i\geq 1}\mu_2((E_i)_x)\\ &= \lim_{n\to\infty}\sum_{i=1}^n \mu_2((E_i)_x),\end{aligned}$$

  which is $\mathcal{A}_1$-measurable from Lemma 4.2 since each $\mu_2((E_i)_x)$ is $\mathcal{A}_1$-measurable. From the MCT, we have

$$\begin{aligned}\int \mu_2\left((\bigcup_{i\geq 1} E_i)_x\right)\mathrm{d}\mu_1 &= \lim_{n\to\infty}\int\sum_{i=1}^n \mu_2((E_i)_x)\,\mathrm{d}\mu_1\\ &= \lim_{n\to\infty}\sum_{i=1}^n\int \mu_2((E_i)_x)\,\mathrm{d}\mu_1\\ &= \lim_{n\to\infty}\sum_{i=1}^n \mu(E_i)\\ &= \mu\left(\bigcup_{i\geq 1} E_i\right).\end{aligned}$$

Therefore, $\mathcal{L}$ is a λ-system containing the collection of rectangles, which is a π-system. From the π-λ Theorem, we obtain $\mathcal{F} \subset \mathcal{L}$. We have now shown that Fubini's Theorem holds for $f = \mathbf{1}_E$, $E \in \mathcal{F}$.

From the linearity of integrals, the theorem holds for all simple functions f.

For $f \geq 0$, $\exists$ simple $f_i \uparrow f$. Applying MCT gives Fubini's theorem for non-negative f.

Finally, for general $f = f^+ - f^-$, we note that $\int |f| \, \mathrm{d}\mu < \infty$ implies $\int f^+ \, \mathrm{d}\mu, \int f^- \, \mathrm{d}\mu < \infty$, and

$$\begin{aligned} \int f^+ \, \mathrm{d}\mu &= \int \int f^+(x,y) \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2 \implies \int f^+(x,y) \, \mathrm{d}\mu_1 < \infty \ \mu_2\text{-a.e.} \\ &= \int \int f^+(x,y) \, \mathrm{d}\mu_2 \, \mathrm{d}\mu_1 \implies \int f^+(x,y) \, \mathrm{d}\mu_2 < \infty \ \mu_1\text{-a.e.} \end{aligned}$$

Similarly for f^- so that

$$\int \int f^+(x,y) \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2 - \int \int f^-(x,y) \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2 = \int \int f(x,y) \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2.$$

The proof is now complete. $\square$

Example 5.2.

$$\begin{array}{cc} & \begin{array}{ccccc} x & \longrightarrow & & & \end{array} \\ y \big\uparrow & \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 1 & -1 & \cdots \\ 0 & 1 & -1 & 0 & \cdots \\ 1 & -1 & 0 & 0 & \cdots \end{pmatrix} \end{array} \tag{13}$$

We have

$$\begin{aligned} \sum_y \sum_x f(x,y) &= \sum_y 0 = 0. \\ \sum_x \sum_y f(x,y) &= 1 + 0 + 0 + \cdots = 1. \end{aligned}$$

This example shows that the conditions in Fubini's Theorem are essentially necessary.

Example 5.3. *Suppose* $((0,1), \mathcal{B}(0,1), \lambda) \times ((0,1), 2^{(0,1)}, \nu)$, *where* $\nu(A) = |A|$ *is the counting measure, which is not* σ*-finite. Let*

$$f(x,y) = \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Then we have

$$\begin{aligned} \int f(x,y) \, \mathrm{d}\nu = 1 \text{ for each } x &\implies \int \int f(x,y) \, \mathrm{d}\nu \, \mathrm{d}\lambda = 1, \\ \int f(x,y) \, \mathrm{d}\lambda = 0 \text{ for each } y &\implies \int \int f(x,y) \, \mathrm{d}\lambda \, \mathrm{d}\nu = 0. \end{aligned}$$

This example shows that σ*-finiteness of the measures is necessary.*

5.4 Independence

Throughout this section, we consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 5.3. *Two events* A *and* B *are independent if* $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. *We write* $A \perp\!\!\!\perp B$.

If $A \perp\!\!\!\perp B$, then it is easy to verify the following:

(i) $A^c \perp\!\!\!\perp B$.

(ii) $A \perp\!\!\!\perp B^c$.

(iii) $A^c \perp\!\!\!\perp B^c$.

I.e., the two σ-algebras $\{\emptyset, \Omega, A, A^c\}$ and $\{\emptyset, \Omega, B, B^c\}$ are "independent".

Definition 5.4. *The sub-σ-algebra $\mathcal{A}_i \subset \mathcal{A}$, $i = 1, 2, \ldots, n$, are independent if $\forall\, A_i \in \mathcal{A}_i$,*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

They are said to be pairwise independent *if $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)\ \forall\, i \neq j$.*

We say that the random variables $X_1, X_2, \ldots, X_n$ are independent if $\sigma(X_i)$, $i = 1, \ldots, n$, are independent, i.e., $\forall\, B_i \in \mathcal{B}$, $i = 1, \ldots, n$,

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

Note that independence $\implies$ pairwise independence but the converse is false. Here are two counterexamples to show that pairwise independence does not imply independence.

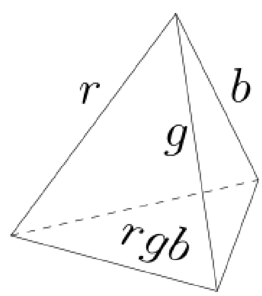

Figure 2: Dice example: pairwise independence does not imply independence.

Example 5.4. *Consider a fair tetrahedron dice that has one red edge, one green edge and one blue edge as shown in Fig. 2. The bottom has all three colors. Let r be the event that the dice when tossed lands on a face with a red boundary. The events b, g and rgb are defined similarly. Then, by symmetry, we have*

$$\mathbb{P}(r) = \mathbb{P}(g) = \mathbb{P}(b) = \frac{1}{2},$$
$$\mathbb{P}(rgb) = \mathbb{P}(rg) = \mathbb{P}(gb) = \mathbb{P}(rb) = \frac{1}{4}.$$

Therefore, these events are pairwise independent but not independent.

Example 5.5. *Consider two six-sided fair dice. Let*

$$A_1 = \{\textit{1st dice is odd}\},$$
$$A_2 = \{\textit{2nd dice is odd}\},$$
$$A_{sum} = \{\textit{sum of the two dice is odd}\}.$$

We have

$$\mathbb{P}(A_1) = \mathbb{P}(A_1) = \mathbb{P}(A_{sum}) = \frac{1}{2},$$
$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_{sum}) = \mathbb{P}(A_2 \cap A_{sum}) = \frac{1}{4},$$
$$\mathbb{P}(A_1 \cap A_2 \cap A_{sum}) = 0.$$

Therefore, the events are pairwise independent but not independent. In particular, $\sigma(A_1)$ is not independent with $\sigma(\{A_2, A_{sum}\})$.

Lemma 5.2. *Suppose the two collections of subsets $\mathcal{E}$ and $\mathcal{C}$ are π-systems and $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, $\forall\, B \in \mathcal{E}$, $C \in \mathcal{C}$. Then $\sigma(\mathcal{E})$ and $\sigma(\mathcal{C})$ are independent.*

Proof. Let $\mathcal{D}_1 = \{D \in \mathcal{A} : \mathbb{P}(D \cap C) = \mathbb{P}(D)\mathbb{P}(C), \forall\, C \in \mathcal{C}\}$. As an exercise, one can check that $\mathcal{D}_1$ is a λ-system. Since $\mathcal{E} \subset \mathcal{D}_1$, from the π-λ theorem we have $\sigma(\mathcal{E}) \subset \mathcal{D}_1$.

Let $\mathcal{D}_2 = \{D \in \mathcal{A} : \mathbb{P}(B \cap D) = \mathbb{P}(B)\mathbb{P}(D), \forall\, B \in \sigma(\mathcal{E})\}$. Similarly $\mathcal{D}_2$ is a λ-system. From above, $\mathcal{C} \subset \mathcal{D}_2$ and by the π-λ theorem, we have $\sigma(\mathcal{C}) \subset \mathcal{D}_2$. Therefore, $\sigma(\mathcal{E}) \perp\!\!\!\perp \sigma(\mathcal{C})$. ☐

By induction, if $\mathcal{B}_i$ for $i = 1, \ldots, n$ are π-systems and are independent, then $\sigma(\mathcal{B}_i)$ are independent.

Corollary 5.1. *The random variables $X_1, X_2, \ldots, X_n$ are independent if*

$$\mathbb{P}(X_1 \le t_1, X_2 \le t_2, \ldots, X_n \le t_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \le t_i).$$

Proof. Since $\{(-\infty, t] : t \in \mathbb{R}\}$ is a π-system that generates $\mathcal{B}(\mathbb{R})$, the corollary follows from Lemma 5.2. ☐

Lemma 5.3. *Suppose that each random variable X_i has pdf f_i, $i = 1, \ldots, n$. Then $X_1, \ldots, X_n$ are independent iff $\exists$ a joint pdf $f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(X_i)$.*

Proof. '$\Leftarrow$':

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \in A_i\}\right) &= \int_{A_1 \times \cdots \times A_n} f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \int_{A_1 \times \cdots \times A_n} \prod_{i=1}^{n} f(x_i)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \prod_{i=1}^{n} \int_{A_i} f(x_i)\,\mathrm{d}x_i \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).
\end{aligned}$$

'$\Rightarrow$': Let $X = (X_1, \ldots, X_n)$. For $A_i \in \mathcal{B}$, $i = 1, \ldots n$, we are given

$$\begin{aligned}
\mathbb{P}(X \in A_1 \times \cdots \times A_n) &= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) \\
&= \prod_{i=1}^{n} \int_{A_i} f_i(x_i)\,\mathrm{d}x_i \\
&= \int_{A_1 \times \cdots \times A_n} \prod_{i=1}^{n} f_i(x_i)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n. \qquad (15)
\end{aligned}$$

We want to show that

$$\mathbb{P}(X \in A) = \int_A \prod_{i=1}^{n} f_i(x_i)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

for all A in the product σ-algebra $\mathcal{B}(\mathbb{R}^n) = \sigma\{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}\}$.

Let $\mathcal{L} = \{A \in \mathcal{E} : \mathbb{P}(X \in A) = \int_A \prod_{i=1}^{n} f_i(x_i)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n\}$. It can be checked that $\mathcal{L}$ is a λ-system and $\mathcal{P} = \{A_1 \times \cdots \times A_n : A_i \subset \mathcal{B}\}$ is a π-system and generates $\mathcal{B}(\mathbb{R}^n)$. Since $\mathcal{P} \subset \mathcal{L}$ from (15), by the π-λ Theorem, we have $\sigma(\mathcal{P}) \subset \mathcal{L}$ and the proof is complete. $\square$

Lemma 5.4. *If $X \perp\!\!\!\perp Y$ and are integrable, then $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$.*

Proof. The distribution of (X, Y) on $\mathbb{R}^2$ is the product measure $\mathbb{P}_X \times \mathbb{P}_Y$. From Fubini's theorem, we have

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy\, \mathrm{d}\mathbb{P}_X \times \mathbb{P}_Y(x, y) = \int_{\mathbb{R}} x\, \mathrm{d}\mathbb{P}_X(x) \int_{\mathbb{R}} y\, \mathrm{d}\mathbb{P}_Y(y) = \mathbb{E}X\mathbb{E}Y.$$

$\square$

Suppose $X \perp\!\!\!\perp Y$. For any measurable functions f and g, $f(X) \perp\!\!\!\perp g(Y)$ since $\sigma(f(X)) \subset \sigma(X)$.

An Analytical Introduction to Probability Theory
6. Borel-Cantelli Lemmas
SIGNAL, NTU `https://personal.ntu.edu.sg/wptay/`

6.1 Introduction

Let $A_1, A_2, \ldots \in \mathcal{A}$ be an infinite sequence of events. Let $N(\omega) = \sum_{i=1}^{\infty} \mathbf{1}_{A_i}(\omega)$. The set of sample points $\omega \in \Omega$ that belong to events in $\{A_1, A_2, \ldots\}$ infinitely often (i.o.) is given by

$$\{A_n \text{ i.o.}\} \triangleq \{\omega : N(\omega) = \infty\} = \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m \triangleq \limsup_{n\to\infty} A_n.$$

Note that $\mathbf{1}_{\{A_n \text{ i.o.}\}} = \limsup_{n\to\infty} \mathbf{1}_{A_n}$.

The set of sample points that belong finitely often (f.o.) to the events in the sequence is

$$\{A_n \text{ f.o.}\} \triangleq \{\omega : N(\omega) < \infty\} = \bigcup_{n\geq 1} \bigcap_{m\geq n} A_m^c \triangleq \liminf_{n\to\infty} A_n^c.$$

Similarly, $\mathbf{1}_{\{A_n \text{ f.o.}\}} = \liminf_{n\to\infty} \mathbf{1}_{A_n^c}$.

Notice that $\{A_n \text{ i.o.}\}$ and $\{A_n \text{ f.o.}\}$ are complements of each other. Thus, $\mathbb{P}(A_n \text{ i.o.}) + \mathbb{P}(A_n \text{ f.o.}) = 1$.

As an example, suppose $X_n(\omega) \to 0$ for all $\omega \in \Omega$, then

$$\exists\, n_0(\omega),\ \text{s.t.}\ X_n(\omega) \leq 1, \forall\, n \geq n_0(\omega).$$

Therefore $\omega \in \{X_n \geq 1\}$ cannot be infinitely often (i.o.).

6.2 Borel-Cantelli Lemmas

Lemma 6.1 (Borel-Cantelli Lemmas)**.**

(i) If $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$.

(ii) If $A_1, A_2, \ldots$ are independent, and $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof.

(i) Let $B_n = \bigcup_{m\geq n} A_m$, then $B_{n+1} \subset B_n$. We have

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\left(\bigcap_{n\geq 1} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) \leq \lim_{n\to\infty} \sum_{m\geq n} \mathbb{P}(A_m),$$

where the second equality and last inequality follow from Lemma 3.1 and Lemma 3.3, respectively. Since $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$, we have $\lim_{n\to\infty} \sum_{m\geq n} \mathbb{P}(A_m) = 0$.

(ii) Let $C_n = \bigcap_{m\geq n} A_m^c$. From independence, we have

$$\begin{aligned}
\mathbb{P}(C_n) = \prod_{m\geq n} \mathbb{P}(A_m^c) &= \prod_{m\geq n} (1 - \mathbb{P}(A_m)) \\
&\leq \prod_{m\geq n} e^{-\mathbb{P}(A_m)} \quad \big(\text{using } 1 - p \leq e^{-p}\big) \\
&= \exp\left(-\sum_{m\geq n} \mathbb{P}(A_m)\right) = 0.
\end{aligned}$$

Therefore, we obtain

$$\mathbb{P}(A_n \text{ f.o.}) = \mathbb{P}\left(\bigcup_{n\geq 1} C_n\right) \leq \sum_{n\geq 1} \mathbb{P}(C_n) = 0 \implies \mathbb{P}(A_n \text{ i.o.}) = 1.$$

$\square$

We can strengthen the second Borel-Cantelli Lemma as follows.

Lemma 6.2. *If $A_1, A_2, \ldots$ are pairwise independent, and $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.*

Proof. Let $N_n = \sum_{k=1}^n \mathbf{1}_{A_k}$. We have

$$\begin{aligned}
\mathbb{E}N_n &= \sum_{k=1}^n \mathbb{P}(A_k), \\
\operatorname{var}(N_n) &= \sum_{k=1}^n \mathbb{P}(A_k)\,(1 - \mathbb{P}(A_k)) \\
&\leq \mathbb{E}N_n. \qquad (16)
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\mathbb{P}\left(N_n \leq \frac{1}{2}\mathbb{E}N_n\right) &\leq \mathbb{P}\left(|N_n - \mathbb{E}N_n| \geq \frac{1}{2}\mathbb{E}N_n\right) \\
&\leq \frac{4}{(\mathbb{E}N_n)^2} \operatorname{var}(N_n) \quad \text{from Chebyshev's inequality} \\
&\leq \frac{4}{\mathbb{E}N_n}.
\end{aligned}$$

Since $N_n \leq N = \sum_{k\geq 1} \mathbf{1}_{A_k}$, we obtain

$$\mathbb{P}\left(N \leq \frac{1}{2}\mathbb{E}N_n\right) \leq \mathbb{P}\left(N_n \leq \frac{1}{2}\mathbb{E}N_n\right) \leq \frac{4}{\mathbb{E}N_n}.$$

Moreover, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}N_n = \lim_{n\to\infty} \sum_{k=1}^n \mathbb{P}(A_k) = \infty, \\
\implies &\lim_{n\to\infty} \mathbb{P}\left(N \leq \frac{1}{2}\mathbb{E}N_n\right) \leq \lim_{n\to\infty} \frac{4}{\mathbb{E}N_n} = 0.
\end{aligned}$$

We claim that

$$\mathbf{1}_{\{N \le \frac{1}{2}\mathbb{E}N_n\}}(\omega) \to \mathbf{1}_{\{N<\infty\}}(\omega),$$

since if $N(\omega) < \infty$, then for n sufficiently large, $\mathbf{1}_{\{N \le \frac{1}{2}\mathbb{E}N_n\}}(\omega) = 1$, otherwise $\mathbf{1}_{\{N \le \frac{1}{2}\mathbb{E}N_n\}}(\omega) = 0$ for all $n \ge 1$. Therefore from DCT, we have

$$\mathbb{P}(A_n \text{ f.o}) = \mathbb{P}(N < \infty) = \lim_{n\to\infty} \mathbb{P}\left(N \le \frac{1}{2}\mathbb{E}N_n\right) = 0,$$

and

$$\mathbb{P}(A_n \text{ i.o}) = 1.$$

□

Remark 6.1. *The condition of pairwise independence in Lemma 6.2 can be strengthened to* $\mathbb{P}(A_i \cap A_j) \le \mathbb{P}(A_i)\mathbb{P}(A_j), \forall\ i \ne j$ *since* (16) *still holds under this condition.*

In the following, we first prove a bound that will be useful in further generalizing the second Borel-Cantelli Lemma.

Lemma 6.3 (Second moment method)**.** *For* $0 \le \rho < 1$ *and* $X \ge 0$ *with* $\mathbb{E}X < \infty$*,*

$$\mathbb{P}(X > \rho\mathbb{E}X) \ge (1-\rho)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}.$$

Proof. Let $A = \{X > \rho\mathbb{E}X\}$. We have

$$\begin{aligned}\mathbb{E}X = \mathbb{E}X\mathbf{1}_A + \mathbb{E}X\mathbf{1}_{A^c} \le \mathbb{E}X\mathbf{1}_A + \rho\mathbb{E}X,\\ (1-\rho)\mathbb{E}X \le \mathbb{E}X\mathbf{1}_A.\end{aligned}$$

From the Cauchy-Schwarz inequality,

$$(1-\rho)^2(\mathbb{E}X)^2 \le (\mathbb{E}X\mathbf{1}_A)^2 \le \mathbb{E}X^2\mathbb{P}(A),$$

and the result follows. □

Lemma 6.4 (Kochen-Stone)**.** *If* $\sum_{n\ge1}\mathbb{P}(A_n) = \infty$*, then*

$$\mathbb{P}(A_n \text{ i.o.}) \ge \limsup_{n\to\infty} \frac{\left(\sum_{k=1}^n \mathbb{P}(A_k)\right)^2}{\sum_{i,j=1}^n \mathbb{P}(A_i \cap A_j)}.$$

Proof. Let $N_n = \sum_{k=1}^n \mathbf{1}_{A_k}$. We have

$$\begin{aligned}\mathbb{E}N_n &= \sum_{i=1}^n \mathbb{P}(A_i),\\ \mathbb{E}N_n^2 &= \sum_{i,j=1}^n P(A_i \cap A_j).\end{aligned}$$

Let $0 < \rho < 1$. Because $\lim_{n\to\infty}\mathbb{E}N_n = \lim_{n\to\infty}\sum_{i=1}^n \mathbb{P}(A_i) = \infty$, we have $\{A_n \text{ f.o.}\} \subset \{N_n \le \rho\mathbb{E}N_n,\ \forall\, n \ge n_0,\ \text{for some } n_0 \ge 1\}$. Therefore, $\{A_n \text{ i.o.}\} \supset \{N_n > \rho\mathbb{E}N_n \text{ i.o.}\}$ and

$$\begin{aligned}\mathbb{P}(A_n \text{ i.o.}) &\ge \mathbb{P}(N_n > \rho\mathbb{E}N_n \text{ i.o.})\\ &\ge \limsup_{n\to\infty}\mathbb{P}(N_n > \rho\mathbb{E}N_n) \text{ (from Fatou's Lemma)}\\ &\ge \limsup_{n\to\infty}(1-\rho)^2\frac{(\mathbb{E}N_n)^2}{\mathbb{E}N_n^2} \text{ (from Lemma 6.3)}.\end{aligned}$$

Taking $\rho \to 0$, we obtain

$$\mathbb{P}(A_n \text{ i.o.}) \geq \limsup_{n\to\infty} \frac{\left(\sum_{k=1}^n \mathbb{P}(A_k)\right)^2}{\sum_{i,j=1}^n \mathbb{P}(A_i \cap A_j)}.$$

$\square$

Lemma 6.5. *For $X_1, X_2, \ldots,$ s.t. $\sum_{n\geq 1} \mathbb{P}(|X_n| \geq \epsilon) < \infty, \forall\ \epsilon > 0$, we have*

$$X_n \to 0 \text{ a.s. as } n \to \infty.$$

Proof. Let

$$\begin{aligned} F &= \left\{\omega \,:\, \limsup_{n\to\infty} |X_n| > 0\right\} \\ &= \bigcup_{m\geq 1} \left\{\omega \,:\, \limsup_{n\to\infty} |X_n| > \frac{1}{m}\right\}. \end{aligned}$$

Let $A_n = \{\omega : |X_n| > \frac{1}{m}\}$. We have $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$. From the first Borel-Cantelli Lemma, we have

$$\mathbb{P}(A_n \text{ i.o.}) = 0.$$

Then, $\mathbb{P}\big(\limsup_{n\to\infty} |X_n| > \frac{1}{m}\big) \leq \mathbb{P}(A_n \text{ i.o.}) = 0$.

$$\begin{aligned} &\implies \mathbb{P}(F) = 0 \\ &\implies \limsup_{n\to\infty} |X_n| = 0 \text{ a.s.} \implies \lim_{n\to\infty} |X_n| = 0 \text{ a.s.}. \end{aligned}$$

$\square$

Corollary 6.1. *If $X_n \xrightarrow{\text{p}} X$, then $\exists$ subsequence $(n(k))_{k\geq 1}$ such that $X_{n(k)} \to X$ a.s.*

Proof. By the definition of convergence in probability, we can choose $(n(k))_{k\geq 1}$ such that $\forall\, \epsilon > 0$, we have

$$\mathbb{P}\big(|X_{n(k)} - X| \geq \epsilon\big) \leq \frac{1}{2^k},\ \forall\, k \geq 1.$$

Summing both sides over $k \geq 1$, we obtain

$$\sum_{k\geq 1} \mathbb{P}\big(|X_{n(k)} - X| \geq \epsilon\big) \leq 1.$$

By Lemma 6.5, we have $|X_{n(k)} - X| \to 0$ a.s. $\square$

Lemma 6.6. $X_n \xrightarrow{\text{p}} X$ *iff for any subsequence $(n(k))_{k\geq 1}$, $\exists$ subsubsequence $(n\,(k(r)))_{r\geq 1}$,* s.t. $X_{n(k(r))} \to X$ *a.s.*

Proof. '$\Rightarrow$': It is obvious by Corollary 6.1.

'$\Leftarrow$': Suppose X_n does not converge in probability to X. Then, $\exists\, \epsilon > 0$ and subsequence $(n(k))_{k\geq 1}$, such that

$$\mathbb{P}\big(|X_{n(k)} - X| \geq \epsilon\big) \geq \epsilon,\ \forall\, k \geq 1.$$

Consequently, $\forall\ (n\,(k(r)))_{r\geq 1}$, $X_{n(k(r))} \not\to X$ a.s., which contradicts the claim. $\square$

Note that Lemma 6.6 implies that the DCT holds with "almostly surely convergence" replaced by "convergence in probability".

6.3 SLLN with Finite 2nd Moments

Lemma 6.7. *Suppose $X_1, X_2, \ldots$ are pairwise independent, $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 \le M < \infty$, $\forall\, i \ge 1$. Let $S_n = \sum_{i=1}^n X_i$. Then $\frac{S_n}{n} \to 0$ a.s. as $n \to \infty$.*

Proof. From Lemma 6.5, it suffices to prove $\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \epsilon \text{ i.o.}\right) = 0$, $\forall\, \epsilon > 0$. By applying Chebyshev's inequality, we obtain

$$\mathbb{P}\left(\frac{|S_n|}{n} \ge \epsilon\right) \le \frac{\mathbb{E}S_n^2}{\epsilon^2 n^2} \le \frac{M}{\epsilon^2 n}.$$

Unfortunately, $\sum_{n\ge 1} 1/n = \infty$ so we cannot obtain the desired conclusion immediately using the Boral Cantelli Lemma. Instead, we use a subsequence "trick" here. Letting $n(k) = k^2$ and summing both sides of above equation over $n(k)$ where $k \ge 1$, we obtain

$$\sum_{k=1}^{\infty} \mathbb{P}\left(\frac{|S_{n(k)}|}{n(k)} \ge \epsilon\right) \le \frac{M}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

By applying Lemma 6.1, we obtain $\dfrac{|S_{n(k)}|}{n(k)} \to 0$ a.s. as $k \to \infty$.

Let $\Delta_k = \max\left\{|S_n - S_{n(k)}| : n(k) < n < n(k+1)\right\}$. For $n(k) \le n < n(k+1)$, we have

$$\begin{aligned} \frac{|S_n|}{n} &\le \frac{|S_{n(k)}|}{n(k)} + \frac{\Delta_k}{n(k)}, \\ \Longrightarrow \limsup_{n\to\infty} \frac{|S_n|}{n} &\le \limsup_{k\to\infty} \frac{|S_{n(k)}|}{n(k)} + \limsup_{k\to\infty} \frac{\Delta_k}{n(k)} = \limsup_{k\to\infty} \frac{\Delta_k}{n(k)}. \end{aligned}$$

The proof is complete if we show $\dfrac{\Delta_k}{n(k)} \to 0$ a.s. as $k \to \infty$. Let $B_j = \left\{\omega : |S_{n(k)+j} - S_{n(k)}| \ge \epsilon n(k)\right\}$, for $1 \le j \le 2k$. We have

$$\begin{aligned} \mathbb{P}(\Delta_k \ge \epsilon n(k)) &= \mathbb{P}\left(\bigcup_{j=1}^{2k} B_j\right) \\ &\le \sum_{j=1}^{2k} \mathbb{P}\big(|S_{n(k)+j} - S_{n(k)}| \ge \epsilon n(k)\big) \\ &\le \sum_{j=1}^{2k} \frac{jM}{\epsilon^2 n(k)^2} = \frac{M}{\epsilon^2 k^3}(2k+1). \end{aligned}$$

Summing both sides over $k \ge 1$, we obtain

$$\sum_{k=1}^{\infty} \mathbb{P}\left(\frac{\Delta_k}{n(k)} \ge \epsilon\right) \le \frac{M}{\epsilon^2} \sum_{k=1}^{\infty} \frac{2k+1}{k^3} < \infty.$$

From Lemma 6.5, we obtain $\dfrac{\Delta_k}{n(k)} \to 0$ a.s. as $k \to \infty$, and the proof is complete. $\square$

For $X \ge 0$, we have

$$\sum_{k=1}^{\infty} \mathbf{1}_{\{X \ge k\}} \le X \le \sum_{k=0}^{\infty} \mathbf{1}_{\{X \ge k\}}.$$

Therefore, for any X, we obtain

$$\sum_{k=1}^{\infty} \mathbb{P}(|X| \geq k) \leq \mathbb{E}|X| \leq \sum_{k=0}^{\infty} \mathbb{P}(|X| \geq k),$$
$$\sum_{k=1}^{\infty} \mathbb{P}(|X| \geq k) \leq \infty \iff \mathbb{E}|X| < \infty.$$

As a side note, if $X \in \mathbb{Z}_+$, we have the following equality:

$$X = \sum_{k=1}^{\infty} \mathbf{1}_{\{X \geq k\}},$$
$$\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).$$

Lemma 6.8. *Suppose $X_1, X_2, \ldots$ are i.i.d. Then,*

$$\lim_{n\to\infty} \frac{X_n}{n} = 0 \text{ a.s.} \iff \mathbb{E}|X_1| < \infty.$$

Proof.

'$\Rightarrow$':

$$\lim_{n\to\infty} \frac{X_n}{n} = 0 \text{ a.s.} \implies \mathbb{P}\left(\frac{|X_n|}{n} \geq 1 \text{ i.o.}\right) = 0.$$

From the second Borel-Cantelli Lemma, we have

$$\sum_{n\geq 1} \mathbb{P}\left(\frac{|X_n|}{n} \geq 1\right) < \infty$$
$$\sum_{n\geq 1} \mathbb{P}(|X_1| \geq n) < \infty$$
$$\mathbb{E}|X_1| < \infty.$$

'$\Leftarrow$':

$$\mathbb{E}\left|\frac{X_1}{\epsilon}\right| < \infty \implies \sum_{n\geq 1} \mathbb{P}(|X_n| \geq n\epsilon) < \infty.$$

The result then follows from Lemma 6.5. $\square$

7.1 SLLN

In this note, we prove the Strong Law of Large Numbers (SLLN). We give three different proofs of SLLN, which are based on the ideas of maximal inequality, truncation, and positivity. First, let us state the SLLN.

Theorem 7.1 (Kolmogorov's SLLN). *Suppose $X_1, X_2, \ldots$ are i.i.d. random variables, $\mathbb{E}|X_1| < \infty$, $\mathbb{E}X_1 = 0$, and $S_n = \sum_{i=1}^n X_i$. We have*

$$\frac{S_n}{n} \to 0 \text{ a.s. as } n \to \infty.$$

7.2 The First Proof: L^1 Maximal Inequality

For a sequence $a_1, a_2, \ldots, a_N$, we say that a_j is an n-leader if $a_j + a_{j+1} + \ldots + a_k > 0$ for some k such that $j \le k \le N$ and $k - j + 1 \le n \le N$.

Lemma 7.1. *Let $G \neq \emptyset$ be the set of indices of all n-leaders of $a_1, \ldots, a_N$. Then*

$$\sum_{j \in G} a_j > 0.$$

Proof. Let

$$\begin{aligned} j_1 &= \min G, \\ k_1 &= \min\{k \ge j_1 : a_{j_1} + \ldots + a_k > 0\}. \end{aligned}$$

Since $j_1 \in G$, $k_1 \le N$ exists and $k_1 - j_1 + 1 \le n$. Then we have $\sum_{j=j_1}^{s-1} a_j < 0$ for $j_1 \le s \le k_1$. Therefore,

$$\sum_{k=s}^{k_1} a_j = \sum_{j=j_1}^{k_1} a_j - \sum_{j=j_1}^{s-1} a_j > 0,$$

This implies that $s \in G$ and hence $[j_1, k_1] \subset G$. Let $G_1 = G \setminus [j_1, k_1] \neq \emptyset$ and repeat the same process until $G_{l+1} = \emptyset$ for some l. We obtain

$$\bigcup_{i=1}^{l} [j_i, k_i] = G,$$

where $[j_i, k_i] \cap [j_m, k_m] = \emptyset,\ \forall\, i \neq m$. Finally, we have

$$\sum_{j \in G} a_j = \sum_{i=1}^{l} \sum_{s=j_i}^{k_i} a_s > 0.$$

$\square$

Theorem 7.2 (L^1 maximal inequality)**.** *Let $X_k,\ k \geq 1$, be i.i.d. random variables. Then, $\forall\, \epsilon > 0$,*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} \frac{S_k}{k} > \epsilon \right) \leq \frac{1}{\epsilon} \mathbb{E} X_1^+.$$

Proof. Fix $\epsilon > 0$ and $N > 1$. For $1 \leq n \leq N$, let

$$A_j = \left\{ \omega \,:\, \max_{\substack{1 \leq k \leq n \\ j+k-1 \leq N}} \frac{S(j, j+k-1)}{k} > \epsilon \right\},$$

where $S(j,k) = X_j + \ldots + X_k$. From the definition, we have

$$A_j = \{\omega \,:\, a_j(\omega) \text{ is } n\text{-leader in}(a_1, a_2, \ldots, a_N)\},$$

where $a_j(\omega) = X_j(\omega) - \epsilon$. Applying Lemma 7.1, we obtain

$$\sum_{j=1}^{N} (X_j - \epsilon) \mathbf{1}_{A_j} > 0,$$

which yields

$$\epsilon \sum_{j=1}^{N-n} \mathbf{1}_{A_j} \leq \epsilon \sum_{j=1}^{N} \mathbf{1}_{A_j} < \sum_{j=1}^{N} X_j \mathbf{1}_{A_j} \leq \sum_{j=1}^{N} X_j^+. \tag{17}$$

Furthermore, for $1 \leq j \leq N - n$, since the X_i's are i.i.d., we have

$$\mathbb{P}(A_j) = \mathbb{P}(A_1) = \mathbb{P}\left(\max_{1 \leq k \leq n} \frac{S_k}{k} > \epsilon \right).$$

Taking expectations in (17), we obtain

$$\begin{aligned} \epsilon(N-n)\mathbb{P}(A_1) &\leq N \mathbb{E} X_1^+ \\ \frac{N-n}{N} \mathbb{P}(A_1) &\leq \frac{1}{\epsilon} \mathbb{E} X_1^+. \end{aligned}$$

By taking $N \to \infty$, the proof is complete. $\square$

Corollary 7.1. *Let $X_k, k \geq 1$ be i.i.d random variables. Then, $\forall\ \epsilon > 0$,*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} \left| \frac{S_k}{k} \right| > \epsilon \right) \leq \frac{1}{\epsilon} \mathbb{E}|X_1|.$$

Proof. Apply the L^1 maximal inequality to (X_k) and $(-X_k)$. $\square$

We are now ready to give the first proof of the SLLN. Fix $\lambda > 0$. Let

$$X_k' = X_k \mathbf{1}_{\{|X_k| \le \lambda\}},$$
$$X_k'' = X_k \mathbf{1}_{\{|X_k| > \lambda\}}.$$

As the threshold λ is fixed for all k, this is known as *fixed truncation*. We let

$$S_n' = X_1' + \ldots + X_n',$$
$$S_n'' = X_1'' + \ldots + X_n'',$$

so that $S_n = S_n' + S_n''$. We have

$$\begin{aligned}
&\mathbb{E}X_k' + \mathbb{E}X_k'' = \mathbb{E}X_k = 0, \\
&S_n = S_n' - n\mathbb{E}X_1' + S_n'' - n\mathbb{E}X_1'', \\
&\left|\frac{S_n}{n}\right| \le \left|\frac{S_n'}{n} - \mathbb{E}X_1 \mathbf{1}_{\{|X_1| \le \lambda\}}\right| + \left|\frac{S_n''}{n} - \mathbb{E}X_1 \mathbf{1}_{\{|X_1| > \lambda\}}\right|.
\end{aligned}$$

Since X_k' has bounded moments, we can apply the SLLN with 2nd moments (Lemma 6.7) to obtain

$$\left|\frac{S_n'}{n} - \mathbb{E}X_1 \mathbf{1}_{\{|X_1| \le \lambda\}}\right| \to 0 \text{ a.s.}$$

Furthermore, we have

$$\begin{aligned}
L = \limsup_{n\to\infty} \left|\frac{S_n}{n}\right| &\le \limsup_{n\to\infty} \left|\frac{S_n''}{n} - \mathbb{E}X_1 \mathbf{1}_{\{|X_1| > \lambda\}}\right| \\
&\le \sup_{n \ge 1} \left|\frac{S_n''}{n} - \mathbb{E}X_1 \mathbf{1}_{\{|X_1| > \lambda\}}\right| \\
&\le \sup_{n \ge 1} \left|\frac{S_n''}{n}\right| + \mathbb{E}X_1 \mathbf{1}_{\{|X_1| > \lambda\}}.
\end{aligned}$$

Since $\mathbb{E}X_1 \mathbf{1}_{\{|X_1| > \lambda\}} \le \mathbb{E}|X_1| < \infty$, we apply DCT (Theorem 4.3) to obtain

$$\lim_{\lambda\to\infty} \mathbb{E}X_1 \mathbf{1}_{\{|X_1| > \lambda\}} = \mathbb{E}X_1 \mathbf{1}_{\{|X_1| = \infty\}} = 0.$$

Therefore, $\forall\, \epsilon > 0$, $\exists\, \lambda_0$ such that

$$\mathbb{E}X_1 \mathbf{1}_{\{|X_1| > \lambda\}} < \epsilon, \ \forall\, \lambda \ge \lambda_0.$$

By using the above result and applying Corollary 7.1, we obtain for all $\lambda \ge \lambda_0$,

$$\begin{aligned}
\mathbb{P}(L > 2\epsilon) &\le \mathbb{P}\left(\sup_{n\ge 1} \left|\frac{S_n''}{n}\right| > \epsilon\right) \\
&\le \lim_{\lambda\to\infty} \frac{1}{\epsilon} \mathbb{E}|X_1| \mathbf{1}_{\{|X_1| > \lambda\}} = 0.
\end{aligned}$$

Taking $\epsilon \to 0$, we have

$$\mathbb{P}(L > 0) = 0 \implies \mathbb{P}(L = 0) = 1 \implies \frac{S_n}{n} \to 0 \text{ a.s. as } n \to \infty.$$

The first proof of SLLN is now complete.

7.3 The Second Proof: Kolmogorov

We next discuss Kolmogorov's proof of SLLN, which was done around 1930.

Theorem 7.3 (Kolmogorov's maximal inequality)**.** *Suppose $X_k, k \geq 1$ are independent random variables and $\mathbb{E}|X_k| < \infty$, then*

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2}\operatorname{var}(S_n),\ \forall\ \lambda > 0.$$

Note that $\operatorname{var}(S_n)$ always exists. If it is infinite, then the bound is trivial.

Proof. Without loss of generality, we assume $\mathbb{E}X_k = 0$. Let

$$\tau = \min\{1 \leq k \leq n \,:\, |S_k| \geq \lambda\},$$

where $\tau = \infty$ if $|S_k| < \lambda$ for all $1 \leq k \leq n$. Define

$$S_\tau = \sum_{k=1}^n S_k \mathbf{1}_{\{\tau=k\}}.$$

Since

$$\lambda^2 \mathbf{1}_{\{\max_{1\leq k\leq n}|S_k|\geq\lambda\}} \leq S_\tau^2,$$

we obtain

$$\lambda^2 \mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \mathbb{E}S_\tau^2.$$

It suffices to prove $\mathbb{E}S_\tau^2 \leq \mathbb{E}S_n^2$. We have

$$\begin{aligned} S_n &= S_\tau + S_n - S_\tau, \\ S_n^2 &= S_\tau^2 + (S_n - S_\tau)^2 + 2S_\tau\left(S_n - S_\tau\right), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}S_\tau\left(S_n - S_\tau\right) &= \mathbb{E}\left[\sum_{k=1}^n \mathbf{1}_{\{\tau=k\}} S_k \left(S_n - S_k\right)\right] \\ &= \sum_{k=1}^n \mathbb{E}\big[\mathbf{1}_{\{\tau=k\}} S_k\big] \mathbb{E}\left(S_n - S_k\right) = 0, \end{aligned}$$

where the penultimate equality follows from independence. We thus have

$$\mathbb{E}S_\tau^2 \leq \mathbb{E}S_n^2,$$

and the proof is complete. □

Theorem 7.4 (Variance convergence criterion)**.** *Suppose $Y_k,\ k \geq 1$ are independent random variables and $\mathbb{E}Y_k = 0$. If $\sum_{k=1}^\infty \operatorname{var}(Y_k) < \infty$, then*

$$\sum_{k=1}^\infty Y_k \text{ converges a.s.}$$

Proof. Let $S_n(\omega) = \sum_{k=1}^{n} Y_k(\omega)$. It suffices to show that $(S_n)_{n\geq 1}$ is Cauchy a.s., i.e.,

$$R_N = \sup_{n,m\geq N} |S_n - S_m| \to 0 \text{ a.s. as } N \to \infty.$$

Let $N \geq N_0 \geq 1$. We have

$$\begin{aligned} R_N &\leq R_{N_0} \\ &\leq \sup_{n\geq N_0} |S_n - S_{N_0}| + \sup_{m\geq N_0} |S_m - S_{N_0}|. \end{aligned}$$

Therefore, $\forall\, \epsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\limsup_{N\to\infty} R_N > \epsilon\right) &\leq 2\mathbb{P}\left(\sup_{n\geq N_0} |S_n - S_{N_0}| > \frac{\epsilon}{2}\right) \\ &\leq \frac{8}{\epsilon^2} \sum_{k=N_0+1}^{\infty} \operatorname{var}(Y_k), \end{aligned}$$

where the last inequality follows from Chebyshev's inequality. Since $\sum_{k=1}^{\infty} \operatorname{var}(Y_k) < \infty$, we have

$$\begin{aligned} &\lim_{N_0\to\infty} \sum_{k=N_0+1}^{\infty} \operatorname{var}(Y_k) = 0 \\ \implies &\mathbb{P}\left(\limsup_{N\to\infty} R_N > \epsilon\right) = 0. \end{aligned}$$

□

Let $\hat{X}_k = X_k \mathbf{1}_{\{|X_k|\leq k\}}$. This is called *moving truncation.* We have

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}\left(\hat{X}_k \neq X_k\right) &= \sum_{k\geq 1} \mathbb{P}(|X_k| > k) \\ &= \sum_{k\geq 1} \mathbb{P}(|X_1| > k) \\ &\leq \mathbb{E}|X_1| < \infty. \end{aligned}$$

From the first Borel-Cantelli lemma (Lemma 6.1), we obtain $\mathbb{P}\left(\hat{X}_k \neq X_k \text{ i.o}\right) = 0$, which means that for a.s. all $\omega \in \Omega$, $\exists\, K(\omega)$ s. t. $X_k(\omega) = \hat{X}_k(\omega), \forall\, k \geq K(\omega)$. Therefore, $\frac{1}{n}\sum_{k=1}^{n} X_k \to 0$ a.s. if and only if $\frac{1}{n}\sum_{k=1}^{n} \hat{X}_k \to 0$ a.s. In the rest of this section, we prove $\frac{1}{n}\sum_{k=1}^{n} \hat{X}_k \to 0$ a.s..

Lemma 7.2. *Suppose X_k, $k \geq 1$ are identically distributed random variables and $\hat{X}_k = X_k \mathbf{1}_{\{|X_k|\leq k\}}$. We have*

$$\sum_{k=1}^{\infty} \frac{\operatorname{var}(\hat{X}_k)}{k^2} \leq 2\mathbb{E}|X_1|.$$

Proof.

$$
\begin{aligned}
\sum_{k=1}^{\infty} \frac{\operatorname{var}(\hat{X}_k)}{k^2} &\le \sum_{k=1}^{\infty} \frac{\mathbb{E}\hat{X}_k^2}{k^2} \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \mathbb{E}\big[X_1^2 \mathbf{1}_{\{|X_1| \le k\}}\big] \\
&= \mathbb{E}\left[X_1^2 \sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{1}_{\{|X_1| \le k\}}\right] \\
&\le 2\mathbb{E}|X_1|,
\end{aligned}
$$

where the last equality comes from Fubini's theorem (Theorem 5.2) and the last inequality is because

$$
\phi(x) = x^2 \sum_{k \ge 1} \frac{1}{k^2} \mathbf{1}_{\{x \le k\}} \le 2x
$$

for all $x \ge 0$. To prove this, we note that

$$
\int_{i-1}^{\infty} \frac{1}{x^2} \, \mathrm{d}x = \frac{1}{i-1} \ge \sum_{k=i}^{\infty} \frac{1}{k^2}.
$$

If $0 \le x \le 1$, we have

$$
\phi(x) = x^2 \sum_{k=1}^{\infty} \frac{1}{k^2} = x^2 \left(1 + \sum_{k=2}^{\infty} \frac{1}{k^2}\right) \le 2x^2 \le 2x.
$$

A similar reasoning applies for $1 < x \le 2$ and $2 < x < \infty$ and the proof is complete. $\square$

Lemma 7.3 (Cesaro's Lemma)**.** *If* $\lim_{n\to\infty} a_n = a$, *then*

$$
\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = a.
$$

Proof. Without loss of generality, we assume $a = 0$. For $k \le n$, we have

$$
\begin{aligned}
\left|\frac{1}{n} \sum_{i=1}^{n} a_i\right| &\le \frac{1}{n} \left|\sum_{i=1}^{k} a_i\right| + \frac{n-k}{n} \sup_{j>k} |a_j|, \\
\Longrightarrow \limsup_{n\to\infty} \frac{1}{n} \left|\sum_{i=1}^{n} a_i\right| &\le \limsup_{j\to\infty} |a_j| = 0.
\end{aligned}
$$

$\square$

Lemma 7.4 (Kronecker's Lemma)**.** *If* $\displaystyle\sum_{k=1}^{\infty} \frac{a_k}{k}$ *converges, then*

$$
\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = 0.
$$

Proof. Let $r_n = \displaystyle\sum_{i=n+1}^{\infty} \frac{a_i}{i}$. We have

$$
\lim_{n\to\infty} r_n = 0.
$$

From Cesaro's lemma (Lemma 7.3), we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} r_k = 0.$$

Furthermore, for $k < n$, we have

$$r_k = \frac{a_{k+1}}{k+1} + \frac{a_{k+2}}{k+2} + \cdots + \frac{a_n}{n} + r_n.$$

Therefore,

$$\sum_{k=0}^{n-1} r_k = a_1 + \ldots + a_n + nr_n,$$
$$\implies \frac{1}{n}\sum_{k=1}^{n} a_k = -r_n + \frac{1}{n}\sum_{k=0}^{n-1} r_k.$$

By using the above results, we obtain

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} a_k = 0.$$

□

We are now ready to give Kolmogorov's proof of the SLLN. From Lemma 7.2, we have

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\frac{1}{k^2}\left(\hat{X}_k - \mathbb{E}\hat{X}_k\right)^2\right] < \infty.$$

From Theorem 7.4, we obtain

$$\sum_{k=1}^{\infty} \frac{1}{k}\left(\hat{X}_k - \mathbb{E}\hat{X}_k\right) \text{ converges a.s.}$$

From Kronecker's lemma (Lemma 7.4), we have

$$\frac{1}{n}\sum_{k=1}^{n}\left(\hat{X}_k - \mathbb{E}\hat{X}_k\right) \to 0 \text{ a.s.} \tag{18}$$

Furthermore, using the DCT, we have

$$\lim_{k\to\infty} \mathbb{E}\hat{X}_k = \mathbb{E}\left[\lim_{k\to\infty} X_1 \mathbf{1}_{\{|X_1|\le k\}}\right] = \mathbb{E}X_1 = 0, \tag{19}$$
$$\implies \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\hat{X}_k \to 0 \text{ a.s..} \tag{20}$$

Therefore, from (18), we obtain

$$\frac{1}{n}\sum_{k=1}^{n} \hat{X}_k \to 0 \text{ a.s.},$$

and the SLLN is proved.

We now give a generalization of Kronecker's lemma, which will be useful later on.

Lemma 7.5 (Generalized Kronecker's Lemma)**.** *Suppose* $0 < b_n \to \infty$ *as* $n \to \infty$*. If* $\sum_{n=1}^{\infty} \frac{a_n}{b_n}$ *converges in* $\mathbb{R}$*, then*

$$\frac{1}{b_n} \sum_{k=1}^{n} a_k \to 0 \text{ as } n \to \infty.$$

Proof. Without loss of generality, we can suppose that $b_n \in \mathbb{Z}_+$, $\forall\, n \geq 1$. Let $b_0 = 0$, $s_0 = 0$ and

$$s_n = \sum_{k=1}^{n} \frac{a_k}{b_k} \to s \in \mathbb{R}.$$

It can be checked that

$$\frac{1}{b_n} \sum_{k=1}^{n} a_k = s_n - \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) s_{k-1} \to 0,$$

since from Cesaro's lemma (note that $b_k - b_{k-1} \in \mathbb{Z}_+$ and $b_n = \sum_{k=1}^{n} (b_k - b_{k-1})$), we have

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) s_{k-1} = s.$$

□

Theorem 7.5 (Generalized variance convergence criterion)**.** *Suppose* $Y_k, k \geq 1$ *are independent random variables,* $\mathbb{E}Y_k = 0$*, and* $0 < b_k \to \infty$ *as* $k \to \infty$*. We have*

$$\sum_{k=1}^{\infty} \frac{\operatorname{var}(Y_k)}{b_k^2} < \infty \implies \frac{1}{b_n} \sum_{k=1}^{\infty} Y_k \to 0 \text{ a.s.}$$

Proof. Apply Theorem 7.4 to Y_k/b_k and the result follows from Lemma 7.5. □

7.4 The Third Proof: Etemadi's Use of Positivity

In the third proof, we present the surprising and elegant proof of Etemadi discovered in 1981, some 50 years after Kolmogorov first proved the SLLN. This proof also strengthens the result to require only pairwise independence.

Theorem 7.6 (Etemadi's SLLN)**.** *Suppose* $X_i,\ i \geq 1$ *are identically distributed and pairwise independent, and* $\mathbb{E}X_i = \mu$*. We have*

$$\frac{S_n}{n} \to \mu \text{ a.s. as } n \to \infty.$$

Proof. We first observe that $X_i = X_i^+ - X_i^-$ and

$$X_i \perp\!\!\!\perp X_j \implies X_i^+ \perp\!\!\!\perp X_j^+,\ X_i^- \perp\!\!\!\perp X_j^-.$$

Therefore, we can without loss of generality assume $X_i \geq 0$. This turns out to be the key of Etemadi's proof. Let

$$\begin{aligned}
\hat{X}_i &= X_i \mathbf{1}_{\{X_i \leq i\}}, \\
\hat{S}_n &= \sum_{i=1}^n \hat{X}_i.
\end{aligned}$$

As shown previously, it suffices to show that $\hat{S}_n/n \to \mu$ a.s. Let $\alpha \in (1,2)$, and $j_n = \lfloor \alpha^n \rfloor$. We have

$$\begin{aligned}
& 1 \leq j_n \leq \alpha^n < j_{n+1} \leq 2j_n, \\
\implies & \frac{1}{j_n} \leq \frac{2}{\alpha^n}.
\end{aligned}$$

For a fixed i, let $n_0 = \min\{n : \alpha^n \geq i\}$. We have

$$\sum_{n: j_n \geq i} \frac{1}{j_n^2} \leq \sum_{n: j_n \geq i} \frac{4}{\alpha^{2n}} = 4 \sum_{k=0}^{\infty} \frac{1}{\alpha^{2k}\alpha^{2n_0}} \leq 4 \sum_{k=0}^{\infty} \frac{1}{i^2} \frac{1}{\alpha^{2k}} = \frac{4}{i^2} \frac{1}{1-\alpha^{-2}}. \tag{21}$$

For $\epsilon > 0$ and using Chebyshev's inequality together with pairwise independence, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \hat{S}_{j_n} - \mathbb{E}\hat{S}_{j_n} \right| \geq \epsilon j_n \right) &\leq \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{j_n^2} \operatorname{var}(\hat{S}_{j_n}) \\
&= \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{j_n^2} \sum_{i=1}^{j_n} \operatorname{var}(\hat{X}_i) \\
&= \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} \operatorname{var}(\hat{X}_i) \sum_{n: j_n \geq i} \frac{1}{j_n^2} \\
&\leq \frac{4}{\epsilon^2} \frac{1}{1-\alpha^{-2}} \sum_{i=1}^{\infty} \frac{\operatorname{var}(\hat{X}_i)}{i^2} \\
&\leq \frac{8}{\epsilon^2} \frac{1}{1-\alpha^{-2}} \mathbb{E}|X_1| < \infty,
\end{aligned}$$

where the penultimate inequality follows from (21), and the last equality from Lemma 7.2. From Lemma 6.5, we obtain

$$\frac{\hat{S}_{j_n} - \mathbb{E}\hat{S}_{j_n}}{j_n} \to 0 \text{ a.s.}$$

Using the same argument as in (20), we have

$$\mathbb{E}\hat{X}_i \to \mathbb{E}X_i = \mu \implies \frac{\mathbb{E}\hat{S}_{j_n}}{j_n} \to \mu.$$

Therefore, we obtain

$$\frac{\hat{S}_{j_n}}{j_n} \to \mu \text{ a.s.}$$

Finally, for any n, there exists k such that

$$j_k \leq n < j_{k+1},$$

and because $\hat{X}_i \geq 0$ for all $i \geq 1$, we have a.s.,

$$\begin{aligned}
&\hat{S}_{j_k} \leq \hat{S}_n \leq \hat{S}_{j_{k+1}}, \\
\Longrightarrow\ & \frac{j_k}{j_{k+1}} \frac{\hat{S}_{j_k}}{j_k} \leq \frac{\hat{S}_n}{n} \leq \frac{j_{k+1}}{j_k} \frac{\hat{S}_{j_{k+1}}}{j_{k+1}}, \\
\Longrightarrow\ & \lim_{n\to\infty} \frac{j_k}{j_{k+1}} \frac{\hat{S}_{j_k}}{j_k} \leq \liminf_{n\to\infty} \frac{\hat{S}_n}{n} \leq \limsup_{n\to\infty} \frac{\hat{S}_n}{n} \leq \lim_{n\to\infty} \frac{j_{k+1}}{j_k} \frac{\hat{S}_{j_{k+1}}}{j_{k+1}}, \\
\Longrightarrow\ & \frac{1}{\alpha}\mu \leq \liminf_{n\to\infty} \frac{\hat{S}_n}{n} \leq \limsup_{n\to\infty} \frac{\hat{S}_n}{n} \leq \alpha\mu.
\end{aligned}$$

Taking $\alpha \to 1$, we obtain

$$\lim_{n\to\infty} \frac{\hat{S}_n}{n} = \mu \text{ a.s.}$$

□

8.1 Glivenko–Cantelli Theorem

Assume that $F(x)$ is a cumulative distribution function (cdf), i.e.,

- $F(x)$ is non-decreasing.
- $F(x)$ is right continuous.
- $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$.

A r.v. X such that $\mathbb{P}(X \le x) = F(x)$ is said to have cdf F. From Carathéodory's Extension Theorem (Theorem 3.1), the cdf of X uniquely determines the distribution $\mathbb{P}_X$ of X.

Then even though $F(x)$ may not be a bijective function, we can always define a pseudo-inverse as

$$F^{-1}(\omega) = \inf\{x : F(x) \ge \omega\} \tag{22}$$

for $\omega \in [0, 1]$. We have the following properties:

(i) If $y < F^{-1}(\omega)$, then $F(y) < \omega$. Otherwise, suppose $F(y) \ge \omega$. Then by the definition (22), $F^{-1}(\omega) \le y$, which is a contradiction.

(ii) If $y > F^{-1}(\omega)$, then $F(y) \ge \omega$ because F is non-decreasing. If ω is a continuity point of F^{-1}, then we have $F(y) > \omega$.

Lemma 8.1. *For any x, we have*

$$\{\omega : \omega \le F(x)\} = \{\omega : F^{-1}(\omega) \le x\}.$$

Proof. If $\omega \le F(x)$, then by definition (22), we have $F^{-1}(\omega) \le x$. On the other hand, if $F^{-1}(\omega) \le x$, then since F is non-decreasing, we have $\omega \le F(y)\ \forall\, y > x$. Since $F(x)$ is right continuous, by letting $y \downarrow x$, we have $\omega \le F(x)$. □

Lemma 8.2. *$F^{-1}(\omega)$, $\omega \in [0, 1]$, is a random variable on $([0, 1], \mathcal{B}([0, 1]), \lambda)$ with cdf F, where λ is the Lebesgue measure.*

Proof. From Lemma 8.1, we have $\lambda(F^{-1}(\omega) \le x) = \lambda(\omega \le F(x)) = F(x)$. □

Suppose that (X_i) are i.i.d. random variables with cdf F, then we define the empirical cdf for $n \ge 1$ as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}}.$$

The SLLN implies that for each x, $F_n(x) \to F(x)$ a. s. as $n \to \infty$.

Theorem 8.1 (Glivenko-Cantelli Theorem)**.**

$$\sup_x |F_n(x) - F(x)| \to 0 \text{ a.s. } \textit{as } n \to \infty.$$

Proof. We reduce the proof to that for the probability space $([0,1], \mathcal{B}[0,1], \lambda)$. Let Λ to be the cdf for λ and suppose Y_i are i.i.d. $\sim \lambda$. We define

$$\Lambda_n(F(x)) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i \le F(x)\}}.$$

If $X_i = F^{-1}(Y_i)$, then from Lemma 8.2, X_i are i.i.d. with cdf F. Therefore, $X_i \le x$ iff $Y_i \le F(x)$, and to prove the theorem, it suffices to show that

$$\sup_y |\Lambda_n(y) - \Lambda(y)| \to 0 \text{ a.s.}$$

Fix $\epsilon > 0$, choose m s.t. $\dfrac{1}{m} \le \dfrac{\epsilon}{2}$. Consider the set

$$E = \left\{ \frac{k}{m} \,:\, k = 0, 1, \ldots, m \right\}.$$

From SLLN, $\Lambda_n(y) \to \Lambda(y)$ for each $y \in E$. Since E is finite, $\exists\, N$ s.t. $\forall\, n \ge N$, $|\Lambda_n(y) - \Lambda(y)| \le \frac{\epsilon}{2}$ $\forall\, y \in E$. For $x \in [0,1]$, we can always find u, v s.t. $u \le x < v$, $u, v \in E$, $v - u = \frac{1}{m}$. Since $\Lambda(y) = y$, we have

$$\begin{aligned} \Lambda_n(x) &\ge \Lambda_n(u) \ge u - \frac{\epsilon}{2} \ge x - \frac{1}{m} - \frac{\epsilon}{2} \ge x - \epsilon, \\ \Lambda_n(x) &\le \Lambda_n(v) \le v + \frac{\epsilon}{2} \le x + \frac{1}{m} + \frac{\epsilon}{2} \le x + \epsilon, \end{aligned}$$

so that

$$|\Lambda_n(x) - \Lambda(x)| \le \epsilon.$$

The theorem is now proved. □

8.2 Kolmogorov's 0-1 Law

Suppose that $(X_i)_{i \ge 1}$ are independent random variables. Consider the σ-algebra $\sigma(X_n, X_{n+1}, \ldots)$, which is the smallest σ-algebra w.r.t. which all X_m, $m \ge n$, are measurable:

$$\sigma(X_n, X_{n+1}, \ldots) = \sigma\left(\bigcup_{m \ge n} \sigma(X_m) \right).$$

Note that the intersection of these σ-algebras is also a σ-algebra,

$$\mathcal{T} = \bigcap_{n \ge 1} \sigma(X_n, X_{n+1}, \ldots),$$

and we call this the tail σ-algebra. The tail σ-algebra contains all events that are not affected by changes in a finite number of random variables X_i.

Let $S_n = \sum_{i=1}^n X_i$. Then $\{\omega : \lim_{n\to\infty} S_n \text{ exists}\} \in \mathcal{T}$. Furthermore, since for any $m \in \mathbb{N}$,

$$\limsup_{n\to\infty} \frac{S_n}{n}(\omega) = \limsup_{m\leq n\to\infty} \frac{S(m,n)}{n},$$

where $S(m,n) = S_n - S_{m-1}$, we have $\{\limsup_{n\to\infty} \frac{S_n}{n} \leq a\} \in \mathcal{T}$ for any $a \in \mathbb{R}$. (To be specific, note that $S(m,n)$ is measurable w.r.t. $\sigma(X_m, X_{m+1}, \ldots)$ for every $n \geq m$. Therefore, $\limsup_{n\to\infty} S(m,n)/n$ is measurable w.r.t. $\sigma(X_m, X_{m+1}, \ldots)$ (cf. Lemma 4.2). This is true for all $m \geq 1$.)

On the other hand, consider the event $\{\limsup_{n\to\infty} S_n \geq a\}$ where X_i is not 0 a.s. It does not belong to $\mathcal{T}$ since it is in $\sigma(X_1, X_2, \ldots)$ but not in $\sigma(X_2, X_3, \ldots)$.

Lemma 8.3 (Grouping Lemma)**.** *Suppose that a set of σ-algebras $\{\mathcal{A}_t : t \in T\}$ indexed by the countable index set T are independent (i.e., any finite subset of σ-algebras $\mathcal{A}_{i_1}, \mathcal{A}_{i_2}, \ldots, \mathcal{A}_{i_n}$ are independent). Then for any finite partition $T = \bigcup_{i=1}^n T_i$ where $T_i \cap T_j = \emptyset \ \forall\, i \neq j$, the σ-algebras $\mathcal{F}_i = \sigma(\bigcup_{t\in T_i} \mathcal{A}_t)$, $i = 1, \ldots, n$ are independent.*

Proof. Let

$$\mathcal{C}_i = \left\{ \bigcap_{t\in F} A_t \ : \ F \text{ is finite subset of } T_i,\ A_t \in \mathcal{A}_t \right\}.$$

We note that $\mathcal{C}_i$ is a π-system. For any finite $F_i \subset T_i$, $A_{t_i} \in \mathcal{A}_{t_i}$ where $t_i \in F_i$, we have

$$\mathbb{P}\left(\bigcap_{i=1}^n \bigcap_{t_i\in F_i} A_{t_i} \right) = \prod_{i=1}^n \mathbb{P}\left(\bigcap_{t_i\in F_i} A_{t_i} \right),$$

thus $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$ are independent. From Lemma 5.2, $\sigma(\mathcal{C}_i)$, $i = 1, \ldots, n$ are independent. For each $t \in T_i$, we have $\mathcal{A}_t \subset \mathcal{C}_i$, hence $\mathcal{F}_i \subset \sigma(\mathcal{C}_i)$, which implies that $\mathcal{F}_i$, $i = 1, \ldots, n$ are independent. ☐

Theorem 8.2 (Kolmogorov's 0-1 law)**.** *Suppose that $(X_i)_{i\geq 1}$ are independent. Let $\mathcal{T}$ be the tail σ-algebra. If $A \in \mathcal{T}$, then $\mathbb{P}(A) = 0$ or 1.*

Proof. We first show that if $A \in \sigma(X_1, X_2, \ldots)$ and $B \in \mathcal{T}$, then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. For any $k \geq 1$, since $\mathcal{T} \subset \sigma(X_{k+1}, X_{k+2}, \ldots)$, by Lemma 8.3, $\sigma(X_1, \ldots, X_k)$ and $\mathcal{T}$ are independent. It then follows that $\bigcup_{k\geq 1} \sigma(X_1, \ldots, X_k)$ and $\mathcal{T}$ are independent. Since they are both π-systems, by Lemma 5.2, we have

$$\sigma(X_1, X_2, \ldots) = \sigma\left(\bigcup_{k\geq 1} \sigma(X_1, \ldots, X_k) \right) \perp\!\!\!\perp \mathcal{T}.$$

Since $A \in \mathcal{T} \subset \sigma(X_1, X_2, \ldots)$, we obtain that A is independent of itself and

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap A) \\ &= \mathbb{P}(A)\mathbb{P}(A) \\ &= \mathbb{P}(A)^2, \end{aligned}$$

and the theorem follows. ☐

Example 8.1. *The event $\{S_n = \sum_{i=1}^n X_i$ converges $\} \in \mathcal{T}$. From the Kolmogorov's 0-1 law, S_n converges with probability 0 or 1. Intuitively, this means that if S_n converges in probability, it should also converge a.s. We will show this result rigorously in a future session.*

8.3 Hewitt-Savage 0-1 Law

We now suppose that $(X_i)_{i\geq 1}$ are i.i.d. A finite permutation $\pi : \{1, 2, \ldots\} \mapsto \{1, 2, \ldots\}$ is such that $\pi(i) \neq i$ for finitely many i. If B is such that $(X_i)_{i\geq 1} \in B \implies (X_{\pi(i)})_{i\geq 1} \in B$ for all finite permutations π, then we say B is invariant w.r.t. finite permutations. Let

$$\mathscr{E} = \{\{(X_i)_{i\geq 1} \in B\} \,:\, B \text{ is invariant}\}.$$

Then $\mathscr{E}$ is called the exchangeable σ-algebra. (Exercise: check that $\mathscr{E}$ is a σ-algebra.)

$\mathscr{E}$ consists of events that are not affected when we permute a finite number of random variables X_i. Clearly, $\mathcal{T} \subset \mathscr{E}$. On the other hand, $\{\limsup_{n\to\infty} S_n \geq a\} \notin \mathcal{T}$ but $\{\limsup_{n\to\infty} S_n \geq a\} \in \mathscr{E}$ so $\mathscr{E}$ is strictly larger than $\mathcal{T}$.

Theorem 8.3 (Hewitt-Savage 0-1 Law)**.** *Assume $(X_i)_{i\geq 1}$ are i.i.d. If $A \in \mathscr{E}$, then $\mathbb{P}(A) = 0$ or 1.*

To prove this, we first introduce a lemma. Let $A \triangle B = (A \backslash B) \cup (B \backslash A)$. We have

$$|\mathbb{P}(A) - \mathbb{P}(B)| = \left| \int \mathbf{1}_A \, \mathrm{d}\mathbb{P} - \int \mathbf{1}_B \, \mathrm{d}\mathbb{P} \right| \leq \int |\mathbf{1}_A(\omega) - \mathbf{1}_B(\omega)| \, \mathrm{d}\mathbb{P} = \mathbb{P}(A \triangle B).$$

Lemma 8.4 (Approximation Lemma)**.** *If $\mathcal{A}$ is a algebra and $B \in \sigma(\mathcal{A})$, then $\exists\ B_n \in \mathcal{A}$ such that $\mathbb{P}(B \triangle B_n) \to 0$ as $n \to 0$.*

Proof. Let $\mathcal{D} = \{B \in \sigma(\mathcal{A}) : \mathbb{P}(B \triangle B_n) \to 0 \text{ as } n \to 0\}$. Since that this is a λ-system and $\mathcal{A} \subset \mathcal{D}$, from the $\pi - \lambda$ theorem, we have $\sigma(\mathcal{A}) \subset \mathcal{D}$. □

Proof of Hewitt-Savage 0-1 Law. Let $X = (X_i)_{i\geq 1}$. We have $A = \{X \in B\}$ for a B that is invariant w.r.t. finite permutations. The σ-algebra $\sigma(X_1, X_2, \ldots)$ is generated by the algebra

$$\mathcal{A} = \Big\{ \{(X_i)_{i\in F} \in C\} \,:\, F \text{ is finite},\ C \in \mathcal{B}(\mathbb{R}^{|F|}) \Big\}.$$

From Lemma 8.4, for any $\epsilon > 0$, $\exists\, A_n \in \sigma(X_1, \ldots, X_n)$ such that $\mathbb{P}(A \triangle A_n) \leq \epsilon$ for all n sufficiently large since $\mathscr{E} \subset \sigma(X_1, X_2, \ldots)$. We can write $A_n = \{(X_1, \ldots, X_n) \in B_n\}$ for some $B_n \in \mathcal{B}(\mathbb{R}^n)$. Now define $A_n' = \{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \in \sigma(X_{n+1}, \ldots, X_{2n})$. From independence, we have $\mathbb{P}(A_n \cap A_n') = \mathbb{P}(A_n)\mathbb{P}(A_n')$.

Let $\pi(X) = (X_{n+1}, \ldots, X_{2n}, X_1, \ldots, X_n, X_{2n+1}, \ldots)$. We have

$$\begin{aligned}
\mathbb{P}(A_n' \triangle A) &= \mathbb{P}(\{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \triangle \{X \in B\}) \\
&= \mathbb{P}(\{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \triangle \{\pi(X) \in \pi(B)\}) \\
&= \mathbb{P}(\{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \triangle \{\pi(X) \in B\}) \ \because \pi(B) = \{\pi(x) \,:\, x \in B\} = B \\
&= \mathbb{P}(\{(X_1, \ldots, X_n) \in B_n\} \triangle \{X \in B\}) \ \because X_i \text{ are i.i.d.} \\
&= \mathbb{P}(A_n \triangle A) \leq \epsilon.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}((A_n' \cap A_n) \triangle A) &\leq \mathbb{P}((A_n \triangle A) \cup (A_n' \triangle A)) \\
&\leq \mathbb{P}(A_n \triangle A) + \mathbb{P}(A_n' \triangle A) \\
&\leq 2\epsilon,
\end{aligned}$$

and since $|\mathbb{P}(A) - \mathbb{P}(A_n \cap A_n')| \leq \mathbb{P}((A_n' \cap A_n) \triangle A)$, we have $|\mathbb{P}(A) - \mathbb{P}(A_n \cap A_n')| \to 0$ as $n \to \infty$. Similarly, we have $|\mathbb{P}(A) - \mathbb{P}(A_n)| \to 0$ and $|\mathbb{P}(A) - \mathbb{P}(A_n')| \to 0$ so that $\mathbb{P}(A_n \cap A_n') = \mathbb{P}(A_n)\mathbb{P}(A_n') \to \mathbb{P}(A)^2$. Therefore, we have $\mathbb{P}(A) = \mathbb{P}(A)^2$. □

8.4 Discussions

Suppose $(X_i)_{i\geq 1}$ are i.i.d., $\mathbb{E}X_1 = 0$, and $\mathbb{E}X_1^2 = 1$. Take $0 < b_n \to \infty$, then $\frac{S_n}{b_n} \to 0$ a.s. if $\sum_{n\geq 1} \frac{1}{b_n^2} < \infty$. For example, if we choose $b_n = \sqrt{n \log n}$, then $\frac{S_n}{\sqrt{n\log n}} \to 0$ a.s., i.e., S_n grows slower than $\sqrt{n \log n}$.

On the other hand, consider $X_i \sim N(0,1)$, then $\frac{S_n}{\sqrt{n}} \sim N(0,1)$. For any $r \in \mathbb{R}$, we have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} \geq r\right) = \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m}\left\{\frac{S_n}{\sqrt{n}} \geq r\right\}\right) \geq \lim_{m\to\infty} \mathbb{P}\left(\frac{S_m}{\sqrt{m}} \geq r\right) > 0.$$

Since $\{\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} \geq r\}$ is a tail event, by Kolmogorov's 0-1 law, we then have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} \geq r\right) = 1,$$

which implies that $\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty$ a.s., i.e., S_n grows faster than $\sqrt{n}$.

The exact rate of growth of S_n is given by the following theorem.

Theorem 8.4 (Law of Iterated Logarithm)**.** *Suppose $(X_i)_{i\geq 1}$ are i.i.d., $\mathbb{E}X_1 = 0$, and $\mathbb{E}X_1^2 = 1$. Let $S_n = \sum_{i=1}^n X_i$. We have*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1 \text{ a.s.}$$

The proof of this theorem is left as a research exercise.

9.1 Weak Convergence

Let (Ω, d) be a metric space with probability measure $\mathbb{P} : \mathcal{B} \mapsto [0,1]$, where $\mathcal{B}$ is the Borel σ-algebra. Since $\mathcal{B}$ defines $\mathbb{P}$, one may want to define the convergence of a sequence of probability distributions $\mathbb{P}_n \to \mathbb{P}$ as $\mathbb{P}_n(B) \to \mathbb{P}(B)$ for all $B \in \mathcal{B}$. However, this definition is too strong as can be seen from this example: Let δ_x be the probability measure with $\delta_x(\{x\}) = 1$. Then we desire that $\delta_{1/n} \to \delta_0$ but $\delta_{1/n}((0,1)) = 1$ and $\delta_0((0,1)) = 0$.

Let $C_b(\Omega)$ be the set of continuous and bounded functions $f : \Omega \mapsto \mathbb{R}$.

Lemma 9.1. *$\mathbb{P}$ on $(\Omega, \mathcal{B})$ is uniquely determined by $\int_\Omega f \,\mathrm{d}\mathbb{P},\ f \in C_b(\Omega)$.*

Proof. We show that $\int_\Omega f \,\mathrm{d}\mathbb{P}_1 = \int_\Omega f \,\mathrm{d}\mathbb{P}_2, \forall f \in C_b(\Omega) \implies \mathbb{P}_1 = \mathbb{P}_2$. For any open set $U \subset \Omega$, we have

$$d(\omega, U^c) = \begin{cases} 0, & \text{if } \omega \in U^c, \\ > 0, & \text{if } \omega \in U. \end{cases}$$

Let $f_m(\omega) = \min\left\{1, m d(\omega, U^c)\right\} \in C_b(\Omega)$. We have

$$\lim_{m\to\infty} f_m(\omega) = \mathbf{1}_U(\omega).$$

From MCT, we obtain

$$\lim_{m\to\infty} \int_\Omega f_m \,\mathrm{d}\mathbb{P}_1 = \int_\Omega \lim_{m\to\infty} f_m \,\mathrm{d}\mathbb{P}_1 = \int_\Omega \mathbf{1}_U(\omega) \,\mathrm{d}\mathbb{P}_1(\omega) = \mathbb{P}_1(U).$$

Since

$$\int_\Omega f_m \,\mathrm{d}\mathbb{P}_1 = \int_\Omega f_m \,\mathrm{d}\mathbb{P}_2,$$

$\mathbb{P}_1(U) = \mathbb{P}_2(U)$ for all open U. Since the open sets generate $\mathcal{B}$ and $\{B \in \mathcal{B} : \mathbb{P}_1(B) = \mathbb{P}_2(B)\}$ is a λ-system, the $\pi - \lambda$ theorem completes the proof. $\square$

Definition 9.1 (Weak convergence or Convergence in distribution)**.** *We say that $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$ if*

$$\int_\Omega f \,\mathrm{d}\mathbb{P}_n \to \int_\Omega f \,\mathrm{d}\mathbb{P},\ \forall f \in C_b(\Omega).$$

Similarly, we say that $X_n \xrightarrow{\mathrm{d}} X$ if $\mathbb{P}_{X_n} \xrightarrow{\mathrm{d}} \mathbb{P}_X$, i.e.,

$$\mathbb{E}f(X_n) = \int_\Omega f \,\mathrm{d}\mathbb{P}_{X_n} \to \int_\Omega f \,\mathrm{d}\mathbb{P}_X = \mathbb{E}f(X),\ \forall f \in C_b(\Omega).$$

In the sequel, we restrict $\Omega = \mathbb{R}$. All results can be extended to $\mathbb{R}^k$ with trivial modifications. We denote $F(t) = \mathbb{P}((\infty, t])$ as the cdf of $\mathbb{P}$ and $F_n(t)$ as the cdf of $\mathbb{P}_n$.

Theorem 9.1. $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P} \iff F_n(t) \to F(t)$ *for all continuity points* t *of* $F(\cdot)$.

Proof. Let Φ be the set of continuity points of $F(t)$.

'$\Rightarrow$': Fix $\epsilon > 0$ and $t \in \Phi$. Let $\varphi_1(x)$ and $\varphi_2(x)$ be the continuous bounded functions shown in Fig. 3, i.e.,

$$\varphi_i(x) = \begin{cases} 1, & \text{if } x \le t - \epsilon, \\ \text{linear}, & \text{if } t - \epsilon < x \le t, \\ 0, & \text{if } x > t. \end{cases}$$

The function φ_2 is defined similarly.

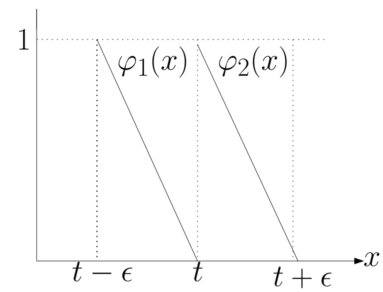

Figure 3: Approximation of $\mathbf{1}_{(-\infty,t]}(x)$.

Then we can approximate $\mathbf{1}_{(-\infty,t]}(x)$ as

$$\mathbf{1}_{(-\infty,t-\epsilon]}(x) \le \varphi_1(x) \le \mathbf{1}_{(-\infty,t]}(x) \le \varphi_2(x) \le \mathbf{1}_{(-\infty,t+\epsilon]}(x).$$

By taking the expectation of the above inequality, we obtain

$$\begin{aligned} F(t-\epsilon) &\le \int \varphi_1 \,\mathrm{d}\mathbb{P} \\ &= \lim_{n\to\infty} \int \varphi_1 \,\mathrm{d}\mathbb{P}_n \\ &\le \liminf_{n\to\infty} F_n(t) \\ &\le \limsup_{n\to\infty} F_n(t) \\ &\le \lim_{n\to\infty} \int \varphi_2 \,\mathrm{d}\mathbb{P}_n \\ &= \int \varphi_2 \,\mathrm{d}\mathbb{P} \\ &\le F(t+\epsilon). \end{aligned}$$

By taking $\epsilon \to 0$, since $t \in \Phi$, we have

$$\lim_{n\to\infty} F_n(t) = F(t).$$

Remark 9.1. *Similar proof steps as above hold if* $C_b(\mathbb{R})$ *is replaced with* $C_b^k(\mathbb{R})$*, the set of* k*-differentiable bounded functions for* $k \ge 1$*.*

'$\Leftarrow$': We first show that $\mathbb{R} \setminus \Phi$ (the set of discontinuity points) is countable. Suppose $x \in \mathbb{R} \setminus \Phi$. Let

$$a(x) = \sup\{F(y) : y < x\},$$
$$b(x) = \inf\{F(y) : y > x\}.$$

Because the set of rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}$, $\exists\, r_x \in (a(x), b(x)),$ s. t. $r_x \in \mathbb{Q}$. Since the intervals $(a(x), b(x))$ are disjoint for all $x \in \mathbb{R} \setminus \Phi$, the mapping $x \mapsto r_x$ is one-to-one. Therefore, $\mathbb{R} \setminus \Phi$ is countable and the claim is proved. This implies that Φ is dense.

Note that $\mathbb{P}((-a, a]^c) = F(-a) + 1 - F(a)$. We then have $\forall\, \epsilon > 0$,

$$\exists \pm M(\epsilon) \in \Phi, \text{ s. t. } \quad \mathbb{P}((-M(\epsilon), M(\epsilon)]^c) \leq \epsilon.$$

Furthermore, we have

$$F_n(M(\epsilon)) \to F(M(\epsilon)),$$
$$F_n(-M(\epsilon)) \to F(-M(\epsilon)).$$

Therefore, $\forall\, n$ sufficiently large, $\mathbb{P}_n\left((-M(\epsilon), M(\epsilon)]^c\right) \leq 2\epsilon$. Choose $-M(\epsilon) = x_1^k \leq x_2^k \leq \ldots \leq x_k^k = M(\epsilon),\ x_i^k \in \Phi$ such that $\lim_{k\to\infty} \max |x_{i+1}^k - x_i^k| = 0$. For $f \in C_b(\mathbb{R})$, let

$$f_k(x) = \sum_{1<i\leq k} f(x_i^k)\mathbf{1}_{\left(x_{i-1}^k, x_i^k\right]}(x) \in C_b(\mathbb{R}).$$

Taking the expectation over $\mathbb{P}_n$, we obtain

$$\begin{aligned}
\int f_k \,\mathrm{d}\mathbb{P}_n &= \sum_{1<i\leq k} f(x_i^k)\left(F_n(x_i^k) - F_n(x_{i-1}^k)\right) \\
&\xrightarrow{n\to\infty} \sum_{1<i\leq k} f(x_i^k)\left(F(x_i^k) - F(x_{i-1}^k)\right) \\
&= \int f_k \,\mathrm{d}\mathbb{P}. \qquad (23)
\end{aligned}$$

Let

$$\eta_k(M(\epsilon)) = \sup_{|x|\leq M(\epsilon)} |f_k(x) - f(x)|\,.$$

Since $f \in C_b(\mathbb{R})$ is continuous, it is uniformly continuous on $[-M(\epsilon), M(\epsilon)]$ and we have $\lim_{k\to\infty} \eta_k(M(\epsilon)) = 0$. We have

$$\begin{aligned}
\left|\int f \,\mathrm{d}\mathbb{P} - \int f_k \,\mathrm{d}\mathbb{P}\right| &\leq 2\|f\|_\infty \mathbb{P}\left((-M(\epsilon), M(\epsilon)]^c\right) + \eta_k(M(\epsilon)) \\
&\leq 2\|f\|_\infty \epsilon + \eta_k(M(\epsilon)), \qquad (24)
\end{aligned}$$

and for n large,

$$\begin{aligned}
\left|\int f \,\mathrm{d}\mathbb{P}_n - \int f_k \,\mathrm{d}\mathbb{P}_n\right| &\leq 2\|f\|_\infty \mathbb{P}_n\left((-M(\epsilon), M(\epsilon)]^c\right) + \eta_k(M(\epsilon)) \\
&\leq 2\|f\|_\infty \epsilon + \eta_k(M(\epsilon)). \qquad (25)
\end{aligned}$$

From (23) to (25), we therefore obtain

$$\begin{aligned}
&\left|\int f \,\mathrm{d}\mathbb{P}_n - \int f \,\mathrm{d}\mathbb{P}\right| \\
\leq &\left|\int f \,\mathrm{d}\mathbb{P}_n - \int f_k \,\mathrm{d}\mathbb{P}_n\right| + \left|\int f_k \,\mathrm{d}\mathbb{P}_n - \int f_k \,\mathrm{d}\mathbb{P}\right| + \left|\int f_k \,\mathrm{d}\mathbb{P} - \int f \,\mathrm{d}\mathbb{P}\right| \to 0,
\end{aligned}$$

as $n \to \infty$, then $k \to \infty$ and finally $\epsilon \to 0$. $\square$

Lemma 9.2. *If* $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$*, g is a continuous mapping, then*

$$\mathbb{P}_n \circ g^{-1} \xrightarrow{\mathrm{d}} \mathbb{P}_n \circ g^{-1},$$

or in other words,

$$X_n \xrightarrow{\mathrm{d}} X \implies g(X_n) \xrightarrow{\mathrm{d}} g(X).$$

Proof. Since $f \in C_b(\mathbb{R}) \implies f \circ g \in C_b(\mathbb{R})$, the result follows. $\square$

Lemma 9.3. *If* $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$*, then* $\exists$ *r.v.s* $Y_n \sim$ *cdf* F_n*,* $Y \sim$ *cdf* F*,* s. t. $Y_n \to Y$ *a.s.*

Proof. From Lemma 8.2, we know that, for the probability space $((0,1), \mathcal{B}(0,1), \lambda)$, where λ is Lebesgue measure, we have r.v.s

$$F_n^{-1}(\omega) = \inf\{y : F_n(y) \geq \omega\} \sim \text{cdf } F_n,$$
$$F^{-1}(\omega) = \inf\{y : F(y) \geq \omega\} \sim \text{cdf } F.$$

Let Ω_0 be the set of continuity points of F^{-1}. A similar argument as that in the proof of Theorem 9.1 shows that Ω_0^c is countable, hence $\lambda(\Omega_0) = 1$. In the following, we show that $F_n^{-1}(\omega) \to F^{-1}(\omega)$ for all $\omega \in \Omega_0$.

For any continuity point y of F such that $y < F^{-1}(\omega)$, we have $F(y) < \omega$ (cf. Week 8), hence for all n sufficiently large, $F_n(y) < \omega$ since $F_n(y) \to F(y)$ from Theorem 9.1. Thus, $F_n^{-1}(\omega) \geq y$ and $\liminf_{n\to\infty} F_n^{-1}(\omega) \geq F^{-1}(\omega)$ by taking $y \to F^{-1}(\omega)$ (note that set of continuity points of F is dense).

For any continuity point y of F such that $y > F^{-1}(\omega)$, $F(y) > \omega$ since $\omega \in \Omega_0$ (cf. Week 8), hence for all n sufficiently large, $F_n(y) > \omega$. Thus, $F_n^{-1}(\omega) \leq y$ and $\limsup_{n\to\infty} F_n^{-1}(\omega) \leq F^{-1}(\omega)$. $\square$

Lemma 9.4. *If for any subsequence* $(n(k))_{k\geq 1}$*,* $\exists$ *subsubsequence* $(n(k(r)))_{r\geq 1}$ s. t. $\mathbb{P}_{n(k(r))} \xrightarrow{\mathrm{d}} \mathbb{P}$*, then*

$$\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}.$$

Proof. Suppose otherwise. Then $\exists\, f \in C_b(\mathbb{R}), \epsilon > 0$, sequence $(n(k))_{k\geq 1}$, such that

$$\left|\int f \,\mathrm{d}\mathbb{P}_{n(k)} - \int f \,\mathrm{d}\mathbb{P}\right| > \epsilon,$$

for all $k \geq 1$. Then, any subsequence of distributions indexed by $(n(k(r)))_{r\geq 1}$ cannot converge, a contradiction. $\square$

Lemma 9.5. $X_n \xrightarrow{\mathrm{p}} X \implies X_n \xrightarrow{\mathrm{d}} X$*.*

Proof. From Corollary 6.1, for any sequence $(n(k))_{k\geq 1}$, $\exists$ subsequence $(n(k(r)))_{r\geq 1}$, s. t. $X_{n(k(r))} \to X$ a.s. Let $f \in C_b(\mathbb{R})$. By DCT, we obtain

$$\lim_{r\to\infty} \mathbb{E}\big[f(X_{n(k(r))})\big] = \mathbb{E}\Big[\lim_{r\to\infty} f(X_{n(k(r))})\Big] = \mathbb{E}[f(X)],$$

Therefore, $X_{n(k(r))} \xrightarrow{\mathrm{d}} X$. From Lemma 9.4, we have $X_n \xrightarrow{\mathrm{d}} X$. $\square$

In general, we can prove $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$ in two steps:

(i) Show that for all sequences $(n(k))_{k\geq 1}$, $\exists\, (n(k(r)))_{r\geq 1}$, s. t. $\mathbb{P}_{n(k(r))}$ converges. This is done using the concept of uniform tightness, which we discuss below.

(ii) Show that all subsequential distribution limits above are the same $\mathbb{P}$. This is done through the use of characteristic functions, which we discuss next week.

9.2 Uniform Tightness

Definition 9.2 (Uniform tightness). *$(\mathbb{P}_n)_{n\geq 1}$ is uniformly tight if $\forall\,\epsilon > 0$, $\exists$ compact $K \subset \mathbb{R}$ such that*

$$\mathbb{P}_n(K) \geq 1-\epsilon,\ \forall\, n \geq 1.$$

Theorem 9.2 (Helly's selection theorem). *If $(\mathbb{P})_{n\geq 1}$ is uniformly tight, then $\exists\ (n(k))_{k\geq 1}$ such that*

$$\mathbb{P}_{n(k)} \xrightarrow{\mathrm{d}} \mathbb{P},\ \textit{for some } \mathbb{P}.$$

Lemma 9.6 (Cantor's Diagonalization). *Suppose A is a countable set, and $f_n : A \mapsto \mathbb{R}$. Then, $\exists\ (n(k))_{k\geq 1}$,* s. t. *$f_{n(k)}(a)$ converges (or goes to $\pm\infty$), $\forall\, a \in A$.*

Proof. Let $A = \{a_1, a_2, \ldots\}$. From Lemma 1.5, we have

$$\begin{aligned}
&\exists\ (n_1(k))_{k\geq 1},\ \text{s.t.}\quad f_{n_1(k)}(a_1) \text{ converges},\\
&\exists\ (n_2(k))_{k\geq 1} \subset (n_1(k))_{k\geq 1},\ \text{s.t.}\quad f_{n_2(k)}(a_2) \text{ converges},\\
&\ldots
\end{aligned}$$

Therefore, for the diagonal sequence $(n_k(k))_{k\geq 1}$, $f_{n_k(k)}(a_l)$ converges $\forall\, l$. □

Proof of Theorem 9.2. Since $\mathbb{Q}$ is a dense subset of $\mathbb{R}$ and countable, by Lemma 9.6, $\exists\ (n(k))_{k\geq 1}$,

$$\text{s.t.}\quad F_{n(k)}(q) \to F(q), \forall\, q \in \mathbb{Q}.$$

Now we extend the definition of F on $\mathbb{Q}$ to $\mathbb{R}$ by defining $F(x) = \inf\big\{F(q) : x < q, q \in \mathbb{Q}\big\}$. We prove that $F(x)$ is a cdf.

- It is obvious $F(x)$ is a non-decreasing function.
- It is right-continuous because

$$\begin{aligned}
\lim_{x_n \downarrow x} F(x_n) &= \lim_{n\to\infty} \inf\{F(q) : x_n < q \in \mathbb{Q}\}\\
&= \inf_{n\geq 1} \inf\{F(q) : x_n < q \in \mathbb{Q}\}\\
&= \inf\{F(q) : x < q \in \mathbb{Q}\}\\
&= F(x),
\end{aligned}$$

  where the second equality follows because $\inf\{F(q) : x_n < q \in \mathbb{Q}\}$ is a decreasing sequence for $x_n \downarrow x$.
- Since $(\mathbb{P}_n)_{n\geq 1}$ is uniformly tight, $\forall\,\epsilon > 0$, $\exists\,[-M, M]$,

$$\text{s.t.}\quad 1 - F_n(M) + F_n(-M) \leq \epsilon,\ \forall\, n.$$

  Choose $r < -M < M < s$, $r, s \in \mathbb{Q}$. We have

$$1 - F(s) + F(r) = \lim_{k\to\infty} \big(1 - F_{n(k)}(s) + F_{n(k)}(r)\big) \leq \epsilon,$$

  which implies

$$\begin{aligned}
\lim_{x\to-\infty} F(x) = 0,\\
\lim_{x\to\infty} F(x) = 1.
\end{aligned}$$

Therefore, $F(x)$ is a cdf.

Let x be a continuity point of F and $a, b \in \mathbb{Q},\ \text{s.t.}\ a < x < b$. We have

$$F(a) = \lim_{k\to\infty} F_{n(k)}(a) \le \liminf_{k\to\infty} F_{n(k)}(x) \le \limsup_{k\to\infty} F_{n(k)}(x) \le \lim_{k\to\infty} F_{n(k)}(b) = F(b).$$

Taking $a \uparrow x$ and $b \downarrow x$, we obtain

$$\lim_{k\to\infty} F_{n(k)}(x) = F(x).$$

From Theorem 9.1, we have

$$\mathbb{P}_{n(k)} \xrightarrow{\mathrm{d}} \mathbb{P}.$$

□

Lemma 9.7. *If $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$, then $(\mathbb{P}_n)_{n\ge 1}$ is uniformly tight.*

Proof. Let $\epsilon > 0$. Choose $M > 0,\ \text{s.t.}\ \mathbb{P}([-M, M]^c) \le \epsilon$. Let

$$f(x) = \begin{cases} 0, & \text{if } |x| \le M, \\ \text{linear}, & \text{if } M < x < 2M, \text{or} -2M < x < -M, \\ 1, & \text{if } |x| \ge 2M. \end{cases}$$

Note that $f \in C_b(\mathbb{R})$. We have

$$\limsup_{n\to\infty} \mathbb{P}_n([-2M, 2M]^c) \le \limsup_{n\to\infty} \int f(x)\,\mathrm{d}\mathbb{P}_n(x) = \int f(x)\,\mathrm{d}\mathbb{P}(x) \le \mathbb{P}([-M, M]^c) \le \epsilon.$$

Therefore, $\exists\, n_0,\ \text{s.t.}\ \forall\, n \ge n_0, \mathbb{P}_n([-2M, 2M]^c) \le 2\epsilon$. For each $n < n_0$, we choose M_n such that

$$\mathbb{P}_n([-M_n, M_n]^c) \le 2\epsilon.$$

Let $M^* = \max\{M_1, M_2, \dots, M_{n-1}, 2M\}$. We then have $\mathbb{P}_n([-M^*, M^*]^c) \le 2\epsilon, \forall\, n \ge 1$. □

10.1 Characteristic Functions

Given a random variable $X : \Omega \mapsto \mathbb{R}$, the characteristic function of X is $\varphi : \mathbb{R} \mapsto \mathbb{C}$ given by

$$\begin{aligned}\varphi(t) &= \mathbb{E}\big[e^{\mathfrak{i}tX}\big], \\ &= \mathbb{E}[\cos(tX)] + \mathfrak{i}\mathbb{E}[\sin(tX)],\end{aligned}$$

where $\mathfrak{i} = \sqrt{-1}$. This characteristic function is always well-defined as cos and sin are bounded and integrable, even if $\mathbb{E}X$ does not exist. If X has pdf $f \in L^1(\mathbb{R})$, then $\varphi(t) = \int e^{itx} f(x)\,\mathrm{d}x$, which is essentially the Fourier transform of f.

Some simple properties:

$$\begin{aligned}\varphi(0) &= 1, \\ \varphi(-t) &= \overline{\varphi(t)}, \\ |\varphi(t)| &\leq 1, \\ |\varphi(t+h) - \varphi(t)| &\leq \mathbb{E}|e^{\mathfrak{i}hX} - 1|.\end{aligned}$$

Since $e^{\mathfrak{i}hX} - 1$ does not depend on t, φ is uniformly continuous in $\mathbb{R}$.

Example 10.1. *The characteristic function of the Gaussian distribution $N(\mu, \sigma^2)$ is*

$$\varphi(t) = e^{\mathfrak{i}\mu t - \frac{\sigma^2 t^2}{2}}.$$

In particular, when $\mu = 0, \sigma^2 = 1$,

$$\varphi(t) = e^{-\frac{t^2}{2}}.$$

Lemma 10.1. *If $\mathbb{E}|X|^r < \infty$ for integer $r \geq 1$, then $\varphi(t) \in C^r(\mathbb{R})$ and $\varphi^{(j)}(t) = \mathbb{E}\big[(\mathfrak{i}X)^j e^{\mathfrak{i}tX}\big]$ for $j \leq r$.*

Proof. We start with a simple fact. For $a \leq b$, we have

$$\begin{aligned}|e^{\mathfrak{i}a} - e^{\mathfrak{i}b}| &= \left|\int_a^b \mathfrak{i}e^{\mathfrak{i}t}\,\mathrm{d}t\right| \\ &\leq \int_a^b |\mathfrak{i}e^{\mathfrak{i}t}|\,\mathrm{d}t \\ &= |b - a|.\end{aligned}$$

The proof of Lemma 10.1 is by induction on r. For $r = 1$, we have $\left|\dfrac{e^{itX} - e^{isX}}{t - s}\right| \leq |X|$. Since $\mathbb{E}|X| < \infty$, from the DCT (Theorem 4.3), we obtain

$$\varphi'(t) = \lim_{s \to t} \mathbb{E}\left[\frac{e^{\mathfrak{i}tX} - e^{\mathfrak{i}sX}}{t - s}\right] = \mathbb{E}\big[\mathfrak{i}Xe^{\mathfrak{i}tX}\big] \in C(\mathbb{R}),$$

where the set inclusion follows from a further application of the DCT. Therefore, $\varphi \in C^1(\mathbb{R})$.

Assume that the lemma holds for r. Then for $r+1$, by assumption, we have

$$\begin{aligned}
&\varphi^{(r)}(t) = \mathbb{E}\big[(\mathrm{i}X)^r e^{\mathrm{i}tX}\big] \in C^r(\mathbb{R}), \\
&\left|\frac{(\mathrm{i}X)^r e^{\mathrm{i}tX} - (\mathrm{i}X)^r e^{\mathrm{i}sX}}{t-s}\right| \leq |X|^{r+1},
\end{aligned}$$

and from the DCT, we have

$$\varphi^{(r+1)}(t) = \lim_{s \to t} \mathbb{E}\left[\frac{(\mathrm{i}X)^r e^{\mathrm{i}tX} - (\mathrm{i}X)^r e^{\mathrm{i}sX}}{t-s}\right] = \mathbb{E}\Big[(\mathrm{i}X)^{(r+1)} e^{\mathrm{i}tX}\Big] \in C(\mathbb{R}),$$

and the proof is complete. □

Suppose $X \perp\!\!\!\perp Y$. We then have $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$. We define

$$\begin{aligned}
\mathbb{P}_X * \mathbb{P}_Y(A) &\triangleq \mathbb{P}_{X+Y}(A) \\
&= \mathbb{E}\big[\mathbf{1}_{\{X+Y \in A\}}(\omega)\big] \\
&\overset{\text{Fubini}}{=} \int\int \mathbf{1}_{\{x+y \in A\}} \,\mathrm{d}\mathbb{P}_X(x)\,\mathrm{d}\mathbb{P}_Y(y) \\
&= \int\int \mathbf{1}_{\{x \in A-y\}} \,\mathrm{d}\mathbb{P}_X(x)\,\mathrm{d}\mathbb{P}_Y(y) \\
&= \int \mathbb{P}_X(A-y)\,\mathrm{d}\mathbb{P}_Y(y),
\end{aligned}$$

where $A - y = \{a - y : a \in A\}$. If X has pdf p_X, then

$$\begin{aligned}
\mathbb{P}_X * \mathbb{P}_Y(A) &= \int\int \mathbf{1}_{\{x+y \in A\}} p_X(x)\,\mathrm{d}x\,\mathrm{d}\mathbb{P}_Y(y) \\
&= \int\int \mathbf{1}_{\{z \in A\}} p_X(z-y)\,\mathrm{d}z\,\mathrm{d}\mathbb{P}_Y(y) \\
&= \int\int_A p_X(z-y)\,\mathrm{d}z\,\mathrm{d}\mathbb{P}_Y(y) \\
&\overset{\text{Fubini}}{=} \int_A\int p_X(z-y)\,\mathrm{d}\mathbb{P}_Y(y)\,\mathrm{d}z,
\end{aligned}$$

hence the pdf of $\mathbb{P}_X * \mathbb{P}_Y$ is $\int p_X(z-y)\,\mathrm{d}\mathbb{P}_Y(y)$, i.e., $\mathbb{P}_X * \mathbb{P}_Y \ll$ Lebesgue measure.

If X and Y have pdfs p_X and p_Y respectively, then the pdf of $\mathbb{P}_X * \mathbb{P}_Y$ is given by

$$p_{X+Y}(z) = \int p_X(z-y) p_Y(y)\,\mathrm{d}y,$$

which is the convolution of the pdfs p_X and p_Y.

A special case is

$$\mathbb{P}^\sigma = \mathbb{P} * \mathcal{N}(0, \sigma^2) \ll \text{Lebesgue measure},$$

therefore, its pdf exists and it can be shown to be given by

$$p^\sigma(x) = \frac{1}{2\pi}\int \varphi(t) e^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2}\,\mathrm{d}t, \tag{26}$$

where φ is the characteristic function of $\mathbb{P}$.

Lemma 10.2 (Uniqueness)**.** *If $\varphi_X(t) = \varphi_Y(t)$, then $\mathbb{P}_X = \mathbb{P}_Y$.*

Proof. Assume $\varphi_X(t) = \varphi_Y(t)$, then from (26), we have $\mathbb{P}_X^\sigma = \mathbb{P}_Y^\sigma$. We have

$$\begin{aligned} X(\omega) + \sigma Z(\omega) &\xrightarrow{\sigma\to 0} X(\omega) \ \text{ a. s.}, \\ Y(\omega) + \sigma Z(\omega) &\xrightarrow{\sigma\to 0} Y(\omega) \ \text{ a. s.}, \end{aligned}$$

and Lemma 9.5 yields

$$\mathbb{P}_X^\sigma \xrightarrow{\mathrm{d}} \mathbb{P}_X, \ \mathbb{P}_Y^\sigma \xrightarrow{\mathrm{d}} \mathbb{P}_Y.$$

Therefore, $\mathbb{P}_X = \mathbb{P}_Y$. □

Lemma 10.3 (Fourier Inversion)**.** *Suppose that $\varphi(t)$ is the characteristic function of $\mathbb{P}$. If $\int |\varphi(t)| \,\mathrm{d}t < \infty$, then $\mathbb{P}$ has pdf*

$$p(x) = \frac{1}{2\pi} \int \varphi(t) e^{-\mathrm{i}tx} \,\mathrm{d}t.$$

Proof. We note that

$$\begin{aligned} &\varphi(t) e^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2} \xrightarrow{\sigma\to 0} \varphi(t) e^{-\mathrm{i}tx}, \ \text{and} \\ &\left| \varphi(t) e^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2} \right| \le |\varphi(t)|. \end{aligned}$$

By the DCT (Theorem 4.3), we then have

$$p^\sigma(x) \xrightarrow{\sigma\to 0} p(x).$$

Since $\mathbb{P}^\sigma \xrightarrow{\mathrm{d}} \mathbb{P}$ and p^σ is the pdf of $\mathbb{P}^\sigma$, we have

$$\int g(x) p^\sigma(x) \,\mathrm{d}x \xrightarrow{\sigma\to 0} \int g(x) \,\mathrm{d}\mathbb{P}(x), \ \forall\, g \in C_b(\mathbb{R}). \tag{27}$$

Now consider a $g \in C_c(\mathbb{R})$.[2] Since $|p^\sigma(x)| \le \frac{1}{2\pi} \int |\varphi(t)| \,\mathrm{d}t < \infty$, from the DCT (Theorem 4.3), we obtain

$$\int g(x) p^\sigma(x) \,\mathrm{d}x \xrightarrow{\sigma\to 0} \int g(x) p(x) \,\mathrm{d}x.$$

Together with (27), we have $\int g(x) \,\mathrm{d}\mathbb{P}(x) = \int g(x) p(x) \,\mathrm{d}x$ and

$$\int_A \mathrm{d}\mathbb{P}(x) = \int_A p(x) \,\mathrm{d}x,$$

for all compact sets A. From Ulam's Theorem (Theorem 3.3), $p(x)$ is the pdf of $\mathbb{P}$. □

We want to find out under what condition convergence of a sequence of characteristic functions $\varphi_n(t)$ implies weak convergence of the corresponding probability distributions $\mathbb{P}_n$. The following example shows that this implication is not always true.

[2] $C_c(\mathbb{R})$ is the set of continuous functions with compact support.

Example 10.2. *The distribution $\mathcal{N}(0,n)$ has characteristic function $\varphi_n(t) = e^{-\frac{nt^2}{2}} \xrightarrow{n\to\infty} 0,\ \forall\, t \neq 0$ and $\varphi_n(0) = 1,\ \forall\, n$. Therefore $\varphi_n(t)$ converges. But for all $x \in \mathbb{R}$,*

$$\mathbb{P}_n((-\infty, x)) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2n}}\right)\right) \to \frac{1}{2},$$

when $n \to \infty$, which implies that this sequence of distributions does not converge weakly.

On the other hand, we have the following converse.

Lemma 10.4. *If $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$ then $\varphi_n(t) \to \varphi(t),\ \forall\, t \in \mathbb{R}$.*

Proof. Obvious because $\varphi_n(t) = \mathbb{E}\left[e^{\mathrm{i}tX_n}\right] = \mathbb{E}[\cos tX_n] + \mathrm{i}\mathbb{E}[\sin tX_n]$, $\cos(t\cdot)$ and $\sin(t\cdot)$ are bounded continuous functions, and $X_n \xrightarrow{\mathrm{d}} X$. □

Lemma 10.5. *Suppose that $(\mathbb{P}_n)_{n\geq 1}$ is uniformly tight on $\mathbb{R}$ and $\varphi_n(t) = \int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_n(x) \xrightarrow{n\to\infty} \varphi(t)$. Then $\varphi(t) = \int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}(x)$ for some probability distribution $\mathbb{P}$ and $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$.*

Proof. From Helly's selection theorem (Theorem 9.2), for every subsequence $N = (n(k))_{k\geq 1}$, there exists a subsequence $(n(k(r)))_{r\geq 1}$ such that $\mathbb{P}_{n(k(r))} \xrightarrow{\mathrm{d}} \mathbb{P}_N$ for some $\mathbb{P}_N$ as $r \to \infty$. Therefore, since $e^{\mathrm{i}tx}$ is bounded and continuous, we have from Definition 9.1 that

$$\int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_{n(k(r))} \to \int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_N.$$

From the lemma assumption, the L.H.S. converges to $\varphi(t)$, therefore $\int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_N = \varphi(t)$. By Lemma 10.2, $\mathbb{P}_N = \mathbb{P}$ is the same for all choices of $N = (n(k))_{k\geq 1}$. The result then follows from Lemma 9.4. □

Theorem 10.1 (Levy's Continuity Theorem)**.** *Suppose $\mathbb{P}_n$ has characteristic function $\varphi_n(t) \to \varphi(t)$, and $\varphi(t)$ is continuous at $t = 0$. Then there exists $\mathbb{P}$ s.t. $\varphi(t)$ is the characteristic function of $\mathbb{P}$ and $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$.*

Proof. From Lemma 10.5, it suffices to show that $(\mathbb{P}_n)$ is uniformly tight. Since $\varphi(0) = \lim_{n\to\infty} \varphi_n(0) = 1$, $\forall\, \epsilon > 0,\ \exists\, \delta > 0$ *s.t.* $|\varphi(t) - 1| < \epsilon$ if $|t| \leq \delta$. Letting $\Re(\cdot)$ denote the real part, we then have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{\delta}\int_0^\delta (1 - \Re(\varphi_n(t)))\,\mathrm{d}t \\
&\overset{\text{DCT}}{=} \frac{1}{\delta}\int_0^\delta (1 - \Re(\varphi(t)))\,\mathrm{d}t \\
&\leq \frac{1}{\delta}\int_0^\delta |1 - \varphi(t)|\,\mathrm{d}t \\
&< \epsilon.
\end{aligned} \tag{28}$$

For each $n \geq 1$, we have

$$\begin{aligned}
\frac{1}{\delta}\int_0^\delta (1 - \Re(\varphi_n(t)))\,\mathrm{d}t &= \frac{1}{\delta}\int_0^\delta \int_{\mathbb{R}} (1 - \cos(tx))\,\mathrm{d}\mathbb{P}_n(x)\,\mathrm{d}t \\
&\overset{\text{Fubini}}{=} \frac{1}{\delta}\int_{\mathbb{R}}\int_0^\delta (1 - \cos(tx))\,\mathrm{d}t\,\mathrm{d}\mathbb{P}_n(x) \\
&= \int_{\mathbb{R}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right)\mathrm{d}\mathbb{P}_n(x) \\
&\geq \int_{|\delta x|\geq \pi} \left(1 - \frac{\sin(\delta x)}{\delta x}\right)\mathrm{d}\mathbb{P}_n(x) \\
&\geq \frac{1}{2}\mathbb{P}_n\left(\left\{x \,:\, |x| \geq \frac{\pi}{\delta}\right\}\right),
\end{aligned}$$

since $\operatorname{sinc} y = \frac{\sin \pi y}{\pi y} \le 1$ for $|y| \le 1$ and $\le 1/2$ for $|y| \ge 1$. From (28), for all n sufficiently large, we have

$$\mathbb{P}_n\Big(\Big\{x \,:\, |x| \ge \frac{\pi}{\delta}\Big\}\Big) \le 4\epsilon.$$

Therefore $(\mathbb{P}_n)$ is uniformly tight. $\square$

An application of Theorem 10.1 is to show that if $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P},\ \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{Q}$, then $\mathbb{P}_n \times \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{P} \times \mathbb{Q}$. For $t = (t_1, t_2)$ and $x = (x_1, x_2)$, we have

$$\begin{aligned}
&\int e^{\mathrm{i}\langle t,x\rangle} \,\mathrm{d}\mathbb{P}_n \times \mathbb{Q}_n(x) \\
&\overset{\text{Fubini}}{=} \int e^{\mathrm{i}t_1x_1} \,\mathrm{d}\mathbb{P}_n(x_1) \int e^{\mathrm{i}t_2x_2} \,\mathrm{d}\mathbb{Q}_n(x_2) \\
&\to \int e^{\mathrm{i}t_1x_1} \,\mathrm{d}\mathbb{P}(x_1) \int e^{\mathrm{i}t_2x_2} \,\mathrm{d}\mathbb{Q}(x_2) \text{ since } \mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P},\ \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{Q} \\
&\overset{\text{Fubini}}{=} \int e^{\mathrm{i}\langle t,x\rangle} \,\mathrm{d}\mathbb{P} \times \mathbb{Q}(x),
\end{aligned}$$

which is the characteristic function of $\mathbb{P} \times \mathbb{Q}$ and is thus continuous at 0. From Theorem 10.1, we then have

$$\mathbb{P}_n \times \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{P} \times \mathbb{Q}.$$

Exercise 10.1. *Show that if $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P},\ \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{Q}$, then $\mathbb{P}_n * \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{P}\mathbb{Q}$. (Hint: Let $g(x,y) = x + y$ and consider $\mathbb{P}_n \times \mathbb{Q}_n \circ g^{-1}$.)*

10.2 Central Limit Theorem for I.I.D. Sequences

Suppose that $f : \mathbb{R} \mapsto \mathbb{R}$ is k-differentiable at $a \in \mathbb{R}$. Then the Taylor series of f evaluated at a is given by

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + o(|x-a|^k).$$

Furthermore, if $f^{(k+1)}$ exists and is continuous in an open interval I containing a, then

$$o(|x-a|^k) \le \sup_I \left| f^{(k+1)} \right| \frac{|x-a|^{k+1}}{(k+1)!}.$$

Theorem 10.2 (CLT for i.i.d. sequences)**.** *Consider i.i.d. random variables $X_1, X_2, \ldots,$ with $\mathbb{E}X_i = \mu$ and $\operatorname{var} X_i = \sigma^2 < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{29}$$

Proof. Without loss of generality, we assume that $\mu = 0, \sigma = 1$. Let

$$\begin{aligned}
\varphi_n(t) &= \mathbb{E}\left[\exp\left(\mathrm{i}t\frac{S_n}{\sqrt{n}}\right)\right] \\
&= \prod_{j=1}^n \mathbb{E}\left[\exp\left(\mathrm{i}t\frac{X_j}{\sqrt{n}}\right)\right] \\
&= \left(\mathbb{E}e^{\mathrm{i}t\frac{X_1}{\sqrt{n}}}\right)^n \\
&= \varphi_1\left(\frac{t}{\sqrt{n}}\right)^n.
\end{aligned}$$

Since $\mathbb{E}X_1^2 < \infty$ we have $\varphi_1(t) \in C^2(\mathbb{R})$ from Lemma 10.1, and its Taylor series is

$$\varphi_1(t) = \varphi_1(0) + \varphi_1'(0)t + \frac{\varphi_1''(0)}{2!} + o(t^2).$$

Since $\varphi_1(0) = 1$, $\varphi_1'(0) = \mathbb{E}\big[\mathrm{i}X_1 e^{i0X_1}\big] = \mathrm{i}\mathbb{E}X_1 = 0$, $\varphi_1''(0) = \mathbb{E}\big[(\mathrm{i}X_1)^2\big] = -1$, we obtain

$$\varphi_1(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Therefore,

$$\begin{aligned}\varphi_n(t) &= \varphi_1\left(\frac{t}{\sqrt{n}}\right)^n \\ &= \left(1 - \frac{t^2}{2n} + o(t^2/n)\right)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}},\end{aligned}$$

the characteristic function of $\mathcal{N}(0,1)$. From Theorem 10.1, we then have

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

□

10.3 Stable Distributions

The above proof of the CLT does not give us much intuition of why the amazing result that is the CLT holds. Some intuition can be had from the following observation: If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. By induction, if $X_i \sim \mathcal{N}(0,1)$ for $i = 1, \dots, n$, we have $\frac{S_n}{\sqrt{n}} \sim \mathcal{N}(0,1)$. Therefore, the normal distribution has a "stability" property.

In general, suppose $X \sim \mathbb{P}$ and $X_1, X_2, \dots$ are i.i.d. copies of X. If

$$X \stackrel{d}{=} \frac{\sum_{j=1}^k X_j - a_k}{b_k}$$

for some sequences of constants (a_k) and (b_k), then we say that $\mathbb{P}$ is a stable distribution or stable law.

It turns out that all stable laws have characteristic functions given by

$$\varphi(t) = \exp(\mathrm{i}tc - b|t|^\alpha(1 + \mathrm{i}\kappa\,\mathrm{sgn}(t)\omega_\alpha(t))),$$

where $-1 \le \kappa \le 1$, $0 < \alpha \le 2$ and

$$\omega_\alpha(t) = \begin{cases} \tan(\pi\alpha/2), & \text{if } \alpha \ne 1, \\ \frac{2}{\pi}\log|t|, & \text{if } \alpha = 1. \end{cases}$$

In particular, if $\alpha = 2$, $\varphi(t) = \exp(\mathrm{i}tc - bt^2) \sim \mathcal{N}(c, 2b)$, and this is the only stable law with finite variance. Therefore, for (29) to converge in distribution, it **must** converge to the normal distribution. Of course, this does not explain why (29) must converge in distribution in the first place.

11.1 Lindeberg's Method

Recall the CLT for an i.i.d. sequence from the last session: Let $X_1, X_2, \ldots$ be an i.i.d. sequence with $\mathbb{E}X_i = 0$ and $\operatorname{var} X_i = 1$. We have

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

In this session, we show another proof of this CLT using Lindeberg's method. Let $G_1, G_2, \ldots$ be i.i.d. $\mathcal{N}(0,1)$, independent of $X_1, X_2, \ldots$. Let

$$T_{m,n} = \frac{1}{\sqrt{n}}(G_1 + \ldots + G_{m-1} + X_m + \ldots + X_n)$$

with

$$\begin{aligned}
T_{n+1,n} =& \frac{1}{\sqrt{n}}(G_1 + \ldots + G_n) \sim \mathcal{N}(0,1),\\
T_{1,n} =& \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n).
\end{aligned}$$

We aim to show that $T_{1,n} \xrightarrow{\mathrm{d}} T_{n+1,n}$ as $n \to \infty$. Let $f \in C_b(\mathbb{R})$ with $c_2 = \sup\left|f^{(2)}\right| < \infty$ and $c_3 = \sup\left|f^{(3)}\right| < \infty$. We have

$$\begin{aligned}
&|\mathbb{E}f(T_{1,n}) - \mathbb{E}f(T_{n+1,n})|\\
&= \left|\sum_{m=1}^{n} (\mathbb{E}f(T_{m,n}) - \mathbb{E}f(T_{m+1,n}))\right|\\
&\leq \sum_{m=1}^{n} |\mathbb{E}f(T_{m,n}) - \mathbb{E}f(T_{m+1,n})|.
\end{aligned}$$

Let

$$U_m = \frac{1}{\sqrt{n}}\left(G_1 + \ldots + G_{m-1} + X_{m+1} + \ldots + X_n\right).$$

Then,

$$\begin{aligned}
T_{m,n} &= U_m + \frac{X_m}{\sqrt{n}},\\
T_{m+1,n} &= U_m + \frac{G_m}{\sqrt{n}}.
\end{aligned}$$

Using Taylor series expansion at U_m (see Section 10.2), for any $\epsilon > 0$, we obtain

$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}} - f''(U_m)\frac{X_m^2}{2n}\right|\mathbf{1}_{\{|X_m|\le\epsilon\sqrt{n}\}} \le \frac{c_3|X_m|^3}{6n^{3/2}}\mathbf{1}_{\{|X_m|\le\epsilon\sqrt{n}\}} \le \frac{c_3\epsilon}{6n}X_m^2,$$
$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}}\right|\mathbf{1}_{\{|X_m|>\epsilon\sqrt{n}\}} \le \frac{c_2X_m^2}{2n}\mathbf{1}_{\{|X_m|>\epsilon\sqrt{n}\}}.$$

By combining the inequality $\left|f''(U_m)\frac{X_m^2}{2n}\right| \le \frac{c_2X_m^2}{2n}$ with the second inequality above, we obtain

$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}} - f''(U_m)\frac{X_m^2}{2n}\right|\mathbf{1}_{\{|X_m|>\epsilon\sqrt{n}\}} \le \frac{c_2X_m^2}{n}\mathbf{1}_{\{|X_m|>\epsilon\sqrt{n}\}}.$$

Therefore, we have

$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}} - f''(U_m)\frac{X_m^2}{2n}\right| \le \frac{c_3\epsilon}{6n}X_m^2 + \frac{c_2\ X_m^2}{n}\mathbf{1}_{\{|X_m|>\epsilon\sqrt{n}\}}.$$

Similarly, we can obtain

$$\left|f(T_{m+1,n}) - f(U_m) - f'(U_m)\frac{G_m}{\sqrt{n}} - f''(U_m)\frac{G_m^2}{2n}\right| \le \frac{c_3\epsilon}{6n}G_m^2 + \frac{c_2G_m^2}{n}\mathbf{1}_{\{|G_m|>\epsilon\sqrt{n}\}}.$$

Furthermore, by definition, we have

$$\mathbb{E}[f'(U_m)X_m] = \mathbb{E}[f'(U_m)]\mathbb{E}[X_m] = 0,$$
$$\mathbb{E}\big[f''(U_m)X_m^2\big] = \mathbb{E}[f''(U_m)]\mathbb{E}\big[X_m^2\big] = \mathbb{E}[f''(U_m)] = \mathbb{E}\big[f''(U_m)G_m^2\big],$$

and

$$|\mathbb{E}[f(T_{m,n})] - \mathbb{E}[f(T_{m+1,n})]| \le \frac{c_3\epsilon}{3n} + \frac{c_2}{n}\left(\mathbb{E}X_m^2\mathbf{1}_{\{|X_m|>\epsilon\sqrt{n}\}} + \mathbb{E}G_m^2\mathbf{1}_{\{|G_m|>\epsilon\sqrt{n}\}}\right).$$

Summing over $1 \le m \le n$, we have

$$|\mathbb{E}[f(T_{1,n})] - \mathbb{E}[f(T_{n+1,n})]| \le \frac{c_3\epsilon}{3} + c_2\left(\mathbb{E}X_1^2\mathbf{1}_{\{|X_1|>\epsilon\sqrt{n}\}} + \mathbb{E}G_1^2\mathbf{1}_{\{|G_1|>\epsilon\sqrt{n}\}}\right).$$

By the DCT (Theorem 4.3), we obtain

$$\mathbb{E}X_1^2\mathbf{1}_{\{|X_1|>\epsilon\sqrt{n}\}} \to \mathbb{E}X_1^2\mathbf{1}_{\{|X_1|=\infty\}} = 0,$$
$$\mathbb{E}G_1^2\mathbf{1}_{\{|G_1|>\epsilon\sqrt{n}\}} \to \mathbb{E}G_1^2\mathbf{1}_{\{|G_1|=\infty\}} = 0,$$

as $n \to \infty$. Therefore, we have $|\mathbb{E}[f(T_{1,n})] - \mathbb{E}[f(T_{n+1,n})]| \to 0$ by taking $n \to \infty$ and $\epsilon \to 0$.

11.2 Lindeberg's CLT

Theorem 11.1 (Lindeberg's CLT)**.** *For each $n \ge 1$, let $(X_{n,m})_{m=1}^n$ be a sequence of independent random variables with $\mathbb{E}X_{n,m} = 0$. Then, $S_n = \sum_{m=1}^n X_{n,m} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ as $n \to \infty$ if*

(i) $\sum_{m=1}^n \mathbb{E}X_{n,m}^2 \to 1$ *as* $n \to \infty$*; and*

(ii) $\forall\,\epsilon > 0$, $\sum_{m=1}^n \mathbb{E}X_{n,m}^2\mathbf{1}_{\{|X_{n,m}|>\epsilon\}} \to 0$ *as* $n \to \infty$.

Remark 11.1. *Let $Y_1, Y_2, \dots$ be i.i.d. $\mathbb{E}Y_1 = 0$, $\mathbb{E}Y_1^2 = 1$, and $X_{n,m} = Y_m/\sqrt{n}$. Then $\sum_{m=1}^n \mathbb{E}X_{n,m}^2 = 1$ and for all $\epsilon > 0$, we have*

$$\sum_{m=1}^n \mathbb{E}X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}|>\epsilon\}} = \mathbb{E}Y_1^2 \mathbf{1}_{\left\{|Y_1|>\epsilon\sqrt{n}\right\}} \xrightarrow{DCT} 0.$$

Therefore, the CLT for i.i.d. sequence follows from Lindeberg's CLT.

Proof. By Chebyshev's inequality, we have

$$\mathbb{P}(|S_n| > M) \leq \frac{1}{M^2} \sum_{m=1}^n \mathbb{E}X_{n,m}^2 \leq \frac{2}{M^2}, \ \forall n \text{ sufficiently large.}$$

Therefore, $(\mathbb{P}_{S_n})_{n\geq 1}$ is uniformly tight. From Lemma 10.5, to prove the theorem is equivalent to showing that the characteristic function $\varphi_{S_n}(t) \to e^{-\frac{t^2}{2}}$ as $n \to \infty$. We have

$$\log \varphi_{S_n}(t) = \log \left(\prod_{m=1}^n \mathbb{E}e^{\mathrm{i}tX_{n,m}} \right) = \sum_{m=1}^n \log\left(1 + \mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right).$$

By Taylor's series, we have the following elementary facts:

$$|\log(1+\xi) - \xi| \leq \xi^2 \text{ for } |\xi| \leq \frac{1}{2}, \tag{30}$$

$$\left| e^{\mathrm{i}a} - \sum_{k=0}^n \frac{(\mathrm{i}a)^k}{k!} \right| \leq \frac{|a|^{n+1}}{(n+1)!} \text{ for } a \in \mathbb{R}. \tag{31}$$

From (31) and $\mathbb{E}X_{n,m} = 0$, we have

$$\begin{aligned} &\left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right| \\ =&\left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1 - \mathrm{i}t\mathbb{E}X_{n,m}\right| \\ \leq&\frac{t^2}{2}\mathbb{E}X_{n,m}^2 \\ \leq&\frac{t^2}{2}\epsilon^2 + \frac{t^2}{2}\mathbb{E}X_{n,m}^2\mathbf{1}_{\{|X_{n,m}|>\epsilon\}}. \end{aligned} \tag{32}$$

Because $\mathbb{E}X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}>\epsilon|\}} \to 0$ as $n \to \infty$, there exists $n_0(m)$ such that for $n \geq n_0(m)$, we have

$$\left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right| \leq t^2\epsilon^2.$$

For $\epsilon \leq \dfrac{1}{t\sqrt{2}}$,

$$\left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right| \leq \frac{1}{2},$$

thus from (30) and (32), we have

$$\begin{aligned} &\sum_{m=1}^n \left|\log\left(1 + (\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1)\right) - \left(\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right)\right| \\ \leq &\sum_{m=1}^n \left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right|^2 \\ \leq &\sum_{m=1}^n \frac{t^4}{4}\left(\mathbb{E}X_{n,m}^2\right)^2 \\ \leq &\frac{t^4}{4} \max_{1\leq m\leq n} \mathbb{E}X_{n,m}^2 \sum_{m=1}^n \mathbb{E}X_{n,m}^2. \end{aligned}$$

We have

$$\max_{1\le m\le n} \mathbb{E}X_{n,m}^2 \le \epsilon^2 + \max_{1\le m\le n} \mathbb{E}X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}|>\epsilon\}} \to \epsilon^2 \text{ as } n\to\infty,$$

where the convergence follows from condition (ii). Combining the above result with condition (i), we obtain

$$\sum_{m=1}^{n} \left|\log\left(1+(\mathbb{E}e^{\mathrm{i}tX_{n,m}}-1)\right) - \left(\mathbb{E}e^{\mathrm{i}tX_{n,m}}-1\right)\right| \le \frac{t^4}{2}\epsilon^2 \text{ as } n\to\infty. \tag{33}$$

Next we show

$$\left|\sum_{m=1}^{n}\left(\mathbb{E}e^{\mathrm{i}tX_{n,m}}-1\right)+\frac{t^2}{2}\right|\to 0,$$

which, due to condition (i), is equivalent to showing that

$$\left|\sum_{m=1}^{n}\left(\mathbb{E}e^{\mathrm{i}tX_{n,m}}-1\right)+\frac{t^2}{2}\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2\right|\to 0.$$

From the Taylor series expansion and using a similar argument in Lindeberg's method, we obtain

$$\begin{aligned}
\left|e^{\mathrm{i}tX_{n,m}}-1-\mathrm{i}tX_{n,m}+\frac{t^2}{2}X_{n,m}^2\right|\mathbf{1}_{\{|X_{n,m}|>\epsilon\}} &\le t^2X_{n,m}^2\mathbf{1}_{\{|X_{n,m}|>\epsilon\}},\\
\left|e^{\mathrm{i}tX_{n,m}}-1-\mathrm{i}tX_{n,m}+\frac{t^2}{2}X_{n,m}^2\right|\mathbf{1}_{\{|X_{n,m}|\le\epsilon\}} &\le \frac{t^3\epsilon}{6}X_{n,m}^2.
\end{aligned}$$

We then obtain

$$\begin{aligned}
&\left|\sum_{m=1}^{n}\left(\mathbb{E}e^{\mathrm{i}tX_{n,m}}-1\right)+\frac{t^2}{2}\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2\right|\\
\le& t^2\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2\mathbf{1}_{\{|X_{n,m}|>\epsilon\}}+\frac{t^3\epsilon}{6}\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2\\
\le& \frac{t^3\epsilon}{6} \text{ as } n\to\infty.
\end{aligned}$$

Taking $\epsilon\to 0$, the theorem is proved. □

Theorem 11.2 (Berry-Essen Theorem)**.** *Let $(X_m)_{m\ge1}$ be a sequence of independent random variables, $\mathbb{E}X_m=0$. Let $\hat{F}_n$ be the cdf of $\frac{S_n}{\sqrt{\sum_{m=1}^n \mathbb{E}X_m^2}}$ and $G\sim\mathcal{N}(0,1)$. We have*

$$\sup_x\left|\hat{F}_n(x)-G(x)\right|\le\frac{10\sum_{m=1}^n\mathbb{E}|X_m|^3}{\left(\sum_{m=1}^n\mathbb{E}X_m^2\right)^{3/2}}.$$

For the special case where (X_m) is an i.i.d. sequence and $\mathbb{E}X_1^2=\sigma^2$, we have

$$\sup_x\left|\hat{F}_n(x)-G(x)\right|\le\frac{10\mathbb{E}|X_1|^3}{\sigma^3\sqrt{n}}.$$

This theorem shows how fast the convergence is. The proof is omitted here as it is quite technical and tedious. Please see Durrett if interested.

11.3 Kolmogorov's Three Series Theorem

Theorem 11.3 (Kolmogorov's Three Series Theorem)**.** *Let $(X_i)_{i\geq 1}$ be a sequence of independent random variables and let $Z_i = X_i \mathbf{1}_{\{|X_i|\leq 1\}}$. Then, $\sum_{i=1}^n X_i$ converges a.s. if and only if*

(i) $\sum_{i\geq 1} \mathbb{P}(|X_i| > 1) < \infty$.

(ii) $\sum_{i\geq 1} \mathbb{E}Z_i$ *converges.*

(iii) $\sum_{i\geq 1} \operatorname{var} Z_i < \infty$.

Proof. "$\Leftarrow$": Using condition (i), we have

$$\sum_{i\geq 1} \mathbb{P}(X_i \neq Z_i) = \sum_{i\geq 1} \mathbb{P}(|X_i| > 1) < \infty.$$

From the Borel-Cantelli Lemma (Lemma 6.1), we have

$$\mathbb{P}(X_i \neq Z_i \text{ i.o.}) = 0.$$

Therefore,

$$\sum_{i\geq 1} X_i \text{ converges } \iff \sum_{i\geq 1} Z_i \text{ converges.}$$

Using condition (ii), we have

$$\sum_{i\geq 1} Z_i \text{ converges } \iff \sum_{i\geq 1} (Z_i - \mathbb{E}Z_i) \text{ converges.}$$

The proof now follows from condition (iii) and the variance convergence criterion (Theorem 7.4).

"$\Rightarrow$": Suppose that $\sum_{i\geq 1} X_i$ converges a.s.
We first show condition (i). Since $\mathbb{P}(|X_i| > 1 \text{ i.o.}) = 0$, the Borel-Cantelli Lemma (Lemma 6.1) yields

$$\sum_{i\geq 1} \mathbb{P}(|X_i| > 1) < \infty.$$

We next show condition (iii) by contradiction. From condition (i), we have

$$\sum_{i\geq 1} X_i \text{ converges a.s. } \implies \sum_{i\geq 1} Z_i \text{ converges a.s.} \tag{34}$$

Therefore,

$$S(m,n) = \sum_{i=m}^{n} Z_i \to 0 \text{ a.s as } m,n \to \infty.$$

Thus, for any $\delta > 0$, we have

$$\mathbb{P}(|S(m,n)| > \delta) \leq \delta \text{ for sufficiently large } m,n. \tag{35}$$

Now assume $\sum_{i\geq 1} \operatorname{var} Z_i = \infty$. Then we have $\sigma^2(m,n) \triangleq \operatorname{var}(S(m,n)) = \sum_{i=m}^{n} \operatorname{var}(Z_i) \to \infty$ as $n \to \infty$ for fixed m. Let

$$T(m,n) = \frac{S(m,n) - \mathbb{E}S(m,n)}{\sigma(m,n)} = \sum_{i=m}^{n} \frac{Z_i - \mathbb{E}Z_i}{\sigma(m,n)}.$$

Let $\widetilde{Z}_i = Z_i - \mathbb{E}Z_i$. We have $\left|\widetilde{Z}_i\right| \leq 2$, and $\left|\frac{\widetilde{Z}_i}{\sigma(m,n)}\right| \to 0$ a.s as $n \to \infty$ because $\sigma(m,n) \to \infty$. For any $\epsilon > 0$, $\exists\, n(m)$ such that for all $n \geq n(m)$, we have $\left|\frac{\widetilde{Z}_i}{\sigma(m,n)}\right| \leq \epsilon$ a.s. Therefore, we obtain

$$\sum_{i=m}^{n} \mathbb{E}\left[\frac{\widetilde{Z}_i^2}{\sigma^2(m,n)}\mathbf{1}_{\left\{\left|\frac{\widetilde{Z}_i}{\sigma(m,n)}\right|>\epsilon\right\}}\right] = 0 \text{ a.s. for } n \geq n(m).$$

Furthermore, we have

$$\sum_{m=1}^{n} \mathbb{E}\left[\frac{\widetilde{Z}_i^2}{\sigma^2(n,m)}\right] = 1.$$

By Lindeberg's CLT, we therefore have

$$T(m,n) \xrightarrow{\text{d}} \mathcal{N}(0,1) \text{ as } m,n \to \infty. \tag{36}$$

But at the same time, we have

$$1 - \delta \leq \mathbb{P}(|S(m,n)| \leq \delta) = \mathbb{P}\left(\left|T(m,n) + \frac{\mathbb{E}S(m,n)}{\sigma(m,n)}\right| \leq \frac{\delta}{\sigma(m,n)}\right).$$

When $m,n \to \infty$, $\sigma(m,n) \to \infty$. Thus, $T(m,n)$ concentrates at a constant, which contradicts (36). Therefore, the assumption that $\sum_{i\geq 1} \operatorname{var} Z_i = \infty$ is false.

Finally, we show condition (ii). From the variance convergence criterion (Theorem 7.4), $\sum_{i\geq 1}(Z_i - \mathbb{E}Z_i)$ converges a.s. Thus, convergence of $\sum_{i\geq 1} \mathbb{E}Z_i$ follows from (34). $\square$

11.4 Levy's Equivalence Theorem

Theorem 11.4 (Levy's Equivalence Theorem)**.** *Let $(X_i)_{i\geq 1}$ be a sequence of independent random variables. Then, $\sum_{i\geq 1} X_i$ converges a.s. $\iff$ converges in probability $\iff$ converges in distribution.*

To prove the Theorem 11.4, we start with a few lemmas.

Lemma 11.1. *Suppose $X_1, X_2, \ldots$ are independent. Let $S_n = \sum_{i\leq n} X_i$. If $\mathbb{P}(|S_n - S_j| \geq a) \leq p < 1$ for all $j \leq n$, then for all $x > a$, we have*

$$\mathbb{P}\left(\max_{1\leq j\leq n} |S_j| \geq x\right) \leq \frac{1}{1-p}\mathbb{P}(|S_n| > x - a).$$

Proof. Let $\tau = \min\{j \leq n : |S_j| \geq x\}$ with $\tau = n+1$ if $|S_j| < x$ for all $j \leq n$. If $\tau = j$, then $|S_j| \geq x$ and

$$\{\omega : |S_n - S_j| < a,\ \tau = j\} \subset \{\omega : |S_n| > x - a,\ \tau = j\}.$$

This gives us

$$\begin{aligned}
\mathbb{P}\left(\max_{j\le n}|S_j|\ge x\right) &= \mathbb{P}(\tau\le n)\\
&= \sum_{j=1}^{n}\mathbb{P}(\tau=j)\\
&\le \frac{1}{1-p}\sum_{j=1}^{n}\mathbb{P}(|S_n-S_j|<a)\mathbb{P}(\tau=j)\\
&= \frac{1}{1-p}\sum_{j=1}^{n}\mathbb{P}(|S_n-S_j|<a,\ \tau=j) \text{ since } \{\tau=j\} \text{ depends only on } X_1,\dots,X_j\\
&\le \frac{1}{1-p}\sum_{j=1}^{n}\mathbb{P}(|S_n|>x-a,\ \tau=j)\\
&= \frac{1}{1-p}\mathbb{P}(|S_n|>x-a).
\end{aligned}$$

□

Lemma 11.2. *$Y_n\to Y$ a.s. iff* $\max_{i\ge n}|Y_i-Y|\xrightarrow{\text{p}}0$.

Proof. We have $\max_{i\ge n}|Y_i-Y|\to 0$ a.s. $\implies \max_{i\ge n}|Y_i-Y|\xrightarrow{\text{p}}0$.

To show the converse, let $M_n=\max_{i\ge n}|Y_i-Y|$, which is a decreasing sequence bounded below by 0. Therefore $M_n\downarrow M$ for some M, which implies that $\forall\,\epsilon>0$, $\mathbb{P}(M>\epsilon)\le\mathbb{P}(M_n>\epsilon)\to 0$ as $n\to\infty$. Therefore, from continuity of $\mathbb{P}$, $\mathbb{P}(M=0)=1$ and $M_n\to 0$ a.s. This is equivalent to saying that $Y_n\to Y$ a.s. □

Lemma 11.3. *(Y_n) converges in probability iff* $\lim_{n,m\to\infty}\mathbb{P}(|Y_m-Y_n|\ge\epsilon)=0$ *for all* $\epsilon>0$.

Proof. The proof in the "⇒" direction is trivial. To prove the converse, we note that the given condition implies that for all $k\ge 1$, $\exists\, n(k)$ such that $\forall\, n,m\ge n(k)$, we have

$$\mathbb{P}\left(|Y_m-Y_n|\ge\frac{1}{2^k}\right)\le\frac{1}{2^k}.$$

We can choose $n(k+1)\ge n(k)$ so that

$$\mathbb{P}\left(|Y_{n(k+1)}-Y_{n(k)}|\ge\frac{1}{2^k}\right)\le\frac{1}{2^k}.$$

Summing over $k\ge 1$, we have

$$\sum_{k\ge 1}\mathbb{P}\left(|Y_{n(k+1)}-Y_{n(k)}|\ge\frac{1}{2^k}\right)\le 1<\infty.$$

The Borel-Cantelli Lemma (Lemma 6.1) implies that $\mathbb{P}\big(|Y_{n(k+1)}-Y_{n(k)}|\ge\frac{1}{2^k}\text{ i.o.}\big)=0$ and therefore $\exists\, l$ such that for all $k\ge l$, $|Y_{n(k+1)}-Y_{n(k)}|<\frac{1}{2^k}$ a.s. and for all $j\ge k$, we have

$$|Y_{n(j)}-Y_{n(k)}|<\sum_{i\ge k}\frac{1}{2^i}\to 0 \text{ as } k\to\infty.$$

Therefore $(Y_{n(k)})_{k\ge 1}$ is Cauchy a.s., and it converges since $\mathbb{R}$ is complete. Hence, $\exists\, Y=\lim_{k\to\infty}Y_{n(k)}$. For any $\epsilon>0$, we can then choose $n(k)$ sufficiently large so that $\forall\, n\ge n(k)$,

$$\begin{aligned}
\mathbb{P}(|Y_n-Y|\ge 2\epsilon) &\le \mathbb{P}\big(|Y_n-Y_{n(k)}|\ge\epsilon\big)+\mathbb{P}\big(|Y_{n(k)}-Y|\ge\epsilon\big)\\
&\le 2\epsilon.
\end{aligned}$$

The lemma is now proved. □

Lemma 11.4. *Suppose that $(\mathbb{P}_{X_n})$ and $(\mathbb{P}_{Y_n})$ are both uniformly tight. Then $(\mathbb{P}_{X_n+Y_n})$ is uniformly tight.*

Proof. Exercise. $\square$

Lemma 11.5. *If $\mathbb{P}\mathbb{Q} = \mathbb{P}$, then $\mathbb{Q}(\{0\}) = 1$.*

Proof. Let $X \sim \mathbb{P}$ and $Y \sim \mathbb{Q}$ be independent with characteristic functions φ_X and φ_Y respectively. Since $\mathbb{P}\mathbb{Q} = \mathbb{P}$, we have $\varphi_X(t)\varphi_Y(t) = \varphi_X(t)$ for all $t \in \mathbb{R}$. Since $\varphi_X(t)$ is continuous and $\varphi_X(0) = 1$, $\exists\, \epsilon > 0$ such that $|\varphi_X(t)| > 0\ \forall\, |t| \leq \epsilon$. This implies that $\varphi_Y(t) = 1$ for such values of t. Therefore $\mathbb{E}[\cos tY] = 1$ but since $\cos(\cdot) \leq 1$, we must have $tY = 0 \mod 2\pi$ a.s.

Take $|s|, |t| \leq \epsilon$ with s/t being irrational. Then for each ω, we have

$$\begin{aligned} tY(\omega) &= 2\pi k,\ k \in \mathbb{Z}, \\ sY(\omega) &= 2\pi m,\ m \in \mathbb{Z}. \end{aligned}$$

If $Y(\omega) \neq 0$, $s/t = m/k$, a contradiction. Therefore $Y = 0$ a.s. $\square$

Proof of Theorem 11.4. We first show that $S_n = \sum_{i=1}^n X_i$ converges in probability implies convergence a.s. Suppose $S_n \xrightarrow{\text{p}} S$, i.e., for all $\epsilon > 0$, $\exists\ n(\epsilon)$ such that $\forall\, n \geq n(\epsilon)$,

$$\mathbb{P}(|S_n - S| > \epsilon) \leq \epsilon. \tag{37}$$

For $k, j \geq n(\epsilon)$, we have

$$\mathbb{P}(|S_k - S_j| \geq 2\epsilon) \leq \mathbb{P}(|S_k - S| \geq \epsilon) + \mathbb{P}(|S_j - S| \geq \epsilon) \leq 2\epsilon$$

From Lemma 11.1, we obtain

$$\begin{aligned} \mathbb{P}\left(\max_{n \leq j \leq k} |S_j - S_n| \geq 4\epsilon\right) &\leq \frac{1}{1 - 2\epsilon}\mathbb{P}(|S_k - S_n| \geq 2\epsilon) \\ &\leq \frac{2\epsilon}{1 - 2\epsilon} \\ &\leq 3\epsilon, \end{aligned}$$

for ϵ sufficiently small. The MCT (Theorem 4.2) then yields

$$\mathbb{P}\left(\max_{j \geq n} |S_j - S_n| \geq 4\epsilon\right) \leq 3\epsilon.$$

Together with (37), we finally have

$$\mathbb{P}\left(\max_{j \geq n} |S_j - S| \geq 5\epsilon\right) \leq 4\epsilon.$$

Lemma 11.2 then implies that $S_j \to S$ a.s. as $j \to \infty$.

We next show that S_n converges in distribution implies convergence in probability. Suppose that $\mathbb{P}_{S_n} \xrightarrow{\text{d}} \mathbb{P}$ for some $\mathbb{P}$. From Lemma 9.7, $(\mathbb{P}_{S_n})_{n \geq 1}$ is uniformly tight. Therefore from Lemma 11.4, $(\mathbb{P}_{S_n - S_k})_{1 \leq k \leq n}$ is uniformly tight. We proceed by contradiction. Suppose $\exists\ \epsilon > 0$ and $(n(l))$, $(m(l))$ with $n(l) \leq m(l)\ \forall\, l$ such that

$$\mathbb{P}\big(|S_{m(l)} - S_{n(l)}| > \epsilon\big) \geq \epsilon. \tag{38}$$

Let $Y_l = S_{m(l)} - S_{n(l)}$, then since $(\mathbb{P}_{Y_l})$ is uniformly tight, from Helly's Selection Theorem (Theorem 9.2), $\exists\ (l(r))$ such that $\mathbb{P}_{Y_{l(r)}} \xrightarrow{\text{d}}$ some distribution $\mathbb{Q}$. Since $S_{m(l(r))} = S_{n(l(r))} + Y_{l(r)}$ and $S_{n(l)}$, $Y_{l(r)}$ are independent, we have

$$\mathbb{P}_{S_{m(l(r))}} = \mathbb{P}_{S_{n(l(r))}} * \mathbb{P}_{Y_{l(r)}}.$$

Taking $r \to \infty$, we then have $\mathbb{P} = \mathbb{P}\mathbb{Q}$ from Exercise 10.1. From Lemma 11.5, $\mathbb{Q}(\{0\}) = 1$, which implies that $\mathbb{P}\big(|Y_{l(r)}| > \epsilon\big) < \epsilon$ for r sufficiently large. This contradicts (38) and the proof is complete. $\square$

11.5 Poisson Convergence

Let $X_{n,m}$, $n, m \geq 1$ be independent Bernoulli r.v.s with $\mathbb{P}(X_{n,m} = 1) = p_{n,m}$. If $p_{n,m} \to 0$ sufficiently fast as $n \to \infty$, then

$$\sum_{m=1}^{n} \operatorname{var} X_{n,m} = \sum_{m=1}^{n} p_{n,m}(1 - p_{n,m}) \to 0.$$

This violates Lindeberg's CLT condition (i), therefore we cannot apply the CLT here. However, we can still obtain a convergence in distribution result.

Theorem 11.5. *Suppose*

(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$ *as* $n \to \infty$, *and*

(ii) $\max_{1 \leq m \leq n} p_{n,m} \to 0$ *as* $n \to \infty$,

then $S_n = \sum_{m=1}^{n} X_{n,m} \xrightarrow{\mathrm{d}} \mathrm{Po}\,(\lambda)$, *where* $\mathrm{Po}\,(\lambda)\,(k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$ *for* $k \in \mathbb{Z}_{\geq 0}$ *is the Poisson distribution.*

As in the proof of Lindeberg's CLT, the proof of this result proceeds via Levy's continuity theorem (Theorem 10.1). We need a preliminary lemma.

Lemma 11.6. *If the complex numbers* z_i, w_i, $i = 1, \ldots, n$, *are such that* $|z_i|, |w_i| \leq \theta$, *then*

$$\left| \prod_{i=1}^{n} z_i - \prod_{i=1}^{n} w_i \right| \leq \theta^{n-1} \sum_{i=1}^{n} |z_i - w_i|.$$

Proof. We prove by induction on n. The result obviously holds for $n = 1$. Suppose it is true for $n - 1$. Then,

$$\begin{aligned}
&\left| \prod_{i=1}^{n} z_i - \prod_{i=1}^{n} w_i \right| \\
&\leq \left| \prod_{i=1}^{n-1} z_i \cdot z_n - \prod_{i=1}^{n-1} w_i \cdot z_n \right| + \left| \prod_{i=1}^{n-1} w_i \cdot z_n - \prod_{i=1}^{n-1} w_i \cdot w_n \right| \\
&\leq \theta \left| \prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} w_i \right| + \theta^{n-1} |z_n - w_n| \\
&\leq \theta^{n-1} \sum_{i=1}^{n-1} |z_i - w_i| + \theta^{n-1} |z_n - w_n| \\
&= \theta^{n-1} \sum_{i=1}^{n} |z_i - w_i|.
\end{aligned}$$

□

Proof of Lemma 11.6. The characteristic function of $\mathrm{Po}\,(\lambda)$ is $\exp(\lambda(e^{\mathrm{i}t} - 1))$ while that for S_n is

$$\mathbb{E} e^{\mathrm{i}tS_n} = \prod_{m=1}^{n} \big(1 + p_{n,m}(e^{\mathrm{i}t} - 1)\big).$$

We have

$$\begin{aligned}\left|\exp(p_{n,m}\bigl(e^{\mathrm{i}t}-1\bigr))\right| &= \exp(p_{n,m}\Re(e^{\mathrm{i}t}-1))\\ &= \exp(p_{n,m}(\cos t-1)) \le 1,\end{aligned}$$

and also

$$\left|1+p_{n,m}(e^{\mathrm{i}t}-1)\right| = |1+p_{n,m}(\cos t-1)+\mathrm{i}\sin t| \le 1,$$

which together with Lemma 11.6 yield

$$\begin{aligned}&\left|\prod_{m=1}^{n}\bigl(1+p_{n,m}\bigl(e^{\mathrm{i}t}-1\bigr)\bigr)-\exp\left(\sum_{m=1}^{n}p_{n,m}(e^{\mathrm{i}t}-1)\right)\right|\\ &\le \sum_{m=1}^{n}\left|\exp\bigl(p_{n,m}\bigl(e^{\mathrm{i}t}-1\bigr)\bigr)-\bigl(1+p_{n,m}\bigl(e^{\mathrm{i}t}-1\bigr)\bigr)\right|\\ &\le \frac{1}{2}\sum_{m=1}^{n}p_{n,m}^2\left|e^{\mathrm{i}t}-1\right|^2\\ &\le 2\max_{1\le m\le n}p_{n,m}\sum_{m=1}^{n}p_{n,m} \to 0 \text{ as } n\to\infty.\end{aligned}$$

Since $\exp\bigl(\sum_{m=1}^{n}p_{n,m}(e^{\mathrm{i}t}-1)\bigr) \to \exp(\lambda(e^{\mathrm{i}t}-1))$ as $n\to\infty$, the result follows from Levy's continuity theorem (Theorem 10.1). □

12.1 Definition

Given a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, let us start off with a simple example that we are familiar with from undergraduate probability courses. Suppose we have a r.v. X and a discrete r.v. $Y \in \{1, 2, \ldots\}$. We can partition $\Omega = \bigcup_{y \geq 1} \Omega_y$, where $\Omega_y = \{\omega : Y(\omega) = y\}$. Then the conditional expectation of X given Y can be defined as

$$\mathbb{E}[X \mid Y = y] = \frac{\mathbb{E}\big[X\mathbf{1}_{\Omega_y}\big]}{\mathbb{P}(\Omega_y)}$$

for each value of y. Note that the expectation is conditioned on a set $\Omega_y = \{Y = y\}$. It depends on the value $Y(\omega) = y$, and is hence a function of ω. Since Ω_y is measurable, this is a measurable function and is hence a *random variable*!

We wish to generalize this definition to all sets in a sub-σ-algebra in $\mathscr{A}$. Furthermore, when we average out the conditional expectation over a set B of feasible Y values, we should get back the expectation over this set:

$$\sum_{y \in B} \mathbb{E}[X \mid Y = y]\mathbb{P}(Y = y) = \sum_{y \in B} \mathbb{E}\big[X\mathbf{1}_{\Omega_y}\big] = \mathbb{E}[X\mathbf{1}_{Y \in B}],$$

where the last inequality follows from Fubini's theorem.

Definition 12.1. *Suppose $\mathbb{E}|X| < \infty$ and the σ-algebra $\mathscr{F} \subset \mathscr{A}$. A random variable $Y : \Omega \mapsto \mathbb{R}$ is a conditional expectation of X given $\mathscr{F}$ if*

(i) $Y^{-1}(B) \in \mathscr{F}$, $\forall\, B \in \mathscr{B}(\mathbb{R})$, i.e., Y is $\mathscr{F}$-measurable (we denote it as $Y \in \mathscr{F}$).

(ii) $\forall\, A \in \mathscr{F}$, $\mathbb{E}[Y\mathbf{1}_A] = \int_A Y \,\mathrm{d}\mathbb{P} = \int_A X \,\mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbf{1}_A]$.

If Y is a conditional expectation of X given $\mathscr{F}$, we write $Y = \mathbb{E}[X \mid \mathscr{F}]$. We also write $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$, where $\sigma(Y)$ is the σ-algebra generated by Y. The notion that expectation is an operator comes from here: $\mathbb{E}[\cdot \mid \mathscr{F}] : L^1(\Omega, \mathscr{A}, \mathbb{P}) \mapsto L^1(\Omega, \mathscr{F}, \mathbb{P})$ is a linear (which will be shown later) transformation.

The **existence** of conditional expectations is given by Radon–Nikodym Theorem (Theorem 4.4). Suppose $X \geq 0$. We can define a measure

$$\mu(A) = \int_A X \,\mathrm{d}\mathbb{P}, \text{ where } A \in \mathscr{F} \text{ and } \mu \ll \mathbb{P}.$$

Since X is integrable, μ is a finite measure. Then there exists $Y = \dfrac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}} \in \mathscr{F}$ such that $\int_A X \,\mathrm{d}\mathbb{P} = \mu(A) = \int_A Y \,\mathrm{d}\mathbb{P}$. The existence of conditional expectations for general $X = X^+ - X^-$ now follows.

We next show that conditional expectations are **unique** almost surely. Suppose that Y and Y' are both versions of $\mathbb{E}[X \mid \mathscr{F}]$ and $\mathbb{P}(Y \neq Y') > 0$, i.e., $\mathbb{P}(Y > Y') > 0$ or $\mathbb{P}(Y < Y') > 0$. Let $A = \{Y > Y'\} \in \mathscr{F}$ and suppose that $\mathbb{P}(A) > 0$. Then, we have

$$\begin{aligned} 0 < \mathbb{E}[(Y - Y')\mathbf{1}_A] &= \mathbb{E}Y\mathbf{1}_A - \mathbb{E}Y'\mathbf{1}_A \\ &= \mathbb{E}X\mathbf{1}_A - \mathbb{E}X\mathbf{1}_A \\ &= 0, \end{aligned}$$

which is a contradiction. A similar argument holds for the case $\mathbb{P}(Y < Y') > 0$. Therefore, $\mathbb{P}(Y \neq Y') = 0$, and $Y = Y'$ a. s.

Example 12.1. *Suppose that the joint pdf of (X, Y) is $f(x, y)$. Let*

$$h(y) = \int g(x) f(x|y)\,\mathrm{d}x = \frac{\int g(x) f(x, y)\,\mathrm{d}x}{\int f(x, y)\,\mathrm{d}x}.$$

We show that $h(Y) = \mathbb{E}[g(X)\,|\,Y]$. Let $A \in \sigma(Y)$. Then $A = \{\omega : Y(\omega) \in B\}$ for some $B \in \mathscr{B}(\mathbb{R})$. We check that

$$\begin{aligned} \mathbb{E}[h(Y)\mathbf{1}_A] &= \int_B \int h(y) f(x, y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_B h(y) \int f(x, y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_B \int g(x) f(x, y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \mathbb{E}[g(X)\mathbf{1}_B(Y)] \\ &= \mathbb{E}[g(X)\mathbf{1}_A], \end{aligned}$$

and the claim is proved.

Example 12.2. *A sensor makes an observation $X \in \mathbb{R}$ and sends a summary $Z = \gamma(X) \in \mathbb{R}$ to a fusion center, where γ is a randomized function. The fusion center uses Z to perform hypothesis testing for*

$$H = \begin{cases} H_0 : & X \sim \mathbb{P}_0, \\ H_1 : & X \sim \mathbb{P}_1. \end{cases}$$

We assume that $\mathbb{P}_0$ and $\mathbb{P}_1$ are absolutely continuous w.r.t. each other. Note that these are the laws of X under different hypotheses.

For $i = 0, 1$, let $\mathbb{P}_{i,Z}$ be the restriction of $\mathbb{P}_i$ on $\sigma(Z)$. Suppose $A \in \sigma(Z)$. We have

$$\begin{aligned} \int_A \frac{\mathrm{d}\mathbb{P}_{1,Z}}{\mathrm{d}\mathbb{P}_{0,Z}}\,\mathrm{d}\mathbb{P}_0 &= \int_A \frac{\mathrm{d}\mathbb{P}_{1,Z}}{\mathrm{d}\mathbb{P}_{0,Z}}\,\mathrm{d}\mathbb{P}_{0,Z} \\ &= \int_A \mathrm{d}\mathbb{P}_{1,Z} \\ &= \int_A \mathrm{d}\mathbb{P}_1 \\ &= \int_A \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\,\mathrm{d}\mathbb{P}_0. \end{aligned}$$

Therefore,

$$\frac{\mathrm{d}\mathbb{P}_{1,Z}}{\mathrm{d}\mathbb{P}_{0,Z}} = \mathbb{E}\left[\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\,\middle|\,\sigma(Z)\right].$$

As a special case, suppose X and Z are continuous r.v.s. Then γ corresponds to a conditional pdf $p(z|x)$ and letting f_i be the pdf of $\mathbb{P}_i$, $i = 0, 1$, we have from Example 12.1,

$$\begin{aligned}\mathbb{E}_0\left[\frac{f_1(x)}{f_0(x)}\,\middle|\, Z = z\right] &= \frac{\int \frac{f_1(x)}{f_0(x)} f_0(x,z)\,\mathrm{d}z}{\int f_0(x,z)\,\mathrm{d}x} \\ &= \frac{\int f_1(x)p(z|x)\,\mathrm{d}x}{\int f_{0,Z}(z)} \\ &= \frac{f_{1,Z}(z)}{f_{0,Z}(z)}.\end{aligned}$$

12.2 Properties

In the following, we list some fundamental properties of conditional expectations, many without proofs. The proofs are left for your exercise.

1. If $\sigma(X) \subset \mathscr{F}$, then $X \in \mathscr{F}$ and by Definition 12.1, $X = \mathbb{E}[X \,|\, \mathscr{F}]$ a.s. As a special case, we have $\mathbb{E}[X \,|\, \mathscr{A}] = X$. If c is a constant, viewed as a r.v., its σ-algebra is the trivial one and so $\mathbb{E}[c \,|\, \mathscr{F}] = c$.
2. If $\sigma(X) \perp\!\!\!\perp \mathscr{F}$, then for $A \in \mathscr{F}$, $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}X\mathbb{E}\mathbf{1}_A = \int_A \mathbb{E}X\,\mathrm{d}\mathbb{P} \Rightarrow \mathbb{E}[X \,|\, \mathscr{F}] = \mathbb{E}X$ a.s. As a special case, if $\mathscr{F} = \{\emptyset, \Omega\}$, we have $\mathbb{E}[X \,|\, \mathscr{F}] = \mathbb{E}X$.
3. Suppose that c is a constant, then $\mathbb{E}[cX + Y \,|\, \mathscr{F}] = c\mathbb{E}[X \,|\, \mathscr{F}] + \mathbb{E}[Y \,|\, \mathscr{F}]$.
4. Suppose $\mathscr{G}$ and $\mathscr{F}$ are σ-algebras with $\mathscr{G} \subset \mathscr{F}$, then $\mathbb{E}[X \,|\, \mathscr{G}] = \mathbb{E}[\mathbb{E}[X \,|\, \mathscr{G}] \,|\, \mathscr{F}] = \mathbb{E}[\mathbb{E}[X \,|\, \mathscr{F}] \,|\, \mathscr{G}]$.

   Proof. Let $A \in \mathscr{G} \subset \mathscr{F}$, then we have

   $$\begin{aligned}\mathbb{E}[\mathbb{E}[\mathbb{E}[X \,|\, \mathscr{F}] \,|\, \mathscr{G}]\mathbf{1}_A] &= \mathbb{E}[\mathbb{E}[X \,|\, \mathscr{F}]\mathbf{1}_A] \text{ by definition of } \mathbb{E}[\cdot \,|\, \mathscr{G}] \\ &= \mathbb{E}X\mathbf{1}_A \text{ by definition of } \mathbb{E}[\cdot \,|\, \mathscr{F}].\end{aligned}$$

   □

   For example, setting $\mathscr{G} = \{\emptyset, \Omega\}$, we obtain $\mathbb{E}[\mathbb{E}[X \,|\, \mathscr{F}]] = \mathbb{E}X$ for all $\mathscr{F}$.
5. If $X \leq Y$ a.s., then $\mathbb{E}[X \,|\, \mathscr{F}] \leq \mathbb{E}[Y \,|\, \mathscr{F}]$.

   Lemma 12.1. *$\mathbb{E}[X \,|\, \mathscr{F}] \leq \mathbb{E}[Y \,|\, \mathscr{F}]$ a.s. iff $\mathbb{E}X\mathbf{1}_A \leq \mathbb{E}Y\mathbf{1}_A$ for all $A \in \mathscr{F}$.*

   Proof. Similar to the uniqueness proof. □
6. Monotone Convergence Theorem. Suppose $\mathbb{E}|X_n| < \infty$, $\forall\, n \geq 1$, $\mathbb{E}|X| < \infty$ and $X_n \uparrow X$ a.s. as $n \to \infty$, then $\mathbb{E}[X_n \,|\, \mathscr{F}] \uparrow \mathbb{E}[X \,|\, \mathscr{F}]$ a.s.

   Proof. Since $X_n \uparrow X$, we have

   $$\mathbb{E}[X_n \,|\, \mathscr{F}] \leq \mathbb{E}[X_{n+1} \,|\, \mathscr{F}] \leq \mathbb{E}[X \,|\, \mathscr{F}]$$

   and there exists

   $$Y \triangleq \lim_{n\to\infty} \mathbb{E}[X_n \,|\, \mathscr{F}] \leq \mathbb{E}[X \,|\, \mathscr{F}].$$

   From Lemma 4.2, Y is $\mathscr{F}$-measurable. Moreover, for each $A \in \mathscr{F}$, we have

   $$\mathbb{E}[X_n \,|\, \mathscr{F}]\mathbf{1}_A \uparrow Y\mathbf{1}_A$$

since $\mathbb{E}[X_n \mid \mathscr{F}] \uparrow Y$. From the MCT, we obtain

$$\mathbb{E}[\mathbb{E}[X_n \mid \mathscr{F}]\mathbf{1}_A] \to \mathbb{E}[Y\mathbf{1}_A].$$

But $\mathbb{E}[\mathbb{E}[X_n \mid \mathscr{F}]\mathbf{1}_A] = \mathbb{E}[X_n\mathbf{1}_A] \xrightarrow{\text{MCT}} \mathbb{E}[X\mathbf{1}_A]$. Therefore, $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$ and $Y = \mathbb{E}[X \mid \mathscr{F}]$ a.s. □

7. Dominated Convergence Theorem. If $|X_n| \le Y$ a. s., $\mathbb{E}Y < \infty$ and $X_n \to X$ a.s., then $\lim_{n\to\infty} \mathbb{E}[X_n \mid \mathscr{F}] = \mathbb{E}[X \mid \mathscr{F}]$.
8. Suppose $X \perp\!\!\!\perp Y$ and $\mathbb{E}|\phi(X, Y)| < \infty$. Let $g(y) = \mathbb{E}[\phi(X, y)]$, then $\mathbb{E}[\phi(X, Y) \mid Y] = g(Y)$.
9. If $\mathbb{E}|X| < \infty$, $\mathbb{E}|XY| < \infty$, and $Y \in \mathscr{F}$, then $\mathbb{E}[XY \mid \mathscr{F}] = Y\mathbb{E}[X \mid \mathscr{F}]$.
10. The usual inequalities apply. We define $\mathbb{P}(A \mid \mathscr{F}) = \mathbb{E}[\mathbf{1}_A \mid \mathscr{F}]$.
    - Markov's inequality: For $a > 0$, $\mathbb{P}(X > a \mid \mathscr{F}) \le \frac{1}{a}\mathbb{E}[X \mid \mathscr{F}]$.
    - Chebyshev's inequality: For $a > 0$, $\mathbb{P}(|X| \ge a \mid \mathscr{F}) \le \frac{1}{a^2}\mathbb{E}\left[X^2 \,\middle|\, \mathscr{F}\right]$.
    - Cauchy-Schwarz inequality: $\mathbb{E}[XY \mid \mathscr{F}]^2 \le \mathbb{E}\left[X^2 \,\middle|\, \mathscr{F}\right]\mathbb{E}\left[Y^2 \,\middle|\, \mathscr{F}\right]$.
    - Jensen's inequality: Given a convex function ϕ with $\mathbb{E}|\phi(X)| < \infty$, we have

    $$\phi(\mathbb{E}[X \mid \mathscr{F}]) \le \mathbb{E}[\phi(X) \mid \mathscr{F}].$$

12.3 L^2 Interpretation

Proposition 12.1. *Consider $L^2(\Omega, \mathscr{F}, \mathbb{P}) = \{X : X \in \mathscr{F},\ \mathbb{E}X^2 < \infty\}$, which is a subspace of $L^2(\Omega, \mathscr{A}, \mathbb{P})$. If $X \in L^2(\Omega, \mathscr{A}, \mathbb{P})$, then*

$$\mathbb{E}[X \mid \mathscr{F}] = \operatorname*{arg\,min}_{Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})} \mathbb{E}(X - Y)^2.$$

Proof. For $Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})$, let $Z = \mathbb{E}[X \mid \mathscr{F}] - Y \in \mathscr{F}$. Then, we have

$$\begin{aligned}\mathbb{E}(X - Y)^2 &= \mathbb{E}\left[(X - \mathbb{E}[X \mid \mathscr{F}] + \mathbb{E}[X \mid \mathscr{F}] - Y)^2\right] \\ &= \mathbb{E}(X - \mathbb{E}[X \mid \mathscr{F}])^2 + \mathbb{E}Z^2 + 2\mathbb{E}[Z(X - \mathbb{E}[X \mid \mathscr{F}])]\end{aligned}$$

and since $Z \in \mathscr{F}$,

$$\begin{aligned}\mathbb{E}[Z(X - \mathbb{E}[X \mid \mathscr{F}])] &= \mathbb{E}[ZX] - \mathbb{E}[\mathbb{E}[ZX \mid \mathscr{F}]] \\ &= \mathbb{E}[ZX] - \mathbb{E}[ZX] \\ &= 0.\end{aligned}$$

Therefore,

$$\mathbb{E}(X - Y)^2 \ge \mathbb{E}(X - \mathbb{E}[X \mid \mathscr{F}])^2,$$

and the proposition is proved. □

We can define $\operatorname{var}(X\mathscr{F}) = \mathbb{E}\left[(X - \mathbb{E}[X \mid \mathscr{F}])^2 \,\middle|\, \mathscr{F}\right] = \mathbb{E}\left[X^2 \,\middle|\, \mathscr{F}\right] - \mathbb{E}[X \mid \mathscr{F}]^2$. One can show (exercise) that

$$\operatorname{var}(X) = \mathbb{E}[\operatorname{var}(X \mid \mathscr{F})] + \operatorname{var}(\mathbb{E}[X \mid \mathscr{F}]).$$

12.4 Martingales

Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. A sequence of sub-$\sigma$-algebras $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \subset \mathscr{F}$ is called a *filtration*. Let the r.v. $M_n \in \mathscr{F}_n$ (i.e., M_n is $\mathscr{F}_n$-measurable). We say that M_n is adapted to $\mathscr{F}_n$.

Definition 12.2 (Martingale)**.** *We say that $(M_n, \mathscr{F}_n)_{n\geq 0}$ is a martingale if $\mathbb{E}|M_n| < \infty$ for all $n \geq 0$ and*

$$\mathbb{E}[M_m \,|\, \mathscr{F}_n] = M_n, \ \forall\, m \geq n. \tag{39}$$

By induction, the condition (39) is equivalent to $\mathbb{E}[M_n \,|\, \mathscr{F}_{n-1}] = M_{n-1}$. We give examples of martingales below.

Example 12.3. *Suppose that $(X_n)_{n\geq 1}$ are independent r.v.s, $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$ and $\mathbb{E}X_n = 0$. Let $S_n = \sum_{i=1}^n X_i$. Then,*

$$\begin{aligned} \mathbb{E}[S_n \,|\, \mathscr{F}_{n-1}] &= \mathbb{E}[S_{n-1} + X_n \,|\, \mathscr{F}_{n-1}] \\ &= S_{n-1} + \mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] \end{aligned}$$

Since $\mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}X_n = 0$, we have

$$\mathbb{E}[S_n \,|\, \mathscr{F}_{n-1}] = S_{n-1}.$$

Therefore, $(S_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale.

Example 12.4. *Suppose that $(X_n)_{n\geq 1}$ are independent r.v.s,* $\mathrm{var}(X_n) = \sigma^2$ *and $\mathbb{E}X_n = 0$. Let $M_0 = 0$, and $M_n = S_n^2 - n\sigma^2$. We have*

$$\begin{aligned} \mathbb{E}[M_n \,|\, \mathscr{F}_{n-1}] &= \mathbb{E}\big[S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 - n\sigma^2 \,\big|\, \mathscr{F}_{n-1}\big] \\ &= S_{n-1}^2 - (n-1)\sigma^2 \\ &= M_{n-1}, \end{aligned}$$

and $(M_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale.

Example 12.5. *Suppose that $(X_i)_{i\geq 1}$ are independent r.v.s., $X_i \geq 0$ and $\mathbb{E}X_i = 1$. Let $M_0 = 1$, and $M_n = \prod_{i=1}^n X_i$. We have*

$$\begin{aligned} \mathbb{E}[M_n \,|\, \mathscr{F}_{n-1}] &= \mathbb{E}[M_{n-1} \cdot X_n \,|\, \mathscr{F}_{n-1}] \\ &= M_{n-1}\mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] \end{aligned}$$

Since $\mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}X_i = 1$, we have

$$\mathbb{E}[S_n \,|\, \mathscr{F}_{n-1}] = M_{n-1},$$

and $(M_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale.

Example 12.6. *Suppose that $(Y_n)_{n\geq 1}$ are i.i.d. and $\phi(\lambda) = \mathbb{E}e^{\lambda Y_1} < \infty$. Let $X_n = \dfrac{e^{\lambda Y_n}}{\phi(\lambda)}$. Then $\mathbb{E}X_n = 1$. Let*

$$M_n = \frac{\exp\left(\lambda \sum_{i=1}^n Y_n\right)}{\phi(\lambda)^n}$$

and from Example 12.5, we have $(M_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale.

Example 12.7. *Given $\mathbb{E}|X| < \infty$ and a filtration $(\mathscr{F}_n)_{n\geq 1}$, let $M_n = \mathbb{E}[X \mid \mathscr{F}_n]$. Then,*

$$\begin{aligned}\mathbb{E}[M_n \mid \mathscr{F}_{n-1}] &= \mathbb{E}[\mathbb{E}[X \mid \mathscr{F}_n] \mid \mathscr{F}_{n-1}] \\ &= \mathbb{E}[X \mid \mathscr{F}_{n-1}] \\ &= M_{n-1}.\end{aligned}$$

Therefore, $(M_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale.

Definition 12.3. *$(A_n)_{n\geq 1}$ is predictable w.r.t. $(\mathscr{F}_n)_{n\geq 0}$ if $A_n \in \mathscr{F}_{n-1}, \forall\, n \geq 1$.*

We call $(\widetilde{M}_n)_{n\geq 0}$ the martingale transform of $(M_n)_{n\geq 0}$ by $(A_n)_{n\geq 1}$ if

(i) $\widetilde{M}_0 = M_0$. (In general, we can choose any integrable $\widetilde{M}_0 \in \mathscr{F}_0$.)

(ii) $\widetilde{M}_n = \widetilde{M}_0 + \sum_{k=1}^n A_k(M_k - M_{k-1})$.

Theorem 12.1 (MTT)**.** *If $(A_n)_{n\geq 1}$ is predictable w.r.t. $(\mathscr{F}_n)_{n\geq 0}$ and bounded, and $(M_n, \mathscr{F}_n)_{n\geq 0}$ is a martingale, then $(\widetilde{M}_n, \mathscr{F}_n)_{n\geq 0}$ is a martingale.*

Proof. It is obvious that $\widetilde{M}_n \in \mathscr{F}_n$ and $\mathbb{E}|\widetilde{M}_n| < \infty$. Furthermore, we have

$$\begin{aligned}\mathbb{E}\left[\widetilde{M}_n - \widetilde{M}_{n-1} \,\middle|\, \mathscr{F}_{n-1}\right] &= \mathbb{E}[A_n(M_n - M_{n-1}) \mid \mathscr{F}_{n-1}] \\ &= A_n \mathbb{E}[M_n - M_{n-1} \mid \mathscr{F}_{n-1}] = 0.\end{aligned}$$

□

Example 12.8. *Suppose we divide time into discrete equal intervals (e.g., days). Let M_n be the price of a stock at time instant $n \geq 1$ and $\mathscr{F}_n$ be the available information up to time n. From the efficient-market hypothesis (assuming no dividends and discount factor), M_n is a "fair" price that has priced in all expected future gains or losses, i.e., $\mathbb{E}[M_m \mid \mathscr{F}_n] = M_n$ for $m \geq n$ and $(M_n, \mathscr{F}_n)$ is a martingale. Let $(A_n)_{n\geq 1}$ be a strategy that decides to hold A_n units of the stock in the time period $[n-1, n)$. Clearly, (A_n) has to be predictable w.r.t. $(\mathscr{F}_n)$. Then $\widetilde{M}_n = \sum_{k\leq n} A_k(M_k - M_{k-1})$ is the change in wealth of this strategy up to time n. The MTT says that the expected wealth of* any *strategy is the same and what you start with, i.e., you cannot beat the market! We will see a much stronger version of this result in Doob's Optional Stopping Theorem in the next session.*

13. Optional Stopping

Review: Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. A sequence of sub-$\sigma$-algebras $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \ldots \subset \mathscr{F}$ is called a filtration. Suppose $M_n \in \mathscr{F}_n$. Then, $(M_n, \mathscr{F}_n)$ is a martingale if

$$\begin{aligned}\mathbb{E}|M_n| < \infty, \ \forall\, n \geq 0,\\ \mathbb{E}[M_n \,|\, \mathscr{F}_{n-1}] = M_{n-1}, \ \forall\, n \geq 0.\end{aligned}$$

13.1 Stopping Times

Definition 13.1. *The random variable τ is called a* stopping time *for $(\mathscr{F}_n)_{n\geq 0}$ if $\tau \in \{0, 1, \ldots\} \cup \{\infty\}$ and $\{\tau \leq n\} \in \mathscr{F}_n, \ \forall\, n \geq 0$.*

From the definition of σ-algebras, we have

$$\begin{aligned}\{\tau > n\} &= \{\tau \leq n\}^{\mathsf{c}} \in \mathscr{F}_n,\\ \{\tau = n\} &= \{\tau \leq n\} \setminus \{\tau \leq n-1\} \in \mathscr{F}_n.\end{aligned}$$

Consider a sequence $(X_n)_{n\geq 0}$. If $\tau < \infty$ a.s., we define

$$X_\tau = \sum_{k=0}^{\infty} \mathbf{1}_{\{\tau = k\}} X_k$$

and we call X_τ a "stopped process". Denote $n \wedge \tau = \min(n, \tau)$ and we have $n \wedge \tau < \infty$ a.s.

Theorem 13.1. *If $(M_n, \mathscr{F}_n)$ is a martingale, then $(M_{n\wedge\tau}, \mathscr{F}_n)$ is a martingale.*

Proof. WLOG, we assume $M_0 = 0$. Let $A_n = \mathbf{1}_{\{\tau \geq n\}} = \mathbf{1}_{\{\tau \leq n-1\}^{\mathsf{c}}}$, where $\{\tau \leq n-1\}^{\mathsf{c}} \in \mathscr{F}_{n-1}$. Therefore, A_n is predictable. The martingale transform $(\sum_{k=1}^n A_k(M_k - M_{k-1}), \mathscr{F}_n)$ is thus a martingale from Theorem 12.1. Furthermore, we have

$$\begin{aligned}\sum_{k=1}^{n} A_k(M_k - M_{k-1}) &= \sum_{k=1}^{n} \mathbf{1}_{\{\tau \geq k\}}(M_k - M_{k-1})\\ &= \sum_{k=1}^{n} \mathbf{1}_{\{\tau \geq k\}} M_k - \sum_{k=0}^{n-1} \mathbf{1}_{\{\tau \geq k+1\}} M_k\\ &= \sum_{k=1}^{n-1} \mathbf{1}_{\{\tau = k\}} M_k + \mathbf{1}_{\{\tau \geq n\}} M_n\\ &= \mathbf{1}_{\{\tau \leq n-1\}} M_\tau + \mathbf{1}_{\{\tau \geq n\}} M_n\\ &= M_{n\wedge\tau},\end{aligned}$$

and the theorem is proved. □

Example 13.1 (Random walk). *Let*

$$X_i = \begin{cases} 1 & w.p.\ \frac{1}{2}, \\ -1 & w.p.\ \frac{1}{2}, \end{cases}$$

for $i \geq 1$ be i.i.d. random variables. Let $S_n = \sum_{i=1}^n X_i$. Then, $(S_n, \mathscr{F}_n)$ is a martingale (see Example 12.3). Let $A, B \in \mathbb{Z}^+$ and $\tau = \inf\{n : S_n = A \text{ or } S_n = -B\}$ (the first time S_n hitting A or $-B$). From Theorem 13.1, $(S_{n\wedge\tau}, \mathscr{F}_n)$ is a martingale. We therefore have

$$\mathbb{E}S_{n\wedge\tau} = \mathbb{E}S_0 = 0.$$

Suppose S_n has not hit $-B$ yet. Then, a sequence of $A+B$ realizations of $X_i = 1$ will make the sum hit A. Let $E_k = \big\{X_i = 1 \,:\, i \in \big[k(A+B), (k+1)(A+B)\big)\big\}$, which are independent for all $k \geq 0$. We have

$$\mathbb{P}(\tau > n(A+B)) \leq \mathbb{P}\left(\bigcap_{k=0}^{n-1} E_k^{\mathsf{c}}\right) = \left(1 - \frac{1}{2^{A+B}}\right)^n.$$

Furthermore,

$$\begin{aligned} \mathbb{E}\tau &= \sum_{k=1}^{\infty} \mathbb{P}(\tau \geq k) \\ &\leq \sum_{n=0}^{\infty} \mathbb{P}(\tau > n(A+B))(A+B) \\ &\leq (A+B) \sum_{n=0}^{\infty} \left(1 - \frac{1}{2^{A+B}}\right)^n \\ &< \infty. \end{aligned}$$

Therefore, $\tau < \infty$ a.s. We also have $|S_{n\wedge\tau}| \leq \max\{A, B\}$. From DCT, we then obtain

$$\begin{aligned} 0 = \lim_{n\to\infty} \mathbb{E}S_{n\wedge\tau} = \mathbb{E}S_\tau &= A\mathbb{P}(S_\tau = A) - B(1 - \mathbb{P}(S_\tau = A)) \\ \implies \mathbb{P}(S_\tau = A) &= \frac{B}{A+B}. \end{aligned}$$

We will see another way to prove $\tau < \infty$ a.s. in Example 15.2.

Example 13.2. *Let $M_n = S_n^2 - n$ and $(M_n, \mathscr{F}_n)$ is a martingale (see Example 12.4) and τ be the same stopping time as in Example 13.1. Then $(M_{n\wedge\tau}, \mathscr{F}_n)$ is a martingale with $\mathbb{E}M_{n\wedge\tau} = \mathbb{E}M_0 = 0$. We have*

$$|M_{n\wedge\tau}| \leq (A \vee B)^2 + \tau,$$

and since τ is integrable (see Example 13.1), the DCT yields

$$\mathbb{E}M_\tau = \lim_{n\to\infty} \mathbb{E}M_{n\wedge\tau} = 0.$$

Therefore, we have $0 = \mathbb{E}M_\tau = \mathbb{E}S_\tau^2 - \mathbb{E}\tau$, and

$$\begin{aligned} \mathbb{E}\tau &= \mathbb{E}S_\tau^2 \\ &= \frac{B}{A+B}A^2 + \frac{A}{A+B}B^2 \\ &= AB. \end{aligned}$$

Example 13.3 (Biased random walk)**.** *Let*

$$X_i = \begin{cases} 1 & w.p.\ p \neq \frac{1}{2}, \\ -1 & w.p.\ q = 1 - p, \end{cases}$$

for $i \geq 1$ be i.i.d. random variables and $\phi(\lambda) = \mathbb{E}e^{\lambda X_1} = pe^{\lambda} + qe^{-\lambda}$. Let

$$M_n = \frac{e^{\lambda S_n}}{\phi(\lambda)^n}.$$

We choose λ such that $\phi(\lambda) = 1 \implies e^{\lambda} = \dfrac{q}{p} \implies M_n = e^{\lambda S_n} = \left(\dfrac{q}{p}\right)^{S_n}$. Then $(M_n, \mathscr{F}_n)$ is a martingale (see Example 12.6), and so is $(M_{n\wedge\tau}, \mathscr{F}_n)$, where τ is the same stopping time as in Example 13.1. We have

$$\begin{aligned} &\mathbb{E}M_{n\wedge\tau} = \mathbb{E}M_0 = 1, \\ &|M_{n\wedge\tau}| \leq (1 \vee q/p)^{A\vee B}. \end{aligned}$$

From DCT, we have

$$\mathbb{E}M_\tau = \lim_{n\to\infty} \mathbb{E}M_{n\wedge\tau} = 1.$$

Denoting $\alpha = \mathbb{P}(S_\tau = A)$, we have

$$\mathbb{E}M_\tau = \alpha \left(\frac{q}{p}\right)^{A} + (1-\alpha)\left(\frac{q}{p}\right)^{-B}.$$

Therefore, we obtain

$$\alpha = \frac{\left(\frac{q}{p}\right)^{B}}{\left(\frac{q}{p}\right)^{A+B} - 1}.$$

13.2 Doob's Optional Stopping Theorem

In the previous examples, we have $\mathbb{E}M_\tau = \mathbb{E}M_0$ using the DCT and special properties of the stopping time τ. We want to know when this is true in general. But first, a counterexample.

Example 13.4 (Counterexample)**.** *Let*

$$X_n = \begin{cases} 2^n & w.p.\ \frac{1}{2}, \\ -2^n & w.p.\ \frac{1}{2}, \end{cases}$$

for $n \geq 0$ be independent random variables. Then $(S_n, \mathscr{F}_n)$ is a martingale. Let $\tau = \min\{k \geq 0 : S_k > 0\}$. When $\tau = k$, we have $S_\tau = S_k = -1 - 2 - 2^2 - \ldots - 2^{k-1} + 2^k = 1$. Therefore, $\mathbb{E}S_\tau = 1 \neq \mathbb{E}S_0 = \mathbb{E}X_0 = 0$.

Let τ be a stopping time w.r.t. $(\mathscr{F}_n)$. The σ-algebra (exercise: verify that it is a σ-algebra)

$$\mathscr{F}_\tau = \{A \in \mathscr{F} \,:\, \{\tau \leq n\} \cap A \in \mathscr{F}_n, \forall\, n \geq 0\}$$

consists of all events that depend on information up to the stopping time τ. We have the following properties:

(i) $\tau \in \mathscr{F}_\tau$ (τ is measurable w.r.t. $\mathscr{F}_\tau$). This is because $\{\tau \le n\} \cap \{\tau \le k\} = \{\tau \le n \wedge k\} \in \mathscr{F}_{n\wedge k} \subset \mathscr{F}_n$ since τ is a stopping time.

(ii) If (M_n) is adapted to $(\mathscr{F}_n)$ and $\tau < \infty$ a.s., then $M_\tau \in \mathscr{F}_\tau$. Proof: $\forall\, B \in \mathcal{B}, \{\tau \le n\} \cap \{M_\tau \in B\} = \bigcup_{k\le n}\{\tau = k\} \cap \{M_k \in B\} \in \mathscr{F}_n$.

(iii) Let τ, ν be stopping times w.r.t. $(\mathscr{F}_n)$. Then $\{\tau < \nu\}, \{\tau = \nu\}, \{\tau > \nu\} \in \mathscr{F}_\tau, \mathscr{F}_\nu$.

(iv) If $A \in \mathscr{F}_\nu$, then $A \cap \{\tau \ge \nu\}$, $A \cap \{\tau > \nu\} \in \mathscr{F}_\tau$.

(v) $\tau(\omega) \le \nu(\omega), \forall\, \omega \in \Omega \implies \mathscr{F}_\tau \subset \mathscr{F}_\nu$.

Theorem 13.2 (Optional Stopping Theorem)**.** *Let* $(M_n, \mathscr{F}_n)$ *be a martingale. Suppose*

(i) $\tau, \nu < \infty$ *are stopping times w.r.t.* $(\mathscr{F}_n)$*,*

(ii) $\mathbb{E}|M_\tau| < \infty$*,*

(iii) $\lim_{n\to\infty} \mathbb{E}|M_n| \mathbf{1}_{\{\tau\ge n\}} = 0$.

Then $\forall\, A \in \mathscr{F}_\nu$*,*

$$\mathbb{E}M_\tau \mathbf{1}_{A\cap\{\tau\ge\nu\}} = \mathbb{E}M_\nu \mathbf{1}_{A\cap\{\tau\ge\nu\}}.$$

An intuitive interpretation is as follows: A martingale $(M_n, \mathscr{F}_n)$ represents a fair game (e.g., M_n is stock price under an efficient market). A stopping time τ represents a strategy to stop the game based only on information up to a given time (e.g., to sell off a stock, execute an option, etc.). By taking $A = \Omega$ and $\nu = 0$, the theorem says that $\mathbb{E}M_\tau = \mathbb{E}M_0$ for any τ satisfying the theorem conditions. Thus, there is no "winning strategy" in an efficient market! Furthermore, given any two competing strategies τ and ν, there is no advantage ν can gain over τ in terms of expected wealth even on events it has full information about.

In Example 13.4, we can check that

$$\mathbb{E}|S_n| \mathbf{1}_{\{\tau\ge n\}} = \mathbb{P}(\tau = n) + (2^{n+1} - 1)\mathbb{P}(\tau \ge n+1) = 1,$$

which violates the third condition in Theorem 13.2.

Proof. For $A \in \mathscr{F}_\nu$, let $A_n = A \cap \{\nu = n\} \in \mathscr{F}_n$. We first show that

$$\mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{\tau\ge n\}} = \mathbb{E}M_n \mathbf{1}_{A_n\cap\{\tau\ge n\}}. \tag{40}$$

We have

$$\begin{aligned}
\mathbb{E}M_n \mathbf{1}_{A_n\cap\{\tau\ge n\}} &= \mathbb{E}M_n \mathbf{1}_{A_n\cap\{\tau=n\}} + \mathbb{E}M_n \mathbf{1}_{A_n\cap\{\tau>n\}} \\
&= \mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{\tau=n\}} + \mathbb{E}\big[\mathbb{E}[M_{n+1} \,|\, \mathscr{F}_n]\, \mathbf{1}_{A_n\cap\{\tau>n\}}\big] \\
&= \mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{\tau=n\}} + \mathbb{E}\big[\mathbb{E}\big[M_{n+1}\mathbf{1}_{A_n\cap\{\tau>n\}} \,\big|\, \mathscr{F}_n\big]\big] \quad \because \mathbf{1}_{A_n\cap\{\tau>n\}} \in \mathscr{F}_n \\
&= \mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{\tau=n\}} + \mathbb{E}M_{n+1} \mathbf{1}_{A_n\cap\{\tau\ge n+1\}}.
\end{aligned}$$

By induction, we have for any $m > n$,

$$\mathbb{E}M_n \mathbf{1}_{A_n\cap\{\tau\ge n\}} = \mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{n\le\tau<m\}} + \mathbb{E}M_m \mathbf{1}_{A_n\cap\{\tau\ge m\}}. \tag{41}$$

We have

$$\left|\mathbb{E}M_m \mathbf{1}_{\{A_n\cap\{\tau\ge m\}\}}\right| \le \mathbb{E}|M_m| \mathbf{1}_{\tau\ge m} \xrightarrow{m\to\infty} 0.$$

Since $\mathbb{E}|M_\tau| < \infty$, the DCT yields

$$\lim_{m\to\infty} \mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{n\le\tau<m\}} = \mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{\tau\ge n\}}.$$

Therefore by letting $m \to \infty$ in (41), we obtain (40). We finally have

$$\begin{aligned}
\mathbb{E}M_\tau \mathbf{1}_{A\cap\{\tau\ge\nu\}} &= \sum_{n\ge 0} \mathbb{E}M_\tau \mathbf{1}_{A\cap\{\nu=n\}\cap\{\tau\ge n\}} \\
&= \sum_{n\ge 0} \mathbb{E}M_\tau \mathbf{1}_{A_n\cap\{\tau\ge n\}} \\
&\overset{(40)}{=} \sum_{n\ge 0} \mathbb{E}M_n \mathbf{1}_{A_n\cap\{\tau\ge n\}} \\
&= \mathbb{E}M_\nu \mathbf{1}_{A\cap\{\tau\ge\nu\}},
\end{aligned}$$

which completes the proof. □

Some commonly used special cases of Theorem 13.2 are the following:

1. Let $\nu = 0$ and $A = \Omega$. Then from Theorem 13.2, we have $\mathbb{E}M_\tau = \mathbb{E}M_0$.
2. If $\tau \le B < \infty$ is bounded a.s., then the conditions in Theorem 13.2 are satisfied.
3. Suppose $\tau < \infty$. If $|M_n| \le K < \infty$ is bounded a.s., then the conditions in Theorem 13.2 are satisfied.

If we have stronger conditions on the martingale, we can obtain similar optional stopping results without appealing to Theorem 13.2. One such result is assuming *bounded* martingale differences.

Lemma 13.1. *Let $(M_n, \mathscr{F}_n)$ be a martingale. If $|M_n - M_{n-1}| \le K < \infty$ a.s. and $\mathbb{E}\tau < \infty$, then $\mathbb{E}M_\tau = \mathbb{E}M_0$.*

Proof. We have

$$|M_{n\wedge\tau} - M_0| = \left| \sum_{k=1}^{n\wedge\tau} (M_k - M_{k-1}) \right| \le K\tau.$$

Then, from DCT, we obtain

$$\mathbb{E}M_\tau = \lim_{n\to\infty} \mathbb{E}M_{n\wedge\tau} = \mathbb{E}M_0,$$

where the last inequality follows from Theorem 13.1. □

Example 13.5. *At each time $n = 1, 2, \ldots$, a monkey chooses one letter randomly and uniformly from the 26 English letters. What is the expected time for the monkey to produce the sequence "ABRACADABRA"?*

Solution: Let T be the first time the monkey produces the desired sequence. One can show that $\mathbb{E}T < \infty$ and hence $T < \infty$ a.s. (hint: see Example 13.1). Suppose at each time j, a new gambler arrives, bet 1 that the j-th letter is "A". If he loses, he leaves. If he wins, he receives a reward of 26, bet all of that on the $(j+1)$-th letter being "B", and so on according to the sequence "ABRACADABRA". When the end of the sequence is reached, he also leaves. Denote M_n^j as the payoff of gambler j at time n. We let $M_n^j = 0$ if $n < j$. we have

$$\mathbb{E}M_n^j \le 26^{n-j+1} < \infty.$$

If the gambler loses before time n, we have

$$\mathbb{E}\big[M_n^j \,\big|\, \mathscr{F}_{n-1}\big] = 0 = M_{n-1}^j.$$

Otherwise, he wins $n-j$ times from j till time $n-1$ and we have

$$\mathbb{E}\big[M_n^j \,\big|\, \mathscr{F}_{n-1}\big] = 26^{n-j+1} \cdot \frac{1}{26} + 0 \cdot \frac{25}{26} = 26^{n-j} = M_{n-1}^j.$$

Therefore, $(M_n^j, \mathscr{F}_n)$ is a martingale. Let $X_0 = 0$ and $X_n = \sum_{j=1}^n (M_n^j - 1)$ for $n \geq 1$, which is the total payoff of all gamblers at time n. It is clear that $(X_n, \mathscr{F}_n)$ is a martingale and $|X_n - X_{n-1}| \leq 26^{11} + 26^4 + 26 < \infty$ (consider the maximum increase or decrease in payoff going from time $n-1$ to n). From Lemma 13.1 (for the case where the martingale has bounded differences), we have

$$\begin{aligned} &\mathbb{E}\big[26^{11} + 26^4 + 26 - T\big] = \mathbb{E}X_T = \mathbb{E}X_0 = 0. \\ \implies &\mathbb{E}T = 26^{11} + 26^4 + 26. \end{aligned}$$

14.1 Submartingales

Definition 14.1. *Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a filtration $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \ldots \subset \mathscr{F}$. Suppose $X_n \in \mathscr{F}_n$ and $\mathbb{E}|X_n| < \infty$. If for all $m \geq n$, $\mathbb{E}[X_m \,|\, \mathscr{F}_n] \geq X_n$, then we say that $(X_n, \mathscr{F}_n)_{n\geq 0}$ is a* submartingale. *If $\mathbb{E}[X_m \,|\, \mathscr{F}_n] \leq X_n$, $(X_n, \mathscr{F}_n)_{n\geq 0}$ is called a* supermartingale.

The same proof as in Theorem 12.1 shows that the martingale transform w.r.t. a predictable sequence (A_n), which is bounded and non-negative, of a submartingale or supermartingale remains as a submartingale or supermartingale, respectively.

Theorem 14.1 (MTT)**.** *If (A_n) is predictable w.r.t. $(\mathscr{F}_n)$, non-negative and bounded, and $(X_n, \mathscr{F}_n)_{n\geq 0}$ is a submartingale (supermartingale), then $(\widetilde{X}_n, \mathscr{F}_n)_{n\geq 0}$ is a submartingale (supermartingale).*

Example 14.1. *In a casino game, let X_n be the amount of money you win at time n if you had bet one dollar at each time, starting with $X_0 = 0$. Recall that a predictable sequence (A_n) is a gambling strategy so that the winnings at time n is*

$$\widetilde{X}_n = \sum_{k=1}^{n} A_k (X_k - X_{k-1}).$$

Suppose that $M_n = X_n - X_{n-1} \in \{-1, 1\}$ has distribution $\mathrm{Bern}\,(p)$*. Your friend claims that a "sure-win" strategy is to choose $A_1 = 1$ and for $n \geq 2$,*

$$A_n = \begin{cases} 2A_{n-1} & \text{if } M_{n-1} = -1, \\ 1 & \text{if } M_{n-1} = 1. \end{cases}$$

Assuming that $p > 0$, $\{M_n = 1 \text{ i.o.}\}$ has probability one. Once $M_n = 1$, the above strategy recoups all your previous losses with a winning of $-1 - 2 - \ldots - 2^{n-1} + 2^n = 1$. Therefore, this strategy seems to suggest you can always beat the house. But if M_n is a supermartingale (i.e., $\mathbb{E}[M_m \,|\, \mathscr{F}_n] \leq M_n$ for $m \geq n$ and this is usually true has the house incorporates some advantages), then $\widetilde{M}_n$ is also a supermartingale. This means no strategy *can beat the house, in an expectation sense. Many gamblers have gone bankrupt using the above strategy!*

Lemma 14.1. *Suppose $f : \mathbb{R} \mapsto \mathbb{R}$ is convex, $(X_n, \mathscr{F}_n)_{n\geq 0}$ is an adaptation with $\mathbb{E}|f(X_n)| < \infty$ for all $n \geq 0$. Then if either*

i) $(X_n, \mathscr{F}_n)_{n\geq 0}$ is a martingale; or

ii) $(X_n, \mathscr{F}_n)_{n\geq 0}$ is a submartingale and f is increasing,

then $(f(X_n), \mathscr{F}_n)_{n\geq 0}$ is a submartingale.

Proof. To prove i), from Jensen's inequality, for $n \leq m$, we have $f(X_n) = f(\mathbb{E}[X_m \,|\, \mathscr{F}_n]) \leq \mathbb{E}[f(X_m) \,|\, \mathscr{F}_n]$. The proof of ii) is similar. $\square$

Theorem 14.2. *A submartingale $(X_n, \mathscr{F}_n)_{n\geq 0}$ can be decomposed a.s. uniquely as $X_n = Y_n + Z_n$, where $(Y_n, \mathscr{F}_n)_{n\geq 0}$ is a martingale and $(Z_n)_{n\geq 0}$ is predictable with $Z_0 = 0$ and $Z_n \leq Z_{n+1}$ a.s.*

Proof. We first construct a decomposition as follows: Let $Z_0 = 0$,

$$Z_n = \sum_{k=1}^{n} \mathbb{E}[X_k - X_{k-1} \,|\, \mathscr{F}_{k-1}]$$

and $Y_n = X_n - Z_n$. By our construction, $Z_n \in \mathscr{F}_{n-1}$ is predictable and since

$$\mathbb{E}[X_k - X_{k-1} \,|\, \mathscr{F}_{k-1}] = \mathbb{E}[X_k \,|\, \mathscr{F}_{k-1}] - X_{k-1} \geq 0,$$

we have $Z_n \leq Z_{n+1}$ a.s. Furthermore, $Z_n - Z_{n-1} = \mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] - X_{n-1}$ and we obtain

$$\mathbb{E}[Y_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}[X_n - Z_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] - Z_n = X_{n-1} - Z_{n-1} = Y_{n-1},$$

showing that Y_n is a martingale.

To show uniqueness, we proceed by induction. The requirement that $Z_0 = 0$ implies that $Y_0 = X_0$ uniquely. Suppose the decomposition is unique a.s. up to X_{n-1}. Then for any decomposition $X_n = Y_n + Z_n$ meeting the criteria,

$$Z_n = \mathbb{E}[Z_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}[X_n - Y_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] - Y_{n-1},$$

because Y_n is a martingale. This implies that Z_n is unique a.s., and hence so is $Y_n = X_n - Z_n$. The proof is now complete. □

14.2 Doob's Inequalities

Theorem 14.3 (Doob's maximal inequality)**.** *Suppose $(X_n, \mathscr{F}_n)_{n\geq 0}$ is a non-negative submartingale. Let $X_n^* = \max_{0\leq k\leq n} X_k$. Then for all $\lambda \geq 0$,*

$$\lambda \mathbb{P}(X_n^* \geq \lambda) \leq \mathbb{E} X_n \mathbf{1}_{\{X_n^* \geq \lambda\}} \leq \mathbb{E} X_n. \tag{42}$$

Proof. Let $\tau = \inf\{k : X_k \geq \lambda\}$ be a stopping time. Then $\{X_n^* \geq \lambda\} = \{\tau \leq n\}$ and

$$\lambda \mathbf{1}_{\{\tau\leq n\}} \leq X_\tau \mathbf{1}_{\{\tau\leq n\}} = \sum_{k=0}^{n} X_k \mathbf{1}_{\{\tau=k\}}. \tag{43}$$

Since $X_k \leq \mathbb{E}[X_n \,|\, \mathscr{F}_k]$ $\forall\, k \leq n$ and $\{\tau = k\} \in \mathscr{F}_k$, $\mathbb{E} X_k \mathbf{1}_{\{\tau=k\}} \leq \mathbb{E} X_n \mathbf{1}_{\{\tau=k\}}$ for all $k \leq n$. Taking expectations in (43), we obtain

$$\begin{aligned}
\lambda \mathbb{P}(\tau \leq n) &\leq \mathbb{E} \sum_{k=0}^{n} X_k \mathbf{1}_{\{\tau=k\}} \\
&\leq \mathbb{E} \sum_{k=0}^{n} X_n \mathbf{1}_{\{\tau=k\}} \\
&= \mathbb{E} X_n \mathbf{1}_{\{\tau\leq n\}} \\
&= \mathbb{E} X_n \mathbf{1}_{\{X_n^* \geq \lambda\}} \\
&\leq \mathbb{E} X_n.
\end{aligned}$$

□

Corollary 14.1. *For $\lambda > 0$ and $p \geq 1$,*

$$\mathbb{P}(X_n^* \geq \lambda) \leq \frac{1}{\lambda^p}\mathbb{E}X_n^p.$$

Proof. From Lemma 14.1, (X_n^p) is a non-negative submartingale. We then apply Theorem 14.3. $\square$

Example 14.2. *Suppose X_i, $i \geq 1$ are independent with $\mathbb{E}X_i = 0$. From Example 12.3, S_n is a martingale. Doob's maximal inequality (or its corollary) recovers Kolmogorov's maximal equality:*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2}\mathbb{E}S_n^2.$$

Let $p \geq 1$. A r.v. $X \in L^p$ if $\|X\|_p = \{\mathbb{E}|X|^p\}^{1/p} < \infty$. Hölder's inequalty says that for $1 \leq p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$ (i.e., $q = \dfrac{p}{p-1}$), then for any r.v.s X, Y,

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Lemma 14.2. *Suppose $X, Y \geq 0$, $\mathbb{E}Y^p < \infty$ for $p > 1$ and for all $\lambda \geq 0$, we have*

$$\lambda \mathbb{P}(X \geq \lambda) \leq \mathbb{E}Y\mathbf{1}_{\{X \geq \lambda\}}.$$

Then,

$$\|X\|_p \leq \frac{p}{p-1}\|Y\|_p. \tag{44}$$

Proof. Let $X_n = X \wedge n$. We use the fact that

$$z^p = p\int_0^z x^{p-1}\,\mathrm{d}x = p\int_0^\infty x^{p-1}\mathbf{1}_{\{z \geq x\}}\,\mathrm{d}x$$

to obtain

$$\begin{aligned}
\mathbb{E}X_n^p &= \mathbb{E}\left[p\int_0^\infty x^{p-1}\mathbf{1}_{\{X_n \geq x\}}\,\mathrm{d}x\right] \\
&\overset{\text{Fubini}}{=} p\int_0^\infty x^{p-1}\mathbb{P}(X_n \geq x)\,\mathrm{d}x \\
&\leq p\int_0^\infty x^{p-2}\mathbb{E}\big[Y\mathbf{1}_{\{X \geq x\}}\big]\,\mathrm{d}x \quad \text{since } \{X_n \geq x\} \subset \{X \geq x\} \\
&\overset{\text{Fubini}}{=} p\mathbb{E}\left[Y\int_0^\infty x^{p-2}\mathbf{1}_{\{X \geq x\}}\,\mathrm{d}x\right] \\
&= \frac{p}{p-1}\mathbb{E}\big[YX^{p-1}\big] \\
&\overset{\text{Hölder}}{\leq} \frac{p}{p-1}\|Y\|_p\|X\|_p^{p-1}.
\end{aligned}$$

Therefore, from Fatou's lemma,

$$\|X\|_p^p \leq \liminf_{n\to\infty}\mathbb{E}X_n^p \leq \frac{p}{p-1}\|Y\|_p\|X\|_p^{p-1},$$

and we obtain (44) (the case $\|X\|_p = 0$ is trivial). $\square$

Theorem 14.4 (Doob's L^p inequality). *If $(X_n, \mathscr{F}_n)$ is a non-negative submartingale, then for all $p > 1$,*

$$\|X_n^*\|_p \leq \frac{p}{p-1}\|X_n\|_p. \tag{45}$$

Proof. We apply Lemma 14.2 with $X = X_n^*$ and $Y = X_n$, together with Doob's maximal inequality (Theorem 14.3). $\square$

Theorems 14.3 and 14.4 require $(X_n, \mathscr{F}_n)$ to be non-negative. For a general submartingale $(X_n, \mathscr{F}_n)$, we can apply these results to $(X_n^+, \mathscr{F}_n)$ since this is a non-negative submartingale from Lemma 14.1.

Theorem 14.4 holds only for $p > 1$. Indeed, there is no corresponding result for $p = 1$ as shown in the following example.

Example 14.3. *Consider the random walk in Example 13.1 with $S_0 = 0$. Take $B = 1$ and $\tau = \inf\{n : S_n = -1\}$. Let $X_n = S_{n\wedge\tau}$. Then using a result in Example 13.1 in the second equality below, we have*

$$\mathbb{E}\left[\max_{m\geq 0} X_m\right] = \sum_{A=1}^{\infty} \mathbb{P}\left(\max_{m\geq 0} X_m \geq A\right) = \sum_{A=1}^{\infty} \frac{1}{A+1} = \infty.$$

The MCT then implies that $\mathbb{E}[\max_{0\leq m\leq n} X_m] \to \infty$ as $n \to \infty$.

14.3 Uniform Integrability

Definition 14.2. *A collection of r.v.s $(X_n)_{n\in N}$ is uniformly integrable if*

$$\sup_{n\in N} \mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} \to 0, \tag{46}$$

as $K \to \infty$.

If $(X_n)_{n\in N}$ is uniformly integrable, then for K sufficiently large, we have $\sup_{n\in N} \mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} \leq 1$ and $\sup_n \mathbb{E}|X_n| \leq K + 1 < \infty$ is uniformly bounded. Clearly, the converse is false.

Example 14.4. *If $|X_n| \leq Y$, $\forall\, n \in N$, and $\mathbb{E}Y < \infty$, then*

$$\mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} \leq \mathbb{E}Y\mathbf{1}_{\{Y>K\}} \to 0,$$

as $K \to \infty$ from DCT. Therefore $(X_n)_{n\in N}$ is uniformly integrable.

Lemma 14.3. *If $X \in L^1$, then $\forall\, \epsilon > 0$, $\exists\, \delta > 0$ such that if $\mathbb{P}(A) \leq \delta$, then $\mathbb{E}|X|\mathbf{1}_A \leq \epsilon$.*

Proof. If $\mathbb{P}(A) \leq \delta$, we have for all $K > 0$,

$$\mathbb{E}|X|\mathbf{1}_A \leq K\mathbb{P}(A) + \mathbb{E}|X|\mathbf{1}_{\{|X|>K\}} \leq K\delta + \mathbb{E}|X|\mathbf{1}_{\{|X|>K\}}.$$

Choose K sufficiently large so that $\mathbb{E}|X|\mathbf{1}_{\{|X|>K\}} \leq \dfrac{\epsilon}{2}$ and set $\delta = \dfrac{\epsilon}{2K}$. Then from above, we have $\mathbb{E}|X|\mathbf{1}_A \leq \epsilon$. $\square$

Proposition 14.1. *Let $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Then $\{\mathbb{E}[X \mid \mathscr{F}'] : \mathscr{F}' \text{ is a } \sigma\text{-algebra} \subset \mathscr{F}\}$ is uniformly integrable.*

Proof. Fix $\epsilon > 0$ and choose $\delta > 0$ as in Lemma 14.3. Pick K large so that $\mathbb{E}|X|/K \le \delta$. Let $Y = \mathbb{E}[X \mid \mathscr{F}']$. From Jensen's inequality, $|Y| \le \mathbb{E}[|X| \mid \mathscr{F}']$, therefore we have

$$\begin{aligned}
\mathbb{E}\big[|Y|\mathbf{1}_{\{|Y|>K\}}\big] &\le \mathbb{E}\big[\mathbb{E}[|X| \mid \mathscr{F}']\mathbf{1}_{\{\mathbb{E}[|X||\mathscr{F}']>K\}}\big] \\
&= \mathbb{E}|X|\mathbf{1}_{\{\mathbb{E}[|X||\mathscr{F}']>K\}} \quad \text{since } \{\mathbb{E}[|X| \mid \mathscr{F}'] > K\} \in \mathscr{F}' \\
&\le \epsilon,
\end{aligned}$$

where the last inequalty follows from Lemma 14.3 as $\mathbb{P}(\mathbb{E}[|X| \mid \mathscr{F}'] > K) \le \mathbb{E}|X|/K \le \delta$. $\square$

Proposition 14.2. *Suppose $\varphi : \mathbb{R} \mapsto \mathbb{R}_+$ is such that $\lim_{x\to\infty} \dfrac{\varphi(x)}{x} = \infty$. (Examples include $\varphi(x) = x^p$, for $p > 1$ and $\varphi(x) = x \log^+ x$.) If $\mathbb{E}\varphi(|X_n|) \le C < \infty$, then $(X_n)_{n\in N}$ is uniformly integrable.*

Proof. Let $\epsilon_K = \sup\{x/\varphi(x) : x \ge K\}$. Note that $\epsilon_K \to 0$ as $K \to \infty$ because for any $\epsilon > 0$, $\exists\, K$ sufficiently large so that $x/\varphi(x) \le \epsilon$ for all $x > K$. Then we have

$$\mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} \le \epsilon_K \mathbb{E}\big[\varphi(|X_n|)\mathbf{1}_{\{|X_n|>K\}}\big] \le C\epsilon_K \to 0,$$

as $K \to \infty$. $\square$

Lemma 14.4. *Suppose $\mathbb{E}|X_n| < \infty$ for all $n \in N$ and $\mathbb{E}|X| < \infty$, then the following are equivalent:*

(i) $X_n \to X$ *in* L^1, *i.e.,* $\mathbb{E}|X_n - X| \to 0$ *as* $n \to \infty$.

(ii) $(X_n)_{n\in N}$ *is uniformly integrable and* $X_n \xrightarrow{\text{p}} X$.

(iii) $X_n \xrightarrow{\text{p}} X$ *and* $\mathbb{E}|X_n| \to \mathbb{E}|X|$.

Proof. We show $(ii) \implies (i) \implies (iii) \implies (ii)$.

$(ii) \implies (i)$: $\forall\, \epsilon > 0$, $K > 0$, we have

$$\begin{aligned}
\mathbb{E}|X_n - X| &\le \epsilon + \mathbb{E}|X_n - X|\mathbf{1}_{\{|X_n-X|>\epsilon\}} \\
&\le \epsilon + \mathbb{E}|X_n|\mathbf{1}_{\{|X_n-X|>\epsilon\}} + \mathbb{E}|X|\mathbf{1}_{\{|X_n-X|>\epsilon\}} \\
&\le \epsilon + 2K\mathbb{P}(|X_n - X| > \epsilon) + \mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} + \mathbb{E}|X|\mathbf{1}_{\{|X|>K\}}
\end{aligned}$$

and

$$\limsup_{n\to\infty} \mathbb{E}|X_n - X| \le \epsilon + \sup_n \mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} + \mathbb{E}|X|\mathbf{1}_{\{|X|>K\}}.$$

Taking $\epsilon \to 0$ and $K \to \infty$ completes the proof.

$(i) \implies (iii)$: From Markov's inequality, for any $\epsilon > 0$, we obtain $\mathbb{P}(|X_n - X| > \epsilon) \le \frac{1}{\epsilon}\mathbb{E}|X_n - X| \to 0$ as $n \to \infty$. We also have

$$|\mathbb{E}|X_n| - \mathbb{E}|X|| \le \mathbb{E}||X_n| - |X|| \le \mathbb{E}|X_n - X| \to 0,$$

as $n \to \infty$.

$(iii) \implies (ii)$: For any $\epsilon > 0$, $\exists\, n_0$ such that $\forall\, n \ge n_0$, $\mathbb{E}|X_n| \le \mathbb{E}|X| + \epsilon/2$. Let

$$\phi_K(x) = \begin{cases} x, & \text{for } x \in [0, K-1], \\ 0, & \text{for } x > K, \\ \text{linear}, & \text{for } x \in [K-1, K]. \end{cases}$$

Then from the DCT, for K sufficiently large,

$$\mathbb{E}|X| - \mathbb{E}\phi_K(|X|) \leq \epsilon.$$

Since ϕ_K is continuous, the DCT also yields $\mathbb{E}\phi_K(|X_n|) \to \mathbb{E}\phi_K(|X|)$ as $n \to \infty$. Therefore, since $x \geq \phi_K(x) + x\mathbf{1}_{\{x>K\}}$ for all $x \geq 0$, we have

$$\begin{aligned}
\mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} &\leq \mathbb{E}|X_n| - \mathbb{E}\phi_K(|X_n|) \\
&\leq \mathbb{E}|X| - \mathbb{E}\phi_K(|X|) + \epsilon \\
&\leq 2\epsilon,
\end{aligned}$$

for all n and K sufficiently large and the proof is complete. $\square$

15.1 Right-Closable Martingale

We can generalize our index set to any totally ordered set $(N, \leq)$ (if $n, m \in N$, then $n \leq m$ or $m \leq n$). We say that $(\mathscr{F}_n)_{n\in N}$ is a filtration if $\mathscr{F}_n \subset \mathscr{F}_m$ for all $n \leq m$ in N.

Definition 15.1. *A submartingale $(X_n, \mathscr{F}_n)_{n\in N}$ is* right-closable *if $X_n \leq \mathbb{E}[X \,|\, \mathscr{F}_n]$ for all $n \in N$, for some $X \in L^1$. (For a martingale, the inequality is replaced with equality.)*

If there exists $n_0 \in N$ such that $n \leq n_0$ for all $n \in N$, then $X_n \leq \mathbb{E}[X_{n_0} \,|\, \mathscr{F}_n]$ for all $n \in N$, and we say that the submartingale is right-closed*.*

Note that a right-closable submartingale or martingale can always be made right-closed by simply adding an upper bound n_0 to N and defining $X_{n_0} = X$.

Example 15.1 (Reverse martingale)**.** *Suppose X_k, $k \geq 1$ are i.i.d. Let $S_n = \sum_{k=1}^n X_k$. Take N to be the set of negative integers and for $n \geq 1$, let*

$$\mathscr{F}_{-n} = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots).$$

Since $\mathscr{F}_{-(n+1)} \subset \mathscr{F}_{-n}$, $(\mathscr{F}_{-n})_{-n\in N}$ is a filtration. By symmetry, we have for all $k = 1, \ldots, n$,

$$\mathbb{E}[X_k \,|\, \mathscr{F}_{-n}] = \mathbb{E}[X_1 \,|\, \mathscr{F}_{-n}].$$

Therefore,

$$S_n = \mathbb{E}[S_n \,|\, \mathscr{F}_{-n}] = \sum_{k=1}^n \mathbb{E}[X_k \,|\, \mathscr{F}_{-n}] = n\mathbb{E}[X_1 \,|\, \mathscr{F}_{-n}].$$

Let $Z_{-n} = \dfrac{S_n}{n} = \mathbb{E}[X_1 \,|\, \mathscr{F}_{-n}]$. Then $(Z_{-n}, \mathscr{F}_{-n})_{-n\in N}$ is a right-closed martingale.

Lemma 15.1. *(i) If $(X_n, \mathscr{F}_n)_{n\in N}$ is a right-closable martingale, then $(X_n)_{n\in N}$ is uniformly integrable.*

(ii) If $(X_n, \mathscr{F}_n)_{n\in N}$ is a right-closable submartingale, then $(X_n \vee a)_{n\in N}$ is uniformly integrable for all $a \in \mathbb{R}$.

Proof. (i) Since the martingale is right-closable, $\exists\, X \in L^1$ such that $X_n = \mathbb{E}[X \,|\, \mathscr{F}_n]$ for all $n \in N$. We have $|X_n| \leq \mathbb{E}[|X| \,|\, \mathscr{F}_n]$ and $\mathbb{E}|X_n| \leq \mathbb{E}|X| < \infty$. Then,

$$\begin{aligned}
\mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} &\leq \mathbb{E}\big[\mathbb{E}[|X| \,|\, \mathscr{F}_n]\mathbf{1}_{\{|X_n|>K\}}\big] \\
&= \mathbb{E}|X|\mathbf{1}_{\{|X_n|>K\}} \quad \because \{|X_n| > K\} \in \mathscr{F}_n \\
&\leq M\mathbb{P}(|X_n| > K) + \mathbb{E}|X|\mathbf{1}_{\{|X|>M\}} \quad \forall\, M > 0 \\
&\overset{\text{Markov}}{\leq} \frac{M}{K}\mathbb{E}|X_n| + \mathbb{E}|X|\mathbf{1}_{\{|X|>M\}}
\end{aligned}$$

and

$$\limsup_{K\to\infty} \mathbb{E}|X_n|\mathbf{1}_{\{|X_n|>K\}} \leq \mathbb{E}|X|\mathbf{1}_{\{|X|>M\}} \xrightarrow{M\to\infty} 0.$$

(ii) Since the submartingale is right-closable, $\exists X \in L^1$ such that $X_n \le \mathbb{E}[X \mid \mathscr{F}_n]$ for all $n \in N$. We have

$$\mathbb{E}\big[X_n \mathbf{1}_{\{X_n > K\}}\big] \le \mathbb{E}\big[X \mathbf{1}_{\{X_n > K\}}\big] \tag{47}$$
$$\mathbb{E}[X_n \vee a] \le E[X \vee a] \quad \text{from Jensen's inequality.} \tag{48}$$

Take $K > |a|$. Then $|X_n \vee a| > K$ iff $X_n \vee a = X_n > K$. Therefore,

$$\begin{aligned}
\mathbb{E}\big[(X_n \vee a)\mathbf{1}_{\{|X_n \vee a| > K\}}\big] &= \mathbb{E}\big[X_n \mathbf{1}_{\{X_n > K\}}\big] \\
&\overset{(47)}{\le} \mathbb{E}\big[X \mathbf{1}_{\{X_n > K\}}\big] \\
&\le M\mathbb{P}(X_n > K) + \mathbb{E}\big[X \mathbf{1}_{\{X > M\}}\big] \quad \forall M > 0 \\
&\overset{\text{Markov}}{\le} \frac{M}{K}\mathbb{E}X_n^+ + \mathbb{E}|X|\mathbf{1}_{\{|X| > M\}} \\
&\le \frac{M}{K}\mathbb{E}X^+ + \mathbb{E}|X|\mathbf{1}_{\{|X| > M\}} \quad \text{from letting } a = 0 \text{ in (48).}
\end{aligned}$$

Taking $K \to \infty$ and then $M \to \infty$, we obtain the desired result.

□

15.2 Doob's Upcrossing Inequality

Consider a submartingale $(X_n, \mathscr{F}_n)_{n \ge 0}$. Let $a < b$. The first times that X_n crosses a downwards and then b upwards are given by the stopping times

$$\tau_1 = \inf\{n \ge 0 \,:\, X_n \le a\}, \;\; \tau_2 = \inf\{n > \tau_1 \,:\, X_n \ge b\}.$$

We repeat this process and define by induction, for $k \ge 2$,

$$\tau_{2k-1} = \inf\{n > \tau_{2k-2} \,:\, X_n \le a\}, \;\; \tau_{2k} = \inf\{n > \tau_{2k-1} \,:\, X_n \ge b\}.$$

Then

$$U_n(a,b) = \sup\{k \,:\, \tau_{2k} \le n\}$$

is the number of upward crossings of the interval $[a,b]$ up to time n.

Theorem 15.1 (Doob's upcrossing inequality)**.** *If $(X_n, \mathscr{F}_n)_{n \ge 0}$ is a submartingale, then for any $a < b$,*

$$(b-a)\mathbb{E}U_n(a,b) \le \mathbb{E}(X_n - a)^+.$$

Proof. Let $Y_n = (X_n - a)^+$, which is a non-negative submartingale by Lemma 14.1. Then $U_n(a,b)$ is equal to the number of upcrossings of $[0, b-a]$ by $Y_0, \dots, Y_n$. For $m \ge 1$, let

$$A_m = \begin{cases} 1, & \text{if } \tau_{2k-1} < m \le \tau_{2k} \text{ for some } k, \\ 0, & \text{otherwise.} \end{cases}$$

Since $\{\tau_{2k-1} < m \le \tau_{2k}\} = \{\tau_{2k-1} \le m-1\} \cap \{\tau_{2k} \le m-1\}^{\mathsf{c}} \in \mathscr{F}_{m-1}$, $(A_m)_{m \ge 1}$ is predictable. Consider the martingale transform

$$\tilde{Y}_n = Y_0 + \sum_{m=1}^{n} A_m (Y_m - Y_{m-1}) \ge (b-a)U_n(a,b), \tag{49}$$

where the inequality follows because each upcrossing of $[0, b-a]$ by (Y_n) results in a gain of at least $b-a$ on the left-hand side. We have

$$\begin{aligned}\mathbb{E}[A_m(Y_m - Y_{m-1}) \,|\, \mathscr{F}_{m-1}] &= A_m \mathbb{E}[Y_m - Y_{m-1} \,|\, \mathscr{F}_{m-1}] \\ &\leq \mathbb{E}[Y_m - Y_{m-1} \,|\, \mathscr{F}_{m-1}],\end{aligned}$$

where the inequality follows because $\mathbb{E}[Y_m \,|\, \mathscr{F}_{m-1}] \geq Y_{m-1}$. Therefore, $\mathbb{E}[A_m(Y_m - Y_{m-1})] \leq \mathbb{E}[Y_m - Y_{m-1}]$ and

$$\begin{aligned}\mathbb{E}\tilde{Y}_n - \mathbb{E}Y_0 &\leq \mathbb{E}Y_n - \mathbb{E}Y_0 \\ \mathbb{E}\tilde{Y}_n &\leq \mathbb{E}Y_n\end{aligned}$$

and from (49),

$$(b-a)\mathbb{E}U_n(a,b) \leq \mathbb{E}Y_n.$$

□

15.3 Submartingale Convergence

Theorem 15.2 (Submartingale Convergence Theorem)**.** *Suppose $(X_n, \mathscr{F}_n)_{-\infty<n<\infty}$ is a submartingale.*

(i) $X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. and $\mathbb{E}X_{-\infty}^+ < \infty$. Let $\mathscr{F}_{-\infty} = \bigcap_n \mathscr{F}_n$. Then $(X_n, \mathscr{F}_n)_{-\infty\leq n<\infty}$ is a submartingale, i.e., $X_{-\infty} \leq \mathbb{E}[X_m \,|\, \mathscr{F}_{-\infty}]$ for all $m > -\infty$.

(ii) If $\mathbb{E}X_n^+ \leq B < \infty$, then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and $\mathbb{E}X_\infty^+ \leq B$.

(iii) If $(X_n^+)_{-\infty<n<\infty}$ is uniformly integrable, then for $X_\infty = \lim_{n\to\infty} X_n$, we have $X_m \leq \mathbb{E}[X_\infty \,|\, \mathscr{F}_m]$ for all $m < \infty$. Let $\mathscr{F}_\infty = \sigma(\bigcup_n \mathscr{F}_n)$. Then $(X_n, \mathscr{F}_n)_{-\infty<n\leq\infty}$ is right-closed.

Proof. Note that X_n converges as $n \to \pm\infty$ iff $\limsup X_n = \liminf X_n$,[3] i.e., X_n diverges only on the event

$$\{\limsup X_n > \liminf X_n\} = \bigcup_{\substack{a<b,\\ a,b\in\mathbb{Q}}} A_{ab},$$

where

$$A_{ab} = \{\limsup X_n \geq b > a \geq \liminf X_n\} \subset \{\lim U_n(a,b) = \infty\}.$$

Therefore to show that $\mathbb{P}(A_{ab}) = 0$, it suffices to show that $\mathbb{E}[\lim U_n(a,b)] < \infty$.

(i) For each $n \geq 1$, $U_n(a,b)$ is the number of upcrossings of the submartingale $Y_1 = X_{-n}, Y_2 = X_{-n+1}, \ldots, Y_n = X_{-1}$. From Doob's upcrossing inequality (Theorem 15.1), we have

$$\mathbb{E}U_n(a,b) \leq \frac{\mathbb{E}(Y_n - a)^+}{b-a} = \frac{\mathbb{E}(X_{-1} - a)^+}{b-a} < \infty.$$

The MCT then gives $\mathbb{E}\lim_{n\to\infty} U_n(a,b) = \lim_{n\to\infty} \mathbb{E}U_n(a,b) < \infty$. Therefore $X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. From Fatou's lemma,

$$\mathbb{E}X_{-\infty}^+ \leq \liminf_{n\to\infty} \mathbb{E}X_{-n}^+ \leq \mathbb{E}X_{-1}^+ < \infty.$$

[3] In this proof, if the convergence is not specified, it is taken to mean either $n \to \infty$ or $n \to -\infty$.

Finally, to show that $(X_n, \mathscr{F}_n)_{-\infty \le n < \infty}$ is a submartingale, we note that $X_{-\infty} \in \mathscr{F}_n$ for all n (Exercise), which implies that $X_{-\infty} \in \mathscr{F}_{-\infty}$. Furthermore, to show $X_{-\infty} \le \mathbb{E}[X_m \,|\, \mathscr{F}_{-\infty}]$ for all $m > -\infty$, it suffices to show that for all $A \in \mathscr{F}_{-\infty}$, we have $\mathbb{E}X_{-\infty}\mathbf{1}_A \le \mathbb{E}X_m\mathbf{1}_A$ (cf. Lemma 12.1).

Consider the submartingale $(X_n, \mathscr{F}_n)_{-\infty \le n \le 0}$, which is right-closed. From Lemma 15.1, $(X_n \vee a, \mathscr{F}_n)_{-\infty \le n \le 0}$ is uniformly integrable for all $a \in \mathbb{R}$. Furthermore, $X_n \vee a \to X_{-\infty} \vee a$ a.s. as $n \to \infty$. From Lemma 14.4, $X_n \vee a \to X_{-\infty} \vee a$ in L^1, hence

$$\mathbb{E}(X_{-\infty} \vee a)\mathbf{1}_A = \lim_{n \to -\infty} \mathbb{E}(X_n \vee a)\mathbf{1}_A \le \mathbb{E}(X_m \vee a)\mathbf{1}_A,$$

for all $m > -\infty$. The last inequality follows because $\mathbb{E}(X_n \vee a)\mathbf{1}_A \le \mathbb{E}(X_m \vee a)\mathbf{1}_A$ for all $n \le m$, which in turn is a consequence of the facts that $(X_n \vee a)$ is a submartingale and $A \in \mathscr{F}_m$. Taking $a \to -\infty$, the MCT gives $\mathbb{E}X_{-\infty}\mathbf{1}_A \le \mathbb{E}X_m\mathbf{1}_A$.

(ii) From Doob's upcrossing inequality,

$$\mathbb{E}U_n(a,b) \le \frac{\mathbb{E}(X_n - a)^+}{b-a} \le \frac{|a| + \mathbb{E}X_n^+}{b-a} \le \frac{|a| + B}{b-a} < \infty.$$

The same argument as in the proof of Item (i) completes the proof of Item (ii).

(iii) We want to show that for all $m < \infty$ and $A \in \mathscr{F}_m$, $\mathbb{E}X_m\mathbf{1}_A \le \mathbb{E}X_\infty\mathbf{1}_A$. From Lemma 14.1, $(X_n \vee a, \mathscr{F}_n)_{n<\infty}$ is a submartingale for all $a \in \mathbb{R}$. For $K > 0$ and $K > a$, $X_n \vee a > K$ iff $X_n \vee a = X_n^+ > K$. Therefore, since (X_n^+) is uniformly integrable, so is $(X_n \vee a)_{n<\infty}$. As $X_n \vee a \to X_\infty \vee a$ a.s. as $n \to \infty$, Lemma 14.4 again yields that $X_n \vee a \to X_\infty \vee a$ in L^1, and hence

$$\mathbb{E}(X_\infty \vee a)\mathbf{1}_A = \lim_{n \to \infty} \mathbb{E}(X_n \vee a)\mathbf{1}_A \ge \mathbb{E}(X_m \vee a)\mathbf{1}_A,$$

for all $m < \infty$. Taking $a \to -\infty$ and applying the MCT give the desired result.

□

Corollary 15.1 (Martingale Convergence Theorem)**.** *Suppose $(X_n, \mathscr{F}_n)_{n<\infty}$ is a martingale.*

(i) If $\sup_n \mathbb{E}|X_n| < \infty$ (equivalently, $\sup_n \mathbb{E}X_n^+ < \infty$ or $\sup_n \mathbb{E}X_n^- < \infty$), then $X_n \to X_\infty$ a.s. and $\mathbb{E}|X_\infty| < \infty$.

(ii) If $(X_n)_{n<\infty}$ is uniformly integrable, then $(X_n, \mathscr{F}_n)_{n \le \infty}$ is a right-closed martingale.

Corollary 15.1 follows from Theorem 15.2 by simply observing that X_n and $-X_n$ are both submartingales. Furthermore, Corollary 15.1 together with Lemma 15.1 says that a martingale is right-closable iff it is uniformly integrable.

Example 15.2. *Consider the simple random walk S_n in Example 13.1 with stopping time τ when S_n hits the boundaries A or $-B$. Let $M_n = S_{n \wedge \tau}$, which is a bounded martingale. From the Martingale Convergence Theorem (Corollary 15.1), M_n converges a.s. For $\omega \in \{\tau = \infty\}$, we have $|M_n(\omega) - M_{n+1}(\omega)| = |S_n(\omega) - S_{n+1}(\omega)| = 1$, hence $M_n(\omega)$ does not converge. Therefore $\mathbb{P}(\tau = \infty) = 0$ and $\mathbb{P}(\tau < \infty) = 1$.*

Corollary 15.2 (Supermartingale Convergence Theorem)**.** *Suppose $(X_n, \mathscr{F}_n)_{n \ge 0}$ is a supermartingale. If $\sup_n \mathbb{E}X_n^- < \infty$, then $X_n \to X_\infty$ a.s., and $\mathbb{E}X_\infty \le \mathbb{E}X_0$.*

Proof. $-X_n$ is submartingale with $\mathbb{E}(-X_n)^+ = \mathbb{E}X_n^-$. □

Theorem 15.3 (Levy's Convergence Theorem)**.** *Given $\mathbb{E}|X| < \infty$ and a filtration $(\mathscr{F}_n)_{n \ge 0}$, let $\mathscr{F}_\infty = \sigma(\bigcup_n \mathscr{F}_n)$. Then $\mathbb{E}[X \,|\, \mathscr{F}_n] \to \mathbb{E}[X \,|\, \mathscr{F}_\infty]$ a.s.*

Proof. From Example 12.7, $X_n = \mathbb{E}[X \mid \mathscr{F}_n]$ is a martingale and $\mathbb{E}|X_n| \leq \mathbb{E}|X| < \infty$. From Corollary 15.1, $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. We next show that $X_\infty = \mathbb{E}[X \mid \mathscr{F}_\infty]$.

Let $A \in \bigcup_n \mathscr{F}_n$. Then there exists m such that $A \in \mathscr{F}_m$ and since X_n is a martingale, we have $\mathbb{E}X_n\mathbf{1}_A = \mathbb{E}X_m\mathbf{1}_A$ for all $n \geq m$. As $(X_n, \mathscr{F}_n)$ is right-closable, it is uniformly integrable, and Lemma 14.4 gives

$$\begin{aligned}
\mathbb{E}X_\infty\mathbf{1}_A &= \lim_{n\to\infty} \mathbb{E}X_n\mathbf{1}_A \\
&= \mathbb{E}X_m\mathbf{1}_A \\
&= \mathbb{E}[\mathbb{E}[X \mid \mathscr{F}_m]\mathbf{1}_A] \\
&= \mathbb{E}X\mathbf{1}_A.
\end{aligned}$$

Since $\bigcup_n \mathscr{F}_n$ is an algebra (because $(\mathscr{F}_n)$ is a filtration), the π-λ theorem completes the proof. ☐

Example 15.3 (Improved SLLN)**.** *For a sequence of r.v.s $(X_n)_{n\geq 1}$, let $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$, $n \geq 1$, be a filtration. Suppose that $\mathbb{E}[X_n \mid \mathscr{F}_{n-1}] = 0$ for all $n \geq 1$. In particular, the r.v.s are not necessarily pairwise independent. Then for $n > m$, we have $\mathbb{E}X_nX_m = \mathbb{E}[\mathbb{E}[X_nX_m \mid \mathscr{F}_m]] = \mathbb{E}[X_m\mathbb{E}[X_n \mid \mathscr{F}_m]] = 0$. Let $S_n = \sum_{k=1}^n X_k$.*

Suppose that $(b_n)_{n\geq 1}$ is a sequence of positive constants increasing to ∞ and $\sum_{n=1}^\infty \dfrac{\mathbb{E}X_n^2}{b_n^2} < \infty$, then $\dfrac{S_n}{b_n} \to 0$ a.s. To see this, let $Y_n = \sum_{k=1}^n \dfrac{X_k}{b_n}$. It is easy to verify that this is a martingale. For $K > 0$, we have

$$\mathbb{E}|Y_n|\mathbf{1}_{\{|Y_n|>K\}} \leq \frac{1}{K}\mathbb{E}|Y_n|^2 = \frac{1}{K}\sum_{k=1}^n \frac{\mathbb{E}X_k^2}{b_k^2}.$$

We then have

$$\sup_n \mathbb{E}|Y_n|\mathbf{1}_{\{|Y_n|>K\}} \leq \frac{1}{K}\sum_{k=1}^\infty \frac{\mathbb{E}X_k^2}{b_k^2} \xrightarrow{K\to\infty} 0.$$

Therefore, (Y_n) is uniformly integrable. From the Martingale Convergence Theorem (Corollary 15.1), Y_n converges a.s. The generalized Kronecker's lemma (Lemma 7.5) then implies that $\dfrac{S_n}{b_n} \to 0$ a.s.

Example 15.4. *Continuing from Example 15.1, $(Z_{-n}, \mathscr{F}_{-n})_{-n\leq 1}$ is a right-closed martingale. The Submartingale Convergence Theorem (Theorem 15.2) says that $\lim_{n\to\infty} Z_{-n} = Z_{-\infty} \in \mathscr{F}_{-\infty} = \bigcap_{n\geq 1} \mathscr{F}_{-n}$. Each event in $\mathscr{F}_{-n}$ is invariant under finite permutations, therefore by the Hewitt-Savage 0-1 Law, it has probability 0 or 1. Hence, $Z_{-\infty}$ is a constant a.s. But $\mathbb{E}Z_{-n} = \mathbb{E}X_1$ for all $n \geq 1$, so $Z_{-\infty} = \mathbb{E}X_1$ a.s., i.e., $\dfrac{S_n}{n} \to \mathbb{E}X_1$ a.s., which is Kolmogorov's SLLN.*

Example 15.5 (Levy's 0-1 Law)**.** *From Theorem 15.3, if $\mathscr{F}_n \uparrow \mathscr{F}_\infty$, then for any $A \in \mathscr{F}_\infty$, we have $\mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_n] \to \mathbf{1}_A$ a.s. The famous probabilist K. L. Chung once commented that this result "is obvious or incredible".*

1. *It is obvious: $\mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_\infty] = \mathbf{1}_A$.*
2. *It is incredible: Consider X_i, $i \geq 1$, independent with $\mathscr{F}_n = \sigma(X_1, \ldots, X_n) \uparrow \mathscr{F}_\infty$. An event $A \in \mathcal{T} = \bigcap_{n\geq 1} \sigma(X_n, X_{n+1}, \ldots)$ the tail σ-algebra is independent of $\mathscr{F}_n$ for all $n \geq 1$ (recall the grouping lemma (Lemma 8.3)). Therefore, $\mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_n] \xrightarrow{n\to\infty} \mathbf{1}_A \in \{0, 1\}$ a.s. from Levy's 0-1 Law, which recovers Kolmogorov's 0-1 Law!*