Probability Theory I
Problem Sets w/ Solutions

Rex C. Liu, Courant Institute of Mathematical Sciences

January 9, 2023

Disclaimer:

These are the problem sets for the course *Probability Theory I (MATH.GA 2911)*, given by professor Paul Bourgade at New York University in Fall 2022. This course is aimed primarily for PhD students. Topics include laws of large numbers, weak convergence, central limit theorems, conditional expectation, martingales and Markov chains. The reference text is [3].

The solutions are given by Rex Liu with help from Xiang Fang (UCSB PSTAT), my classmates Andrew Zhang, Asher Miao, Michael Ye, and references from [1], [2], [4]. If you see any mistakes or think that the presentation is unclear and could be improved, please send an email to: cl5682@nyu.edu. All comments and suggestions are appreciated.

Contents

Chapter 1

Questions

1.1 Measure Theory

1.1.1 Construction of Measures

1. Let $(\mathcal{G}_\alpha)_{\alpha \in A}$ be an arbitrary family of σ-fields defined on an abstract space Ω, with A possibly uncountable. Show that $\bigcap_{\alpha \in A} \mathcal{G}_\alpha$ is also a σ-field.

2. Let $\varnothing \subsetneq A \subsetneq B \subsetneq \Omega$ (these are strict inclusions). What is the σ-field generated by $\{A, B\}$?

3. Let $\mathcal{F}$, $\mathcal{G}$ be σ-fields for the same Ω. Is $\mathcal{F} \cup \mathcal{G}$ a σ-field?

4. For $\Omega = \mathbb{N}$ and $n \geq 0$, let $\mathcal{F}_n = \sigma\left(\{\{0\}, \ldots, \{n\}\}\right)$. Show that $(\mathcal{F}_n)_{n \geq 0}$ is a non-decreasing sequence but that $\bigcup_{n \geq 0} \mathcal{F}_n$ is a not a σ-field.

5. Let Ω be an infinite set (countable or not). Let $\mathcal{A}$ be the set of subsets of Ω that are either finite or with finite complement in Ω. Prove that $\mathcal{A}$ is a field but not a σ-field.

6. (a) Prove whether the following sets are countable or not.

      i. All intervals in $\mathbb{R}$ with rational endpoints.

      ii. All circles in the plane with rational radii and centers on the diagonal $x = y$.

      iii. All sequences of integers whose terms are either 0 or 1.

   (b) Can you build an infinite, countable σ-field?

7. A monotone class is a collection $\mathcal{M}$ of sets closed under both monotone increasing and monotone decreasing (i.e. if $A_i \in \mathcal{M}$ and either $A_i \uparrow A$ or $A_i \downarrow A$ then $A \in \mathcal{M}$). Prove that if $\mathcal{A} \subset \mathcal{M}$ with $\mathcal{A}$ a field and $\mathcal{M}$ a monotone class, then $\sigma(\mathcal{A}) \subset \mathcal{M}$.

8. Let $\mathbb{P}$ be a probability measure on Ω, endowed with a σ-field $\mathcal{A}$.

(a) What is the meaning of the following events, where all A_n's are elements of $\mathcal{A}$?

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1}\bigcap_{k\geq n} A_k, \quad \limsup_{n\to\infty} A_n = \bigcap_{n\geq 1}\bigcup_{k\geq n} A_k$$

(b) Prove that $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$ are in $\mathcal{A}$.

(c) In the special case $\Omega = \mathbb{R}$, for any $p \geq 1$, let

$$A_{2p} = \left[-1, 2 + \frac{1}{2p}\right), \quad A_{2p+1} = \left(-2 - \frac{1}{2p+1}, 1\right].$$

What are $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$?

(d) Prove that the following always holds:

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mathbb{P}\left(A_n\right), \quad \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \geq \limsup_{n\to\infty} \mathbb{P}\left(A_n\right)$$

9. The symmetric difference of two events A and B, denoted $A\triangle B$, is the event that precisely one of them occurs: $A\triangle B = (A\cup B)\backslash(A\cap B)$.

(a) Write a formula for $A\triangle B$ that only involves the operations of union, intersection and complement, but no set difference.

(b) Define $d(A,B) = \mathbb{P}(A\triangle B)$. Show that for any three events A, B, C,

$$d(A,B) + d(B,C) - d(A,C) = 2\left(\mathbb{P}\left(A\cap B^c\cap C\right) + \mathbb{P}\left(A^c\cap B\cap C^c\right)\right)$$

(c) Assume $A\subset B\subset C$. Prove that $d(A,C) = d(A,B) + d(B,C)$.

10. Prove the *Bonferroni inequalities*: if $A_i \in \mathcal{A}$ is a sequence of events, then

(a) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j}\mathbb{P}(A_i\cap A_j)$,

(b) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j}\mathbb{P}(A_i\cap A_j) + \sum_{i<j<k}\mathbb{P}(A_i\cap A_j\cap A_k)$

1.1.2 Integration

1. Let $\mathcal{A}$ be a σ-algebra, $\mathbb{P}$ a probability measure and $(A_n)_{n\geq 1}$ a sequence of events in $\mathcal{A}$ which converges to A. Prove that

(a) $A \in \mathcal{A}$;

(b) $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

2. Suppose a distribution function F is given by

$$F(x) = \frac{1}{4}\mathbb{1}_{[0,\infty)}(x) + \frac{1}{2}\mathbb{1}_{[1,\infty)}(x) + \frac{1}{4}\mathbb{1}_{[2,\infty)}(x)$$

   What is the probability of the following events, $(-1/2, 1/2)$, $(-1/2, 3/2)$, $(2/3, 5/2)$, $(3, \infty)$?

3. Let μ be the Lebesgue measure on $\mathbb{R}$. Build a sequence of functions $(f_n)_{n\geq 0}$, $0 \leq f_n \leq 1$, such that $\int f_n \,\mathrm{d}\mu \to 0$ but for any $x \in \mathbb{R}$, $(f_n(x))_{n\geq 0}$ does not converge.

4. Let X be a random variable in $\mathrm{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Let $(A_n)_{n\geq 0}$ be a sequence of events in $\mathcal{A}$ such that $\mathbb{P}(A_n) \xrightarrow[n\to\infty]{} 0$. Prove that $\mathbb{E}(X\mathbb{1}_{A_n}) \xrightarrow[n\to\infty]{} 0$.

5. Let $(d_n)_{n\geq 0}$ be a sequence in $(0,1)$, and $K_0 = [0,1]$. We define iteratively $(K_n)_{n\geq 0}$ in the following way. From K_n, which is the union of closed disjoint intervals, we define K_{n+1} by removing from each interval of K_n an open interval, centered at the middle of the previous one, with length d_n times the length of the previous one. Let $K = \bigcap_{n\geq 0} K_n$ (K is called a Cantor set).

   (a) Prove that K is an uncountable compact set, with empty interior, and whose points are all accumulation points.

   (b) What is the Lebesgue measure of K?

6. Let X be a nonnegative random variable. Prove that $\mathbb{E}(X) < +\infty$ if and only if

$$\sum_{n\in\mathbb{N}} \mathbb{P}(X \geq n) < \infty$$

7. **Convergence in Measure**

   Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. and $(f_n)_{n\geq 1}$, $f : \Omega \to \mathbb{R}$ measurable (for the Borel σ-field on $\mathbb{R}$). We say that $(f_n)_{n\geq 1}$ converges in measure to f if for any $\varepsilon > 0$ we have

$$\mu(|f_n - f| > \varepsilon) \xrightarrow[n\to\infty]{} 0$$

   (a) Show that $\int |f - f_n| \,\mathrm{d}\mu \to 0$ implies that f_n converges to f in measure. Is the reciprocal true?

   (b) Show that if $f_n \to f$ μ-almost surely, then $f_n \to f$ in measure. Is the reciprocal true?

   (c) Show that if $f_n \to f$ in measure, there exists a subsequence of $(f_n)_{n\geq 1}$ which converges μ-almost surely.

   (d) **A stronger dominated convergence theorem**

   We assume that $f_n \to f$ in measure and $|f_n| \leq g$ for some integrable $g : \Omega \to \mathbb{R}$, for any $n \geq 1$.

i. Show that $|f| \leq g$ μ-a.s.

ii. Deduce that $\int |f_n - f| \, \mathrm{d}\mu \to 0$.

8. Consider a probability space $(\Omega, \mathcal{A}, \mu)$ and $(A_n)_n$ a sequence in $\mathcal{A}$. Let $f : \Omega \to \mathbb{R}$ be measurable (for the Borel σ-field on $\mathbb{R}$) such that $\int_\Omega |\mathbb{1}_{A_n} - f| \, \mathrm{d}\mu \to 0$ as $n \to \infty$. Prove that there exists $A \in \mathcal{A}$ such that $f = \mathbb{1}_A$ μ-a.s., i.e. $\mu(f = \mathbb{1}_A) = 1$.

1.1.3 Transformations, Product Spaces, Distributions and Expectations

1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Prove that if $A \cap B = \varnothing$ and A, B are independent, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$.

2. Let X be a nonnegative random variable with null expectation. Prove that it is 0 almost surely.

3. Calculate $\mathbb{E}(X)$ for the following probability measures $\mathbb{P}^X$.

   (a) $\mathbb{P}^X$ has Gaussian density $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$, for some $\sigma > 0$ and $\mu \in \mathbb{R}$;

   (b) $\mathbb{P}^X$ has exponential dentity $\lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}$ for some $\lambda > 0$;

   (c) $\mathbb{P}^X = p\delta_a + q\delta_b$ where $p + q = 1$, $p, q \geq 0$ and $a, b \in \mathbb{R}$;

   (d) $\mathbb{P}^X$ is the Poisson distribution: $\mathbb{P}^X(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}$ for any integer $n \geq 0$, for some $\lambda > 0$.

4. Let X be a standard Gaussian random variable. What is the density of $1/X^2$?

5. Let X be uniformly distributed on $[0, 1]$ and $\lambda > 0$. Show that $-\lambda^{-1} \log X$ has the same distribution as an exponential random variable with parameter λ.

6. A samouraï wants to create a triangle with a (rigid) spaghetti. With his saber, he cuts this spaghetti on two places, chosen uniformly and independently along this traditional pasta. What is the probability that he can create a triangle with sides these three pieces of spaghetti?

7. Assume that $X_1, X_2, \dots$ are independent random variables uniformly distributed on $[0, 1]$. Let $Y^{(n)} = n \inf\{X_i, 1 \leq i \leq n\}$. Prove that it converges weakly to an exponential random variable, i.e. for any continuous bounded function $f : \mathbb{R}^+ \to \mathbb{R}$,

$$\mathbb{E}\left(f\left(Y^{(n)}\right)\right) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^+} f(u) e^{-u} \, \mathrm{d}u$$

8. Let n and m be random numbers chosen independently and uniformly on $[\![1, N]\!]$. What are $\Omega, \mathcal{A}$ and $\mathbb{P}$ (which all implicitly depend on N) ? Prove that

$$\mathbb{P}(n \wedge m = 1) \xrightarrow[N\to\infty]{} \zeta(2)^{-1}$$

where

$$\zeta(2) = \prod_{p\in\mathcal{P}}(1-p^{-2})^{-1} = \sum_{n\geq 1} n^{-2} = \frac{\pi^2}{6}$$

(you don't have to prove these equalities). Here $\mathcal{P}$ is the set of prime numbers and $n \wedge m = 1$ means that their greatest common divisor is 1.

9. Let $\varepsilon > 0$ and X be uniformly distributed on $[0,1]$. Prove that, almost surely (i.e. the following event has probability 1), there exists only a finite number of rationals $\frac{p}{q}$, with $p \wedge q = 1$, such that
$$\left|X - \frac{p}{q}\right| < \frac{1}{q^{2+\varepsilon}}$$

10. You toss a coin repeatedly and independently. The probability to get a head is p, a tail is $1-p$. Let A_k be the following event: k or more consecutive heads occur amongst the tosses numbered $2^k, \dots, 2^{k+1}-1$. Prove that $\mathbb{P}(A_k \text{ i.o.}) = 1$ if $p \geq 1/2$, and 0 otherwise.

    Remark. i.o. stands for "infinitely often", and A_k i.o. is the event $\bigcap_{n\geq 1}\bigcup_{m\geq n} A_m$.

1.2 Weak Convergence

1. Let X be a random variable with density $f_X(x) = (1-|x|)\mathbb{1}_{(-1,1)}(x)$. Show that its characteristic function is
$$\phi_X(u) = \frac{2(1-\cos u)}{u^2}$$

2. (a) Prove that $\hat{\mu}$ is real-valued if and only if μ is symmetric, i.e. $\mu(A) = \mu(-A)$ for any Borel set A.

    (b) If X and Y are i.i.d., prove that $X - Y$ has a symmetric distribution.

3. Let X_λ be a real random variable, with Poisson distribution with parameter λ. Calculate the characteristic function of X_λ. Conclude that $(X_\lambda - \lambda)/\sqrt{\lambda}$ converges in distribution to a standard Gaussian, as $\lambda \to \infty$.

4. Assume that the sequence of random variables $(X_n)_{n\geq 1}$ satisfies $\mathbb{E}X_n \to 1$ and $\mathbb{E}X_n^2 \to 1$. Prove that $(X_n)_{n\geq 1}$ converges in distribution. What is the limit?

5. Let $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$ be real random variables, with X_n and Y_n independent for any $n \geq 1$, and assume that X_n converges in distribution to X and Y_n to Y, with X and Y independent defined on the same probability space. Prove that $X_n + Y_n$ converges in distribution to $X + Y$.

6. Let X, Y be independent and assume that for some constant α we have $\mathbb{P}(X + Y = \alpha) = 1$. Prove that X and Y are both constant random variables.

7. Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with standard Cauchy distribution and let $M_n = \max(X_1, \dots, X_n)$. Prove that $(nM_n^{-1})_{n\geq 1}$ converges in distribution and identify the limit.

8. Let X, Y be i.i.d., with characteristic functions denoted φ_X, φ_Y, and suppose $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) = 1$. Assume also that $X + Y$ and $X - Y$ are independent.

   (a) Prove that
   $$\varphi_X(2u) = (\varphi_X(u))^3 \varphi_X(-u)$$

   (b) Prove that X is a standard Gaussian random variable.

9. For any $d \geq 1$, we admit that there is only one probability measure μ on $\mathcal{S}_d$, (the $(d-1)$-th dimensional sphere embedded in $\mathbb{R}^d$) that is uniform, in the following sense: for any isometry $A \in \mathrm{O}(d)$ (the orthogonal group in $\mathbb{R}^d$), and any continuous function $f : \mathcal{S}_d \to \mathbb{R}$,
   $$\int_{\mathcal{S}_d} f(x)\,\mathrm{d}\mu(x) = \int_{\mathcal{S}_d} f(Ax)\,\mathrm{d}\mu(x)$$
   Let $X = (X_1, \dots, X_d)$ be a vector of independent centered and reduced Gaussian random variables.

   (a) Prove that the random variable $U = X/\|X\|_{L^2}$ is uniformly distributed on the sphere.

   (b) Prove that, as $d \to \infty$, the main part of the globe is concentrated close to the Equator, i.e. for any $\varepsilon > 0$,
   $$\int_{x\in\mathcal{S}_d, |x_1|<\varepsilon} \mathrm{d}\mu(x) \to 1$$

1.3 Independent Sums

1.3.1 Convolutions, Laws of Large Numbers

1. Prove that if a sequence of real random variables (X_n) converge in distribution to X, and (Y_n) converges in distribution to a constant c, then $X_n + Y_n$ converges in distribution to $X + c$.

2. Assume that (X, Y) has joint density
   $$ce^{-(1+x^2)(1+y^2)},$$
   where c is properly chosen. Are X and Y Gaussian random variables? Is (X, Y) a Gaussian vector?

3. Let $(X_i)_{i\geq 1}$ be a sequence of independent random variables, with X_i uniform on $[-i, i]$. Let $S_n = X_1 + \ldots + X_n$. Prove that $S_n/n^{3/2}$ converges in distribution and describe the limit.

4. Find the probability distribution μ of a $\mathbb{Z}$-valued random variable which is symmetric[1], not integrable, but such that its characteristic function is differentiable at 0.

5. For any probability measure μ supported on $\mathbb{R}_+$, one defines the Laplace transform as

$$\mathscr{L}_\mu(\lambda) = \int_0^\infty e^{-\lambda x}\,\mathrm{d}\mu(x),\ \lambda \geq 0$$

   (a) Prove that $\mathscr{L}_\mu$ is well-defined, continuous on $\mathbb{R}_+$ and $\mathscr{C}^\infty$ on $\mathbb{R}_+^*$.

   (b) Prove that $\mathscr{L}_\mu$ characterizes the probability measure μ supported on $\mathbb{R}_+$.

   (c) Assume that for a sequence $(\mu_n)_{n\geq 1}$ of probability measure supported on $\mathbb{R}_+$, one has $\mathscr{L}_{\mu_n}(\lambda) \to \ell(\lambda)$ for any $\lambda \geq 0$, and ℓ is right-continuous at 0. Prove that $(\mu_n)_{n\geq 1}$ is tight, and that it converges weakly to a measure μ such that $\ell = \mathscr{L}_\mu$.

6. The goal of this problem is to prove the iterated logarithm law, first for Gaussian random variables. In other words, for $X_1, X_2 \dots$ i.i.d. standard Gaussian random variables, denoting $S_n = X_1 + \cdots + X_n$, we have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right) = 1 \tag{1.1}$$

   (a) Prove that

$$\mathbb{P}(X_1 > \lambda) \underset{\lambda\to\infty}{\sim} \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}}.$$

   In the following questions we denote $f(n) = \sqrt{2n\log\log n}$, $\lambda > 1$, $c, \alpha > 0$,

$$\begin{aligned}
A_k &= \left\{ S_{\lfloor \lambda^k \rfloor} \geq cf\left(\lambda^k\right) \right\}, \\
C_k &= \left\{ S_{\lfloor \lambda^{k+1} \rfloor} - S_{\lfloor \lambda^k \rfloor} \geq cf\left(\lambda^{k+1} - \lambda^k\right) \right\}, \\
D_k &= \left\{ \sup_{n\in[\![\lambda^k, \lambda^{k+1}]\!]} \frac{S_n - S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} \geq \alpha \right\}
\end{aligned}$$

   (b) Prove that for any $c > 1$ we have $\sum_{k\geq 1} \mathbb{P}(A_k) < \infty$ and

$$\limsup_{k\to\infty} \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} \leq 1 \text{ a.s.}$$

   (c) Prove that for any $c < 1$ we have $\sum_{k\geq 1} \mathbb{P}(C_k) = \infty$ and

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

[1] $\mu(\{i\}) = \mu(\{-i\})$ for any $i \in \mathbb{Z}$.

(d) Let $\varepsilon > 0$ and choose $c = 1 - \varepsilon/10$. Prove that almost surely the following inequality holds for infinitely many k:

$$\frac{S_{\lfloor \lambda^{k+1} \rfloor}}{f(\lambda^{k+1})} \geq c \frac{f(\lambda^{k+1} - \lambda^k)}{f(\lambda^{k+1})} - (1+\varepsilon) \frac{f(\lambda^k)}{f(\lambda^{k+1})}$$

(e) By choosing a large enough λ in the previous inequality, prove that almost surely

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \geq 1$$

(f) Prove that for any $n \in [\![\lambda^k, \lambda^{k+1}]\!]$ we have

$$\frac{S_n}{f(n)} \leq \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lfloor \lambda^k \rfloor)} + \frac{S_n - S_{\lfloor \lambda^k \rfloor}}{f(\lfloor \lambda^k \rfloor)}$$

(g) Prove that

$$\mathbb{P}(D_k) \underset{k\to\infty}{\sim} 2\mathbb{P}\left(X_1 \geq \frac{\alpha f(\lambda^k)}{\sqrt{\lambda^{k+1} - \lambda^k}} \right) \underset{k\to\infty}{\sim} \frac{c}{\sqrt{\log \lambda}} \left(\frac{1}{k} \right)^{\frac{\alpha^2}{\lambda - 1}}$$

(h) Prove that for $\alpha^2 > \lambda - 1$, almost surely

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \leq \limsup_{n\to\infty} \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} + \alpha$$

(i) By choosing appropriate λ and α, prove that almost surely

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \leq 1$$

(j) State a result similar to (1.1) for i.i.d. uniformly bounded random variables. Which steps in the above proof need to be modified to prove this universality result? How?

1.3.2 Central Limit Theorem

1. Assume $(\Omega, \mathcal{A}, \mathbb{P})$ is such that Ω is countable and $\mathcal{A} = 2^\Omega$. Prove that convergence in probability and convergence almost sure are the same.

2. Let $(X_i)_{i\geq 1}$ be i.i.d. Gaussian with mean 1 and variance 3. Show that

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2} = \frac{1}{4} \quad \text{a.s.}$$

3. Let f be a continuous function on $[0,1]$. Calculate the asymptotics, as $n \to \infty$, of

$$\int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

4. The goal of this exercise is to prove that any function, continuous on an interval of $\mathbb{R}$, can be approximated by polynomials, arbitrarily close for the L^∞ norm (this is the Bernstein-Weierstrass theorem). Let f be a continuous function on $[0,1]$. The n-th Bernstein polynomial is

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

   (a) Let $S_n(x) = B^{(n,x)}/n$, where $B^{(n,x)}$ is a binomial random variable with parameters n and x: $B^{(n,x)} = \sum_{\ell=1}^{n} X_i$ where the X_i's are independent and $\mathbb{P}(X_i = 1) = x$, $\mathbb{P}(X_i = 0) = 1 - x$. Prove that $B_n(x) = \mathbb{E}(f(S_n(x)))$.

   (b) Prove that $\|B_n - f\|_{L^\infty([0,1])} \to 0$ as $n \to \infty$.

5. Calculate

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$$

6. Let $\alpha > 0$ and, given $(\Omega, \mathcal{A}, \mathbb{P})$, let $(X_n, n \geq 1)$ be a sequence of independent real random variables with law $\mathbb{P}(X_n = 1) = \frac{1}{n^\alpha}$ and $\mathbb{P}\left(X_n = 0\right) = 1 - \frac{1}{n^\alpha}$. Prove that $X_n \to 0$ in $\mathcal{L}^1$, but that almost surely

$$\limsup_{n\to\infty} X_n = \begin{cases} 1 & \text{if } \alpha \leq 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$$

7. A sequence of random variables $(X_i)_{i\geq 1}$ is said to be completely convergent to X if for any $\varepsilon > 0$, we have $\sum_{i\geq 1} \mathbb{P}(|X_i - X| > \varepsilon) < \infty$. Prove that if the X_i's are independent then complete convergence implies almost sure convergence.

8. Let $(X_n)_{n\geq 1}$ be a sequence of random variables, on the same probability space, with $\mathbb{E}(X_\ell) = \mu$ for any ℓ, and a weak correlation in the following sense: $\mathrm{Cov}(X_k, X_\ell) \leq f(|k - \ell|)$ for all indexes k, ℓ, where the sequence $(f(m))_{m\geq 0}$ converges to 0 as $m \to \infty$. Prove that $\left(n^{-1} \sum_{k=1}^{n} X_k\right)_{n\geq 1}$ converges to μ in L^2.

9. The goal of this exercise is to prove the Erdős-Kac theorem: if $w(m)$ denotes the number of distinct prime factors of m and k is a random variable uniformly distributed on $[\![1, n]\!]$, then the following convergence in distribution holds:

$$\frac{w(k) - \log\log n}{\sqrt{\log\log n}} \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

(a) Prove that if $(X_n)_{n\geq 1}$ converges in distribution to $\mathcal{N}(0,1)$ and $\sup_{n\geq 1}\mathbb{E}[X_n^{2k}]<\infty$ for any $k\in\mathbb{N}$, then

$$\lim_{n\to\infty}\mathbb{E}[X_n^k]=\mathbb{E}\left[\mathcal{N}(0,1)^k\right]$$

for any $k\in\mathbb{N}$.

(b) Prove that for any $x\in\mathbb{R}$ and $d\geq 1$ we have

$$\left|e^{\mathrm{i}x}-\sum_{\ell=0}^{d}\frac{(\mathrm{i}x)^\ell}{\ell!}\right|\leq\frac{|x|^{d+1}}{(d+1)!}$$

(c) Assume that

$$\lim_{n\to\infty}\mathbb{E}[X_n^k]=\mathbb{E}\left[\mathcal{N}(0,1)^k\right]$$

for any $k\in\mathbb{N}$. Prove that X_n converges in distribution to X.

(d) Let $w_y(m)$ be the number of prime factors of m which are smaller than y. Let $(B_p)_{p\text{ prime}}$ be independent random variables such that $\mathbb{P}(B_p=1)=1-\mathbb{P}(B_p=0)=1/p$. Denote

$$W_y=\sum_{p\leq y}B_p,\quad \mu_y=\sum_{p\leq y}\frac{1}{p},\quad \sigma_y^2=\sum_{p\leq y}\left(\frac{1}{p}-\frac{1}{p^2}\right)$$

Prove that if $y=n^{\mathrm{o}(1)}$, then for any $k\in\mathbb{N}$ we have

$$\mathbb{E}\left[\left(\frac{w_y(k)-\mu_y}{\sigma_y}\right)^d\right]-\mathbb{E}\left[\left(\frac{W_y-\mu_y}{\sigma_y}\right)^d\right]\xrightarrow[n\to\infty]{}0$$

(e) Conclude.

1.4 Dependent Random Variables

1.4.1 Conditioning, Radon-Nikodym Theorem

1. Let X and Y be independent Gaussian random variables with null expectation and variance 1. Show that $\frac{X+Y}{\sqrt{2}}$ and $\frac{X-Y}{\sqrt{2}}$ are also independent $\mathcal{N}(0,1)$.

2. Let $(X_n)_{n\geq 0}$ be a sequence of i.i.d random variables, with uniform distribution on $[0,1]$. Let $Y_n=(X_n)^n$.

   (a) Calculate the distribution of Y_n.

   (b) Show that $(Y_n)_{n\geq 0}$ converges to 0 in probability.

   (c) Show that $(Y_n)_{n\geq 0}$ converges in L^1.

(d) Show that almost surely $(Y_n)_{n\geq 0}$ does not converge.

3. Let $(X_n)_{n\geq 1}$ be i.i.d. Bernoulli random variables with parameter $p \in (0,1)$, i.e. $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p$. Let N be a Poisson random variable with parameter $\lambda > 0$, i.e. for any $k \geq 0$ we have $\mathbb{P}(N = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. Assume N is independent from $(X_n)_{n\geq 1}$. Let $P = \sum_{i=1}^{N} X_i$, $F = N - P$.

   (a) What is the joint distribution of (P, N)?

   (b) Prove that P and F are independent.

4. (*the number of buses stopping till time t*) Let $(X_n)_{n\geq 1}$ be i.i.d, random variables on $(\Omega, \mathcal{A}, \mathbb{P})$, X_1 being an exponential random variable with parameter 1. Define $T_0 = 0, T_n = X_1 + \cdots + X_n$, and for any $t > 0$,
$$N_t = \max\{n \geq 0 \mid T_n \leq t\}$$

   (a) For any $n \geq 1$, calculate the joint distribution of $(T_1, \ldots, T_n)$.

   (b) Deduce the distribution of N_t, for arbitrary t.

5. Let $(X_n)_{n\geq 0}$ be real, independent, random variables on $(\Omega, \mathcal{A}, \mathbb{P})$.

   (a) Prove that the radius of convergence R of the random series $\sum_{n\geq 0} X_n z^n$ is almost surely constant.

   (b) Assume also that the X_n's have the same distribution. Prove that $R = 0$ almost surely if $\mathbb{E}\left[\log(|X_0|)_+\right] = \infty$, and $R \geq 1$ a.s. if $\mathbb{E}\left[\log(|X_0|)_+\right] < \infty$.

6. Prove that there is no probability measure on $\mathbb{N}$ such that for any $n \geq 1$, the probability of the set of multiples of n is $1/n$.

7. Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}, \mathcal{H}$ sub σ-fields of $\mathcal{F}$ such that $\sigma(\mathcal{G}, \mathcal{H}) = \mathcal{F}$. Find counterexamples to the following assertions:

   (a) If $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$ then X and Y are independent.

   (b) If $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{H}] = 0$ then $X = 0$.

   (c) If X and Y are independent then so are $\mathbb{E}[X \mid \mathcal{G}]$ and $\mathbb{E}[Y \mid \mathcal{G}]$.

8. Let Y be an integrable random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{G}$ a sub σ-field of $\mathcal{A}$. Show that $|\mathbb{E}(Y \mid \mathcal{G})| \leq \mathbb{E}(|Y| \mid \mathcal{G})$ (almost surely).

9. Let Y be an integrable random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{G}$ a sub σ-field of $\mathcal{A}$. Suppose that $\mathcal{H} \subset \mathcal{G}$ is a sub σ-field of $\mathcal{G}$. Show that $\mathbb{E}(\mathbb{E}(Y \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(Y \mid \mathcal{H})$ (almost surely).

10. Let $(X_n)_{n\geq 0}$ be defined on $(\Omega, \mathcal{A}, \mathbb{P})$. Assume this sequence converges in probability (under $\mathbb{P}$) to X. Let $\mathbb{Q}$ be another probability measure on $(\Omega, \mathcal{A})$ assumed to be absolutely continuous w.r.t. $\mathbb{P}$. Prove that $X_n \to X$ in probability under $\mathbb{Q}$.

1.4.2 Conditional Expectation

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(A_n)_{n\geq 1}$ be a sequence of independent events. We denote $a_n = \mathbb{P}(A_n)$ and define $b_n = a_1 + \cdots + a_n$, $S_n = \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}$. Assuming $b_n \to \infty$ as $n \to \infty$, prove that S_n/b_n converges almost surely.

2. Let $X_1, \dots, X_n$ be i.i.d. integrable random variables, and $S = \sum_{i=1}^n X_i$. Calculate $\mathbb{E}[S \mid X_1]$ and $\mathbb{E}[X_1 \mid S]$.

3. For fixed $a, b > 0$, let (X, Y) be a $\mathbb{N} \times \mathbb{R}_+$-valued random variable such that
$$\mathbb{P}(X = n, Y \leq t) = b \int_0^t \frac{(ay)^n}{n!} e^{-(a+b)y}\, \mathrm{d}y.$$
For $h : \mathbb{R}_+ \to \mathbb{R}$ continuous and bounded, calculate $\mathbb{E}[h(Y) \mid X]$. Calculate $\mathbb{E}\left[\frac{Y}{X+1}\right]$. Calculate $\mathbb{P}(X = n \mid Y)$. Calculate $\mathbb{E}[X \mid Y]$.

4. Let (X_1, X_2) be a Gaussian vector with mean (m_1, m_2) and non-degenerate covariance matrix $(C_{ij})_{1\leq i,j\leq 2}$. Prove that
$$\mathbb{E}[X_1 \mid X_2] = m_1 + \frac{C_{12}}{C_{22}}(X_2 - m_2).$$

5. Let X be a random variable such that $\mathbb{P}(X > t) = \exp(-t)$ for any $t \geq 0$. Let $Y = \min(X, s)$, where $s > 0$ is fixed. Prove that, almost surely,
$$\mathbb{E}[X \mid Y] = Y\mathbb{1}_{Y<s} + (1 + s)\mathbb{1}_{Y=s}.$$

6. Let $(X_n)_{n\geq 1}$ be a sequence of nonnegative random variables on $(\Omega, \mathcal{A}, \mathbb{P})$, and $(\mathcal{F}_n)_{n\geq 0}$ a sequence of sub σ-fields of $\mathcal{F}$. Assume that $\mathbb{E}(X_n \mid \mathcal{F}_n)$ converges to 0 in probability.
   (a) Show that X_n converges to 0 in probability.
   (b) Show that the reciprocal is wrong.

7. Let μ and ν be two probability measures such that $\mu \ll \nu$ and $\nu \ll \mu$ (usually abbreviated $\mu \sim \nu$). Let $X = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.
   (a) Prove that $\nu(X = 0) = 0$.
   (b) Prove that $\frac{1}{X} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ almost surely (for μ or ν).

8. On the same probability space, let X, Y be positive random variables such that $\mathbb{E}[X \mid Y] = Y$ and $\mathbb{E}[Y \mid X] = X$ (almost surely). Prove that $X = Y$ almost surely.

1.4.3 Markov Chains

1. Let $X_1, X_2 \ldots$ be i.i.d. Bernoulli random variables ($\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$) and $S_n = \sum_{i=1}^n X_i$. Which of the following sequences are Markovian? If Markovian, give the transition matrix.

   (a) $(S_n^2 - n)_{n\geq 0}$. (b) $(S_{2n})_{n\geq 0}$. (c) $(|S_n|)_{n\geq 0}$.

2. Consider a Markov chain X with state space $\{0, 1, \ldots, n\}$ and transition matrix

$$\pi(0,k) = \frac{1}{2^{k+1}},\ 0 \leq k \leq n-1,\ \pi(0,n) = \frac{1}{2^n}$$
$$\pi(k,k-1) = 1,\ 1 \leq k \leq n-1,\ \pi(n,n) = \pi(n,n-1) = \frac{1}{2}.$$

   (a) Prove that the chain has a unique invariant probability measure μ and calculate it.

   (b) Prove that for any $0 \leq x_0 \leq n-1$, $\pi^{(x_0+1)}(x_0, \cdot) = \mu$.

   (c) Prove that for any $0 \leq x_0 \leq n$, $\pi^{(n)}(x_0, \cdot) = \mu$.

   (d) For any $t \geq 1$, calculate

$$d(t) := \frac{1}{2}\sum_{x=0}^{n}\left|\pi^{(t)}(n,x) - \mu(x)\right|,$$

   and plot $t \mapsto d(t)$.

3. For fixed $p, q \in [0,1]$, consider a Markov chain X with two states $\{1, 2\}$, with transition matrix

$$\pi = (\pi(i,j))_{1\leq i,j\leq 2} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

   (a) For which p, q is the chain irreducible? Aperiodic (see Lemma 4.14 and Remark 4.10 in Varadhan)?

   (b) What are the invariant probability measures of X?

   (c) Compute $\pi^{(n)}$, $n \geq 1$.

   (d) When X is irreducible, for this invariant probability measure μ, calculate

$$d_1(n) := \frac{1}{2}\left(|\mathbb{P}_1(X_n = 1) - \mu(1)| + |\mathbb{P}_1(X_n = 2) - \mu(2)|\right)$$
$$d_2(n) := \frac{1}{2}\left(|\mathbb{P}_2(X_n = 1) - \mu(1)| + |\mathbb{P}_2(X_n = 2) - \mu(2)|\right)$$

   where $\mathbb{P}_x$ means the chain starts at x.

4. Let T be a stopping time for a filtration $(\mathcal{F}_n)_{n\geq 1}$. Prove that $\mathcal{F}_T$ is a σ-field.

5. Let S and T be stopping times for a filtration $(\mathcal{F}_n)_{n\geq 1}$. Prove that $\max(S,T)$ and $\min(S,T)$ are stopping times.

6. Let $S \leq T$ be two stopping times and $A \in \mathcal{F}_S$. Define $U(\omega) = S(\omega)$ if $\omega \in A$, $U(\omega) = T(\omega)$ if $\omega \notin A$. Prove that U is a stopping time.

7. An ant walks on a round clock, starting at 0. At each second, it walks either clockwise or counterclockwise, with probability $1/2$ to a neighbouring number, and through independent steps. Let X be the last number visited by the ant. Prove it is equidistributed on $\{1, 2, \ldots, 11\}$.

8. Let $X_1, X_2, \ldots$ be i.i.d., $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and $S_n = X_1 + \cdots + X_n$. Prove that the following random variable converges in distribution as $n \to \infty$, and identify the limit:

$$\left(\sum_{k=1}^{n} e^{S_k}\right)^{\frac{1}{\sqrt{n}}}$$

9. Consider a Markov chain X with state space $\mathbb{N}$ and transition matrix

$$\pi(0,0) = r_0,\ \pi(0,1) = p_0, \text{and } \forall i \geq 1,\ \pi(i, i-1) = q_i,\ \pi(i,i) = r_i,\ \pi(i, i+1) = p_i,$$

with $p_0, r_0 > 0$, $p_0 + r_0 = 1$ and for all $i \geq 1$, $p_i, q_i > 0$, $p_i + q_i + r_i = 1$. Prove that the chain is irreducible, aperiodic. Give a necessary and sufficient condition for the chain to have an invariant probability measure.

10. Let $(G, \cdot)$ be a group, μ a probability measure on G and X the Markov chain such that $\pi(g, h \cdot g) = \mu(h)$. We call such a process X a random walk on G with jump kernel μ.

    (a) Explain why the usual random walk on $\mathbb{Z}^d$ is such process. Same question for the usual random walk on $(\mathbb{Z}/n\mathbb{Z})^d$, $n \geq 1$.

    (b) Consider the following shuffling of a deck of $n \geq 2$ cards: pick two such distinct cards uniformly at random and exhange their positions in the deck. Show that this is also an example of a random walk on a group.

    (c) Let $\mathcal{H} = \{h_1 \cdot h_2 \cdot \cdots \cdot h_n, \mu(h_i) > 0, 1 \leq i \leq n, n \in \mathbb{N}\}$. Discuss irreductibility of X depending on $\mathcal{H}$.

    (d) If X is irreducible on finite G, what are the invariant probability measures? What if G is not finite?

    (e) Make some search to define a reversible Markov chain. In the context of this exercise, show that X is reversible if and only if $\mu(h) = \mu(h^{-1})$ for any $h \in G$.

    (f) Give an example of an irreducible random walk on a group which is not reversible.

1.5 Martingales

1. Let $(X_n)_{n\geq 1}$ be independent such that $\mathbb{E}(X_i) = m_i$, $\mathrm{Var}(X_i) = \sigma_i^2$, $i \geq 1$. Let $S_n = \sum_{i=1}^n X_i$ and $\mathcal{F}_n = \sigma(X_i, 1 \leq i \leq n)$.

   (a) Find sequences $(b_n)_{n\geq 1}$, $(c_n)_{n\geq 1}$ of real numbers such that $(S_n^2 + b_n S_n + c_n)_{n\geq 1}$ is a $(\mathcal{F}_n)_{n\geq 1}$-martingale.

   (b) Assume moreover that there is a real number λ such that $e^{\lambda X_i} \in \mathrm{L}^1$ for any $i \geq 1$. Find a sequence $\left(a_n^{(\lambda)}\right)_{n\geq 1}$ such that $\left(e^{\lambda S_n - a_n^{(\lambda)}}\right)_{n\geq 1}$ is a $(\mathcal{F}_n)_{n\geq 1}$-martingale.

2. Let $(X_k)_{k\geq 0}$ be i.i.d. random variables, $\mathcal{F}_m = \sigma(X_1, \ldots, X_m)$ and $Y_m = \prod_{k=1}^m X_k$. Under which conditions is $(Y_m)_{m\geq 1}$ a $(\mathcal{F}_m)_{m\geq 1}$-submartingale, supermartingale, martingale?

3. Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration and $(X_n)_{n\geq 0}$ be a sequence of integrable random variables with $\mathbb{E}\left(X_n \mid \mathcal{F}_{n-1}\right) = 0$, and assume X_n is $\mathcal{F}_n$-measurable for every n. Let $S_n = \sum_{k=0}^n X_k$. Show that $(S_n)_{n\geq 0}$ is a $(\mathcal{F}_n)_{n\geq 0}$-martingale.

4. Let $a > 0$ be fixed, $(X_i)_{i\geq 1}$ be iid, $\mathbb{R}^d$-valued random variables, uniformly distributed on the ball $\mathrm{B}(0, a)$. Set $S_n = x + \sum_{i=1}^n X_i$.

   (a) Let f be a superharmonic function. Show that $(f(S_n))_{n\geq 1}$ defines a supermartingale.

   (b) Prove that if $d \leq 2$ any nonnegative superharmonic function is constant. Does this result remain true when $d \geq 3$?

5. In the following exercise you can use the following fact: If a martingale is a.s. bounded by a deterministic constant, it converges almost surely.

   Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of random variables, and assume (Y_n) converges in distribution to a limiting Y. Also, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the sequence of independent random variables $X := (X_n)_{n\in\mathbb{N}^*}$ is defined, and we assume that the sequence of partial sums $(S_n)_{n\in\mathbb{N}}$ (i.e. $S_0 = 0$ and $S_n := \sum_{j=1}^n X_j$) converges in distribution. Set $(\mathcal{F}_n)$ the natural filtration of X and $\Phi_n(t) = \mathbb{E}(\exp(\mathrm{i}tS_n))$ for $t \in \mathbb{R}$.

   (a) Establish that $(\Phi_{Y_n}(\cdot))_{n\geq 1}$ converges uniformly on every compact set, i.e. show that for any $a > 0$, $\max_{t\in[-a,a]} |\Phi_{Y_n}(t) - \Phi_Y(t)| \to 0$ as $n \to \infty$. Establish moreover that there exists $a > 0$ such that for any $n \geq 1$, $\min_{t\in[-a,a]} |\Phi_{Y_n}(t)| \geq 1/2$.

   (b) Show that there exists $t_0 > 0$ such that if $t \in [-t_0, t_0]$, then $(\exp(\mathrm{i}tS_n)/\Phi_n(t))_{n\geq 0}$ is a $(\mathcal{F}_n)-$martingale (i.e. both its real and imaginary parts are martingales).

   (c) Prove that we can choose $t_0 > 0$ such that for any $t \in [-t_0, t_0]$, $\lim_{n\to\infty} \exp(\mathrm{i}tS_n)$ exists $\mathbb{P}$-a.s.

(d) Set

$$C = \left\{ (t, \omega) \in [-t_0, t_0] \times \Omega : \lim_{n\to\infty} \exp(\mathrm{i} t S_n(\omega)) \text{ exists} \right\}$$

Prove that C is measurable, i.e. in the product of $\mathcal{B}([-t_0, t_0])$ with $\mathcal{F}$.

(e) Establish that $\int_{-t_0}^{t_0} \mathbb{1}_C(t, \omega)\mathbb{P}(\mathrm{d}\omega)\mathrm{d}t = 2t_0$.

(f) Prove that $\lim_{n\to\infty} S_n$ exists $\mathbb{P}$-a.s.

Chapter 2

Solutions

2.1 Measure Theory

2.1.1 Construction of Measures

1. Let $(\mathcal{G}_\alpha)_{\alpha \in A}$ be an arbitrary family of σ-fields defined on an abstract space Ω, with A possibly uncountable. Show that $\bigcap_{\alpha \in A} \mathcal{G}_\alpha$ is also a σ-field.

   Proof. We show by definition: (i) For any $\mathcal{G}_\alpha$, $\Omega \in \mathcal{G}_\alpha$, so that $\Omega \in \bigcap_{\alpha \in A} \mathcal{G}_\alpha$; (ii) If some set $S \in \bigcap_{\alpha \in A} \mathcal{G}_\alpha$, then $S \in \mathcal{G}_\alpha$ for any $\alpha \in A$, so that $S^{\complement} \in \mathcal{G}_\alpha, \forall \alpha \in A$. Hence, $S^{\complement} \in \bigcap_{\alpha \in A} \mathcal{G}_\alpha$; (iii) Similarly, if $A_1, A_2, \ldots \in \bigcap_{\alpha \in A} \mathcal{G}_\alpha$, since $\mathcal{G}_\alpha$ is σ-field, one has $\bigcap_i A_i \in \mathcal{G}_\alpha, \forall \alpha \in A$. Hence, $\bigcap_i A_i \in \bigcap_{\alpha \in A} \mathcal{G}_\alpha$. □

2. Let $\varnothing \subsetneq A \subsetneq B \subsetneq \Omega$ (these are strict inclusions). What is the σ-field generated by $\{A, B\}$?

   Solution. In general, one has $\sigma(\{A, B\}) = \sigma(\{A \backslash B, B \backslash A, A \cap B, \Omega \backslash (A \cup B)\})$.

   Here, since $\varnothing \subsetneq A \subsetneq B \subsetneq \Omega$, one has

$$\begin{aligned}\sigma(\{A, B\}) &= \sigma(\{A, B \backslash A, \Omega \backslash B\}) \\ &= \left\{\varnothing, \Omega, A, A^{\complement}, B, B^{\complement}, B \backslash A, A \cup B^{\complement}\right\}\end{aligned}$$

   □

3. Let $\mathcal{F}, \mathcal{G}$ be σ-fields for the same Ω. Is $\mathcal{F} \cup \mathcal{G}$ a σ-field?

   Solution. Not necessarily. For example, let $\Omega = \{a, b, c\}$, and define σ-fields A, B that $A = \{\{a\}, \{b, c\}, \varnothing, \Omega\}$, $B = \{\{b\}, \{a, c\}, \varnothing, \Omega\}$. However,

$$A \cup B = \{\{a\}, \{b, c\}, \{b\}, \{a, c\}, \varnothing, \Omega\}$$

   is not a σ-field as it is not closed under union: $\{a\} \cup \{b\} = \{a, b\} \notin A \cup B$. □

4. For $\Omega = \mathbb{N}$ and $n \geq 0$, let $\mathcal{F}_n = \sigma(\{\{0\}, \dots, \{n\}\})$. Show that $(\mathcal{F}_n)_{n\geq 0}$ is a non-decreasing sequence but that $\bigcup_{n\geq 0} \mathcal{F}_n$ is a not a σ-field.

   Proof. We first show that, for any $n \geq 0$, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

   Let $A = \{\{0\}, \dots, \{n\}\}$ and $B = \{\{0\}, \dots, \{n\}, \{n+1\}\}$. Pick arbitrarily $a \in A$. For $a \in \mathcal{F}_n$, then, since $a \in B$, $a \in \mathcal{F}_{n+1}$. Also, $a^{\complement} \in \mathcal{F}_n \implies a^{\complement} \in \mathcal{F}_{n+1}$. Then, for countably many $a_1, a_2, \dots \in A$, $\bigcup_i a_i \in \mathcal{F}_n \implies \bigcup_i a_i \in \mathcal{F}_{n+1}$. Thus, every element of $\mathcal{F}_n$ generated by A is contained in $\mathcal{F}_{n+1}$ (including $\varnothing$ and $\Omega = \mathbb{N}$), so $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$.

   However, $\bigcup_{n\geq 0} \mathcal{F}_n$ is not a σ-field. Consider $\{2\}, \{4\}, \dots, \{2i\}, \dots$, each of which is in $\bigcup_{n\geq 0} \mathcal{F}_n$, but $\bigcup_i \{2i\}$ is not contained in any of the $\mathcal{F}_n$, so $\bigcup_{n\geq 0} \mathcal{F}_n$ is not closed under countable union. □

5. Let Ω be an infinite set (countable or not). Let $\mathcal{A}$ be the set of subsets of Ω that are either finite or with finite complement in Ω. Prove that $\mathcal{A}$ is a field but not a σ-field.

   Proof. The set $\mathcal{A}$ is a field: clearly the null and full sets are in it, $\mathcal{A}$ is stable by immediate definition, and if A and B are in $\mathcal{A}$, then $A \cup B$ is either finite (when A and B are finite) or its complement $A^c \cap B^c$ is finite (when A or B is infinite with finite complement), so that $A \cup B \in \mathcal{A}$.

   However, $\mathcal{A}$ is not a σ-field. Indeed, let $\omega_1, \omega_2, \dots$ be an infinite sequence of disjoint elements in Ω. Then $\{\omega_2, \omega_4, \omega_6, \dots\}$ is the countable union of elements in $\mathcal{A}$ (the $\{\omega_{2i}\}$'s), but it is neither finite nor with finite complement, thus it cannot be in $\mathcal{A}$. □

6. (a) Prove whether the following sets are countable or not.

      i. All intervals in $\mathbb{R}$ with rational endpoints.

      ii. All circles in the plane with rational radii and centers on the diagonal $x = y$.

      iii. All sequences of integers whose terms are either 0 or 1.

   (b) Can you build an infinite, countable σ-field?

   Proof. (a) Countable; uncountable; uncountable.

   (b) Consider $(\Omega, \mathcal{A})$ with $\mathcal{A}$ a σ-field. For any $x \in \Omega$, define

$$\dot{x} = \bigcap_{A \in \mathcal{A} : x \in A} A.$$

   Then $\{\dot{x}, x \in \Omega\}$ (eliminating repetitions) form a partition of Ω (uses stability of the field by complement).

   If $\mathcal{A}$ is countable, $\dot{x}$ is an element of $\mathcal{A}$, as a countable intersetion of elements in $\mathcal{A}$. Thus $\mathcal{A}$ contains all sets of type $\dot{x}$, and all countable unions of such sets. In fact, $\mathcal{A}$ consists exactly in such unions (easy to prove).

If the number of atoms is finite, then $\mathcal{A}$ is finite. If the number of atoms is infinite, then the number of their countable unions is uncountable (cf a map to Cantor diagonalization), a contradiction. We have proved that $\mathcal{A}$ is either finite or uncountable.

$\square$

7. A monotone class is a collection $\mathcal{M}$ of sets closed under both monotone increasing and monotone decreasing (i.e. if $A_i \in \mathcal{M}$ and either $A_i \uparrow A$ or $A_i \downarrow A$ then $A \in \mathcal{M}$). Prove that if $\mathcal{A} \subset \mathcal{M}$ with $\mathcal{A}$ a field and $\mathcal{M}$ a monotone class, then $\sigma(\mathcal{A}) \subset \mathcal{M}$.

   Proof. Clearly, any field which is a monotone class must be a σ-field. Moreover, we will prove below the following:

   Lemma 2.1.1. *The intersection $m(\mathcal{A})$ of all monotone classes containing containing a field $\mathcal{A}$ is both a field and a monotone class.*

   Proof of the Lemma. The fact that $m(\mathcal{A})$ is a monotone class is straightforward from the definition. We now show that $m(\mathcal{A})$ is a field. Inclusion of null and full sets is trivial.

   For the stability by complement: Consider $\mathcal{B} = \{A : A^c \in m(\mathcal{A})\}$. As $m(\mathcal{A})$ is monotone, so is $\mathcal{B}$. Hence $\mathcal{B}$ is a monotone class containing $\mathcal{A}$, hence $m(\mathcal{A}) \subset \mathcal{B}$, which implies stability by complement. For stability by union, let

$$\mathcal{G}_1 = \{A : A \cup B \in m(\mathcal{A}) \text{ for all } B \in \mathcal{A}\}.$$

   Then $\mathcal{G}_1$ is a monotone class (simply checked by careful writing based on definitions) containing $\mathcal{A}$ hence $m(\mathcal{A}) \subset \mathcal{G}_1$. Now let

$$\mathcal{G}_2 = \{B : B \cup A \in m(\mathcal{A}) \text{ for all } A \in m(\mathcal{A})\}.$$

   Then $\mathcal{G}_2$ is a monotone class (again just by a careful direct writing). We have $\mathcal{A} \subset \mathcal{G}_2$ because for $B \in \mathcal{A}$ and $A \in m(\mathcal{A}) \cup \mathcal{G}_1$ we have $A \cup B \in m(\mathcal{A})$ by definition of $\mathcal{G}_1$. From $\mathcal{A} \subset \mathcal{G}_2$ we deduce that $m(\mathcal{A}) \subset \mathcal{G}_2$, so for any $A, B \in m(\mathcal{A})$ we have $A \cup B \in m(\mathcal{A})$ by definition of $\mathcal{G}_2$. $\square$

   Hence $m(\mathcal{A})$ is a σ-algebra. Since $\mathcal{A} \subset m(\mathcal{A})$ this implies $\sigma(\mathcal{A}) \subset m(\mathcal{A})$, which concludes the proof by noting $m(\mathcal{A}) \subset \mathcal{M}$. $\square$

8. Let $\mathbb{P}$ be a probability measure on Ω, endowed with a σ-field $\mathcal{A}$.

   (a) What is the meaning of the following events, where all A_n's are elements of $\mathcal{A}$?

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k, \quad \limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$$

Solution. $\liminf$: there is an $n \geq 1$ s.t. for all $k \geq n$, A_k happens; i.e. from some point on, every event happens. $\limsup$: for each $n \geq 1$, there is a $k \geq n$ s.t. A_k happens; i.e. infinitely many of events A_k happens. □

(b) Prove that $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$ are in $\mathcal{A}$.

Proof. Note that every $A_n \in \mathcal{A}$, so $\bigcap_{k\geq n} A_k$, as countable intersection, is in $\mathcal{A}$. Also, $\bigcup_{n\geq 1}\bigcap_{k\geq n} A_k$, as countable union, is in $\mathcal{A}$. Similarly, $\bigcap_{n\geq 1}\bigcup_{k\geq n} A_k \in \mathcal{A}$. □

(c) In the special case $\Omega = \mathbb{R}$, for any $p \geq 1$, let

$$A_{2p} = \left[-1, 2 + \frac{1}{2p}\right), \quad A_{2p+1} = \left(-2 - \frac{1}{2p+1}, 1\right].$$

What are $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$?

Solution. By definition, one has

$$\begin{aligned}\liminf_{n\to\infty} A_n &= (\cap_{k\geq 1} A_k) \bigcup (\cap_{k\geq 2} A_k) \bigcup (\cap_{k\geq 3} A_k) \bigcup \dots \\ &= [-1,1] \cup [-1,1] \cup \dots = \boxed{[-1,1]}\end{aligned}$$

and similarly,

$$\begin{aligned}\limsup_{n\to\infty} A_n &= (\cup_{k\geq 1} A_k) \bigcap (\cup_{k\geq 2} A_k) \bigcap (\cup_{k\geq 3} A_k) \bigcap \dots \\ &= \left(-2-1, 2+\frac{1}{2}\right) \bigcup \left(-2-\frac{1}{3}, 2+\frac{1}{2}\right) \bigcup \left(-2-\frac{1}{3}, 2+\frac{1}{4}\right) \dots \\ &= \boxed{[-2,2]}\end{aligned}$$

□

(d) Prove that the following always holds:

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n), \quad \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \geq \limsup_{n\to\infty} \mathbb{P}(A_n)$$

Proof. Note that $\left\{\bigcap_{k\geq n} A_k\right\}_{n=1}^{\infty}$ is a sequence which increases to $\liminf A_n$, so

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} A_k\right)$$

by continuity of $\mathbb{P}$ from below. And since $\bigcap_{k\geq n} A_k \subseteq A_n, \forall n$, one has

$$\mathbb{P}\left(\bigcap_{k\geq n} A_k\right) \leq \mathbb{P}(A_n) \implies \liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} A_k\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n)$$

Hence, $\mathbb{P}(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} \mathbb{P}(A_n)$. Then, similarly, one has

$$\begin{aligned}
\mathbb{P}\left(\limsup_n A_n\right) =& \mathbb{P}\left(\bigcap_{n>1}\bigcup_{k\geq n} A_k\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geq n} A_k\right) \\
&\geq \lim_{n\to\infty}\sup_{k\geq n} \mathbb{P}(A_k) \qquad \text{since } \mathbb{P}\left(\bigcup_{k\geq n} A_k\right) \geq \mathbb{P}(A_t), \forall t \geq n \\
&= \limsup_{n\to\infty} \mathbb{P}(A_n)
\end{aligned}$$

□

9. The symmetric difference of two events A and B, denoted $A\triangle B$, is the event that precisely one of them occurs: $A\triangle B = (A\cup B)\backslash(A\cap B)$.

   (a) Write a formula for $A\triangle B$ that only involves the operations of union, intersection and complement, but no set difference.

   (b) Define $d(A, B) = \mathbb{P}(A\triangle B)$. Show that for any three events A, B, C,

   $$d(A,B) + d(B,C) - d(A,C) = 2\left(\mathbb{P}(A\cap B^c\cap C) + \mathbb{P}(A^c\cap B\cap C^c)\right)$$

   (c) Assume $A \subset B \subset C$. Prove that $d(A, C) = d(A, B) + d(B, C)$.

Proof. (a) $A\triangle B = (A\cap B^c)\cup(B\cap A^c)$

(b) Note that $A\cap B^c$ and $B\cap A^c$ are disjoint so $\mathbb{P}(A\triangle B) = \mathbb{P}(A\cap B^c) + \mathbb{P}(B\cap A^c) = \mathbb{P}(A\cap B^c\cap C) + \mathbb{P}(B\cap A^c\cap C^c) + \mathbb{P}(A\cap B^c\cap C) + \mathbb{P}(B\cap A^c\cap C^c)$. Injecting this formula and the analogues for $\mathbb{P}(A\triangle C), \mathbb{P}(B\triangle C)$ gives the answer.

(c) Immediate from the previous question as the sets on the RHS are empty.

□

10. Prove the *Bonferroni inequalities*: if $A_i \in \mathcal{A}$ is a sequence of events, then

   (a) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j}\mathbb{P}(A_i\cap A_j)$,

   (b) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j}\mathbb{P}(A_i\cap A_j) + \sum_{i<j<k}\mathbb{P}(A_i\cap A_j\cap A_k)$

Proof. (a) For any $x\in\Omega$ we claim that

$$\mathbb{1}_{x\in\bigcup_{i=1}^n A_i} \geq \sum_i \mathbb{1}_{x\in A_i} - \sum_{i<j}\mathbb{1}_{x\in A_i\cap A_j}.$$

Indeed, if x is in exactly m sets A_i 's, for $m = 0$ the above relation is $0 \geq 0$ and if $m \geq 1$ this is

$$1 \geq m - \binom{m}{2}$$

which is true. Integrating the first inequality w.r.t. $\mathbb{P}$ gives the result.

(b) The same type of reasoning holds, stopping the binomial expansion at third order.

□

2.1.2 Integration

1. Let $\mathcal{A}$ be a σ-algebra, $\mathbb{P}$ a probability measure and $(A_n)_{n\geq 1}$ a sequence of events in $\mathcal{A}$ which converges to A. Prove that

   (a) $A \in \mathcal{A}$;

   (b) $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Proof. (a) By definition, the sequence of events converges if

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$$

and A is equal to this common set. We have shown in part 8b of HW1 that

$$\limsup_{n\to\infty} A_n, \liminf_{n\to\infty} A_n \in \mathcal{A}$$

Hence, A is in $\mathcal{A}$.

(b) Note that we have shown in part 8d of HW1 that

$$\mathbb{P}\left[\limsup_{n\to\infty} A_n\right] \geq \limsup_{n\to\infty} \mathbb{P}\left[A_n\right], \quad \mathbb{P}\left[\liminf_{n\to\infty} A_n\right] \leq \liminf_{n\to\infty} \mathbb{P}\left[A_n\right]$$

Note also that in general one has $\liminf_{n\to\infty} \mathbb{P}\left[A_n\right] \leq \limsup_{n\to\infty} \mathbb{P}\left[A_n\right]$.

Following $A = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$, one has

$$\mathbb{P}[A] \leq \liminf_{n\to\infty} \mathbb{P}\left[A_n\right] \leq \limsup_{n\to\infty} \mathbb{P}\left[A_n\right] \leq \mathbb{P}[A]$$

Hence,

$$\liminf_{n\to\infty} \mathbb{P}\left[A_n\right] = \limsup_{n\to\infty} \mathbb{P}\left[A_n\right] = \lim_{n\to\infty} \mathbb{P}\left[A_n\right] = \mathbb{P}[A]$$

□

2. Suppose a distribution function F is given by

$$F(x) = \frac{1}{4}\mathbb{1}_{[0,\infty)}(x) + \frac{1}{2}\mathbb{1}_{[1,\infty)}(x) + \frac{1}{4}\mathbb{1}_{[2,\infty)}(x)$$

What is the probability of the following events, $(-1/2, 1/2)$, $(-1/2, 3/2)$, $(2/3, 5/2)$, $(3, \infty)$?

Solution. By definition, one has

$$\begin{aligned}
\mathbb{P}\left[X \in \left(-\frac{1}{2}, \frac{1}{2}\right)\right] &= F\left(\frac{1}{2}\right) - F\left(-\frac{1}{2}\right) = \frac{1}{4} \\
\mathbb{P}\left[X \in \left(-\frac{1}{2}, \frac{3}{2}\right)\right] &= F\left(\frac{3}{2}\right) - F\left(-\frac{1}{2}\right) = \frac{3}{4} \\
\mathbb{P}\left[X \in \left(\frac{2}{3}, \frac{5}{2}\right)\right] &= F\left(\frac{5}{2}\right) - F\left(\frac{2}{3}\right) = \frac{3}{4} \\
\mathbb{P}\left[X \in (3, \infty)\right] &= F(\infty) - F(3) = 0
\end{aligned}$$

□

3. Let μ be the Lebesgue measure on $\mathbb{R}$. Build a sequence of functions $(f_n)_{n\geq 0}$, $0 \leq f_n \leq 1$, such that $\int f_n \, \mathrm{d}\mu \to 0$ but for any $x \in \mathbb{R}$, $(f_n(x))_{n\geq 0}$ does not converge.

Solution. We first define

$$g_n^{(i)} := \mathbb{1}_{\left[\frac{n-2^k}{2^k}+i, \frac{n-2^k+1}{2^k}+i\right)}, \quad k = \lfloor \log_2(n) \rfloor, i \in \mathbb{Z}$$

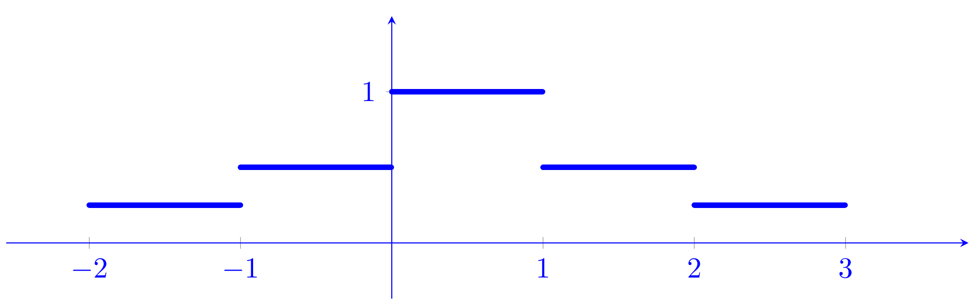
That is,

$$\begin{aligned}
g_1^{(0)} &= \mathbb{1}_{[0,1)}, g_2^{(0)} = \mathbb{1}_{[0,1/2)}, g_3^{(0)} = \mathbb{1}_{[1/2,1)}, \dots \\
g_1^{(1)} &= \mathbb{1}_{[1,2)}, g_2^{(1)} = \mathbb{1}_{[1,3/2)}, g_3^{(1)} = \mathbb{1}_{[3/2,2)}, \dots
\end{aligned}$$

and so on. Let

$$f_n := \sum_{i\in\mathbb{Z}} \frac{1}{2^{|i|}} g_n^{(i)}$$

For instance, f_1 is visualized as



Comment. *Can find an easier counter-example.*

Then, one has

$$\int f_n \,\mathrm{d}\mu = \left(\frac{n-2^k+1}{2^k} - \frac{n-2^k}{2^k}\right)\cdot\left(1+2\left(\frac{1}{2}+\frac{1}{4}+\dots\right)\right)$$
$$= \frac{3}{2^{\lfloor \log_2(n)\rfloor}} \xrightarrow[n\to\infty]{} 0,$$

but $(f_n(x))_{n\geq 0}$ does not converge anywhere on $\mathbb{R}$: WLOG, consider $x \in [0,1)$. Note that the interval where $f_n = 1$ is always sliding away by increasing n, so that $(f_n(x))_{n\geq 0}$ does not converge anywhere on $x \in [0,1)$. Similarly, $(f_n(x))_{n\geq 0}$ does not converge anywhere on $x \in [i, i+1), \forall i \in \mathbb{Z}$. Hence, for any $x \in \mathbb{R}$, $(f_n(x))_{n\geq 0}$ does not converge. □

4. Let X be a random variable in $\mathrm{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Let $(A_n)_{n\geq 0}$ be a sequence of events in $\mathcal{A}$ such that $\mathbb{P}(A_n) \xrightarrow[n\to\infty]{} 0$. Prove that $\mathbb{E}(X\mathbb{1}_{A_n}) \xrightarrow[n\to\infty]{} 0$.

   Solution. Note that $\mathbb{E}(|X|) \leq \infty$. Define $A_M := \{|X| \leq M\}, M = 1,2,3,\dots$. Note that $(|X|\mathbb{1}_{A_M})_M$ is increasing and it pointwise-converges to $|X|$. By the monotone convergence theorem, one has
   $$\mathbb{E}\left[|X|\mathbb{1}_{A_{M_0}}\right] \xrightarrow[n\to\infty]{} \mathbb{E}[|X|] < \infty$$
   implying that there exists some M_0 such that
   $$\mathbb{E}\left[|X|\left(1-\mathbb{1}_{A_{M_0}}\right)\right] < \varepsilon/2$$
   Now, since $\mathbb{P}(A_n) \xrightarrow[n\to\infty]{} 0$, we can always find some A in the sequence that $\mathbb{P}[A] < \delta$ for any $\delta > 0$. We let $\delta < \frac{\varepsilon}{2M_0}$, then
   $$\begin{aligned}\mathbb{E}[X\mathbb{1}_A] \leq \mathbb{E}[|X|\mathbb{1}_A] &= \mathbb{E}[|X|\mathbb{1}_A(1-\mathbb{1}_{A_{M_0}})] + \mathbb{E}[|X|(\mathbb{1}_{A\cap A_{M_0}})]\\ &\leq \mathbb{E}[|X|(1-\mathbb{1}_{A_{M_0}})] + M_0\mathbb{P}(A)\\ &< \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon, \quad \forall\varepsilon > 0\end{aligned}$$
   □

5. Let $(d_n)_{n\geq 0}$ be a sequence in $(0,1)$, and $K_0 = [0,1]$. We define iteratively $(K_n)_{n\geq 0}$ in the following way. From K_n, which is the union of closed disjoint intervals, we define K_{n+1} by removing from each interval of K_n an open interval, centered at the middle of the previous one, with length d_n times the length of the previous one. Let $K = \bigcap_{n\geq 0} K_n$ (K is called a Cantor set).

   (a) Prove that K is an uncountable compact set, with empty interior, and whose points are all accumulation points.

(b) What is the Lebesgue measure of K?

Proof. (a) Construct a map from the set of all binary sequence to every element k in K: if the first term of the sequence is 0, then $k \in [0, (1-d_0)/2]$, otherwise $k \in [(1+d_0)/2, 1]$; if the second term of the sequence is 0, then $k \in [0, ((1-d_0)/2 - d_1)/2]$ or $k \in [(1+d_0)/2, (1+d_0)/2 + ((1-d_0)/2 - d_1)/2]$; and so on.

That is, if the ith term of the sequence is 0, for the ith operation, k belongs to the left sub-interval, and if the ith term is 1, k belongs to the right sub-interval. Note that the map is injective (if two sequence vary by at least one term, on that operation, one sequence maps to some point on the left sub-interval and the other maps to one on the other), then since the number of binary sequence is uncountable (having the same cardinality with $2^{\mathbb{N}}$), K with $\mathrm{Card}(K) \geq \mathrm{Card}(2^{\mathbb{N}})$ is also uncountable.

Note that K is bounded and is an intersection of a sequence of closed sets (or is the complement of a union of open interval), so K is compact.

Suppose K has non-empty interior, then there exists some interval such that we should do the opration that removes an open interval centered at the middle of that interval. After the operation, that interval is not in K, contradiction!

Comment. *Not clear. Maybe it wasn't at the middle of the intervals containing it.*

Let $k \in K$. Then $k \in K_n$ for every $n \geq 0$. For each K_n, consider

$$r_n := \min\left\{\frac{1}{2^n}, \max \text{distance to the endpoints of the sub-interval } k \text{ is in}\right\}$$

Then, one can always find a point in $(k - r_n, k + r_n)$. As $n \to \infty$, the sequence converges to k.

(b) Note that the Lebesgue measure of K is equal to the intervals of the total length, given that the removing length is d_n times the length of the previous one:

$$\prod_{i=0}^{\infty}(1 - d_i)$$

$\square$

6. Let X be a nonnegative random variable. Prove that $\mathbb{E}(X) < +\infty$ if and only if

$$\sum_{n\in\mathbb{N}} \mathbb{P}(X \geq n) < \infty$$

Proof. Let $n \in \mathbb{N}$. For $x \in [n-1, n]$, one has

$$\mathbb{P}(|X| \geq n) \leq \int_{n-1}^{n} \mathbb{P}(|X| \geq x)\,\mathrm{d}x \leq \mathbb{P}(|X| \geq n-1),$$

and since X is nonnegative, one has

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) \leq \int_0^{\infty} \mathbb{P}(X \geq x)\,\mathrm{d}x \leq \sum_{n=1}^{\infty} \mathbb{P}(X \geq n-1)$$

Note that

$$\int_0^{\infty} \mathbb{P}(X \geq x)\,\mathrm{d}x = \int_0^{\infty}\int_x^{\infty} f(y)\,\mathrm{d}y\mathrm{d}x = \int_0^{\infty}\int_0^{y} f(y)\,\mathrm{d}y\mathrm{d}x = \int_0^{\infty} yf(y)dy = \mathbb{E}[X]$$

Comment. *Better to write it without the probability density function.*

Hence,

$$\sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n) \leq \mathbb{E}(|X|) \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n),$$

implying that $\mathbb{E}(X) < +\infty$ if and only if $\sum_{n\in\mathbb{N}} \mathbb{P}(X \geq n) < \infty$. $\square$

7. **Convergence in Measure**

   Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. and $(f_n)_{n\geq 1}, f : \Omega \to \mathbb{R}$ measurable (for the Borel σ-field on $\mathbb{R}$). We say that $(f_n)_{n\geq 1}$ converges in measure to f if for any $\varepsilon > 0$ we have

   $$\mu(|f_n - f| > \varepsilon) \xrightarrow[n\to\infty]{} 0$$

   (a) Show that $\int |f - f_n|\,\mathrm{d}\mu \to 0$ implies that f_n converges to f in measure. Is the reciprocal true?

   (b) Show that if $f_n \to f$ μ-almost surely, then $f_n \to f$ in measure. Is the reciprocal true?

   (c) Show that if $f_n \to f$ in measure, there exists a subsequence of $(f_n)_{n\geq 1}$ which converges μ-almost surely.

   (d) **A stronger dominated convergence theorem**

   We assume that $f_n \to f$ in measure and $|f_n| \leq g$ for some integrable $g : \Omega \to \mathbb{R}$, for any $n \geq 1$.

   i. Show that $|f| \leq g$ μ-a.s.

   ii. Deduce that $\int |f_n - f|\,\mathrm{d}\mu \to 0$.

Proof. (a) One shall use the *Chebyshev's Inequality.*

Theorem 2.1.2 (Chebyshev)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let f be a nonnegative measurable function defined on X. Then for all $\lambda \in \mathbb{R}$, $\lambda > 0$ we have*

$$\mu\{x \in X : f(x) \geq \lambda\} \leq \frac{1}{\lambda} \int_X f(x) \, \mathrm{d}\mu$$

Proof. Define the set X_λ by

$$X_\lambda = \{x \in X : f(x) \geq \lambda\}$$

Consider the function $\varphi = \lambda \mathbb{1}_{X_\lambda}$ which is a simple function and $0 \leq \varphi \leq f$. Then, one has

$$\lambda \mu(X_\lambda) = \int_X \varphi(x) \, \mathrm{d}\mu \leq \int_X f(x) \, \mathrm{d}\mu,$$

and following $\lambda > 0$ we are done. □

Here we apply Chebyshev's Inequality on $|f - f_n|$ and take $\lambda = \varepsilon$: as $\int |f - f_n| \, \mathrm{d}\mu \to 0$, one has

$$\frac{1}{\varepsilon} \int_X |f - f_n| \, \mathrm{d}\mu \geq \mu(|f_n - f| > \varepsilon) \xrightarrow[n \to \infty]{} 0$$

The reverse is not necessarily true. Consider μ as the Lebesgue measure, and define

$$f_n(x) := n \cdot \mathbb{1}_{(0,1/n)}(x), \quad x \in (0, 1)$$

then f_n tends to 0 in measure ($1/n \to 0$) but the integral is 1.

(b) By definition, there exists a measurable set N with $\mu(N) = 0$ such that

$$\lim_{n \to \infty} f_n(\omega) = f(\omega), \quad \forall \omega \in N^{\complement}$$

If $\{|f_n - f| > \varepsilon\} = N$, then we are done. Otherwise, $\forall \omega \in N^{\complement}$,

$$0 = \lim_{n \to \infty} f_n(\omega) - f(\omega) = \lim_{n \to \infty} (f_n(\omega) - f(\omega)) = \lim_{n \to \infty} |f_n - f|$$

Hence, for $\varepsilon > 0$, one has

$$\lim_{n \to \infty} \mu(|f_n - f| > \varepsilon) = \mu(\varnothing) = 0$$

Comment. *Cannot make this jump.*

The reverse is not necessarily true. Similar to exercise 3, consider the interval $[0,1]$ and divide it successively into $2,3,4,\ldots$ parts and enumerate the intervals in succession. That is, $I_1=\left[0,\frac{1}{2}\right], I_2=\left(\frac{1}{2},1\right], I_3=\left[0,\frac{1}{3}\right], I_4=\left(\frac{1}{3},\frac{2}{3}\right], I_5=\left(\frac{2}{3},1\right]$ and so on. If $f_n(x)=\mathbb{1}_{I_n}(x)$, then f_n tends to 0 in measure $(1/n\to 0)$ but not almost everywhere (as it is always sliding).

(c) Let $\varepsilon=2^{-k}$, $k\in\mathbb{N}$. Choose n_k such that

$$\mu(|f_n-f|>2^{-k})\leq 2^{-k}$$

for all $n\geq n_k$. WLOG, suppose $n_{k+1}\geq n_k, \forall k\in\mathbb{N}$. Define

$$A_k:=\{x\in X:|f_{n_k}(x)-f(x)|>2^{-k}\}$$

Following the *Borel-Cantelli lemma*, as

$$\sum_{k\geq 1}\mu(A_k)\leq\sum_{k=1}^{\infty}2^{-k}<\infty,$$

one has

$$\mu\left(\limsup_{k\to\infty}A_k\right)=0,$$

implying that

$$\lim_{k\to\infty}f_{n_k}(x)=f(x)$$

(d) i. WLOG, suppose f_n is nonnegative. Since for a convergent sequence the $\lim$ and the $\limsup$ coincide, one has $f=\limsup f_n$ μ-a.s., and since $\limsup f_n\leq g$ $(g\geq f_n,\forall f_n)$ we are done.

ii. By *Fatou's Lemma* (for nonnegative Lebesgue measurable functions), one has

$$\int_\Omega g-\int_\Omega f=\int_\Omega(g-f)\leq\liminf_{n\to\infty}\int_\Omega(g-f_n)=\int_\Omega g-\limsup_{n\to\infty}\int_\Omega f_n$$

implying that $\int_\Omega f\geq\limsup_{n\to\infty}\int_\Omega f_n$. And similarly, one has

$$\int_\Omega g+\int_\Omega f=\int_\Omega(g+f)\leq\liminf_{n\to\infty}\int_\Omega(g+f_n)=\int_\Omega g+\liminf_{n\to\infty}\int_\Omega f_n,$$

showing that $\int_\Omega f\leq\liminf_{n\to\infty}\int_\Omega f_n$. Hence,

$$\lim_{n\to\infty}\int_\Omega f_n=\int_\Omega f\quad\iff\quad\int|f_n-f|\,\mathrm{d}\mu\to 0$$

Comment. *This $\iff$ is not true in general. Need to do some argument for $|f_n - f|$.*

$\square$

8. Consider a probability space $(\Omega, \mathcal{A}, \mu)$ and $(A_n)_n$ a sequence in $\mathcal{A}$. Let $f : \Omega \to \mathbb{R}$ be measurable (for the Borel σ-field on $\mathbb{R}$) such that $\int_\Omega |\mathbb{1}_{A_n} - f| \, \mathrm{d}\mu \to 0$ as $n \to \infty$. Prove that there exists $A \in \mathcal{A}$ such that $f = \mathbb{1}_A$ μ-a.s., i.e. $\mu(f = \mathbb{1}_A) = 1$.

   Proof. By definition, $\mathbb{1}_{A_n}$ converges in L^1 space, and by uniqueness of limit, one has

   $$\lim_{n\to\infty} \mathbb{1}_{A_n} = f, \quad \mu\text{−a.s.}$$

   Comment. *L^1 convergence does not imply a.s. convergence but implies a.s. convergence on a subsequence.*

   We want to show that $\limsup \mathbb{1}_{A_n} = \mathbb{1}_{\limsup A_n}$ and $\liminf \mathbb{1}_{A_n} = \mathbb{1}_{\liminf A_n}$ (since then, as $\lim_{n\to\infty} \mathbb{1}_{A_n}$ exists, one has $f = \mathbb{1}_{\limsup A_n} = \mathbb{1}_{\liminf A_n} \overset{\text{injective}}{=} \mathbb{1}_{\lim A_n} = \mathbb{1}_A$ exists).

   Note that $\mathbb{1}_{\bigcup_n A_n} = \sup_n \mathbb{1}_{A_n}$ and $\mathbb{1}_{\bigcap_n A_n} = \inf_n \mathbb{1}_{A_n}$, so that

   $$\begin{aligned} \mathbb{1}_{\liminf A_n} &= \sup_{n\ge1} \mathbb{1}_{\bigcap_{k=n}^\infty A_k} = \sup_{n\ge1} \inf_{k\ge n} \mathbb{1}_{A_k} = \liminf_n \mathbb{1}_{A_n}, \\ \mathbb{1}_{\limsup A_n} &= \inf_{n\ge1} \mathbb{1}_{\bigcup_{k=n}^\infty A_k} = \inf_{n\ge1} \sup_{k\ge n} \mathbb{1}_{A_k} = \limsup_n \mathbb{1}_{A_n} \end{aligned}$$

   Hence, $A = \lim_n A_n \in \mathcal{A}$. $\square$

2.1.3 Transformations, Product Spaces, Distributions and Expectations

1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Prove that if $A \cap B = \varnothing$ and A, B are independent, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$.

   Proof. By definition, one has

   $$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

   where $\mathbb{P}(A \cap B) = \mathbb{P}(\varnothing) = 0$. Hence, $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$ (otherwise $\mathbb{P}(A \cap B) \neq 0$). $\square$

2. Let X be a nonnegative random variable with null expectation. Prove that it is 0 almost surely.

   Proof. By definition, one has

   $$\mathbb{E}(X) = \int_\Omega X \, \mathrm{d}\mathbb{P} = 0$$

   Consider a sequence of events A_n such that

   $$A_n := \left\{\omega \in \Omega : X(\omega) > \frac{1}{n}\right\}, \quad n = 1, 2, \ldots$$

Since X is nonnegative, i.e. $\mathbb{P}(\{\omega \in \Omega : X(\omega) < 0\}) = 0$, one has

$$0 = \int_\Omega X \, \mathrm{d}\mathbb{P} \geq \int_{A_n} X \, \mathrm{d}\mathbb{P} \geq \mathbb{P}(A_n) \cdot \frac{1}{n}$$

so that (as a probability is nonnegative) $\mathbb{P}(A_n) = 0, \forall n \in \mathbb{N}$. Hence, by *continuity of measure*,

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \neq 0\}) = \mathbb{P}\left(\bigcup_n A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) = 0$$

implying that $\mathbb{P}(\{\omega \in \Omega : X(\omega) = 0\}) = 1$, i.e. it is 0 almost surely. □

3. Calculate $\mathbb{E}(X)$ for the following probability measures $\mathbb{P}^X$.
   (a) $\mathbb{P}^X$ has Gaussian density $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$, for some $\sigma > 0$ and $\mu \in \mathbb{R}$;
   (b) $\mathbb{P}^X$ has exponential density $\lambda e^{-\lambda x} \mathbb{1}_{x\geq 0}$ for some $\lambda > 0$;
   (c) $\mathbb{P}^X = p\delta_a + q\delta_b$ where $p + q = 1$, $p, q \geq 0$ and $a, b \in \mathbb{R}$;
   (d) $\mathbb{P}^X$ is the Poisson distribution: $\mathbb{P}^X(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}$ for any integer $n \geq 0$, for some $\lambda > 0$.

Solution. (a) By the change-of-variables formula, one has

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/(2\sigma^2)} \, \mathrm{d}y + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/(2\sigma^2)} \, \mathrm{d}y
\end{aligned}$$

where the first integrand is symmetric by the y axis and the second is a Gaussian density with $\mu = 0$

$$= 0 + \mu \cdot 1 = \boxed{\mu}$$

(b) Similarly, one has

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \cdot \lambda e^{-\lambda x} \mathbb{1}_{x\geq 0} \, \mathrm{d}x \\
&= \lambda \int_0^{\infty} x \cdot e^{-\lambda x} \, \mathrm{d}x \\
&= \lambda \cdot \left(-\frac{x e^{-\lambda x}}{\lambda} - \int -\frac{e^{-\lambda x}}{\lambda} \, \mathrm{d}x \right) \\
&= \lambda \cdot \left(-\frac{x e^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2} \right)_0^{\infty} \\
&= \lambda \cdot \left(0 + \frac{1}{\lambda^2} \right) = \boxed{\frac{1}{\lambda}}
\end{aligned}$$

(c) Note that δ is the *Dirac measure* such that

$$\delta_x(A) = \mathbb{1}_A(x) = \begin{cases} 0, & x \notin A; \\ 1, & x \in A. \end{cases}$$

Hence, one has $\mathbb{P}(X = a) = p$, $\mathbb{P}(X = b) = q$, and $\mathbb{P}(X = x) = 0, \forall x \neq a, b$, so that

$$\mathbb{E}[X] = \boxed{ap + bq}$$

(d) Given the probability mass function, one has

$$\begin{aligned} \mathbb{E}[X] &= \sum_{n=0}^{\infty} n \cdot e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \boxed{\lambda} \end{aligned}$$

□

4. Let X be a standard Gaussian random variable. What is the density of $1/X^2$?

Solution. Let $Y = 1/X^2$, then the probability distribution function of Y is given as

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) && y > 0 \\ &= \mathbb{P}\left(|X| \geq \frac{1}{\sqrt{y}}\right) \\ &= 2\mathbb{P}\left(X \leq -\frac{1}{\sqrt{y}}\right) && \mathcal{N}(0,1) \text{ is symmetric by } 0 \\ &= 2F_X\left(-\frac{1}{\sqrt{y}}\right) \\ &= \sqrt{\frac{2}{\pi}} \int_{-\infty}^{-1/\sqrt{y}} e^{-t^2/2} \, \mathrm{d}t \end{aligned}$$

Hence, the density of $Y = 1/X^2$ is given by

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} \mathbb{1}_{(0,\infty)}(y) \\ &= \sqrt{\frac{2}{\pi}} \frac{e^{-1/2y}}{2y^{3/2}} \mathbb{1}_{(0,\infty)}(y) \\ &= \boxed{\frac{e^{-1/2y}}{\sqrt{2\pi} y^{3/2}} \mathbb{1}_{(0,\infty)}(y)} \end{aligned}$$

□

5. Let X be uniformly distributed on $[0,1]$ and $\lambda > 0$. Show that $-\lambda^{-1}\log X$ has the same distribution as an exponential random variable with parameter λ.

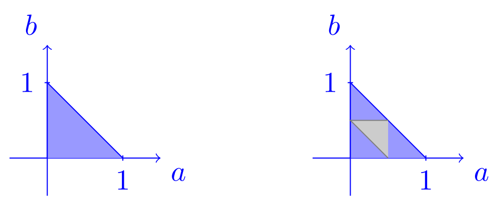
   Proof. Let $Y = -\lambda^{-1}\log X$, then the probability distribution function of Y is given as

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) && y \ge 0 \\
&= \mathbb{P}\left(-\lambda^{-1}\log X \le y\right) \\
&= \mathbb{P}\left(\log X \ge -\lambda y\right) \\
&= \mathbb{P}\left(X \ge e^{-\lambda y}\right) \\
&\text{Note that } e^{-\lambda y} \le 1 \text{ when } y \ge 0. \\
&= \left(1 - e^{-\lambda y}\right)\mathbb{1}_{[0,\infty)}(y) && X \sim \mathcal{N}(0,1)
\end{aligned}
$$

   which is the same PDF as an exponential random variable. □

6. A samouraï wants to create a triangle with a (rigid) spaghetti. With his saber, he cuts this spaghetti on two places, chosen uniformly and independently along this traditional pasta. What is the probability that he can create a triangle with sides these three pieces of spaghetti?

   Solution. Let the length of the left piece be a, the middle piece be b, and then the right piece is $1-a-b$. Since $a+b \le 1$, and the cuts are uniformly chosen, the probability space can be visualized as follows: every case is a point uniformly distributed in the shaded triangle (left).



   In order to create a triangle, one has to satisfy

$$
\begin{cases}
a+b > 1-a-b \\
|a-b| < 1-a-b \\
a+1-a-b > b \\
|a-1+a+b| < b \\
b+1-a-b > a \\
|b-1+a+b| < a
\end{cases}
\implies
\begin{cases}
a+b > 1/2 \\
a, b < 1/2
\end{cases}
$$

   which is represented by the gray area shown in the right graph. As uniformly distributed, the probability to create a triangle is the ratio of the gray area to the total area, i.e. $\boxed{1/4}$. □

7. Assume that $X_1, X_2, \dots$ are independent random variables uniformly distributed on $[0,1]$. Let $Y^{(n)} = n\inf\{X_i, 1 \le i \le n\}$. Prove that it converges weakly to an exponential random variable, i.e. for any continuous bounded function $f : \mathbb{R}^+ \to \mathbb{R}$,

$$\mathbb{E}\left(f\left(Y^{(n)}\right)\right) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^+} f(u)e^{-u}\,\mathrm{d}u$$

Proof. Consider $I_n = \inf\{X_i, 1 \le i \le n\}$. For $0 < t < 1$, one has

$$\{I_n \ge t\} = \bigcap_{i=1}^{n}\{X_i \ge t\} \overset{\text{i.i.d}}{\Longrightarrow} \mathbb{P}(I_n \ge t) = (1-t)^n$$

Then, for $0 < s < 1$, one has

$$\begin{aligned} F_{Y^{(n)}}(s) &= \mathbb{P}(Y^{(n)} \le s) \\ &= \mathbb{P}(nI_n \le s) \\ &= \mathbb{P}(I_n \le s/n) \\ &= 1 - \mathbb{P}(I_n > s/n) \\ &= 1 - (1 - s/n)^n \xrightarrow[n\to\infty]{} 1 - e^{-s} \end{aligned}$$

which gives us $\lim_{n\to\infty} f_{Y^{(n)}}(s) = e^{-s}$, i.e. for any continuous bounded function $f : \mathbb{R}^+ \to \mathbb{R}$,

$$\mathbb{E}\left(f\left(Y^{(n)}\right)\right) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^+} f(u)e^{-u}\,\mathrm{d}u$$

□

8. Let n and m be random numbers chosen independently and uniformly on $[\![1, N]\!]$. What are $\Omega, \mathcal{A}$ and $\mathbb{P}$ (which all implicitly depend on N) ? Prove that

$$\mathbb{P}(n \wedge m = 1) \xrightarrow[N\to\infty]{} \zeta(2)^{-1}$$

where

$$\zeta(2) = \prod_{p\in\mathcal{P}} \left(1 - p^{-2}\right)^{-1} = \sum_{n\ge1} n^{-2} = \frac{\pi^2}{6}$$

(you don't have to prove these equalities).

Remark. $\mathcal{P}$ is the set of prime numbers and $n \wedge m = 1$ means that their greatest common divisor is 1.

Proof. Here the Ω is the sample space as the set of all possible outcomes that

$$\Omega = \{(1,1),(1,2),\dots,(2,1),(2,2),\dots,(N,N-1),(N,N)\}$$

where the pair $(i,j) := \{n = i, m = j\}$.

$\mathcal{A}$, as a set of events, is the power set of Ω, i.e $\mathcal{A} = 2^{\Omega}$.

Since n and m are chosen independently and uniformly, the $\mathbb{P}$ is a function that maps each event to the number of outcomes in that event divided by N^2, i.e. $\mathbb{P}(A) = \mathrm{Card}(A)/N^2$.

Note that the probability that any integer is divisible by a prime number (or simply just an integer) p is $1/p$, since every pth integer is divisible by p.

Note also that $n \wedge m = 1$ if and only if no prime number p divides both n and m. Since n and m are independently chosen, for any $p \in \mathcal{P}$, one has $\mathbb{P}[p \nmid n, p \nmid m] = (1/p)(1/p) = 1/p^2$. Since the divisibility for each prime is independent[1], one has

$$\mathbb{P}(n \wedge m = 1) \approx \prod_{p \in \mathcal{P} \cap [\![1,N]\!]} \left(1 - \frac{1}{p^2}\right) \xrightarrow[N\to\infty]{} \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p^2}\right) = \zeta(2)^{-1}$$

Note that $\approx$ holds for large N, as the probability $1/p$ is only accurate for $p|N$ (or $N \to \infty$). $\square$

A rigorous proof. We can choose $\Omega = [\![1,N]\!]^2, \mathcal{A} = \{A : A \subset \Omega\}$ and define $\mathbb{P}$ through $\mathbb{P}(A) = |A|/|\Omega|$. We denote the random variable $(n,m) = (n(\omega), m(\omega))$. Let $A_p = \{p \mid n \text{ and } p \mid m\}$ and $f_N = \#\{p \le N, p \in \mathcal{P}\}$. Then, by inclusion-exclusion,

$$\begin{aligned}
&\mathbb{P}(n \wedge m = 1)\\
&= \mathbb{P}\left(\cap_{p\le N} A_p^c\right) = 1 - \mathbb{P}\left(\cup_{p \le N} A_p\right)\\
&= 1 - \sum_{p_1 \le N} \mathbb{P}\left(A_{p_1}\right) + \sum_{p_1 < p_2 \le N} \mathbb{P}\left(A_{p_1} \cap A_{p_2}\right) + \dots + (-1)^{f_N} \mathbb{P}\left(A_{p_1} \cap A_{p_2} \cap \dots A_{p_{f_N}}\right)\\
&= 1 - \sum_{p_1 \le N} \left(\frac{\lfloor N/p_1 \rfloor}{N}\right)^2 + \sum_{p_1 < p_2 \le N} \left(\frac{\lfloor N/(p_1 p_2) \rfloor}{N}\right)^2 - \dots
\end{aligned}$$

When replacing the integral parts by fractions, we obtain

$$\begin{aligned}
&1 - \sum_{p_1 \le N} p_1^{-2} + \sum_{p_1 < p_2 \le N} (p_1 p_2)^{-2} - \dots + (-1)^{f_N} \left(p_1 \dots p_{f_N}\right)^{-2}\\
&= \left(1 - p_1^{-2}\right) \dots \left(1 - p_{f_N}\right)^{-2}
\end{aligned}$$

[1] It can be shown by letting $A_p := \{X \text{ can be divided by } p\}$, then $\mathbb{P}(X = n) = 1/N, n = 1, \dots, N$, $\mathbb{P}(A_p) = \lfloor N/p \rfloor / N$ and $\mathbb{P}(A_q) = \lfloor N/q \rfloor / N$. Since p, q are primes, one has $\mathbb{P}(A_p \cap A_q) = \mathbb{P}(A_{pq}) = \lfloor N/(pq) \rfloor / N$, so that $\mathbb{P}(A_p \cap A_q) \approx \mathbb{P}(A_p) \cdot \mathbb{P}(A_q)$ for N is large.

Comment. *Ok, but as $N \to \infty$ that independence for each pair is not something that can be used. It's just an insight.*

which converges to the expected result as $N \to \infty$. Hence we now just need to bound the error when replacing integral parts with fractions. Top bound this error we use $\left|a^2 - \lfloor a \rfloor^2\right| \leq (a - \lfloor a \rfloor) \cdot 2a$ for any $a > 0$ so that

$$\left| \left(\frac{\lfloor N/(p_1 p_2 \dots p_k) \rfloor}{N} \right)^2 - \left(\frac{(N/(p_1 p_2 \dots p_k))}{N} \right)^2 \right| \leq \frac{2}{N p_1 p_2 \dots p_k}$$

and the global error is at most

$$N^{-1} \left(\sum_{p_1 \leq N} p_1^{-1} + \sum_{p_1 < p_2 \leq N} (p_1 p_2)^{-1} + \dots + (p_1 \dots p_{f_N})^{-1} \right)$$
$$= N^{-1} \left(1 + p_1^{-1}\right) \dots \left(1 + p_{f_N}^{-1}\right) \implies 0,$$

where the last convergence relies on

$$\log \left[\left(1 + p_1^{-1}\right) \dots \left(1 + p_{f_N}^{-1}\right) \right] \sim \sum_{p \leq N} 1/p \sim \log \log N$$

$\square$

9. Let $\varepsilon > 0$ and X be uniformly distributed on $[0, 1]$. Prove that, almost surely (i.e. the following event has probability 1), there exists only a finite number of rationals $\frac{p}{q}$, with $p \wedge q = 1$, such that
$$\left| X - \frac{p}{q} \right| < \frac{1}{q^{2+\varepsilon}}$$

Proof. WLOG, suppose $q \in \mathbb{N}$. Consider the sequence of events that

$$A_q := \left\{ \exists p \text{ s.t. } p \wedge q = 1 \text{ and } \left| X - \frac{p}{q} \right| < \frac{1}{q^{2+\varepsilon}} \right\}$$

and let

$$B_{p,q} := \left\{ \left| X - \frac{p}{q} \right| < \frac{1}{q^{2+\varepsilon}} \right\}$$

Then, since X is uniform on $[0, 1]$, one has

$$\mathbb{P}(A_q) \leq \mathbb{P}\left(\bigcup_{p=1}^{q} B_{p,q} \right) \leq \sum_{p=1}^{q} \frac{2}{q^{2+\varepsilon}} = \frac{2}{q^{1+\varepsilon}}$$

Hence, the series

$$\sum_{q=1}^{\infty} \mathbb{P}(A_q) = \sum_{q=1}^{\infty} \frac{1}{q^{1+\varepsilon}} < \infty \impliedby \varepsilon > 0$$

By the *Borel-Cantelli lemma*, one has

$$\mathbb{P}\left(\limsup_{q\to\infty} A_q\right) = 0,$$

implying that the probability that infinitely many of A_q occur is 0, i.e. there exists only a finite number of rationals $\frac{p}{q}$, with $p \wedge q = 1$, such that

$$\left|X - \frac{p}{q}\right| < \frac{1}{q^{2+\varepsilon}}$$

□

10. You toss a coin repeatedly and independently. The probability to get a head is p, a tail is $1-p$. Let A_k be the following event: k or more consecutive heads occur amongst the tosses numbered $2^k, \ldots, 2^{k+1}-1$. Prove that $\mathbb{P}(A_k \text{ i.o.}) = 1$ if $p \geq 1/2$, and 0 otherwise.

Remark. i.o. stands for "infinitely often", and A_k i.o. is the event $\bigcap_{n\geq 1}\bigcup_{m\geq n} A_m$.

Proof. Note that there are 2^k tosses between $2^k, \ldots, 2^{k+1}-1$, so that there are $2^k - k + 1$ blocks of k consecutive tosses. For each block of k consecutive tosses, the probability of all k tosses being all heads is p^k, so that

$$\begin{aligned}
&\mathbb{P}[\{k \text{ or more consecutive heads occur}\}] \\
=&\,\mathbb{P}\left[\bigcup_{i=1}^{2^k-k+1}\{\text{the } i\text{th block of } k \text{ consecutive tosses are all heads}\}\right] \\
\leq&\sum_{i=1}^{2^k-k+1}\mathbb{P}[\{\text{the } i\text{th block of } k \text{ consecutive tosses are all heads}\}] \\
=&\sum_{i=1}^{2^k-k+1} p^k = \left(2^k - k + 1\right)p^k \leq (2p)^k,
\end{aligned}$$

which follows that

$$p < \frac{1}{2} \implies \sum_{k=1}^{\infty}\mathbb{P}(A_k) \leq \sum_{k=1}^{\infty}(2p)^k = \frac{2p}{1-2p} < \infty$$

By the *Borel-Cantelli lemma*, one has $\mathbb{P}(A_k \text{ i.o.}) = 0$.

On the other hand, note that, for every $i \neq j$, the set of tosses numbered $2^i, \ldots, 2^{i+1}-1$ and the set numbered $2^j, \ldots, 2^{j+1}-1$ are disjoint, so that the event A_k are mutually independent.

There are $\left\lfloor \frac{2^k}{k} \right\rfloor$ disjoint blocks of k consecutive tosses. so that one has

$$\begin{aligned}
&\mathbb{P}[\{k \text{ or more consecutive heads occur}\}] \\
&\geq \mathbb{P}\left[\left\{\text{at least one of the } \left\lfloor \tfrac{2^k}{k} \right\rfloor \text{ disjoint blocks of } k \text{ consecutive tosses are all heads}\right\}\right] \\
&= 1 - \mathbb{P}\left[\left\{\text{none of the } \left\lfloor \tfrac{2^k}{k} \right\rfloor \text{ disjoint blocks of } k \text{ consecutive tosses are all heads}\right\}\right] \\
&= 1 - \left(1 - p^k\right)^{\left\lfloor \frac{2^k}{k} \right\rfloor} \xrightarrow[k\to\infty]{} 1, \quad \text{for } p > \frac{1}{2}
\end{aligned}$$

since for $p > 1/2$, one has

$$\left\lfloor \frac{2^k}{k} \right\rfloor \log\left(1 - p^k\right) \leq -\frac{2^k}{k}p^k = -\frac{(2p)^k}{k} \xrightarrow[k\to\infty]{} -\infty \implies \left(1 - p^k\right)^{\left\lfloor \frac{2^k}{k} \right\rfloor} \xrightarrow[k\to\infty]{} 0$$

Then, by the converse of the *Borel-Cantelli lemma*[2],

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty \implies \mathbb{P}(A_k \text{ i.o.}) = 1$$

Finally, consider $p = 1/2$. One has

$$\log(1 - \mathbb{P}(A_k)) \leq -\frac{1}{k} \implies \mathbb{P}(A_k) \geq 1 - e^{-\frac{1}{k}},$$

implying that

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) \geq \sum_{k=1}^{\infty} \frac{1}{k} = \infty \implies \mathbb{P}(A_k \text{ i.o.}) = 1$$

□

2.2 Weak Convergence

1. Let X be a random variable with density $f_X(x) = (1 - |x|)\mathbb{1}_{(-1,1)}(x)$. Show that its characteristic function is

$$\phi_X(u) = \frac{2(1 - \cos u)}{u^2}$$

[2] The converse of the lemma states that, if the sum of the probabilites of the *independent* events A_n is infinite, then the set of all outcomes that are repeated infintely many times must occur with probability one.

Proof. By definition, one has

$$\begin{aligned}
\phi_X(u) &= \int_{\mathbb{R}} e^{\mathrm{i}ux} f_X(x)\,\mathrm{d}x \\
&= \int_{-1}^{1} e^{\mathrm{i}ux}(1-|x|)\,\mathrm{d}x \\
&= \int_{-1}^{0} e^{\mathrm{i}ux}(1+x)\,\mathrm{d}x + \int_{0}^{1} e^{iux}(1-x)\,\mathrm{d}x \\
&= \left[-\frac{\mathrm{i}\cdot(ux+u+\mathrm{i})\,\mathrm{e}^{\mathrm{i}ux}}{u^2}\right]_{-1}^{0} + \left[\frac{\mathrm{i}\cdot(ux-u+\mathrm{i})\,\mathrm{e}^{\mathrm{i}ux}}{u^2}\right]_{0}^{1} \\
&= \frac{\mathrm{i}\sin(u)-\cos(u)-\mathrm{i}u+1}{u^2} - \frac{\mathrm{i}\sin(u)+\cos(u)-\mathrm{i}u-1}{u^2} \\
&= \frac{2(1-\cos u)}{u^2}
\end{aligned}$$

as desired. □

2. (a) Prove that $\hat{\mu}$ is real-valued if and only if μ is symmetric, i.e. $\mu(A)=\mu(-A)$ for any Borel set A.

Proof. Assume that μ is a probability measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Assign a random variable X to μ. Note that such X exists by letting $X(\omega) := \inf\{z : \mu((-\infty,z]) \geq \omega\}$[3].

As $\mathcal{B}(\mathbb{R})$ can be generated by $S := \{(-\infty,a] : a\in\mathbb{R}\}$ and μ is finite[4], one has

$$\begin{aligned}
\mu(A)=\mu(-A),\forall A\in\mathcal{B}(\mathbb{R}) &\iff \mu_X(A)=\mu_X(-A),\forall A\in S \\
&\iff \mu_X(A)=\mu_{-X}(A),\forall A\in S, \\
&\iff X \text{ and } -X \text{ have the same distribution} \\
&\iff X \text{ and } -X \text{ have the same characteristic function}^5
\end{aligned}$$

By definition, $\hat{\mu}(t) := \int_{\mathbb{R}} e^{\mathrm{i}tx}\,\mathrm{d}\mu = \int_{\mathbb{R}}\cos tx\,\mathrm{d}\mu + \mathrm{i}\cdot\int_{\mathbb{R}}\sin tx\,\mathrm{d}\mu$. Note that, as $\sin x$ is odd but $\cos x$ is even, $\hat{\mu}_{-X}(t)=\hat{\mu}_X(-t)$ is the complex conjugate of $\hat{\mu}_X(t)$.

($\Longrightarrow$) Now, if $\hat{\mu}$ is real-valued, then $\hat{\mu}(t)=\int_{\mathbb{R}}\cos tx\,\mathrm{d}\mu$. Hence,

$$\hat{\mu}_{-X}(t) = \int_{\mathbb{R}}\cos(t(-x))\,\mathrm{d}\mu = \int_{\mathbb{R}}\cos tx\,\mathrm{d}\mu = \hat{\mu}_X(t)$$

($\Longleftarrow$) If $\hat{\mu}_X(t)=\hat{\mu}_{-X}(t)$, one has $\mathrm{Im}(\hat{\mu}_X(t))=0$, so that $\hat{\mu}$ is real-valued. □

[3] See reference here
[4] See reference here.
[5] Proved in Varadhan's note, pp.20 - 21.

(b) If X and Y are i.i.d., prove that $X - Y$ has a symmetric distribution.

Proof. As X and Y are i.i.d, one has

$$\begin{aligned}\phi_{X-Y}(t) &= \phi_X(t)\phi_{-Y}(t)\\ &= \phi_X(t)\phi_Y(-t)\\ &= \phi_X(t)\phi_X(-t)\end{aligned}$$

where $\phi_X(t) = \mathbb{E}\left[e^{\mathrm{i}tX}\right] = \mathbb{E}\left[\cos(tX)\right] + \mathrm{i}\cdot\mathbb{E}\left[\sin(tX)\right]$, so that

$$\begin{aligned}\phi_{X-Y}(t) &= \phi_X(t)\phi_X(-t)\\ &= \left(\mathbb{E}\left[\cos(tX)\right] + \mathrm{i}\cdot\mathbb{E}\left[\sin(tX)\right]\right)\left(\mathbb{E}\left[\cos(-tX)\right] + \mathrm{i}\cdot\mathbb{E}\left[\sin(-tX)\right]\right)\end{aligned}$$

where $\mathrm{Im}(\phi_{X-Y}(t)) = \mathbb{E}\left[\sin(tX)\right]\mathbb{E}\left[\cos(-tX)\right] + \mathbb{E}\left[\sin(-tX)\right]\mathbb{E}\left[\cos(tX)\right]$. Note that $\forall x \in \mathbb{R}$, one has $\cos x = \cos(-x)$ and $\sin x = -\sin(-x)$, which implies that

$$\mathrm{Im}(\phi_{X-Y}(t)) = 0$$

Following part 2a, one has $\mathbb{P}(X - Y) = \mathbb{P}(-(X - Y))$, i.e. $X - Y$ has a symmetric distribution. □

3. Let X_λ be a real random variable, with Poisson distribution with parameter λ. Calculate the characteristic function of X_λ. Conclude that $(X_\lambda - \lambda)/\sqrt{\lambda}$ converges in distribution to a standard Gaussian, as $\lambda \to \infty$.

Proof. Note that $\mathbb{P}_{X_\lambda}(k) = \frac{\lambda^k}{k!}e^{-\lambda}$ and that $e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}$. One has

$$\begin{aligned}\phi_{X_\lambda}(t) &= \sum_{k=0}^{\infty} e^{\mathrm{i}tk}\frac{\lambda^k}{k!}e^{-\lambda}\\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(e^{\mathrm{i}t}\lambda\right)^k}{k!}\\ &= e^{-\lambda}e^{\lambda e^{\mathrm{i}t}} = \boxed{e^{\lambda(e^{\mathrm{i}t}-1)}}\end{aligned}$$

Let $Z = (X_\lambda - \lambda)/\sqrt{\lambda}$. Then,

$$\begin{aligned}\phi_Z(t) &= \mathbb{E}\left[e^{\mathrm{i}tZ}\right]\\ &= \mathbb{E}\left[e^{\mathrm{i}t(X_\lambda-\lambda)/\sqrt{\lambda}}\right]\\ &= \frac{\mathbb{E}\left[e^{\mathrm{i}(t/\sqrt{\lambda})X}\right]}{e^{\mathrm{i}t\sqrt{\lambda}}}\\ &= \frac{e^{\lambda(e^{\mathrm{i}(t/\sqrt{\lambda})}-1)}}{e^{\mathrm{i}t\sqrt{\lambda}}} = \exp\left\{\lambda(e^{\mathrm{i}(t/\sqrt{\lambda})} - 1) - \mathrm{i}t\sqrt{\lambda}\right\}\end{aligned}$$

Note that

$$\begin{aligned} e^{\mathrm{i}(t/\sqrt{\lambda})} &= \sum_{n=0}^{\infty} \frac{(\mathrm{i}(t/\sqrt{\lambda}))^n}{n!} \\ &= 1 + \mathrm{i}t/\sqrt{\lambda} - t^2/2\lambda + \sum_{n=3}^{\infty} \frac{(\mathrm{i}(t/\sqrt{\lambda}))^n}{n!} \end{aligned}$$

so that

$$\begin{aligned} \phi_Z(t) &= \exp\left\{\lambda(e^{\mathrm{i}(t/\sqrt{\lambda})} - 1) - \mathrm{i}t\sqrt{\lambda}\right\} \\ &= \exp\left\{\lambda\left(1 + \mathrm{i}t/\sqrt{\lambda} - t^2/2\lambda + \sum_{n=3}^{\infty} \frac{(\mathrm{i}(t/\sqrt{\lambda}))^n}{n!} - 1\right) - \mathrm{i}t\sqrt{\lambda}\right\} \\ &= \exp\left\{\mathrm{i}t\sqrt{\lambda} - t^2/2 + \lambda\sum_{n=3}^{\infty} \frac{(\mathrm{i}(t/\sqrt{\lambda}))^n}{n!} - \mathrm{i}t\sqrt{\lambda}\right\} \\ &= \exp\left\{-t^2/2 + \sum_{n=3}^{\infty} \frac{(\mathrm{i}t)^n}{\lambda^{n/2-1}n!}\right\} \xrightarrow[\lambda\to\infty]{} e^{-t^2/2} \end{aligned}$$

which is the characteristic function of $\mathcal{N}(0,1)$. By *Lévy's continuity theorem*, one has $(X_\lambda - \lambda)/\sqrt{\lambda}$ converges in distribution to a standard Gaussian, as $\lambda \to \infty$. □

4. Assume that the sequence of random variables $(X_n)_{n\geq 1}$ satisfies $\mathbb{E}X_n \to 1$ and $\mathbb{E}X_n^2 \to 1$. Prove that $(X_n)_{n\geq 1}$ converges in distribution. What is the limit?

Proof. We first prove the following lemmas.

Lemma 2.2.1. $X_n \xrightarrow{L^r} X \implies X_n \xrightarrow{\mathbb{P}} X, r \geq 1$.

Proof. By *Markov's inequality*, one has

$$\begin{aligned} \mathbb{P}\big(|X_n - X| \geq \epsilon\big) &= \mathbb{P}\big(|X_n - X|^r \geq \epsilon^r\big) && \text{(since } r \geq 1\text{)} \\ &\leq \frac{\mathbb{E}|X_n - X|^r}{\epsilon^r} \xrightarrow[n\to\infty]{} 0 && \text{(by Markov's inequality).} \end{aligned}$$

□

Lemma 2.2.2. $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathcal{D}} X$.

Proof. Followed by the *stronger dominated convergence theorem* proved in HW2. □

We now prove that $(X_n)_{n\geq 1}$ converges in mean square to X where $\mathbb{E}[X] = 1, \mathbb{E}[X^2] = 1$.

$$\begin{aligned} \mathbb{E}\left[(X_n - X)^2\right] &= \mathbb{E}\left[X_n^2\right] - 2\mathbb{E}[X]\mathbb{E}\left[X_n\right] + \mathbb{E}[X^2] \\ &= \mathbb{E}\left[X_n^2\right] - 2\mathbb{E}\left[X_n\right] + 1 \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

Hence, by the lemmas, $X_n \xrightarrow{L^2} X \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathcal{D}} X$.

Note that $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = 0$, so that $\mathbb{P}^X = \delta_1$ is the limit of $(X_n)_{n\geq 1}$. □

5. Let $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$ be real random variables, with X_n and Y_n independent for any $n \geq 1$, and assume that X_n converges in distribution to X and Y_n to Y, with X and Y independent defined on the same probability space. Prove that $X_n + Y_n$ converges in distribution to $X + Y$.

   Proof. We first show that

   Proposition 2.2.3. *If X and Y are independent, then $\phi_{X+Y} = \phi_X \phi_Y$.*

   Proof. Given X and Y are independent, one has $\mathbb{E}(f(X)\cdot g(Y)) = \mathbb{E}(f(X))\cdot\mathbb{E}(g(Y))$. □

   Let $Z_n := X_n + Y_n$. Then, as assumed that $X_n \perp\!\!\!\perp Y_n, \forall n$, one has

$$\begin{aligned} &\phi_{Z_n}(t) = \phi_{X_n}(t)\phi_{Y_n}(t), \forall n \\ \implies &\lim_n \phi_{Z_n}(t) = \lim_n \left[\phi_{X_n}(t)\phi_{Y_n}(t)\right] \\ &\qquad = \lim_n \phi_{X_n}(t) \cdot \lim_n \phi_{Y_n}(t) \qquad \text{since } X_n, Y_n \text{ converge}^{6} \\ &\qquad = \phi_X(t)\cdot\phi_Y(t) = \phi_Z(t) \end{aligned}$$

   Then, by *Lévy's continuity theorem*, $X_n + Y_n$ converges in distribution to $X + Y$. □

6. Let X, Y be independent and assume that for some constant α we have $\mathbb{P}(X + Y = \alpha) = 1$. Prove that X and Y are both constant random variables.

   Proof. Note that $X \perp\!\!\!\perp Y \implies X \perp\!\!\!\perp (\alpha - Y)$ as α is a constant. Note also that

$$\mathbb{P}(X + Y = \alpha) = 1 \implies \mathbb{P}(X = \alpha - Y) = 1,$$

   i.e. $X = \alpha - Y$ a.s. Therefore, one has

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(X \leq x,\ X \leq x) \\ &= \mathbb{P}(X \leq x,\ \alpha - Y \leq x) \\ &= \mathbb{P}(X \leq x)\mathbb{P}(\alpha - Y \leq x) \\ &= [\mathbb{P}(X \leq x)]^2, \quad \forall x \in \mathbb{R} \end{aligned}$$

   which gives us $\mathbb{P}(X \leq x) = 0$ or $1, \forall x \in \mathbb{R}$.

   Now we show that this implies that X is a constant almost surely. Let $F(x) = \mathbb{P}(X \leq x)$. Note that F is non-decreasing and right-continuous, so that

$$\sup\{x \in \mathbb{R} : F(x) = 0\} = \inf\{x \in \mathbb{R} : F(x) = 1\} =: \beta$$

[6] Weak convergence of measure.

is a constant. Then, for any $\epsilon > 0$, one has

$$\mathbb{P}(X \leq \beta + \epsilon) = 1 \tag{2.1}$$

and

$$\mathbb{P}(X \leq \beta - \epsilon) = 0 \implies \mathbb{P}(X > \beta - \epsilon) = 1 \tag{2.2}$$

It follows that $\mathbb{P}(|X - \beta| \leq \epsilon) = 1, \forall \epsilon > 0$, so $X = \beta$ a constant a.s. and hence $Y = \alpha - \beta$ a constant a.s. ☐

7. Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with standard Cauchy distribution and let $M_n = \max(X_1, \ldots, X_n)$. Prove that $(nM_n^{-1})_{n\geq 1}$ converges in distribution and identify the limit.

Proof. Let $Y_n := nM_n^{-1}, n \geq 1$. Then,

$$\begin{aligned}
F_{Y_n}(y) &= \mathbb{P}(nM_n^{-1} \leq y) \\
&= \mathbb{P}(n/y \leq M_n) \\
&= 1 - \mathbb{P}(M_n \leq n/y) \\
&= 1 - \mathbb{P}(X_1, X_2, \ldots, X_n \leq n/y) \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}(X_i \leq n/y) \\
&= 1 - F_{X_1}^n(n/y)
\end{aligned}$$

Note that

$$\begin{aligned}
\lim_{n\to\infty} F_{Y_n}(y) &= 1 - \lim_{n\to\infty} F_{X_1}^n(n/y) \\
&= 1 - \lim_{n\to\infty} \left(\frac{1}{2} + \frac{\arctan(n/y)}{\pi}\right)^n \\
&= \boxed{\left(1 - e^{-y/\pi}\right) \cdot \mathbb{1}_{[0,\infty)}(y)} = F_Y(y)
\end{aligned}$$

Hence, Y_n converges in distribution. ☐

8. Let X, Y be i.i.d., with characteristic functions denoted φ_X, φ_Y, and suppose $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) = 1$. Assume also that $X + Y$ and $X - Y$ are independent.

   (a) Prove that
   $$\varphi_X(2u) = (\varphi_X(u))^3 \varphi_X(-u)$$

   (b) Prove that X is a standard Gaussian random variable.

Proof. (a) We first show that

Proposition 2.2.4. *If X has characteristic function $\varphi_X(t)$, then $\varphi_{a+bX}(t) = e^{\mathrm{i}at}\varphi(bt)$.*

Proof. $\mathbb{E}\left[e^{\mathrm{i}t(a+bX)}\right] = e^{\mathrm{i}ta}\mathbb{E}\left[e^{\mathrm{i}tbX}\right] = e^{\mathrm{i}at}\varphi(bt)$. □

Following 2.2.3 and 2.2.4, one has

$$\begin{aligned}
\varphi_X(2u) &= \varphi_{2X}(u) \\
&= \varphi_{(X+Y)+(X-Y)}(u) \\
&= \varphi_{(X+Y)}(u)\varphi_{(X-Y)}(u) \\
&= \varphi_{(X)}(u)\varphi_{(Y)}(u)\varphi_{(X)}(u)\varphi_{(Y)}(-u) \\
&= (\varphi_X(u))^3\varphi_X(-u)
\end{aligned}$$

(b) Observe that $\varphi(\cdot)$ never vanishes. If it did at some point u_0, then following the above equation it also vanishes at $u_0/2$, and by induction at $u_0/2^n, \forall n \in \mathbb{N}$[7], which contradicts to the fact that $\lim_n \varphi(u_0/2^n) = \varphi(0) = 1$.

We can define now that $\rho(u) := \varphi(u)/\varphi(-u)$ and following the above equation one has

$$\varphi_X(-2u) = (\varphi_X(-u))^3\varphi_X(u)$$

Taking the quotient of these two equations, one has $\rho(2u) = \rho^2(u)$, and hence by induction,

$$\rho(u) = \rho^{2^n}(u/2^n) = [1 + o(1/2^n)]^{2^n} \xrightarrow[n\to\infty]{} 1, \quad \forall u \in \mathbb{R},$$

which implies that $\rho \equiv 1$ and hence $\varphi(u) = \varphi(-u) \implies \varphi(2u) = \varphi^4(u)$. Iterating the above argument, one has

$$\varphi(u) = \varphi^{4^n}(u/2^n) = \left[1 - (1/2)(u/2^n)^2 + o\left((u/2^n)^2\right)\right]^{4^n} \xrightarrow[n\to\infty]{} e^{-u^2/2}, \quad \forall u \in \mathbb{R},$$

which is the characteristic function of the standard normal distribution, so that[8] $X \sim \mathcal{N}(0,1)$.

□

9. For any $d \geq 1$, we admit that there is only one probability measure μ on $\mathcal{S}_d$, (the $(d-1)$-th dimensional sphere embedded in $\mathbb{R}^d$) that is uniform, in the following sense: for any isometry $A \in \mathbb{O}(d)$ (the orthogonal group in $\mathbb{R}^d$), and any continuous function $f : \mathcal{S}_d \to \mathbb{R}$,

$$\int_{\mathcal{S}_d} f(x)\,\mathrm{d}\mu(x) = \int_{\mathcal{S}_d} f(Ax)\,\mathrm{d}\mu(x)$$

Let $X = (X_1, \ldots, X_d)$ be a vector of independent centered and reduced Gaussian random variables.

[7] We are using here the important fact that every characteristic function is (uniformly) continuous, which can be proved by the dominated convergence theorem.

[8] Proved in Varadhan's note, pp.20 - 21.

(a) Prove that the random variable $U = X/\|X\|_{L^2}$ is uniformly distributed on the sphere.

(b) Prove that, as $d \to \infty$, the main part of the globe is concentrated close to the Equator, i.e. for any $\varepsilon > 0$,

$$\int_{x\in\mathcal{S}_d, |x_1|<\varepsilon} \mathrm{d}\mu(x) \to 1$$

Proof. (a) [9]Note that the standard normal distribution is invariant under orthogonal transformation: this is because $A \in \mathbb{O}(d)$ is by definition $AA^\top = I$. Consider X and $Y := AX$. One has for each $i \in 1, 2, \ldots, d$,

$$Y_i = \sum_{k=1}^{d} a_{ik} X_k$$

where its mean is obviously 0 and $\mathrm{Var}(Y_i) = \sum_{k=1}^d a_{ik}^2 \mathrm{Var}(X_k) = 1$, hence $Y_i \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, I_d)$.

Note also that $A \in \mathbb{O}(d)$ preserves the length of the vector, That is, $\|X\|_{L^2} = \|AX\|_{L^2}$. Consider U and $V := AU = AX/\|AX\|_2$. One has

$$V = AX/\|X\|_2 \sim X/\|X\|_2 = U,$$

implying that U is invariant under orthogonal transformation. Since U is by definition on the unit sphere in $\mathbb{R}^d$, one has

$$\int_{\mathcal{S}_d} f(u)\,\mathrm{d}\mu(u) = \int_{\mathcal{S}_d} f(v)\,\mathrm{d}\mu(v) = \int_{\mathcal{S}_d} f(Au)\cdot \det\left|\frac{\partial v}{\partial u}\right| \mathrm{d}\mu(u)$$

where

$$v_i = \sum_{i=1}^{n} a_{ij} u_j \implies \frac{\partial v_i}{\partial u_j} = a_{ij} \implies \det\left|\frac{\partial v}{\partial u}\right| = \det A = 1$$

Thus, the equation holds as desired, so U is uniformly distributed on the sphere.

(b) [10]It is pointed out in the question that it is suffice to show that $x_1 \to 0$ as $d \to \infty$, because then the unit vector $\vec{x} = (x_1, \ldots, x_d)$ will be near the equator relative to the first basis vector.

We pick a unit vector randomly and uniformly. Then it only has to satisfy

$$x_1^2 + \ldots + x_d^2 = 1, \quad \mathbb{E}[x_1^2] = 1/d$$

[9]See reference here, here, and here.
[10]See reference here.

By Markov's inequality, one has

$$\begin{aligned}\mathbb{P}\left(x_1^2 \geq \frac{1}{\sqrt{d}}\right) \leq \frac{\mathbb{E}(x_1^2)}{\frac{1}{\sqrt{d}}} = \frac{1}{\sqrt{d}} &\implies \mathbb{P}\left(|x_1| \geq \frac{1}{d^{1/4}}\right) \leq \frac{1}{\sqrt{d}} \\ &\implies \mathbb{P}\left(|x_1| < \frac{1}{d^{1/4}}\right) \geq 1 - \frac{1}{\sqrt{d}}\end{aligned}$$

For every given ε, let $d > 1/\varepsilon^4$, then

$$\int_{x \in \mathcal{S}_d, |x_1| < \varepsilon} \mathrm{d}\mu(x) \geq \mathbb{P}\left(|x_1| < \frac{1}{d^{1/4}}\right) \geq 1 - \frac{1}{\sqrt{d}} \xrightarrow[d \to \infty]{\varepsilon \to 0} 1$$

$\square$

2.3 Independent Sums

2.3.1 Convolutions, Laws of Large Numbers

1. Prove that if a sequence of real random variables (X_n) converge in distribution to X, and (Y_n) converges in distribution to a constant c, then $X_n + Y_n$ converges in distribution to $X + c$.

   Proof. Note that the CDF of $Z = X + c$ is

   $$F_Z(z) = \mathbb{P}(X + c \leq z) = F_X(z - c)$$

   and thus z is a continuity point of F_Z if and only if $z - c$ is a continuity point of F_X. Consider z at which F_Z is continuous (so F_X is continuous at $z - c$), then

   $$\begin{aligned}\mathbb{P}(X_n + Y_n \leq z) &= \mathbb{P}(X_n + Y_n \leq z, Y_n \geq c) + \mathbb{P}(X_n + Y_n \leq z, Y_n < c) \\ &\leq \mathbb{P}(X_n \leq z - c) + \mathbb{P}(Y_n < c)\end{aligned}$$

   Since $Y_n \xrightarrow{\mathcal{D}} c$, one has $\lim_{n\to\infty} \mathbb{P}[Y_n < c] = 0$. With the assumption that $X_n \xrightarrow{\mathcal{D}} x$, one has

   $$\limsup_{n\to\infty} \mathbb{P}(X_n + Y_n \leq z) \leq \mathbb{P}(X \leq z - c) = F_Z(z) \tag{2.3}$$

   On the other hand,

   $$\begin{aligned}\mathbb{P}(X_n + Y_n > z) &= \mathbb{P}(X_n + Y_n > z, Y_n > c) + \mathbb{P}(X_n + Y_n > z, Y_n \leq c) \\ &\leq \mathbb{P}(Y_n > c) + \mathbb{P}(X_n > z - c) \\ &= \mathbb{P}(Y_n > c) + 1 - \mathbb{P}(X_n \leq z - c) \\ \iff 1 - \mathbb{P}(X_n + Y_n > z) &\geq \mathbb{P}(X_n \leq z - c) - \mathbb{P}(Y_n > c) \\ \implies \mathbb{P}(X_n + Y_n \leq z) &\geq \mathbb{P}(X_n \leq z - c) - \mathbb{P}(Y_n > c)\end{aligned}$$

along with the fact that $\lim_{n\to\infty} \mathbb{P}[Y_n > c] = 0$ and $X_n \xrightarrow{\mathcal{D}} x$, one has

$$\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \le z) \ge \mathbb{P}(X \le z - c) = F_Z(z) \tag{2.4}$$

Hence, 2.3 and 2.4 imply that

$$F_Z(z) \le \liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \le z) \le \limsup_{n\to\infty} \mathbb{P}(X_n + Y_n \le z) \le F_Z(z),$$

i.e. $F_Z(z) = \lim_{n\to\infty} \mathbb{P}(X_n + Y_n \le z), Z = x + c$, whenever F_Z is continuous as z, so $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$. □

2. Assume that (X, Y) has joint density

$$ce^{-(1+x^2)(1+y^2)},$$

where c is properly chosen. Are X and Y Gaussian random variables? Is (X, Y) a Gaussian vector?

Proof. No, X and Y are not Gaussian random variables. Note that x and y are symmetric in the joint density, so that it suffices to show that the marginal density of x does not follow a Gaussian random variable. One has

$$f_X(x) = \int_{\mathbb{R}} ce^{-(1+x^2)(1+y^2)}\,\mathrm{d}y = \left[\frac{\sqrt{\pi}\,\mathrm{c}e^{-x^2-1}\,\mathrm{erf}\left(\sqrt{x^2+1}\,y\right)}{2\sqrt{x^2+1}}\right]_{-\infty}^{\infty} = \frac{\sqrt{\pi}\,\mathrm{c}e^{-x^2-1}}{\sqrt{x^2+1}}$$

Suppose X is Gaussian that $X \sim \mathcal{N}(\mu, \sigma^2)$. By definition, X has the density

$$f_X(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{\sqrt{\pi}\,\mathrm{c}e^{-x^2-1}}{\sqrt{x^2+1}}, \quad \forall x \in \mathbb{R},$$

implying that

$$\frac{e}{\pi c\sqrt{2}}\sqrt{x^2+1} = \sigma \exp\left\{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 - x^2\right\},$$

but it is impossible, by taking any fixed μ, σ, for the equation to hold for all $x \in \mathbb{R}$. Hence, X is not a Gaussian random variable, and similarly Y is also not.

Since by definition, Gaussian vector requires "every linear combination of its components" to be Gaussian, which is false, simply for that $1 \cdot X + 0 \cdot Y = X$ is not Gaussian. Hence, (X, Y) is not a Gaussian vector. □

3. Let $(X_i)_{i\ge1}$ be a sequence of independent random variables, with X_i uniform on $[-i, i]$. Let $S_n = X_1 + \ldots + X_n$. Prove that $S_n/n^{3/2}$ converges in distribution and describe the limit.

Proof. We first show that *Lyapunov's condition* is sufficient for the CLT to hold[11].

Lemma 2.3.1 (Lyapunov's). *Suppose $\{X_1, \ldots, X_n, \ldots\}$ is a sequence of independent random variables, each with finite expected value μ_i and variance σ_i^2. Define $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If for some $\delta > 0$, Lyapunov's condition*

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0$$

is satisfied, then a sum of $\frac{X_i - \mu_i}{s_n}$ converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Proof. Since we have proved *Lindeberg's Theorem*[12], it suffices to show that *Lyapunov's* condition implies Lindeberg's condition.

Now, fix any $\varepsilon, \delta > 0$. For any random variable $|X| > \varepsilon$, one has

$$X^2 = \frac{|X|^{2+\delta}}{|X|^\delta} \le \frac{|X|^{2+\delta}}{\varepsilon^\delta}$$

Thus for any r.v. X one has

$$\mathbb{E}\left[X^2 \mathbb{1}_{|X|>\varepsilon}\right] \le \frac{\mathbb{E}\left[|X|^{2+\delta}\right]}{\varepsilon^\delta},$$

so that (replacing ε with εs_n, X with $X_j - \mu_j$)

$$\begin{aligned} \frac{1}{s_n^2} \sum_{j=1}^n \mathbb{E}\left[(X_j - \mu_j)^2 \cdot \mathbf{1}_{\{|X_j - \mu_j| > \varepsilon s_n\}}\right] &\le \frac{1}{s_n^2} \frac{1}{(\varepsilon s_n)^\delta} \sum_{i=1}^n \mathbb{E}\left[|X_i - \mu_i|^{2+\delta}\right] \\ &= \frac{1}{\varepsilon^\delta} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\left[|X_i - \mu_i|^{2+\delta}\right] \xrightarrow[n\to\infty]{\text{Lyapunov's}} 0 \end{aligned}$$

which implies Lindeberg's condition:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=1}^n \mathbb{E}\left[(X_j - \mu_j)^2 \cdot \mathbf{1}_{\{|X_j - \mu_j| > \varepsilon s_n\}}\right] = 0$$

□

[11] Exercise 3.18 in text (Varadhan's notes).

[12] See Varadhan's notes in pp. 51 - 53, and also in p. 50 that it emphasizes that the proof "of the central limit theorem we do *not* need the random variables $\{X_j\}$ to have identical distributions."

Note that for each i, one has $\mathbb{E}[X_i] = 0$, $\mathrm{Var}(X_i) = (i-(-i))^2/12 = i^2/3$. Hence, for each n, one has $\mathbb{E}[S_n] = 0$ and $\mathrm{Var}(S_n) = \sum_i i^2/3 = n(n+1)(2n+1)/18$. We first check Lyapunov's condition: taking $\delta = 2$, one has

$$\begin{aligned}
\frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[|X_i - \mu_i|^{2+\delta}\right] &= \frac{1}{s_n^4} \sum_{i=1}^{n} \mathbb{E}\left[|X_i|^4\right] \\
&= \frac{1}{\mathrm{Var}(S_n)^2} \sum_{i=1}^{n} \frac{i^5 - (-i)^5}{5 \cdot 2i} \\
&= \frac{324}{5} \frac{\frac{1}{5}n^5 + \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n}{n^2(n+1)^2(2n+1)^2} \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

So, following Lyapunov's CLT, one has

$$\begin{aligned}
&\frac{1}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \sum_{i=1}^{n} (X_i - \mu_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \\
\implies &\frac{1}{\sqrt{n(n+1)(2n+1)/18}} S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \\
\implies &\sqrt{\frac{18}{(1+1/n)(2+1/n)}} \frac{S_n}{n^{3/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)
\end{aligned}$$

where $k = \sqrt{\frac{18}{(1+1/n)(2+1/n)}} \xrightarrow[n\to\infty]{} 3$ a constant. Hence, $S_n/n^{3/2}$ converges in distribution. Since $k \cdot S_n/n^{3/2} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, one has $S_n/n^{3/2} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{9}\right)$. □

4. Find the probability distribution μ of a $\mathbb{Z}$-valued random variable which is symmetric[13], not integrable, but such that its characteristic function is differentiable at 0.

Proof. Take a $\mathbb{Z}$-valued random variable X s.t.

$$\mu(X = k) = \mu(X = -k) = \begin{cases} \frac{C}{k^2 \log k}, & k \geq 2, k \in \mathbb{N} \\ 0, & k = 0, 1 \end{cases}$$

for some well-chosen $C > 0$. Such C exists as

$$1 = \int_1^\infty \frac{1}{k^2} > \sum_2^\infty \frac{1}{k^2 \log k} > 0 \implies \exists C > 0 \text{ s.t. } C \sum_2^\infty \frac{1}{k^2 \log k} = \frac{1}{2}$$

(i) Note that X is symmetric by definition.

[13] $\mu(\{i\}) = \mu(\{-i\})$ for any $i \in \mathbb{Z}$.

(ii) X is not integrable (i.e. $\int |f| < \infty$) as $\mathbb{E}|X|$ is infinite:

$$\mathbb{E}|X| = 2C \sum_{k=2}^{\infty} \frac{k}{k^2 \log k} = 2C \sum_{k=2}^{\infty} \frac{1}{k \log k} = \infty$$

since

$$\sum_{k=2}^{\infty} 2^k a_{2^k} = \sum_{k=2}^{\infty} \frac{2^k}{2^k \log 2^k} = \sum_{k=2}^{\infty} \frac{1}{\log 2} \frac{1}{k}$$

diverges and then following the Cauchy Condensation Test.

(iii) By definition, the characteristic function is given as

$$\varphi_X(t) = \sum_{k \geq 2} \frac{C}{k^2 \log k} e^{itk} + \frac{C}{k^2 \log k} e^{-itk}$$

It suffices to show that $e^{itk}, k \in \mathbb{Z} \backslash \{0, \pm 1\}$ is differentiable at $t = 0$:

$$e^{itk} = \cos(tk) + i \sin(tk) \implies \lim_{h \to 0} \frac{e^{ikh} - 1}{h} = ik$$

as both sin and cos are differentiable at 0.

Comment. *No it is not suffice.*

□

5. For any probability measure μ supported on $\mathbb{R}_+$, one defines the Laplace transform as

$$\mathscr{L}_\mu(\lambda) = \int_0^\infty e^{-\lambda x} \, \mathrm{d}\mu(x), \ \lambda \geq 0$$

   (a) Prove that $\mathscr{L}_\mu$ is well-defined, continuous on $\mathbb{R}_+$ and $\mathscr{C}^\infty$ on $\mathbb{R}_+^*$.

   Proof. Note that μ is a positive finite measure supported on $\mathbb{R}_+$. For $\lambda \geq 0$, one has $e^{-\lambda x}$ monotonely decreasing on $\mathbb{R}_+$ (but positive). Hence,

$$\int_0^\infty |e^{-\lambda x}| \, \mathrm{d}\mu(x) \leq \int_0^\infty 1 \, \mathrm{d}\mu(x) \leq 1$$

   So $\mathscr{L}_\mu(\lambda)$ is integrable on $\mathbb{R}_+$ for any probability measure μ, thus well-defined.

   Now, note that $f(\lambda, x) = e^{-\lambda x}$ is measurable for each $\lambda \geq 0$ and $\lambda \mapsto e^{-x\lambda}$ is continuous for each $x \geq 0$. To show $\mathscr{L}_\mu$ is continuous at $\lambda_0 \in \mathbb{R}_+^*$, it suffices to show that for each sequence $(\lambda_n)_{n \geq 1}$ with $\lim_{n \to \infty} \lambda_n = \lambda_0$ one has $\lim_{n \to \infty} \mathscr{L}_\mu(\lambda_n) = \mathscr{L}_\mu(\lambda_0)$. Note that this follows from the dominated convergence theorem applied to $f_n(\lambda) = f(x, \lambda_n)$, since one has

$$\lim_{n\to\infty} f_n(\lambda) = \lim_{n\to\infty} f(x, \lambda_n) = f(x, \lambda_0)$$

by continuity of $\lambda \mapsto f(x, \lambda)$, and one also has $f(x, \lambda_n) \leq 1$ for each n and x.

Note that $\partial^\alpha e^{-\lambda x}$ is bounded for $\alpha = 1, 2, \ldots$ as $e^{-\lambda}$ is a Schwartz function. Following the dominated convergence theorem, one has

$$\int_0^\infty \partial^\alpha e^{-\lambda x} \,\mathrm{d}\mu = \partial^\alpha \int_0^\infty e^{-\lambda x} \,\mathrm{d}\mu, \quad \forall \alpha \in \mathbb{N}$$

Hence, $\mathscr{L}_\mu \in \mathscr{C}^\infty$ on $\mathbb{R}_+^*$. $\square$

(b) Prove that $\mathscr{L}_\mu$ characterizes the probability measure μ supported on $\mathbb{R}_+$.

Proof. Since we have proved that $\mathscr{L}_\mu$ is infinitely differentiable, one has, for positive λ,

$$\mathscr{L}^{(k)}(\lambda) = (-1)^k \int_0^\infty y^k e^{-\lambda y} \mu(\mathrm{d}y).$$

Therefore, for positive x and λ,

$$\begin{aligned}\sum_{k=0}^{[\lambda x]} \frac{(-1)^k}{k!} \lambda^k \mathscr{L}^{(k)}(\lambda) &= \int_0^\infty \sum_{k=0}^{[\lambda x]} e^{-\lambda y} \frac{(\lambda y)^k}{k!} \mu(\mathrm{d}y) \\ &= \int_0^\infty G_{\lambda y}\left(\frac{x}{y}\right) \mu(\mathrm{d}y)\end{aligned}$$

where

$$G_\lambda(t) = \sum_{k=0}^{[\lambda t]} e^{-\lambda} \frac{\lambda^k}{k!}$$

is exactly the distribution function of Y_λ/λ where $Y_\lambda \sim \operatorname{Pois}(\lambda)$. Applying Chebyshev's inequality, one has

$$\mathbb{P}\left[\left|\frac{Y_\lambda - \lambda}{\lambda}\right| \geq \varepsilon\right] \leq \frac{\lambda}{\lambda^2 \varepsilon^2} \to 0, \quad \lambda \to \infty$$

This implies that

$$\lim_{\lambda\to\infty} G_\lambda(t) = \begin{cases} 1 & \text{if } t > 1 \\ 0 & \text{if } t < 1. \end{cases}$$

Now, fix $x > 0$. If[14] $0 \leq y < x$, then $G_{\lambda y}(x/y) \to 1$ as $\lambda \to \infty$; if $y > x$, the limit is 0. If $\mu\{x\} = 0$, the integrand $G_{\lambda y}(x/y)$ thus converges as $\lambda \to \infty$ to $\mathbb{1}_{[0,x]}(y)$ except on a

[14] If $y = 0$, the integrand in $y^k e^{-\lambda y}$ is 1 for $k = 0$ and 0 for $k \geq 1$; hence for $y = 0$, the integrand $\sum_{k=0}^{[\lambda x]} e^{-\lambda y} \frac{(\lambda y)^k}{k!}$ is 1.

set of μ-measure 0. The bounded convergence theorem then gives

$$\lim_{\lambda\to\infty}\sum_{k=0}^{[\lambda x]}\frac{(-1)^k}{k!}\lambda^k\mathscr{L}^{(k)}(\lambda)=\mu[0,x]=F(x) \tag{2.5}$$

Thus $\mathscr{L}(\lambda)$ determines the value of F at x if $x>0$ and $\mu\{x\}=0$, which covers all but countably many values of x in $[0,\infty)$. Since F is right-continuous, F itself and hence μ are determined through 2.5 by $\mathscr{L}(\lambda)$. $\square$

(c) Assume that for a sequence $(\mu_n)_{n\geq1}$ of probability measure supported on $\mathbb{R}_+$, one has $\mathscr{L}_{\mu_n}(\lambda)\to\ell(\lambda)$ for any $\lambda\geq0$, and ℓ is right-continuous at 0. Prove that $(\mu_n)_{n\geq1}$ is tight, and that it converges weakly to a measure μ such that $\ell=\mathscr{L}_\mu$.

Proof. By Fubini's theorem, one has

$$\begin{aligned}\frac{2}{u}\int_0^u(1-\mathscr{L}_\mu(\lambda))\,\mathrm{d}\lambda&=\int_0^\infty\frac{2}{u}\int_0^u(1-e^{-\lambda x})\,\mathrm{d}\lambda\,\mu_n(\mathrm{d}x)\\&=\int_0^\infty\frac{2}{u}\frac{(ux-1)+e^{-ux}}{x}\,\mu_n(\mathrm{d}x)\\&\geq2\int_{2/u}^\infty\frac{1}{u}\frac{(ux-1)+e^{-ux}}{x}\,\mu_n(\mathrm{d}x)\\&=2\int_{2/u}^\infty1+\frac{e^{-ux}-1}{ux}\,\mu_n(\mathrm{d}x)\\&\geq2\int_{2/u}^\infty1+\frac{e^{-ux}-1}{2}\,\mu_n(\mathrm{d}x)\\&\geq2\int_{2/u}^\infty\frac{1}{2}\,\mu_n(\mathrm{d}x)=\mu_n(x:x\in(2/u,\infty))\end{aligned}$$

Since $\mathscr{L}_\mu$ is continuous on $\mathbb{R}_+$, there is for positive ε a u for which

$$2u^{-1}\int_0^u(1-\mathscr{L}_\mu(\lambda))\mathrm{d}\lambda<\varepsilon$$

Since $\mathscr{L}_{\mu_n}$ converges to ℓ, the bounded convergence theorem implies that there exists an n_0 such that $2u^{-1}\int_0^u(1-\mathscr{L}_{\mu_n}(\lambda))\,\mathrm{d}\lambda<2\varepsilon$ for $n\geq n_0$. If $a=2/u$, then $\mu_n[x:x\geq a]<2\varepsilon$ for $n\geq n_0$. Increasing a if necessary will ensure that this inequality also holds for the finitely many n preceding n_0. Therefore, $\{\mu_n\}$ is tight. Following *Helly's selection theorem*, there exists some subsequence $\{\mu_{n_k}\}$ that converges weakly to some μ.

Now, $\mu_n\Rightarrow\mu$ will follow if it is shown that each subsequence $\{\mu_{n_k}\}$ that converges weakly at all converges weakly to μ. But if $\mu_{n_k}\Rightarrow\nu$ as $k\to\infty$, following part (b) one must have ν and μ coincide as μ is uniquely determined by $\mathscr{L}_\mu$, and thus $\ell=\mathscr{L}_\mu$. $\square$

6. (*law of the iterated logarithm*) The goal of this problem is to prove the iterated logarithm law, first for Gaussian random variables. In other words, for $X_1, X_2 \dots$ i.i.d. standard Gaussian random variables, denoting $S_n = X_1 + \cdots + X_n$, we have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right) = 1 \tag{2.6}$$

(a) Prove that

$$\mathbb{P}(X_1 > \lambda) \underset{\lambda\to\infty}{\sim} \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}}$$

Proof. Let $x = \lambda + t/\lambda$. Following the change of variable, for every positive λ, one has

$$\mathbb{P}(X > \lambda) = \frac{1}{\sqrt{2\pi}} \int_\lambda^{+\infty} \mathrm{e}^{-x^2/2}\mathrm{d}x = \frac{1}{\lambda\sqrt{2\pi}} \mathrm{e}^{-\lambda^2/2} \int_0^{+\infty} \mathrm{e}^{-t}\mathrm{e}^{-t^2/(2\lambda^2)}\mathrm{d}t$$

When $\lambda \to \infty$, $\mathrm{e}^{-t^2/(2\lambda^2)} \to 1$ and hence $\int_0^{+\infty} \mathrm{e}^{-t}\mathrm{e}^{-t^2/(2\lambda^2)}\mathrm{d}t \to 1$ (as the density function is dominated, the limit can be taken out of the integral by DCT), so that

$$\mathbb{P}(X_1 > \lambda) \underset{\lambda\to\infty}{\sim} \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}}$$

□

In the following questions we denote $f(n) = \sqrt{2n\log\log n}$, $\lambda > 1$, $c, \alpha > 0$,

$$\begin{aligned} A_k &= \left\{S_{\lfloor\lambda^k\rfloor} \geq cf\left(\lambda^k\right)\right\}, \\ C_k &= \left\{S_{\lfloor\lambda^{k+1}\rfloor} - S_{\lfloor\lambda^k\rfloor} \geq cf\left(\lambda^{k+1} - \lambda^k\right)\right\}, \\ D_k &= \left\{\sup_{n\in[\![\lambda^k,\lambda^{k+1}]\!]} \frac{S_n - S_{\lfloor\lambda^k\rfloor}}{f(\lambda^k)} \geq \alpha\right\} \end{aligned}$$

(b) Prove that for any $c > 1$ we have $\sum_{k\geq1} \mathbb{P}(A_k) < \infty$ and

$$\limsup_{k\to\infty} \frac{S_{\lfloor\lambda^k\rfloor}}{f(\lambda^k)} \leq 1 \text{ a.s.}$$

Proof. Let $c = 1 + \varepsilon$ for some $\varepsilon > 0$. Following part 6a, with the fact that $S_n \sim \mathcal{N}(0, n)$, one has (as $(1+\varepsilon)f(n)/\sqrt{n} \to \infty$)

$$\begin{aligned} \mathbb{P}(S_n \geq (1+\varepsilon)f(n)) &= \mathbb{P}(S_n/\sqrt{n} \geq (1+\varepsilon)f(n)/\sqrt{n}) \\ &\sim \frac{1}{(1+\varepsilon)\sqrt{2\log\log n}\sqrt{2\pi}} \frac{1}{(\log n)^{(1+\varepsilon)^2}} \leq \frac{1}{(\log n)^{(1+\varepsilon)^2}} \end{aligned}$$

Replace n with $\lfloor \lambda^k \rfloor$, one has

$$\mathbb{P}(A_k) \leq \frac{1}{(k \log \lambda)^{(1+\varepsilon)^2}}$$

Hence, the series $\sum_{k\geq 1} \mathbb{P}(A_k)$ converges by comparison test to $\sum_n \frac{1}{n^{1+\varepsilon}}$. Then, following Borel–Cantelli, one has

$$\mathbb{P}\left(\limsup_{k\to\infty} \frac{S_{\lfloor \lambda^k \rfloor}}{f\left(\lambda^k\right)} \geq 1+\varepsilon\right) = 0, \quad \forall \varepsilon > 0$$

i.e. $\limsup_{k\to\infty} \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} \leq 1$ a.s.. □

(c) Prove that for any $c < 1$ we have $\sum_{k\geq 1} \mathbb{P}(C_k) = \infty$ and

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

Proof. Note that C_k is pairwise independent since they are sums of different X_i, where X_i are i.i.d. Assume λ is large enough so that $f\left(\lambda^{k+1} - \lambda^k\right) \in \mathbb{R}$. Now, set $Y_k := S_{\lfloor \lambda^{k+1} \rfloor} - S_{\lfloor \lambda^k \rfloor}$, so that $Y_k \sim \mathcal{N}\left(0, \left\lfloor \lambda^{k+1} \right\rfloor - \left\lfloor \lambda^k \right\rfloor\right)$. For $\varepsilon > 0$, one has

$$\begin{aligned}
\mathbb{P}\left(Y_k \geq c\left(f\left(\lambda^{k+1} - \lambda^k\right)\right)\right) &\sim \mathbb{P}\left(Y_k \geq c\left(f\left(\left\lfloor \lambda^{k+1} \right\rfloor - \left\lfloor \lambda^k \right\rfloor\right)\right)\right) \\
&\sim \frac{\exp\left(-(1-\varepsilon)^2 \log\left(\log\left(\lfloor \lambda^{k+1} \rfloor - \lfloor \lambda^k \rfloor\right)\right)\right)}{2(1-\varepsilon)\sqrt{\pi \log\log\left(\lfloor \lambda^{k+1} \rfloor - \lfloor \lambda^k \rfloor\right)}} \\
&\geq \frac{\alpha(k+1)^{-(1+\varepsilon)^2}}{\sqrt{\log(k+1)}} \\
&\geq \frac{\beta}{(k+1)\log(k+1)}
\end{aligned}$$

So by comparison test to the series $\sum_n \frac{1}{n}$, one has $\sum_{k\geq 1} \mathbb{P}(C_k) = \infty$. Since C_k are mutually independent, by (the second) Borel-Cantelli, $\mathbb{P}(C_k \text{ i.o.}) = 1$. □

(d) Let $\varepsilon > 0$ and choose $c = 1 - \varepsilon/10$. Prove that almost surely the following inequality holds for infinitely many k:

$$\frac{S_{\lfloor \lambda^{k+1} \rfloor}}{f(\lambda^{k+1})} \geq c\frac{f(\lambda^{k+1} - \lambda^k)}{f(\lambda^{k+1})} - (1+\varepsilon)\frac{f(\lambda^k)}{f(\lambda^{k+1})}$$

Proof. As $f(\lambda^{k+1}) > 0$, it suffices to show that

$$S_{\lfloor \lambda^{k+1} \rfloor} \geq cf(\lambda^{k+1} - \lambda^k) - (1+\varepsilon) f(\lambda^k)$$

Following the result of part 6c, one has

$$S_{\lfloor \lambda^{k+1} \rfloor} \geq cf(\lambda^{k+1} - \lambda^k) + S_{\lfloor \lambda^k \rfloor} \quad \text{i.o.}$$

Now it suffices to show that

$$S_{\lfloor \lambda^k \rfloor} \geq -(1+\varepsilon) f(\lambda^k),$$

but this follows from that S_n is symmetric by 0, so that $\liminf_{k\to\infty} \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} \geq -1$ a.s. □

(e) By choosing a large enough λ in the previous inequality, prove that almost surely

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \geq 1$$

Proof. Note that we λ is large enough, one has (since n increases much faster than $\log \log n$)

$$\frac{f(\lambda^{k+1} - \lambda^k)}{f(\lambda^{k+1})} \to \sqrt{\frac{\lambda - 1}{\lambda}}, \quad \frac{f(\lambda^k)}{f(\lambda^{k+1})} \to \frac{1}{\sqrt{\lambda}}$$

Thus, when λ is large enough, the above inequality becomes

$$\frac{S_{\lfloor \lambda^{k+1} \rfloor}}{f(\lambda^{k+1})} \geq (1-\gamma)\sqrt{\frac{\lambda - 1}{\lambda}} - (1+\varepsilon)\frac{1}{\sqrt{\lambda}}$$

Taking $\gamma \to 0$ and $\lambda \to \infty$, one has

$$\mathbb{P}\left(\limsup_n \frac{S_{\lfloor \lambda^{k+1} \rfloor}}{f(\lambda^{k+1})} \geq (1-\gamma)\sqrt{\frac{\lambda - 1}{\lambda}} - (1+\varepsilon)\frac{1}{\sqrt{\lambda}} = 1 \right) = 1$$

□

(f) Prove that for any $n \in [\![\lambda^k, \lambda^{k+1}]\!]$ and $S_n > 0$ we have

$$\frac{S_n}{f(n)} \leq \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lfloor \lambda^k \rfloor)} + \frac{S_n - S_{\lfloor \lambda^k \rfloor}}{f(\lfloor \lambda^k \rfloor)}$$

Proof. Note that $f(\lfloor\lambda^k\rfloor) \leq f(n)$ since f is monotonely increasing and $n \geq \lfloor\lambda^k\rfloor$, which implies that

$$\begin{aligned} \frac{1}{f(n)} &\leq \frac{1}{f(\lfloor\lambda^k\rfloor)} \\ \implies \frac{S_n}{f(n)} &\leq \frac{S_n}{f(\lfloor\lambda^k\rfloor)} \\ \implies \frac{S_n}{f(n)} &\leq \frac{S_{\lfloor\lambda^k\rfloor}}{f(\lfloor\lambda^k\rfloor)} + \frac{S_n - S_{\lfloor\lambda^k\rfloor}}{f(\lfloor\lambda^k\rfloor)} \end{aligned}$$

□

(g) Prove that

$$\mathbb{P}(D_k) \underset{k\to\infty}{\sim} 2\mathbb{P}\left(X_1 \geq \frac{\alpha f(\lambda^k)}{\sqrt{\lambda^{k+1}-\lambda^k}}\right) \underset{k\to\infty}{\sim} \frac{c}{\sqrt{\log k}}\left(\frac{1}{k}\right)^{\frac{\alpha^2}{\lambda-1}}$$

Proof. Note that the *reflection principle* of a random walk on $\mathbb{Z}$ gives us

$$\mathbb{P}\left(\max_{1\leq k\leq n} X_k \geq b\right) = 2\mathbb{P}(X_n \geq b)$$

Plugging in $\mathbb{P}(D_k)$, one has

$$\mathbb{P}\left(\max_{n\in[\![\lambda^k,\lambda^{k+1}]\!]} \frac{S_n - S_{\lfloor\lambda^k\rfloor}}{f(\lambda^k)} \geq \alpha\right) = 2\mathbb{P}\left(\frac{S_{\lfloor\lambda^{k+1}\rfloor} - S_{\lfloor\lambda^k\rfloor}}{f(\lambda^k)} \geq \alpha\right) \tag{2.7}$$

where $S_{\lfloor\lambda^{k+1}\rfloor} - S_{\lfloor\lambda^k\rfloor} \sim \mathcal{N}(0, \lambda^{k+1}-\lambda^k) \sim X_1 \cdot \sqrt{\lambda^{k+1}-\lambda^k}$, so that

$$2\mathbb{P}\left(\frac{S_{\lfloor\lambda^{k+1}\rfloor} - S_{\lfloor\lambda^k\rfloor}}{f(\lambda^k)} \geq \alpha\right) = 2\mathbb{P}\left(X_1 \geq \frac{\alpha f(\lambda^k)}{\sqrt{\lambda^{k+1}-\lambda^k}}\right)$$

Now, following part 6a, one has

$$\begin{aligned} 2\mathbb{P}\left(X_1 \geq \frac{\alpha f(\lambda^k)}{\sqrt{\lambda^{k+1}-\lambda^k}}\right) &= 2\mathbb{P}\left(X_1 \geq \sqrt{\frac{2\lambda^k \log\log\lambda^k}{\lambda^{k+1}-\lambda^k}} \cdot \alpha\right) \\ &\underset{k\to\infty}{\sim} \frac{1}{\sqrt{\frac{\log k + \log\log\lambda}{\lambda-1}}\alpha\sqrt{\pi}} e^{-\frac{\alpha^2(\log\log\lambda^k)}{\lambda-1}} \\ &\underset{k\to\infty}{\sim} c_1 \cdot \frac{1}{\sqrt{\log k}} \cdot \frac{1}{(k\log\lambda)^{\alpha^2/(\lambda-1)}} \\ &\underset{k\to\infty}{\sim} \frac{c}{\sqrt{\log k}}\left(\frac{1}{k}\right)^{\frac{\alpha^2}{\lambda-1}} \end{aligned}$$

for some constants c_1 and c. □

(h) Prove that for $\alpha^2 > \lambda - 1$, almost surely

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \leq \limsup_{n\to\infty} \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} + \alpha$$

Proof. Following part 6g, if $\alpha^2 > \lambda - 1$, then $\mathbb{P}(D_k \text{ i.o.}) = 0$ by Borel-Cantelli, so that

$$\mathbb{P}\left\{ \sup_{n\in[\lambda^k,\lambda^{k+1}]\cap z} \frac{S_n - S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} < \alpha \right\} = 1$$

for k sufficiently large. Now fix n large enough that $n \in [\lambda^k, \lambda^{k+1}]$ where k is large enough that the above holds. We obtain, provided $S_n > 0$ that the following holds almost surely (for an appropriately fixed n and all sufficiently large k)

$$\frac{S_n}{f(n)} < \frac{S_n}{f(\lambda^k)} < \alpha + \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)}$$

Let $k \to \infty$. One has

$$\frac{S_n}{f(n)} \leq \alpha + \limsup_k \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} \quad \text{a.s.}$$

The result in part 6e implies there is a subsequence $(n_j)_{j\geq 1}$ such that $S_{n_j} > 0$ for each j, and since the above holds for every n_j (and moreover, for any $S_n > 0$), we may let $n \to \infty$, implying that

$$\limsup_n \frac{S_n}{f(n)} \leq \alpha + \limsup_k \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} \quad \text{a.s.}$$

□

(i) By choosing appropriate λ and α, prove that almost surely

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \leq 1$$

Proof. Following part 6h and part 6b above, one has

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \leq \limsup_{n\to\infty} \frac{S_{\lfloor \lambda^k \rfloor}}{f(\lambda^k)} + \alpha \leq 1 + \alpha$$

where $\alpha > \sqrt{\lambda - 1}$, where $\lambda > 1$ can be arbitrarily close to 1. Hence,

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \leq 1 + \alpha, \quad \forall \alpha > 0$$

leading us to the desired result. □

(j) State a result similar to (2.6) for i.i.d. uniformly bounded random variables. Which steps in the above proof need to be modified to prove this universality result? How?

Solution. Let $\{\phi_n(x)\}$ be a uniformly bounded orthonormal system of realvalued functions on the interval $[0,1]$. Then there exists a subsequence $\{\phi_{n_k}(x)\}$ and a real-valued function $f(x), \int_0^1 f^2(x)dx = 1, 0 \leqq f(x) \leqq B$, where B is the uniform bound of $\{\phi_n(x)\}$, such that for any arbitrary sequence $\{a_k\}$ of real numbers satisfying

$$A_N = \left(a_1^2 + a_2^2 + \cdots + a_N^2\right)^{1/2} \to \infty \text{ as } N \to \infty,$$
$$M_N = o\left(A_N (\log\log A_N)^{-1/2}\right) \text{ where } M_N = \max_{k \leqq N} |a_k|$$

we have

$$\limsup \frac{S_N(x)}{\left(2A_N^2 \log\log A_N\right)^{1/2}} = f(x) \quad \text{where } S_N(x) = \sum_{k=1}^{N} a_k \phi_{n_k}(x)$$

□

2.3.2 Central Limit Theorem

1. Assume $(\Omega, \mathcal{A}, \mathbb{P})$ is such that Ω is countable and $\mathcal{A} = 2^\Omega$. Prove that convergence in probability and convergence almost sure are the same.

Proof. We have proved in class that in general one has convergence a.s. implies convergence in probability (briefly: $0 = \mathbb{P}\{\limsup\{|X_n - X| > \varepsilon\} \geq \limsup \mathbb{P}\{|X_n - X| > \varepsilon\}, \forall \varepsilon > 0$).

Now we show that convergence in probability implies convergence a.s. when Ω is countable.

Since $\mathcal{A} = 2^\Omega$, the singletons are measureable. Let $\{\omega_n : n \in \mathbb{N}\}$ be the set of elements whose singletons have positive probability. It suffices to show that if $X_n \xrightarrow{\mathbb{P}} X$, then $X_n(\omega_i) \to X(\omega_i)$ for each $i \in \mathbb{N}$.

Fix $i \in \mathbb{N}, \varepsilon > 0$ and assume that $X_n \xrightarrow{\mathbb{P}} X$. Then there is an N s.t. $\mathbb{P}(\{|X_n - X| \geq \varepsilon\}) < \mathbb{P}(\omega_i)$ whenever $n \geq N$. This implies that if $n \geq N$, then $|X_n(\omega_i) - X(\omega_i)| < \varepsilon$. By definition, $X_n \xrightarrow{\text{a.s.}} X$. □

2. Let $(X_i)_{i \geq 1}$ be i.i.d. Gaussian with mean 1 and variance 3. Show that

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2} = \frac{1}{4} \quad \text{a.s.}$$

Proof. Following the *strong law of large numbers*, for $(X_i)_{i\geq 1}$ being i.i.d., one has

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}[X_i] = 1$$

Next, as $(X_i)_{i\geq 1}$ are i.i.d., $(X_i^2)_{i\geq 1}$ are also i.i.d., with $\mathbb{E}[X_i^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2 = 3 + 1 = 4$. Again, following the SLLN, one has

$$\frac{X_1^2 + \cdots + X_n^2}{n} \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}[X_i^2] = 4$$

Since $f(x) = 1/x$ is continuous a.e., following the *continuous mapping theorem* (CMT), one has

$$\frac{n}{X_1^2 + \cdots + X_n^2} \xrightarrow[n\to\infty]{\text{a.s.}} \frac{1}{4}$$

Also, if $X_n \to X$ and $Y_n \to Y$ a.s., one has $X_n Y_n \to XY$ a.s., since if $X_n(\omega) \to X(\omega)$ and $Y_n(\omega) \to Y(\omega)$ for some ω, then $X_n(\omega)Y_n(\omega) \to X(\omega)Y(\omega)$. By contraposition, one has $\{X_n Y_n \not\to XY\} \subseteq \{X_n \not\to X\} \cup \{Y_n \not\to Y\}$, which is a negligible set. Hence, one has

$$\frac{X_1 + \cdots + X_n}{n} \cdot \frac{n}{X_1^2 + \cdots + X_n^2} = \frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2} \xrightarrow[n\to\infty]{\text{a.s.}} \frac{1}{4}$$

□

3. Let f be a continuous function on $[0,1]$. Calculate the asymptotics, as $n \to \infty$, of

$$\int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

Solution. Suppose $(X_i)_{i\geq 1}$ is a sequence of i.i.d random variables with uniform distribution in $[0,1]$. Then, since X_i has density $1 \cdot \mathbb{1}_{[0,1]}(x_i), \forall i$, one has by definition

$$\int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) \mathrm{d}x_1 \ldots \mathrm{d}x_n = \mathbb{E}\left[f\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)\right]$$

Following SLLN, one has

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}[X_i] = \frac{1}{2}$$

By CMT, as f is continuous, one has

$$f\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \xrightarrow[n\to\infty]{\text{a.s.}} f\left(\frac{1}{2}\right)$$

Since f is continuous on $[0,1]$ and thus bounded ($\max f$ exists on compact space), by the dominated convergence theorem, one has almost surely

$$\begin{aligned}\lim_{n\to\infty}\int_{[0,1]^n} f\left(\frac{x_1+\cdots+x_n}{n}\right)\mathrm{d}x_1\ldots\mathrm{d}x_n &= \lim_{n\to\infty}\mathbb{E}\left[f\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\right]\\ &= \mathbb{E}\left[\lim_{n\to\infty} f\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\right]\\ &= \mathbb{E}\left[f\left(\frac{1}{2}\right)\right] = \boxed{f\left(\frac{1}{2}\right)}\end{aligned}$$

□

4. (*Bernstein-Weierstrass*) The goal of this exercise is to prove that any function, continuous on an interval of $\mathbb{R}$, can be approximated by polynomials, arbitrarily close for the L^∞ norm. Let f be a continuous function on $[0,1]$. The n-th Bernstein polynomial is

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

   (a) Let $S_n(x) = B^{(n,x)}/n$, where $B^{(n,x)}$ is a binomial random variable with parameters n and x: $B^{(n,x)} = \sum_{\ell=1}^{n} X_i$ where the X_i's are independent and $\mathbb{P}(X_i = 1) = x$, $\mathbb{P}(X_i = 0) = 1-x$. Prove that $B_n(x) = \mathbb{E}(f(S_n(x)))$.

   (b) Prove that $\|B_n - f\|_{L^\infty([0,1])} \to 0$ as $n\to\infty$.

Proof. (a) Note that $B^{(n,x)}$ is the sum of i.i.d. $X_i \sim \mathrm{Ber}(x)$, thus $B^{(n,x)} \sim \mathrm{Bin}\,(n,x)$ with

$$\mathbb{P}\left[B^{(n,x)} = k\right] = \binom{n}{k}x^k(1-x)^{n-k}$$

So, by definition, one has

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

(b) Intuitively, following SLLN and CMT, $S_n(x)\to x \implies \mathbb{E}[f(S_n(x))]\to f(x)$. Rigorously, we show that

$$\|B_n - f\|_{L^\infty([0,1])} = \sup_x |f(x) - B_n(x)| \le \delta(\varepsilon) + \frac{2M}{n\varepsilon^2}$$

where $M := \sup_x |f(x)|$ and $\delta(\varepsilon) := \sup[|f(x)-f(y)| : |x-y|\le\varepsilon]$. Since then by the uniform continuity of f, one has $\lim_{\varepsilon\to 0}\delta(\varepsilon) = 0$, and taking $\varepsilon = n^{-1/3}$ leads us to the desired result.

Note that we can bound $|f(S_n(x)) - f(x)|$ by $\delta(\varepsilon)$ on the set $\{|S_n(x) - x| < \varepsilon\}$ and by $2M$ on $\{|S_n(x) - x| < \varepsilon\}^{\complement}$. Note also that $B^{(n,x)} \sim \text{Bin}(n, x)$ has $\mathbb{E}[B^{(n,x)}] = nx$ and $\text{Var}(B^{(n,x)}) = nx(1-x)$, so that $S_n(x)$ has $\mathbb{E}[S_n(x)] = x$ and $\text{Var}(B^{(n,x)}) = x(1-x)/n$. Hence, following Chebyshev's inequality, one has

$$\begin{aligned}
|f(x) - B_n(x)| &= |f(x) - \mathbb{E}(f(S_n(x)))| \\
&\leq \mathbb{E}\left[|f(x) - f(S_n(x))|\right] \quad \text{expectation preserves order} \\
&\leq \delta(\varepsilon)\mathbb{P}[|S_n(x) - x| < \varepsilon] + 2M\mathbb{P}[|S_n(x) - x| \geq \varepsilon] \quad \text{bounds} \\
&\leq \delta(\varepsilon) + 2M\frac{x(1-x)}{n\varepsilon^2} \quad \text{Chebyshev} \\
&\leq \delta(\varepsilon) + \frac{2M}{n\varepsilon^2} \quad \text{for } x \in [0,1]
\end{aligned}$$

□

5. Calculate

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$$

Solution. The limit is $\boxed{1/2}$.

Observe that $e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \mathbb{P}(P_n \leq n)$ where $P_n \sim \text{Pois}\,(n)$. By the property[15] of Poisson distribution, one has $P_n = X_1 + \ldots + X_n$ where $(X_n)_{n\geq 1}$ is a sequence of i.i.d. $\text{Pois}\,(1)$ r.v. Following the *central limit theorem*, one has $Y_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n - n)$ converging in distribution to $\mathcal{N}(0,1)$. Hence, one has

$$\begin{aligned}
\mathbb{P}(P_n \leq n) &= \mathbb{P}\left(X_1 + \cdots + X_n \leq n\right) \\
&= \mathbb{P}\left(X_1 + \cdots + X_n - n \leq 0\right) \\
&= \mathbb{P}\left(\frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \leq 0\right) \\
&= \mathbb{P}(Y_n \leq 0) = \mathbb{P}(\mathcal{N}(0,1) \leq 0) = \boxed{\frac{1}{2}}
\end{aligned}$$

□

6. Let $\alpha > 0$ and, given $(\Omega, \mathcal{A}, \mathbb{P})$, let $(X_n, n \geq 1)$ be a sequence of independent real random variables with law $\mathbb{P}(X_n = 1) = \frac{1}{n^\alpha}$ and $\mathbb{P}\left(X_n = 0\right) = 1 - \frac{1}{n^\alpha}$. Prove that $X_n \to 0$ in $\mathcal{L}^1$, but that almost surely

$$\limsup_{n\to\infty} X_n = \begin{cases} 1 & \text{if } \alpha \leq 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$$

[15] Sum of independent Poisson distribution is Poisson, with parameter as the sum of parameters.

Proof. First, we check that $\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|\right) = \lim_{n\to\infty} \mathbb{E}\left(|X_n|\right) = 0$:

$$\mathbb{E}(|X_n|) = 1 \cdot \frac{1}{n^\alpha} + 0 \cdot \left(1 - \frac{1}{n^\alpha}\right) = \frac{1}{n^\alpha} \xrightarrow[n\to\infty]{} 0, \quad \forall \alpha > 0$$

since $n^\alpha \to \infty$ no matter how small $\alpha > 0$ is.

Then, note that

$$\sum_{n=1}^{\infty} \mathbb{P}[X_n = 1] = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} \begin{cases} = \infty & \text{if } \alpha \leq 1 \text{ (by comparison test with Harmonic series)} \\ < \infty & \text{if } \alpha > 1 \text{ (by integral test on } \int_1^\infty \frac{1}{x^\alpha}\,\mathrm{d}x) \end{cases}$$

Since X_n are independent, one can apply the second Borel-Cantelli to the first case and the first Borel-Cantelli to the second case and get

$$\mathbb{P}\left(\limsup_{n\to\infty} X_n = 1\right) = \begin{cases} 1 & \text{if } \alpha \leq 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$$

When $\alpha > 1$, since for any i, X_i is 1 or 0 a.s., one has $\sup_{m\geq n} X_m = 1$ or 0 a.s., so by definition

$$1 = \mathbb{P}\left(\left\{\limsup_{n\to\infty} X_n = 1\right\}^\complement\right) = \mathbb{P}\left(\limsup_{n\to\infty} X_n \neq 1\right) = \mathbb{P}\left(\limsup_{n\to\infty} X_n = 0\right)$$

Hence, one has almost surely

$$\limsup_{n\to\infty} X_n = \begin{cases} 1 & \text{if } \alpha \leq 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$$

□

7. A sequence of random variables $(X_i)_{i\geq 1}$ is said to be completely convergent to X if for any $\varepsilon > 0$, we have $\sum_{i\geq 1} \mathbb{P}(|X_i - X| > \varepsilon) < \infty$. Prove that if the X_i's are independent then complete convergence implies almost sure convergence.

   Proof. Following Borel-Cantelli, one has

   $$\mathbb{P}\left(\limsup_{n\to\infty} |X_i - X| > \varepsilon\right) = 0, \quad \forall \varepsilon > 0,$$

   which is in fact the definition of almost sure convergence.

Moreover, if one needs to show the equivalence of definition of almost sure convergence:

$$\begin{aligned}
0 &= \mathbb{P}\left(\limsup_{n\to\infty} |X_i - X| > \frac{1}{n}\right) \\
&= \mathbb{P}\left(\omega : \forall i, \exists j > i, \text{s.t.} \, |X_j(\omega) - X(\omega)| \geq \frac{1}{n}\right) \\
\Longrightarrow \; 1 &= \mathbb{P}\left(\omega : \exists i, \forall j > i, \text{s.t.} \, |X_j(\omega) - X(\omega)| < \frac{1}{n}\right) \\
&= \mathbb{P}\left(\omega : \lim_{i\to\infty} |X_j(\omega) - X(\omega)| = 0\right) \\
&= \mathbb{P}\left(\omega : \lim_{i\to\infty} X_j(\omega) = X(\omega)\right)
\end{aligned}$$

Remark. It does not require X_i to be independent. I believe the intention of this exercise is instead to show independence and almost sure convergence imply complete convergence.

□

8. Let $(X_n)_{n\geq 1}$ be a sequence of random variables, on the same probability space, with $\mathbb{E}(X_\ell) = \mu$ for any ℓ, and a weak correlation in the following sense: $\mathrm{Cov}(X_k, X_\ell) \leq f(|k-\ell|)$ for all indexes k, ℓ, where the sequence $(f(m))_{m\geq 0}$ converges to 0 as $m \to \infty$. Prove that $\left(n^{-1}\sum_{k=1}^n X_k\right)_{n\geq 1}$ converges to μ in L^2.

Proof. One should assume that $f(0), f(1), f(2), \ldots$ are finite. We shall prove that

$$\begin{aligned}
& \lim_{n\to\infty} \mathbb{E}\left[\left(\left(n^{-1}\sum_{k=1}^n X_k\right) - \mu\right)^2\right] = 0 \\
\Longleftrightarrow \quad & \lim_{n\to\infty} \mathbb{E}\left[\left(n^{-1}\sum_{k=1}^n X_k\right)^2 - 2\mu\left(n^{-1}\sum_{k=1}^n X_k\right) + \mu^2\right] = 0 \\
\Longleftrightarrow \quad & \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] - 2\mu^2 + \mu^2 \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

So it suffices to show that

$$\begin{aligned}
&\frac{1}{n^2}\mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] \xrightarrow[n\to\infty]{} \mu^2 \\
\iff\quad &\frac{1}{n^2}\mathbb{E}\left[\sum_{k=1}^{n} X_k^2 + 2\sum_{1\le i<j\le n} X_i X_j\right] \xrightarrow[n\to\infty]{} \mu^2 \\
\iff\quad &\frac{1}{n^2}\left(\sum_{k=1}^{n} \mathbb{E}\left[X_k^2\right] + 2\sum_{1\le i<j\le n} \mathbb{E}\left[X_i X_j\right]\right) \xrightarrow[n\to\infty]{} \mu^2
\end{aligned}$$

where $\mathbb{E}\left[X_k^2\right] = \mathrm{Var}(X) + \mu^2 = f(0) + \mu^2$ and $\mathbb{E}\left[X_i X_j\right] = \mathrm{Cov}(X_i, X_j) + \mu^2 = f(|i - j|) + \mu^2$. Hence, one has

$$\begin{aligned}
&\frac{1}{n^2}\left(\sum_{k=1}^{n} \mathbb{E}\left[X_k^2\right] + 2\sum_{1\le i<j\le n} \mathbb{E}\left[X_i X_j\right]\right) \\
&= \frac{1}{n^2}\left(n(f(0) + \mu^2) + 2\sum_{i=1}^{n-1}(n-i)(f(i) + \mu^2)\right) \\
&= \frac{1}{n^2}\left(nf(0) + 2\sum_{i=1}^{n-1}(n-i)f(i) + n\mu^2 + 2\sum_{i=1}^{n-1}(n-i)\mu^2\right) \\
&= \frac{f(0)}{n} + 2\sum_{i=1}^{n-1}\frac{(n-i)}{n}\frac{f(i)}{n} + \mu^2
\end{aligned}$$

Note that $f(i) \to 0$ when $i \to \infty$, so that for every $\varepsilon > 0$, there is a $k < \infty$ s.t. $|f(j)| < \varepsilon, \forall j > k$, so that

$$\sum_{i=1}^{n-1}\frac{(n-i)}{n}\frac{f(i)}{n} \le \sum_{i=1}^{k}\frac{f(i)}{n} + \sum_{j=k+1}^{n-1}\frac{\varepsilon}{n} \le \sum_{i=1}^{k}\frac{f(i)}{n} + \varepsilon \xrightarrow[n\to\infty]{k<\infty} 0$$

Therefore,

$$\frac{1}{n^2}\mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] = \frac{f(0)}{n} + 2\sum_{i=1}^{n-1}\frac{(n-i)}{n}\frac{f(i)}{n} + \mu^2 \xrightarrow[n\to\infty]{} \mu^2$$

as desired. $\square$

9. (*Erdős-Kac*) The goal of this exercise is to prove that if $w(m)$ denotes the number of distinct prime factors of m and k is a random variable uniformly distributed on $[\![1, n]\!]$, then the following convergence in distribution holds:

$$\frac{w(k) - \log\log n}{\sqrt{\log\log n}} \xrightarrow[n\to\infty]{} \mathcal{N}(0,1)$$

(a) Prove that if $(X_n)_{n\geq 1}$ converges in distribution to $\mathcal{N}(0,1)$ and $\sup_{n\geq 1}\mathbb{E}[X_n^{2k}] < \infty$ for any $k \in \mathbb{N}$, then

$$\lim_{n\to\infty} \mathbb{E}\left[X_n^k\right] = \mathbb{E}\left[\mathcal{N}(0,1)^k\right]$$

for any $k \in \mathbb{N}$.

Proof. Consider $k = 1$. Since $\sup_n \mathbb{E}[|X_n|^{1+\varepsilon}] < \infty$ where here we have $\varepsilon = 1$, $(X_n)_n$ is uniformly integrable. So that one has

$$\lim_{\alpha\to\infty} \sup_n \int_{|X_n|>\alpha} |X_n| \,\mathrm{d}\mathbb{P} = \lim_{\alpha\to\infty} \sup_n \mathbb{E}[|X_n| 1_{|X_n|>\alpha}] = 0$$

Then, X is integrable and

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] = \mathbb{E}\left[\mathcal{N}(0,1)\right]$$

Comment. *The first equality may not hold since you may not have* $\mathbb{E}\left[\lim_{n\to\infty} X_n\right]$ *at all. They may lie in different probability spaces.*

For $k > 1$, following the continuous mapping theorem, one has $(X_n^k)_{n\geq 1}$ converges in distribution to $\mathcal{N}(0,1)^k$, and then it follows similarly, as $\sup_{n\geq 1}\mathbb{E}[X_n^{2k}] < \infty$, from above. $\square$

(b) Prove that for any $x \in \mathbb{R}$ and $d \geq 1$ we have

$$\left| e^{\mathrm{i}x} - \sum_{\ell=0}^{d} \frac{(\mathrm{i}x)^\ell}{\ell!} \right| \leq \frac{|x|^{d+1}}{(d+1)!}$$

Proof. We first show by induction that

$$e^{\mathrm{i}x} = \sum_{\ell=0}^{d} \left[\frac{(\mathrm{i}x)^\ell}{\ell!}\right] + \frac{(\mathrm{i}x)^{d+1}}{d!} \int_0^1 (1-u)^d e^{\mathrm{i}ux} \,\mathrm{d}u \tag{2.8}$$

By *fundamental theorem of calculus*, one has $e^{\mathrm{i}x} = 1 + (\mathrm{i}x)\int_0^1 e^{\mathrm{i}ux}\,\mathrm{d}u$, so it is true for $d = 0$. Assume inductively that 2.8 is true for $d-1, d \geq 1$. We show that it is true for d. Integrating by parts with

$$\begin{aligned} U &= e^{\mathrm{i}ux}, & V &= -\frac{(1-u)^d}{d}, \\ \mathrm{d}U &= \mathrm{i}xe^{\mathrm{i}ux}\,\mathrm{d}u, & \mathrm{d}V &= (1-u)^{d-1}\,\mathrm{d}u, \end{aligned}$$

gives

$$\begin{aligned} &\frac{(\mathrm{i}x)^d}{(d-1)!}\int_0^1 (1-u)^d e^{\mathrm{i}ux}\,\mathrm{d}u \\ &= \frac{(\mathrm{i}x)^d}{(d-1)!}\left[-e^{\mathrm{i}ux}\frac{(1-u)^d}{d}\bigg|_{u=0}^{u=1} + \frac{(\mathrm{i}x)}{d}\int_0^1 (1-u)^d e^{\mathrm{i}ux}\,\mathrm{d}u\right] \\ &= \frac{(\mathrm{i}x)^d}{d!} + \frac{(\mathrm{i}x)^{d+1}}{d!}\int_0^1 (1-u)^d e^{\mathrm{i}ux}\,\mathrm{d}u \end{aligned}$$

Hence, one has

$$\begin{aligned} e^{\mathrm{i}x} &= \sum_{\ell=0}^{d-1}\left[\frac{(\mathrm{i}x)^\ell}{\ell!}\right] + \frac{(\mathrm{i}x)^d}{(d-1)!}\int_0^1 (1-u)^{d-1}e^{\mathrm{i}ux}\,\mathrm{d}u \\ &= \sum_{\ell=0}^{d-1}\left[\frac{(\mathrm{i}x)^\ell}{\ell!}\right] + \frac{(\mathrm{i}x)^d}{d!} + \frac{(\mathrm{i}x)^{d+1}}{d!}\int_0^1 (1-u)^d e^{\mathrm{i}ux}\,\mathrm{d}u \\ &= \sum_{\ell=0}^{d}\left[\frac{(\mathrm{i}x)^\ell}{\ell!}\right] + \frac{(\mathrm{i}x)^{d+1}}{d!}\int_0^1 (1-u)^d e^{\mathrm{i}ux}\,\mathrm{d}u \end{aligned}$$

which completes the inductive step.

Now, it follows that

$$\begin{aligned} \left|e^{\mathrm{i}x} - \sum_{\ell=0}^{d}\frac{(\mathrm{i}x)^\ell}{\ell!}\right| &= \left|\frac{(\mathrm{i}x)^{d+1}}{d!}\int_0^1 (1-u)^d e^{\mathrm{i}ux}\,\mathrm{d}u\right| \\ &= \left|\frac{(\mathrm{i}x)^{d+1}}{d!}\right| \cdot \left|\int_0^1 (1-u)^d e^{\mathrm{i}ux}\,\mathrm{d}u\right| \\ &\leq \frac{|x|^{d+1}}{d!}\cdot\frac{1}{d+1} = \frac{|x|^{d+1}}{(d+1)!} \end{aligned}$$

as desired. $\square$

(c) Assume that

$$\lim_{n\to\infty} \mathbb{E}\left[X_n^k\right] = \mathbb{E}\left[\mathcal{N}(0,1)^k\right]$$

for any $k \in \mathbb{N}$. Prove that X_n converges in distribution to X.

Proof. Let $\alpha_k := \mathbb{E}[X^k] = \int_{\mathbb{R}} x^k \mu(\mathrm{d}x)$. It suffices to show that the probability measure μ is unique with the moments $\alpha_1, \alpha_2, \ldots$, since then the distribution of the convergence of X_n is uniquely determined to be $X \sim \mathcal{N}(0,1)$.

Note that for a standard normal, its moments are $0, 1!!, 0, 3!!, 0, 5!!, 0, 7!!, \ldots$, so $\alpha_k \leq k!$ is finite of all orders, implying that $\alpha_k s^k/k! \to 0$ for some positive s. Let $\beta_k = \int_{-\infty}^{\infty} |x|^k \mu(dx)$ be the absolute moments. We first show that

$$\frac{\beta_k r^k}{k!} \xrightarrow[k\to\infty]{} 0 \tag{2.9}$$

for some positive r. Choose $0 < r < s$. Since $\alpha_k s^k/k! \to 0$, one has $2kr^{2k-1} < s^{2k}$ for large k. Since $|x|^{2k-1} \leq 1 + |x|^{2k}$,

$$\frac{\beta_{2k-1} r^{2k-1}}{(2k-1)!} \leq \frac{r^{2k-1}}{(2k-1)!} + \frac{\beta_{2k} s^{2k}}{(2k)!}$$

for large k. Hence 2.9 holds as k goes to infinity through odd values, and $\beta_k = \alpha_k$ for k even, so it holds for all k.

From part 9b, one has

$$\left| e^{\mathrm{i}tx} \left(e^{\mathrm{i}hx} - \sum_{k=0}^{n} \frac{(\mathrm{i}hx)^k}{k!} \right) \right| \leq \frac{|hx|^{n+1}}{(n+1)!},$$

and therefore the characteristic function φ of μ satisfies

$$\left| \varphi(t+h) - \sum_{k=0}^{n} \frac{h^k}{k!} \int_{-\infty}^{\infty} (\mathrm{i}x)^k e^{\mathrm{i}tx} \mu(dx) \right| \leq \frac{|h|^{n+1} \beta_{n+1}}{(n+1)!}.$$

Observe that the integral in the above equation is $\varphi^{(k)}(t)$, the k-th derivative of φ. By 2.9,

$$\varphi(t+h) = \sum_{k=0}^{\infty} \frac{\varphi^{(k)}(t)}{k!} h^k, \quad |h| \leq r.$$

If ν is another probability measure with moments α_k and characteristic function $\psi(t)$, the same argument gives

$$\psi(t+h) = \sum_{k=0}^{\infty} \frac{\psi^{(k)}(t)}{k!} h^k, \quad |h| \leq r.$$

Take $t = 0$; since $\varphi^{(k)}(0) = i^k \alpha_k = \psi^{(k)}(0)$, φ and ψ agree in $(-r, r)$ and hence have identical derivatives there. Taking $t = r - \varepsilon$ and $t = -r + \varepsilon$ in the above two equations shows that φ and ψ also agree in $(-2r + \varepsilon, 2r - \varepsilon)$ and hence in $(-2r, 2r)$. But then they must by the same argument agree in $(-3r, 3r)$ as well, and so on. Thus φ and ψ coincide, and by the uniqueness theorem for characteristic functions, so do μ and ν. □

(d) Let $w_y(m)$ be the number of prime factors of m which are smaller than y. Let $(B_p)_{p \text{ prime}}$ be independent random variables such that $\mathbb{P}(B_p = 1) = 1 - \mathbb{P}(B_p = 0) = 1/p$. Denote

$$W_y = \sum_{p \le y} B_p, \quad \mu_y = \sum_{p \le y} \frac{1}{p}, \quad \sigma_y^2 = \sum_{p \le y} \left(\frac{1}{p} - \frac{1}{p^2} \right)$$

Prove that if $y = n^{o(1)}$, then for any $d \in \mathbb{N}$ we have

$$\mathbb{E}\left[\left(\frac{w_y(k) - \mu_y}{\sigma_y} \right)^d \right] - \mathbb{E}\left[\left(\frac{W_y - \mu_y}{\sigma_y} \right)^d \right] \xrightarrow[n \to \infty]{} 0$$

Proof. Let

$$\delta_p(m) = \begin{cases} 1 & p \mid m \\ 0 & p \nmid m \end{cases}$$

Then,

$$w_y(m) = \sum_{p \le y} \delta_p(m)$$

Note that $\mathbb{E}\left[W_y^d\right]$ is the sum

$$\sum_{u=1}^{d} {\sum}' \frac{d!}{d_1! \cdots d_u!} \frac{1}{u!} {\sum}'' \mathbb{E}\left[B_{p_1}^{d_1} \cdots B_{p_u}^{d_u} \right], \tag{2.10}$$

where $\sum'$ extends over the u tuples $(d_1, \ldots, d_u)$ of positive integers satisfying $d_1 + \ldots + d_u = d$, and $\sum''$ extends over the u tuples $(p_1, \ldots, p_u)$ of distinct primes not exceeding y. Since B_p assumes only the values 0 and 1, from the independence of the B_p and the fact that the p_i are distinct, it follows that the summand in 2.10 is

$$\mathbb{E}\left[B_{p_1} \cdots B_{p_u}\right] = \frac{1}{p_1 \cdots p_u} \tag{2.11}$$

By definition, $\mathbb{E}_n\left[w_y^d\right]$[16] is just 2.10 with the summand replaced by $\mathbb{E}_n\left[\delta_{p_1}^{d_1} \cdots \delta_{p_u}^{d_u}\right]$. Since $\delta_p(m)$ assumes only the values 0 and 1, given that the p_i are distinct, it follows that this

[16] $\mathbb{E}_n$ here means

$$\mathbb{E}_n[f] = n^{-1} \sum_{m=1}^{n} f(m),$$

and we assume that $\mathbb{E}\left[\left(\frac{w_y(k) - \mu_y}{\sigma_y} \right)^d \right]$ is calculated in this way.

summand is

$$\mathbb{E}_n\left[\delta_{p_1}\cdots\delta_{p_u}\right] = \frac{1}{n}\left\lfloor \frac{n}{p_1\cdots p_u}\right\rfloor \tag{2.12}$$

But 2.11 and 2.12 differ by at most $1/n$, and hence $\mathbb{E}\left[W_y^d\right]$ and $\mathbb{E}_n\left[w_y^d\right]$ differ by at most the sum 2.10 with the summand replaced by $1/n$. Therefore,

$$\left|\mathbb{E}\left[W_y^d\right] - \mathbb{E}_n\left[w_y^d\right]\right| \leq \frac{1}{n}\left(\sum_{p\leq y} 1\right)^d \leq \frac{y^d}{n} \tag{2.13}$$

Now

$$\mathbb{E}\left[(W_y-\mu_y)^d\right] = \sum_{j=0}^{d}\binom{d}{j}\mathbb{E}\left[W_y^j\right](-\mu_y)^{d-j},$$

and $\mathbb{E}_n\left[(w_y-\mu_y)^d\right]$ has the analogous expansion. Comparing the two expansions term for term and applying 2.13 shows that

$$\left|\mathbb{E}\left[(W_y-\mu_y)^d\right] - \mathbb{E}_n\left[(w_y-\mu_y)^d\right]\right| \leq \sum_{j=0}^{d}\binom{d}{j}\frac{y^j}{n}\mu_y^{d-j} = \frac{1}{n}(y+\mu_y)^d$$

Since $\mu_y \leq y$, and since $y^d/n \to 0$ by the assumption $y = n^{o(1)}$, one has

$$\mathbb{E}\left[\left(\frac{w_y(k)-\mu_y}{\sigma_y}\right)^d\right] - \mathbb{E}\left[\left(\frac{W_y-\mu_y}{\sigma_y}\right)^d\right] \xrightarrow[n\to\infty]{} 0$$

as desired. □

(e) Conclude.

Proof. We first want to show that

$$\frac{w(k)-\log\log n}{\sqrt{\log\log n}} \xrightarrow[n\to\infty]{} \mathcal{N}(0,1)$$

is unaffected if the range of p is further restricted with $w_y(m)$.

Proof. By *Mertens' second theorem*, one has the estimate that

$$\mu_y = \sum_{p\leq y}\frac{1}{p} = \log\log y + \mathrm{O}(1)$$

This satisfies (for example let $\log y = \log n/\log\log n$) that $y\to\infty$ slowly enough that $\log y/\log n \to 0$ but fast enough that

$$\sum_{y<p\leq n}\frac{1}{p} = \mathrm{o}(\log\log n)^{1/2} \tag{2.14}$$

Now, let $\mathbb{P}_n$ be the probability measure that places mass $1/n$ at each of $1, 2, \ldots, n$. Recall that δ_p is defined as 1 or 0 according as the prime p divides m or not. Note that if $p_1, \ldots, p_u$ are distinct primes, then $\forall i, p_i \mid m$ iff $\prod_i p_i \mid m$, so that one has

$$\mathbb{P}_n\left[m : \delta_{p_i}(m) = 1, \forall i \in [\![1, u]\!]\right] = \frac{1}{n}\left\lfloor \frac{n}{\prod_{i=1}^{u} p_i} \right\rfloor$$

In particular when $u = 1$,

$$\mathbb{E}_n\left[\sum_{p>y} \delta_p\right] = \sum_{y<p\leq n} \mathbb{P}_n\left[m : \delta_p(m) = 1\right] \leq \sum_{y<p\leq n} \frac{1}{p}$$

By 2.14 and Markov's inequality,

$$\mathbb{P}_n\left[m : |w(m) - w_y(m)| \geq \varepsilon(\log\log n)^{1/2}\right] \to 0$$

Therefore the desired result is unaffected if $w_y(m)$ is substituted for $w(m)$. □

Now compare $w_y(m)$ with the corresponding sum $W_y = \sum_{p\leq y} B_p$. The mean and variance of S_n are

$$\mu_y = \sum_{p\leq y} \frac{1}{p}, \quad \sigma_y^2 = \sum_{p\leq y} \frac{1}{p}\left(1 - \frac{1}{p}\right),$$

and each is $\log\log n + o(\log\log n)^{1/2}$ by 2.14. Thus, it suffices to show that

$$\mathbb{P}_n\left[m : \frac{w_y(m) - \mu_y}{\sigma_y} \leq x\right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du.$$

Since the B_p are bounded, following part 9c, it suffices to show the moments converge to $\mathcal{N}(0,1)$. Note that B_p should be replaced by $B_p - p^{-1}$ to center it. Thus the d-th moment of $(W_y - \mu_y)/\sigma_y$ converges to that of $\mathcal{N}(0,1)$. However, it is already proved in part 9d that, as $n \to \infty$, one has

$$\mathbb{E}\left[\left(\frac{w_y(k) - \mu_y}{\sigma_y}\right)^d\right] - \mathbb{E}\left[\left(\frac{W_y - \mu_y}{\sigma_y}\right)^d\right] \xrightarrow[n\to\infty]{} 0,$$

so we are done. □

2.4 Dependent Random Variables

2.4.1 Conditioning, Radon-Nikodym Theorem

1. Let X and Y be independent Gaussian random variables with null expectation and variance 1. Show that $\frac{X+Y}{\sqrt{2}}$ and $\frac{X-Y}{\sqrt{2}}$ are also independent $\mathcal{N}(0,1)$.

Proof. Note that $X \sim \mathcal{N}(0,1) \implies X/\sqrt{2} \sim \mathcal{N}(0,1/2)$ and similar for Y. Since the characteristic function uniquely determines the distribution, it suffices to show that $\frac{X+Y}{\sqrt{2}}$ and $\frac{X-Y}{\sqrt{2}}$ both have[17] $\varphi(t) = e^{-t^2/2}$. Then,

$$\begin{aligned}
\varphi_{\frac{X+Y}{\sqrt{2}}}(t) &= \varphi_{\frac{X}{\sqrt{2}}}(t)\varphi_{\frac{Y}{\sqrt{2}}}(t) \qquad \text{since } X \perp\!\!\!\perp Y \\
&= \exp\left(it\mu_{\frac{X}{\sqrt{2}}} - \frac{\sigma^2_{\frac{X}{\sqrt{2}}} t^2}{2}\right) \exp\left(it\mu_{\frac{Y}{\sqrt{2}}} - \frac{\sigma^2_{\frac{Y}{\sqrt{2}}} t^2}{2}\right) \\
&= \exp\left(-\frac{(1/2+1/2)t^2}{2}\right) = e^{-t^2/2}
\end{aligned}$$

and since $Y/\sqrt{2}$ is symmetric by 0, we have $Y/\sqrt{2} \sim -Y/\sqrt{2} \sim \mathcal{N}(0,1/2)$:

$$\begin{aligned}
\varphi_{\frac{X-Y}{\sqrt{2}}}(t) &= \varphi_{\frac{X}{\sqrt{2}}}(t)\varphi_{\frac{-Y}{\sqrt{2}}}(t) \qquad \text{since } X \perp\!\!\!\perp Y \\
&= \exp\left(it\mu_{\frac{X}{\sqrt{2}}} - \frac{\sigma^2_{\frac{X}{\sqrt{2}}} t^2}{2}\right) \exp\left(it\mu_{\frac{-Y}{\sqrt{2}}} - \frac{\sigma^2_{\frac{-Y}{\sqrt{2}}} t^2}{2}\right) \\
&= \exp\left(-\frac{(1/2+1/2)t^2}{2}\right) = e^{-t^2/2}
\end{aligned}$$

Hence, both $U := \frac{X+Y}{\sqrt{2}}$ and $V := \frac{X-Y}{\sqrt{2}}$ follow $\mathcal{N}(0,1)$. Now, to show that they are independent, recall the density of a bivariate normal distribution (Z,W) is

$$f_{Z,W}(z,w) = \frac{1}{2\pi\sigma_Z\sigma_W\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{z-\mu_Z}{\sigma_Z}\right)^2 - 2\rho\left(\frac{z-\mu_Z}{\sigma_Z}\right)\left(\frac{w-\mu_W}{\sigma_W}\right) + \left(\frac{w-\mu_W}{\sigma_W}\right)^2\right]}$$

where $\rho_{U,V} = \frac{\mathrm{Cov}(U,V)}{\sigma_U\sigma_V} = \mathbb{E}[UV] = (\mathbb{E}[X^2] - \mathbb{E}[Y^2])/2 = 0$, so

$$f_{U,V}(u,v) = \frac{1}{2\pi}e^{-\frac{1}{2}(u^2+v^2)} = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}} = f_U(u)f_V(v)$$

implying that $U \perp\!\!\!\perp V$. □

[17] From the fact that for $\mathcal{N}(\mu,\sigma^2)$ we have

$$\begin{aligned}
\varphi(k) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left(x - (\mu + ik\sigma^2)\right)^2 - \frac{k^2\sigma^2}{2} + ik\mu\right\} \mathrm{d}x \\
&= \exp\left\{-\frac{k^2\sigma^2}{2} + ik\mu\right\} \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left(x - (\mu + ik\sigma^2)\right)^2\right\} \mathrm{d}x \\
&= \exp\left\{-\frac{k^2\sigma^2}{2} + ik\mu\right\}
\end{aligned}$$

2. Let $(X_n)_{n\geq 0}$ be a sequence of i.i.d random variables, with uniform distribution on $[0,1]$. Let $Y_n = (X_n)^n$.

   (a) Calculate the distribution of Y_n.

   Solution. When $n=0$, $Y_0 = 1$ a constant. When $n \geq 1$, note that, for $t \geq 0$, the CDF of Y_n is

$$F_{Y_n}(t) = \mathbb{P}\{Y_n \leq t\} = \mathbb{P}\left(X_n \leq t^{1/n}\right) = \begin{cases} t^{1/n} & \text{if } 0 \leq t \leq 1 \\ 1 & \text{if } t \geq 1 \end{cases}$$

   and $F_{Y_n}(t) = 0$ if $t \leq 0$. Moreover, we can get the density

$$f_{Y_n}(t) = \begin{cases} \frac{1}{n} t^{\frac{1}{n}-1} & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise.} \end{cases} \qquad n = 1, 2, \ldots$$

   □

   (b) Show that $(Y_n)_{n\geq 0}$ converges to 0 in probability.

   Proof. We show that $\lim_{n\to\infty} \mathbb{P}\big(|Y_n - 0| > \varepsilon\big) = \lim_{n\to\infty} \mathbb{P}\big(Y_n > \varepsilon\big) = 0$ for all $\varepsilon > 0$. It is obvious for $\varepsilon \geq 1$, so we set $0 < \varepsilon < 1$. Then,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\big(Y_n > \varepsilon\big) &= \lim_{n\to\infty} \big(1 - \mathbb{P}\big(Y_n \leq \varepsilon\big)\big) \\ &= 1 - \lim_{n\to\infty} \varepsilon^{1/n} = 1 - 1 = 0\end{aligned}$$

   Hence, $(Y_n)_{n\geq 0}$ converges to 0 in probability. □

   (c) Show that $(Y_n)_{n\geq 0}$ converges in L^1.

   Proof. We show that Y_n converges in mean towards 0, i.e.

$$\lim_{n\to\infty} \mathbb{E}\left(|Y_n - 0|\right) = \lim_{n\to\infty} \mathbb{E}\left(Y_n\right) = 0$$

   One has

$$\lim_{n\to\infty} \mathbb{E}\left(Y_n\right) = \lim_{n\to\infty} \int_0^1 f_{Y_n}(t) t \,\mathrm{d}t = \lim_{n\to\infty} \frac{1}{n+1} = 0$$

   Hence, Y_n converges in mean towards 0. □

   (d) Show that almost surely $(Y_n)_{n\geq 0}$ does not converge.

   Proof. Suppose $(Y_n)_{n\geq 0}$ converges a.s. to Y, then $Y = 0$, since if not, $(Y_n)_{n\geq 0}$ will not converge to 0 in probability, which contradicts to part 2b. Then, by definition, one has

$$\mathbb{P}\Big(\limsup_{n\to\infty} \big\{\omega \in \Omega : |Y_n(\omega)| > \varepsilon\big\}\Big) = 0 \quad \text{for all } \varepsilon > 0$$

Note that X_n are by definition pairwise independent, and so are Y_n. Let $A_n := \{\omega \in \Omega : |Y_n(\omega)| > \varepsilon\}$ taking $0 < \varepsilon < 1$ be a sequence of independent events. Then

$$\sum_{n=0}^{\infty} \mathbb{P}(A_n) = 1 + \sum_{n=1}^{\infty} \left(1 - \varepsilon^{1/n}\right) = \infty^{18}$$

By the second Borel-Cantelli, one has

$$\mathbb{P}\Big(\limsup_{n\to\infty} \big\{\omega \in \Omega : |Y_n(\omega)| > \varepsilon\big\}\Big) = 1 \quad \forall\, 0 < \varepsilon < 1$$

which contradicts with above. Hence, $(Y_n)_{n\geq 0}$ does not converge a.s. □

3. Let $(X_n)_{n\geq 1}$ be i.i.d. Bernoulli random variables with parameter $p \in (0,1)$, i.e. $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p$. Let N be a Poisson random variable with parameter $\lambda > 0$, i.e. for any $k \geq 0$ we have $\mathbb{P}(N = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. Assume N is independent from $(X_n)_{n\geq 1}$. Let $P = \sum_{i=1}^{N} X_i$, $F = N - P$.

   (a) What is the joint distribution of (P, N)?

   Solution. Note that the sum of N i.i.d. $\mathrm{Ber}(p)$'s follows $\mathrm{Bin}(N, p)$. By definition, the PMF of the distribution of (P, N) is

$$\begin{aligned} p_{P,N}(k,n) &= \mathbb{P}\left(P = k, N = n\right) \\ &= \mathbb{P}\left(P = k \mid N = n\right)\mathbb{P}\left(N = n\right) \\ &= \binom{n}{k} p^k \left(1-p\right)^{n-k} \cdot e^{-\lambda}\frac{\lambda^n}{n!} \end{aligned}$$

   where $0 \leq k \leq n$, $k \in \mathbb{Z}$, $n = 0, 1, 2, \ldots$, and $p_{P,N}(k,n) = 0$ otherwise. □

   (b) Prove that P and F are independent.

   Proof. Following part 3a, one has

$$\begin{aligned} p_{P,N}(k,n) &= \mathbb{P}\left(P = k, N = n\right) \\ &= \mathbb{P}\left(P = k, F = n-k\right) = p_{P,F}(k, n-k) \\ &= \frac{1}{k!} e^{-\lambda p}\left(\lambda p\right)^k \cdot \frac{1}{(n-k)!} e^{-\lambda(1-p)}\left(\lambda\left(1-p\right)\right)^{n-k} \\ &= p_{\mathrm{Pois}\,(\lambda p)}(k) \cdot p_{\mathrm{Pois}\,(\lambda(1-p))}(n-k) \end{aligned}$$

[18] This is because, for $f(x) = 1 - x^{1/n}$, we have $f'(1) = -1/n$ which is the tangent line at $x = 1$, so that $1 - x^{1/n} > (1-x)/n \implies$

$$\sum_{n=1}^{\infty}\left(1 - \varepsilon^{1/n}\right) > \left(\sum \frac{1}{n}\right)(1-x) = \infty$$

where $k = 0, 1, 2, \ldots, n-k = 0, 1, 2, \ldots$ and $p_{P,F}(k, n-k) = 0$ otherwise. Note that the value of k and $n-k$ do not depend on each other, i.e. $\mathbb{1}_{0\le k\le n} = \mathbb{1}_{k\ge 0} \cdot \mathbb{1}_{n-k\ge 0}$.

We check that $P \sim \operatorname{Pois}(\lambda p)$:

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{N} X_i = k\right) &= \sum_{l=k}^{\infty} \mathbb{P}\left((N = l) \cap \left(\sum_{i=1}^{l} X_i = k\right)\right) \\
&= \sum_{l=k}^{\infty} \mathbb{P}(N = l)\, \mathbb{P}\left(\sum_{i=1}^{l} X_i = k\right) \quad \text{since } N \perp\!\!\!\perp X_i, \forall i \\
&= \sum_{l=k}^{\infty} e^{-\lambda} \frac{\lambda^l}{l!} \binom{l}{k} p^k (1-p)^{l-k} \\
&= \frac{e^{-\lambda} p^k \lambda^k}{k!} \sum_{l=k}^{\infty} \frac{\lambda^{l-k}(1-p)^{l-k}}{(l-k)!} \\
&= \frac{e^{-\lambda} p^k \lambda^k}{k!} e^{\lambda(1-p)} = \frac{e^{-\lambda p}(\lambda p)^k}{k!}
\end{aligned}$$

where $k = 0, 1, 2, \ldots$ and $\mathbb{P}(P = k) = 0$ otherwise.

Also, check that $F \sim \operatorname{Pois}(\lambda(1-p))$:

$$\begin{aligned}
\mathbb{P}(N - P = f) &= \sum_{n-k=f} \mathbb{P}(P = k, N = n) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(P = k, N = k + f) \\
&= \sum_{k=0}^{\infty} \binom{k+f}{k} p^k (1-p)^f \cdot e^{-\lambda} \frac{\lambda^{k+f}}{(k+f)!} \\
&= \frac{\lambda^f (1-p)^f e^{-\lambda}}{f!} \sum_{k=0}^{\infty} p^k \frac{\lambda^k}{k!} \\
&= \frac{\lambda^f (1-p)^f e^{-\lambda}}{f!} e^{\lambda p} = \frac{e^{-\lambda(1-p)}(\lambda(1-p))^f}{f!}
\end{aligned}$$

where $f = 0, 1, 2, \ldots$ and $\mathbb{P}(F = f) = 0$ otherwise.

Hence, $p_{P,F}(k, n-k) = p_P(k) \cdot p_F(n-k)$, implying that $P \perp\!\!\!\perp F$. □

4. (*the number of buses stopping till time t*) Let $(X_n)_{n\ge 1}$ be i.i.d, random variables on $(\Omega, \mathcal{A}, \mathbb{P})$, X_1 being an exponential random variable with parameter 1. Define $T_0 = 0$, $T_n = X_1 + \cdots + X_n$, and for any $t > 0$,
$$N_t = \max\{n \ge 0 \mid T_n \le t\}$$

(a) For any $n \geq 1$, calculate the joint distribution of $(T_1, \ldots, T_n)$.

(b) Deduce the distribution of N_t, for arbitrary t.

Solution. (a) Note that $T_1 = X_1$ follows $\text{Exp}(1)$, i.e. $f_{T_1}(t) = e^{-t}\mathbb{1}_{[0,\infty)}(t)$. Since $X_1 \perp\!\!\!\perp X_2$, one has, for $0 \leq t_1 \leq t_2$,

$$f_{T_1,T_2}(t_1,t_2) = f_{T_1}(t_1)f_{X_2}(t_2 - t_1) = e^{-t_1} \cdot e^{-(t_2-t_1)} = e^{-t_2}$$

Inductively, one has, for $0 \leq t_1 \leq t_2 \ldots \leq t_n$,

$$f_{T_1,T_2,\ldots,T_n}(t_1,t_2,\ldots,t_n) = f_{T_1,T_2,\ldots,T_{n-1}}(t_1,t_2,\ldots,t_{n-1})f_{X_n}(t_n - t_{n-1}) = e^{-t_n}$$

Hence, the joint distribution of $(T_1, \ldots, T_n)$ has PDF

$$f_{T_1,T_2,\ldots,T_n}(t_1,t_2,\ldots,t_n) = \begin{cases} e^{-t_n} & 0 \leq t_1 \leq t_2 \ldots \leq t_n \\ 0 & \text{otherwise} \end{cases}$$

(b) Recall that the sum of i.i.d exponential random variables follow an Erlang distribution with parameter $\lambda = 1$ and n: $f_{T_n}(t) = \frac{t^{n-1}e^{-t}}{(n-1)!}$, so that, for $0 \leq t_1 \leq \cdots \leq t_n \leq t$,

$$\begin{aligned} f_{T_1\cdots T_n|T_{n+1}}(t_1,\cdots,t_n|t) &= \frac{f_{T_1,\cdots,T_{n+1}}(t_1,\cdots,t_n,t)}{f_{T_{n+1}}(t)} \\ &= \frac{e^{-t}}{t^n e^{-t}/n!} \\ &= n!\left(\frac{1}{t}\right)^n \end{aligned}$$

which is the PDF of the order statistics $(U_{(1)},\cdots,U_{(n)})$ where $U_1,\cdots,U_n \overset{\text{iid}}{\sim} \text{Unif}(0,t)$. Therefore, one has

$$\begin{aligned} \mathbb{P}(N_t = n) &= \mathbb{P}(T_n \leq t, T_{n+1} > t) \\ &= \int_t^\infty \mathbb{P}(T_n \leq t \mid T_{n+1} = u) f_{T_{n+1}}(u)\,\mathrm{d}u \\ &= \int_t^\infty \mathbb{P}(U_{(n)} \leq t) f_{T_{n+1}}(u)\,\mathrm{d}u \end{aligned}$$

where $U_{(n)}$ is the n-th maximum of a sample $(U_1,\cdots,U_n)$ of size n from $\text{Unif}(0,u)$. As $0 < t < u$, one has $\mathbb{P}(U_{(n)} \leq t) = (t/u)^n$, so that

$$\begin{aligned} \mathbb{P}(N_t = n) &= \int_t^\infty \left(\frac{t}{u}\right)^n \frac{1}{n!}u^n e^{-u}\,\mathrm{d}u \\ &= \frac{t^n}{n!}\int_t^\infty e^{-u}du \\ &= \frac{t^n}{n!}e^{-t} \end{aligned}$$

where $n = 1, 2, \dots$ And a trivial case for $n = 0$ is that $\mathbb{P}(N_t = 0) = \mathbb{P}(T_0 = 0 \le t, T_1 = X_1 > t) = e^{-t}$, which also satisfies the equation above.

$\square$

5. Let $(X_n)_{n\ge 0}$ be real, independent, random variables on $(\Omega, \mathcal{A}, \mathbb{P})$.

   (a) Prove that the radius of convergence R of the random series $\sum_{n\ge 0} X_n z^n$ is almost surely constant.

   Proof. By definition that $R := \sup\{z : \sum |X_n| z^n < \infty\} = \left(\limsup_{n\to\infty} |X_n|^{1/n}\right)^{-1}$, it suffices to show that $Y := \limsup_{n\to\infty} |X_n|^{1/n}$ is almost surely constant. Now, for each n, let $\mathcal{F}_n$ be the σ-algebra generated by $(X_k)_{k\ge n}$, and let $\mathcal{F}_\infty = \bigcap_n \mathcal{F}_n$. Then Y is $\mathcal{F}_\infty$-measurable. By Kolmogorov's zero-one law, one has $\mathbb{P}(A) = 0$ or 1, $\forall A \in \mathcal{F}_\infty$. Hence, Y is a.s. constant, and so is R. $\square$

   (b) Assume also that the X_n's have the same distribution. Prove that $R = 0$ almost surely if $\mathbb{E}\left[\log(|X_0|)_+\right] = \infty$, and $R \ge 1$ a.s. if $\mathbb{E}\left[\log(|X_0|)_+\right] < \infty$.

   Proof. If $\mathbb{E}\left[\log(|X_0|)_+\right] = \infty$, then $\forall \varepsilon < \infty$, one has $\sum_{n\ge 1} \mathbb{P}(\log |X_n|_+ > \varepsilon n) = \infty$. Since X_i are i.i.d., following the second Borel-Cantelli, one has $\mathbb{P}(|X_n| > e^{\varepsilon n} \text{ i.o.}) = 1$. Hence, $R := \sup\{z : \sum |X_n| z^n < \infty\} = 0$.

   On the other hand, if $\mathbb{E}\left[\log(|X_0|)_+\right] < \infty$, then $\forall \varepsilon > 0$, one has $\mathbb{E}\left[\frac{\log |X_n|_+}{\varepsilon}\right] < \infty$. Since in general $\mathbb{E}[|X|] = \int_0^\infty \mathbb{P}(|X| \ge t)\,\mathrm{d}t$, one has

   $$\int_0^\infty \mathbb{P}\left(\frac{\log(|X_n|)_+}{\varepsilon} \ge t\right) \mathrm{d}t < \infty$$

   By integral test and non-negativity, one has

   $$\sum_{n=1}^\infty \mathbb{P}\left(\left|\frac{\log(|X_n|)_+}{n}\right| > \varepsilon\right) < \infty$$

   By Borel-Cantelli, one has $\frac{\log(|X_n|)_+}{n} \xrightarrow{a.s.} 0$, implying that $|X_n| \le e^{n\varepsilon}$ for large n. Hence, if $|z| < 1$, by choosing $\varepsilon > 0$ s.t. $e^{\varepsilon}|z| < 1$, we have $|X_n z^n| \le e^{n\varepsilon}|z|^n$ for n large enough, and $\sum e^{n\varepsilon}|z|^n < \infty$ implies that $\sum |X_n z^n|$ converges for $|z| < 1$, so $R \ge 1$[19]. $\square$

6. Prove that there is no probability measure on $\mathbb{N}$ such that for any $n \ge 1$, the probability of the set of multiples of n is $1/n$.

   Proof. Suppose there is, and let it be μ with $\mu(\mathbb{N}) = 1$ on the σ-field $\mathcal{A}$. Suppose further that, for $n = 1, 2, \dots,$

   $$A[n] := n\mathbb{N} = \{0, n, 2n, 3n, \dots\} \in \mathcal{A}, \quad \mu(A[n]) = \frac{1}{n}$$

[19] In fact, one can show that $R = 1$ a.s. by proving $R \le 1$ as well.

Then $\mathbb{N}\backslash A[n] \in \mathcal{A}$ with $\mu(\mathbb{N}\backslash A[n]) = (n-1)/n$. Let $p_1, p_2, \dots, p_k$ be distinct primes, $k \in \mathbb{N}$. One has

$$A[p_1] \cap A[p_2] \cap \dots \cap A[p_k] = A[p_1 p_2 \dots p_k]$$
$$\mu\big(A[p_1] \cap A[p_2] \cap \dots \cap A[p_k]\big) = \frac{1}{p_1 p_2 \dots p_k}$$

Note that $\mu(\mathbb{N}\backslash A[p_i]) = 1 - 1/p_i$, so by inclusion-exclusion, one has

$$\begin{aligned}
&\mu\left(\left(A[p_1] \cup A[p_2] \cup \dots \cup A[p_k]\right)^{\complement}\right) \\
&= \mu(\mathbb{N}) - \sum_{i=1}^{n} \mu(A[p_i]) + \sum_{1 \le i < j \le n} \mu(A[p_i] \cap A[p_j]) - \cdots \\
&= \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)
\end{aligned}$$

Since in general one has (μ has continuity from below)

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

the above result holds also for $k = \infty$. WLOG, let $p_1 = 2, p_2 = 3, p_3 = 5, \dots$ enumerate all the primes in order. Also, let $Q_0 := \{1\}, Q_1 := \{1, 2, 4, \dots\}$ and in general $Q_i := \{n \in \mathbb{N} : n$ not divisible by any primes except $p_1, \dots, p_i\}$. Note that $Q_0 \subset Q_1 \subset Q_2 \subset \dots$ One has

$$\mu(Q_0) = \prod_{j=1}^{\infty} \left(1 - \frac{1}{p_j}\right) = 0, \mu(Q_1) = \prod_{j=2}^{\infty} \left(1 - \frac{1}{p_j}\right) = 0, \dots$$

implying that $\mu(Q_j) = 0$ for any $j \in \mathbb{N}$. Note that $\mathbb{N}$ is the set of natural numbers divisible by all primes, i.e. $\mathbb{N} = \bigcup_{i=1}^{\infty} Q_i$, but following continuity from below, one has

$$\mu(\mathbb{N}) = \mu\left(\bigcup_{i=1}^{\infty} Q_i\right) = \lim_{i \to \infty} \mu(Q_i) = 0$$

contradicting with the fact that $\mu(\mathbb{N}) = \mu(\Omega) = 1$. $\square$

7. Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}, \mathcal{H}$ sub σ-fields of $\mathcal{F}$ such that $\sigma(\mathcal{G}, \mathcal{H}) = \mathcal{F}$. Find counterexamples to the following assertions:

   (a) If $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$ then X and Y are independent.

   (b) If $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{H}] = 0$ then $X = 0$.

(c) If X and Y are independent then so are $\mathbb{E}[X \mid \mathcal{G}]$ and $\mathbb{E}[Y \mid \mathcal{G}]$.

Solution. (a) For example, let X be a discrete random variable s.t. $\mathbb{P}[X = 1] = \mathbb{P}[X = 0] = \mathbb{P}[x = -1] = 1/3$, then $\mathbb{E}[X] = 0$. Also, let $Y := X^2$. They satisfy that $\mathbb{E}[X|Y = 0] = \mathbb{E}[X|Y = 1] = 0$, so that $\mathbb{E}[X|Y] = 0$, but $Y = X^2$ depends on X by assumption.

(b) Let $\Omega = \{1, 2, 3, 4\}$ and $\mathcal{G} = \sigma(\{1, 2\}, \{3, 4\}), \mathcal{H} = \sigma(\{1, 3\}, \{2, 4\})$. Then $\sigma(\mathcal{G}, \mathcal{H}) = \mathcal{F} = 2^\Omega$. Let $\mathbb{P}(\{i\}) = 1/4, i = 1, 2, 3, 4$ and let X be an R.V. s.t. $X(\{1\}) = X(\{4\}) = 1, X(\{2\}) = X(\{3\}) = -1$. Then, one has $\mathbb{E}(X \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{G}) = 0$, but $X \neq 0$ by definition.

(c) Toss a fair coin twice, so that $\Omega = \{HH, HT, TH, TT\}$. Let $A = \{HT, HH\}$ be the event that the first flip is head, $B = \{TH, HH\}$ be the event that the second flip is head, $C = \{HH, TT\}$ being that the two flips have the same result. Then A and B are independent, but $A \mid C$ and $B \mid C$ are not, since knowing C will inform the second flip after the first one. Now, we can construct $X = \mathbb{1}_A, Y = \mathbb{1}_B$ and $\mathcal{G} = \sigma(C) = \{\varnothing, C, C^\complement, \Omega\}$. One has $X \perp\!\!\!\perp Y$ but $\mathbb{E}[X \mid \mathcal{G}] \not\perp\!\!\!\perp \mathbb{E}[Y \mid \mathcal{G}]$.

□

8. Let Y be an integrable random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{G}$ a sub σ-field of $\mathcal{A}$. Show that $|\mathbb{E}(Y \mid \mathcal{G})| \leq \mathbb{E}(|Y| \mid \mathcal{G})$ (almost surely).

Proof. We check that $|x|$ is convex: for any $0 \leq t \leq 1$ and all $x_1, x_2 \in \mathbb{R}$, we have

$$f(tx_1 + (1-t)x_2) = |tx_1 + (1-t)x_2| \leq |tx_1| + |(1-t)x_2| = tf(x_1) + (1-t)f(x_2)$$

Therefore, by Jensen's inequality[20], for $\phi(x) = |x|$, one has $|\mathbb{E}(Y \mid \mathcal{G})| \leq \mathbb{E}(|Y| \mid \mathcal{G})$ almost surely. □

9. Let Y be an integrable random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{G}$ a sub σ-field of $\mathcal{A}$. Suppose that $\mathcal{H} \subset \mathcal{G}$ is a sub σ-field of $\mathcal{G}$. Show that $\mathbb{E}(\mathbb{E}(Y \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(Y \mid \mathcal{H})$ (almost surely).

Proof. Note that by definition, $\mathbb{E}(Y \mid \mathcal{H})$ is $\mathcal{H}$-measurable and

$$\int_H \mathbb{E}(Y \mid \mathcal{H}) \, \mathrm{d}\mathbb{P} = \int_H Y \, \mathrm{d}\mathbb{P} \qquad \text{for all} \quad H \in \mathcal{H}$$

Since $\mathcal{H} \subset \mathcal{G}$, one has

$$\int_H \mathbb{E}(Y \mid \mathcal{G}) \, \mathrm{d}\mathbb{P} = \int_H Y \, \mathrm{d}\mathbb{P} \qquad \text{for all} \quad H \in \mathcal{H}$$

Hence, the above equations imply

$$\int_H \mathbb{E}(Y \mid \mathcal{H}) \, \mathrm{d}\mathbb{P} = \int_H \mathbb{E}(Y \mid \mathcal{G}) \, \mathrm{d}\mathbb{P}$$

[20] Varadhan's notes, page 80.

which by definition means $\mathbb{E}\left(\mathbb{E}(Y \mid \mathcal{G}) \mid \mathcal{H}\right) = \mathbb{E}\left(Y \mid \mathcal{H}\right)$. ☐

10. Let $(X_n)_{n\geq 0}$ be defined on $(\Omega, \mathcal{A}, \mathbb{P})$. Assume this sequence converges in probability (under $\mathbb{P}$) to X. Let $\mathbb{Q}$ be another probability measure on $(\Omega, \mathcal{A})$ assumed to be absolutely continuous w.r.t. $\mathbb{P}$. Prove that $X_n \to X$ in probability under $\mathbb{Q}$.

Proof. Following *Radon-Nikodym Theorem*, there exists a measurable function f s.t. for every $A \in \mathcal{A}$ one has

$$\mathbb{Q}(A) = \int_A f \,\mathrm{d}\mathbb{P}$$

Since $\mathbb{Q}$ is a probability measure, thus finite, $\int f \,\mathrm{d}\mathbb{P} \leq C$ for some $C > 0$. Now, let $\varepsilon > 0$ and $A_n := \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$. Then, for some $c > 0$, one has

$$\begin{aligned}
\mathbb{Q}(A_n) &= \int_{A_n} f \mathrm{d}\mathbb{P} \\
&= \int_{A_n} f \mathbb{1}_{\{f\leq c\}} \mathrm{d}\mathbb{P} + \int_{A_n} f \mathbb{1}_{\{f> c\}} \mathrm{d}\mathbb{P} \\
&\leq c \cdot \mathbb{P}(A_n) + \int_{A_n} f \mathbb{1}_{\{f> c\}} \mathrm{d}\mathbb{P}
\end{aligned}$$

Since X_n converges in probability under $\mathbb{P}$ to X, we have $\mathbb{1}_{A_n} \to 0$ $\mathbb{P}$-almost surely. Following dominated convergence, we have $\int_{A_n} f \mathbb{1}_{\{f>c\}} \mathrm{d}\mathbb{P} \to 0$ when $n \to \infty$. Hence,

$$\lim_{n\to\infty} \mathbb{Q}(A_n) \leq \lim_{n\to\infty} \left(c \cdot \mathbb{P}(A_n) + \int_{A_n} f \mathbb{1}_{\{f>c\}} \right) = c \lim_{n\to\infty} \mathbb{P}(A_n) = 0,$$

implying that $X_n \to X$ in probability under $\mathbb{Q}$. ☐

2.4.2 Conditional Expectation

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(A_n)_{n\geq 1}$ be a sequence of independent events. We denote $a_n = \mathbb{P}(A_n)$ and define $b_n = a_1 + \cdots + a_n$, $S_n = \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}$. Assuming $b_n \to \infty$ as $n \to \infty$, prove that S_n/b_n converges almost surely.

Proof. Let $Q_n = S_n/b_n$. *Step 1.* We start by proving $Q_n \to 1$ in probability. For every $\epsilon > 0$, Chebyshev's inequality gives that

$$\mathbb{P}\left(|Q_n - 1| > \epsilon\right) = \mathbb{P}\left(\left| \frac{S_n}{b_n} - 1 \right| > \epsilon \right) \leq \frac{1}{\epsilon^2} \cdot \mathbb{E}\left(\frac{S_n - b_n}{b_n} \right)^2$$

Notice that

$$\begin{aligned}\mathbb{E}\left(S_n-b_n\right)^2&=\mathbb{E}\left[\sum_{j=1}^{n}\left(\mathbb{1}_{A_j}-\mathbb{P}\left(A_j\right)\right)\right]^2\\&=\sum_{j=1}^{n}\mathbb{E}\left[\mathbb{1}_{A_j}-\mathbb{P}\left(A_j\right)\right]^2=\sum_{j=1}^{n}\left[\mathbb{P}\left(A_j\right)-\mathbb{P}\left(A_j\right)^2\right]\end{aligned}$$

where the cross terms vanish since $\mathbb{E}\left[\left(\mathbb{1}_{A_i}-\mathbb{P}\left(A_i\right)\right)\left(\mathbb{1}_{A_j}-\mathbb{P}\left(A_j\right)\right)\right]=\mathbb{E}\left[\mathbb{1}_{A_i}-\mathbb{P}\left(A_i\right)\right]$. $\mathbb{E}\left[1_{A_i}-\mathbb{P}\left(A_i\right)\right]=0$ by pairwise independence of $\left(A_n\right)_{n\geq 1}$. Therefore,

$$\mathbb{E}\left(\frac{S_n-b_n}{b_n}\right)^2\leq\frac{\sum_{j=1}^{n}\mathbb{P}\left(A_j\right)-\sum_{j=1}^{n}\mathbb{P}\left(A_j\right)^2}{b_n^2}\leq\frac{b_n-\frac{1}{n}b_n^2}{b_n^2}=\frac{1}{b_n}-\frac{1}{n}\to 0$$

because $b_n\to\infty$ by assumption, and the second inequality follows from Cauchy-Schwartz inequality. Therefore, $Q_n\xrightarrow{\mathbb{P}}1$.

Step 2. We carefully chose a subsequence of $\left(Q_n\right)_{n\geq 1}$ that converges to 1, a.e. Let

$$n_k=\inf\left\{n\geq 1:\sum_{j=1}^{n}\mathbb{P}\left(A_j\right)\geq k^2\right\}$$

It follows that

$$\mathbb{P}\left(\left|Q_{n_k}-1\right|>\epsilon\right)\leq\frac{1}{\epsilon^2}\left(\frac{1}{b_{n_k}}-\frac{1}{n_k}\right)<\frac{1}{\epsilon^2k^2}$$

so $\sum_{k=1}^{\infty}\mathbb{P}\left(\left|Q_{n_k}-1\right|>\epsilon\right)<\infty$. The first Borel-Cantelli lemma gives that

$$\mathbb{P}\left(\left|Q_{n_k}-1\right|>\epsilon\text{ i.o.}\right)=0$$

for every $\epsilon>0$, which means $Q_{n_k}\to 1$, a.e.

Step 3. The a.e. convergence can actually be passed to the whole sequence. For every n with $n_k<n<n_{k+1}$ for some k, we have $k^2\leq b_{n_k}\leq b_n<(k+1)^2\leq b_{n_{k+1}}<(k+2)^2$, which implies

$$\begin{aligned}\frac{S_n}{b_n}&\leq\frac{S_{n_{k+1}}}{b_{n_k}}=\frac{b_{n_{k+1}}}{b_{n_k}}\cdot\frac{S_{n_{k+1}}}{b_{n_{k+1}}}\leq\frac{(k+2)^2}{k^2}\cdot\frac{S_{n_{k+1}}}{b_{n_{k+1}}}\\\frac{S_n}{b_n}&\geq\frac{S_{n_k}}{b_{n_{k+1}}}=\frac{S_{n_k}}{b_{n_k}}\cdot\frac{b_{n_k}}{b_{n_{k+1}}}\geq\frac{k^2}{(k+2)^2}\cdot\frac{S_{n_k}}{b_{n_k}}\end{aligned}$$

That is, $\frac{k^2}{(k+2)^2}\cdot Q_{n_k}\leq Q_n\leq\frac{(k+2)^2}{k^2}\cdot Q_{n_{k+1}}$. Let $k\to\infty$ then $n\to\infty$ as well, and the result follows because the lower and upper bound for Q_n both converge to 1 a.e. □

2. Let $X_1, \ldots, X_n$ be i.i.d. integrable random variables, and $S = \sum_{i=1}^n X_i$. Calculate $\mathbb{E}[S \mid X_1]$ and $\mathbb{E}[X_1 \mid S]$.

   Solution. Let $T = \sum_{i=2}^n X_i$, then $S = T + X_1$. Since X_j are i.i.d., one has $T \perp\!\!\!\perp X_1$. Therefore, one has

$$\begin{aligned}\mathbb{E}[S \mid X_1] &= \mathbb{E}[T + X_1 \mid X_1] \\ &= \mathbb{E}[T] + X_1 \\ &= \sum_{i=2}^n \mathbb{E}[X_i] + X_1 \\ &= \boxed{(n-1)\mathbb{E}[X_1] + X_1}\end{aligned}$$

   Now, for any $1 \leq j \leq n$, since X_j are i.i.d., $\mathbb{E}(X_j \mid S)$ has the same value. Therefore, one has

$$\begin{aligned}\mathbb{E}(X_1 \mid S) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i \mid S) \\ &= \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^n X_i \,\middle|\, S\right) \\ &= \frac{1}{n}\mathbb{E}(S \mid S) \\ &= \boxed{\frac{S}{n}}\end{aligned}$$

   □

3. For fixed $a, b > 0$, let (X, Y) be a $\mathbb{N} \times \mathbb{R}_+$-valued random variable such that

$$\mathbb{P}(X = n, Y \leq t) = b\int_0^t \frac{(ay)^n}{n!} e^{-(a+b)y}\,\mathrm{d}y.$$

   For $h : \mathbb{R}_+ \to \mathbb{R}$ continuous and bounded, calculate $\mathbb{E}[h(Y) \mid X]$. Calculate $\mathbb{E}\left[\frac{Y}{X+1}\right]$. Calculate $\mathbb{P}(X = n \mid Y)$. Calculate $\mathbb{E}[X \mid Y]$.

   Solution. Given $\mathbb{P}(X = n, Y \leq t)$, one has

$$f_{X,Y}(n,t) = b\frac{(at)^n}{n!}e^{-(a+b)t}, \quad n \in \mathbb{N}, t > 0,$$

   so that

$$\begin{cases} f_X(n) = \mathbb{P}(X = n) = \int_{\mathbb{R}_+} f_{X,Y}(n,t)\,\mathrm{d}t = \frac{b \cdot a^n}{(a+b)^{n+1}} \\ f_Y(t) = \sum_{n \in \mathbb{N}} f_{X,Y}(n,t) = b \cdot e^{-bt} \sim \mathrm{Exp}\,(b) \end{cases}$$

Then, one has

$$f_{Y|X}(t,n) = \frac{f_{X,Y}(n,t)}{f_X(n)} = \frac{t^n}{n!}(a+b)^{n+1}e^{-(a+b)t} \sim \Gamma(n+1, a+b)$$

with mean $\mathbb{E}[Y \mid X = n] = \frac{n+1}{a+b}$. Then, by definition, one has

$$\begin{aligned}
\mathbb{E}[h(Y) \mid X = n] &= \int_{\mathbb{R}_+} h(t) f_{Y|X}(n,t)\,\mathrm{d}t \\
&= \int_{\mathbb{R}_+} h(t)\frac{t^n}{n!}(a+b)^{n+1}e^{-(a+b)t}\,\mathrm{d}t \\
\implies \mathbb{E}[h(Y) \mid X] &= \boxed{\int_{\mathbb{R}_+} h(t)\frac{t^X}{X!}(a+b)^{X+1}e^{-(a+b)t}\,\mathrm{d}t}
\end{aligned}$$ [21]

Let $g(x,y) = \frac{y}{x+1}$. By tower property, one has

$$\begin{aligned}
\mathbb{E}[g(X,Y)] &= \mathbb{E}[\mathbb{E}(g(X,Y) \mid X)] \\
&= \mathbb{E}\left[\frac{1}{X+1}\mathbb{E}[Y \mid X]\right] \\
&= \mathbb{E}\left[\frac{1}{X+1}\frac{X+1}{a+b}\right] \\
&= \boxed{\frac{1}{a+b}}
\end{aligned}$$

[21] Informally, one can further simplify this as

$$\begin{aligned}
\mathbb{E}[h(Y) \mid X = n] &= \frac{(a+b)^n}{n!}\int_{\mathbb{R}+} h(t)t^n(a+b)e^{-(a+b)t}\,\mathrm{d}t \\
&= \frac{(a+b)^n}{n!}\mathbb{E}[h(Z)Z^n], \quad Z \sim \mathrm{Exp}\,(a+b) \\
\implies \mathbb{E}[h(Y) \mid X] &= \frac{(a+b)^X}{X!}\mathbb{E}\left[h(Z)Z^X\right], \quad Z \sim \mathrm{Exp}\,(a+b)
\end{aligned}$$

but the $\mathbb{E}$ in the RHS should be the expectation of only Z but not X, which may lead to misunderstandings for the notation. Therefore, I omit it in the formal answer.

By definition, one has

$$\begin{aligned}
\mathbb{P}(X = n \mid Y = t) &= \lim_{\epsilon \to 0} \frac{\mathbb{P}(X = n, t < Y \leq t + \epsilon)}{\mathbb{P}(t < Y \leq t + \epsilon)} \\
&= \lim_{\epsilon \to 0} \frac{b \int_t^{t+\epsilon} \frac{(ay)^n}{n!} e^{-(a+b)y} \, \mathrm{d}y}{\int_t^{t+\epsilon} b e^{-by} \, \mathrm{d}y} \\
&= \lim_{\epsilon \to 0} \frac{\frac{1}{\epsilon} b \int_t^{t+\epsilon} \frac{(ay)^n}{n!} e^{-(a+b)y} \, \mathrm{d}y}{\frac{1}{\epsilon} \int_t^{t+\epsilon} b e^{-by} \, \mathrm{d}y} \\
&= \frac{b \cdot \frac{(at)^n}{n!} e^{-(a+b)t}}{b e^{-bt}} \qquad \text{by L'Hôpital's rule and F.T.o.C} \\
&= \frac{(at)^n}{n!} e^{-at} \sim \text{Pois}\,(at),
\end{aligned}$$

so that $\mathbb{P}(X = n \mid Y) = \boxed{\frac{(aY)^n}{n!} e^{-aY}}$. Following from above, one has

$$f_{X|Y}(n, t) = \frac{(at)^n}{n!} e^{-at},$$

so that

$$\begin{aligned}
\mathbb{E}[X \mid Y = t] &= \sum_{n=0}^{\infty} n \cdot \frac{(at)^n}{n!} e^{-at} \\
&= e^{-at} (at) \sum_{n=1}^{\infty} \frac{(at)^{n-1}}{(n-1)!} = at \\
\implies \mathbb{E}[X \mid Y] &= \boxed{aY}
\end{aligned}$$

□

4. Let (X_1, X_2) be a Gaussian vector with mean (m_1, m_2) and non-degenerate covariance matrix $(C_{ij})_{1 \leq i,j \leq 2}$. Prove that

$$\mathbb{E}[X_1 \mid X_2] = m_1 + \frac{C_{12}}{C_{22}} (X_2 - m_2).$$

Proof. Let $X_1 \sim \mathcal{N}(m_1, \sigma_X^2)$ and $X_2 \sim \mathcal{N}(m_2, \sigma_Y^2)$. By definition, $C_{12} = \rho \sigma_X \sigma_Y$ and $C_{22} =$

σ_Y^2 where ρ is the correlation between X_1 and X_2.

$$\begin{aligned}
\mathbb{E}[X_1 \mid X_2 = y] &= \int_{-\infty}^{\infty} x f_{X_1|X_2}(x \mid y)\,\mathrm{d}x = \int_{-\infty}^{\infty} x \frac{f_{X_1,X_2}(x,y)}{f_{X_2}(y)}\,\mathrm{d}x \\
&= \frac{\int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_X^2} + \frac{(y-m_2)^2}{\sigma_Y^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_Y\sigma_X}\right]\right)\mathrm{d}x}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_X^2} + \frac{(y-m_2)^2}{\sigma_Y^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_Y\sigma_X}\right]\right)\mathrm{d}x} \\
&= \frac{\int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{x-m_1}{\sigma_X} - \rho\frac{y-m_2}{\sigma_Y}\right]^2\right)\mathrm{d}x}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{x-m_1}{\sigma_X} - \rho\frac{y-m_2}{\sigma_Y}\right]^2\right)\mathrm{d}x} \\
&= \frac{\int_{-\infty}^{\infty} \left(x + m_1 + \sigma_X\rho\frac{y-m_2}{\sigma_Y}\right) \exp\left(-\frac{1}{2(1-\rho^2)}\frac{x^2}{\sigma_X^2}\right)\mathrm{d}x}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\rho^2)}\frac{x^2}{\sigma_X^2}\right)\mathrm{d}x} \\
&= m_1 + \sigma_X\rho\frac{y-m_2}{\sigma_Y} = m_1 + \frac{C_{12}}{C_{22}}(y - m_2)
\end{aligned}$$

Therefore, one has

$$\mathbb{E}[X_1 \mid X_2] = m_1 + \frac{C_{12}}{C_{22}}(X_2 - m_2)$$

□

5. Let X be a random variable such that $\mathbb{P}(X > t) = \exp(-t)$ for any $t \geq 0$. Let $Y = \min(X, s)$, where $s > 0$ is fixed. Prove that, almost surely,

$$\mathbb{E}[X \mid Y] = Y\mathbb{1}_{Y<s} + (1+s)\mathbb{1}_{Y=s}.$$

Proof. Note that $X \sim \mathrm{Exp}\,(1)$ and $Y = X \cdot \mathbb{1}_{\{X<s\}} + s \cdot \mathbb{1}_{\{X \geq s\}}$. Then, one has

$$\begin{cases} \mathbb{E}[X|Y = x < s] = x & (1) \\ \mathbb{E}[X|Y = s] = \mathbb{E}[X|X > s] \end{cases}$$

Now, since

$$f_{X|X>s}(x|X > s) = \frac{f_X(x)}{\mathbb{P}(X > s)} = \frac{e^{-x}}{e^{-s}} = e^{s-x}, \quad \forall x > s,$$

one has

$$\mathbb{E}[X|Y = s] = \mathbb{E}[X|X > s] = \int_s^{\infty} xe^{s-x}\,\mathrm{d}x = s + 1 \qquad (2)$$

Hence, combining (1) and (2), one has $\mathbb{E}[X \mid Y] = Y\mathbb{1}_{Y<s} + (1+s)\mathbb{1}_{Y=s}$. □

6. Let $(X_n)_{n\geq 1}$ be a sequence of nonnegative random variables on $(\Omega, \mathcal{A}, \mathbb{P})$, and $(\mathcal{F}_n)_{n\geq 0}$ a sequence of sub σ-fields of $\mathcal{F}$. Assume that $\mathbb{E}(X_n \mid \mathcal{F}_n)$ converges to 0 in probability.

   (a) Show that X_n converges to 0 in probability.

   Proof. Suppose X_n does not converge to 0 in probability. Then there exists $\delta > 0$ such that $\limsup_n \mathbb{P}(X_n \geq \delta) > 0$, and so there is a subsequence X_{n_k} such that

   $$\mathbb{P}(X_{n_k} \geq \delta) \geq \varepsilon > 0, \forall k,$$

   for some $\varepsilon > 0$. Let $A_k := \{\mathbb{E}(X_{n_k} \mid \mathcal{F}_n) < \varepsilon\}$. By definition,

   $$\begin{aligned} \varepsilon \geq \int_{A_k} \mathbb{E}(X_{n_k} \mid \mathcal{F}_{n_k}) \,\mathrm{d}\mathbb{P} &= \int_{A_k} X_{n_k} \,\mathrm{d}\mathbb{P} \\ \implies \int_{A_k} X_{n_k} \,\mathrm{d}\mathbb{P} \geq \int_{A_k, X_{n_k} \geq \delta} X_{n_k} \,\mathrm{d}\mathbb{P} &\geq \delta \mathbb{P}(A_k, X_{n_k} \geq \delta) \end{aligned}$$

   Since $\mathbb{P}(A_k) \to 1$ and $\mathbb{P}(X_{n_k} \geq \delta, A_k) = \mathbb{P}(X_{n_k} \geq \delta) + \mathbb{P}(A_k) - \mathbb{P}(A_k \cup X_{n_k} \geq \delta)$, one has, for k sufficiently large,

   $$\alpha \mathbb{P}(X_{n_k} \geq \delta) \leq \mathbb{P}(X_{n_k} \geq \delta, A_k), \quad \forall \alpha \in [0, 1)$$

   Hence,

   $$\varepsilon \geq \delta \alpha \mathbb{P}(X_{n_k} \geq \delta) \geq \delta \alpha \varepsilon \implies \alpha \delta \leq 1$$

   But $\alpha\delta \leq 1$ cannot hold for every $\alpha \in [0, 1)$ and a fixed $\delta > 0$, contradiction! □

   (b) Show that the reciprocal is wrong.

   Solution. For example, let $(X_n)_{n\geq 1}$ be a sequence of nonnegative random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ s.t. $\mathbb{P}(X_n = n) = 1/n$ and $\mathbb{P}(X_n = 0) = (n-1)/n$. Take $\mathcal{F}_n = \{\varnothing, \Omega\}$ for every $n \geq 1$. Then $X_n \to 0$ in probability, but $\mathbb{E}(X_n \mid \mathcal{F}_n) = \mathbb{E}(X_n) = 1, n \geq 1$. □

7. Let μ and ν be two probability measures such that $\mu \ll \nu$ and $\nu \ll \mu$ (usually abbreviated $\mu \sim \nu$). Let $X = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.

   (a) Prove that $\nu(X = 0) = 0$.

   Proof. Since $\mu \sim \nu$, it suffices to show $\mu(X = 0) = 0$, but one has

   $$\mu(X = 0) = \int_{X=0} X \,\mathrm{d}\nu = 0,$$

   and so we're done. □

   (b) Prove that $\frac{1}{X} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ almost surely (for μ or ν).

   Proof. We shall use the lemma that

Lemma 2.4.1. *Suppose that ν is a σ-finite measure and μ, λ are σ-finite measures on $(X, \mathcal{M})$ such that $\nu \ll \mu$ and $\mu \ll \lambda$. Then $\nu \ll \lambda$ and*

$$\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \quad \lambda\text{-a.s.}$$

Proof of Lemma. By considering $\nu^+$and ν^-separately, we may assume that $\nu \geq 0$. The equation $\int g d\nu = \int g(d\nu/d\mu)d\mu$ is true when $g = \mathbb{1}_E$ by definition of $d\nu/d\mu$. It is therefore true for simple functions by linearity, then for nonnegative measurable functions by the monotone convergence theorem, and finally for functions in $L^1(\nu)$ by linearity again. Replacing ν, μ by μ, λ and setting $g = \mathbb{1}_E(d\nu/d\mu)$, we obtain

$$\nu(E) = \int_E \frac{d\nu}{d\mu}d\mu = \int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}d\lambda$$

for all $E \in \mathcal{M}$, whence $(d\nu/d\lambda) = (d\nu/d\mu)(d\mu/d\lambda)$ λ-a.s. □

Since then, we can take $\nu = \lambda$ in the lemma and get

$$\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) = \frac{\mathrm{d}\nu}{\mathrm{d}\nu} = 1 \qquad \nu\text{-a.s.}$$

Since $\mu \sim \nu$, one also has

$$\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) = \frac{\mathrm{d}\mu}{\mathrm{d}\mu} = 1 \qquad \mu\text{-a.s.}$$

Hence, $\frac{1}{X} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ almost surely (for μ or ν). □

8. On the same probability space, let X, Y be positive random variables such that $\mathbb{E}[X \mid Y] = Y$ and $\mathbb{E}[Y \mid X] = X$ (almost surely). Prove that $X = Y$ almost surely.

   Proof. First note that, since $\mathbb{E}(X \mid Y) = Y$ almost surely, for every $c > 0$,

   $$\mathbb{E}(X - Y; Y \leq c) = \mathbb{E}(\mathbb{E}(X \mid Y) - Y; Y \leq c) = 0,$$

   and that, as $[Y \leq c] = [X > c, Y \leq c] \cup [X \leq c, Y \leq c]$ is a disjoint union, one has

   $$\mathbb{E}(X - Y; Y \leq c) = \mathbb{E}(X - Y; X > c, Y \leq c) + \mathbb{E}(X - Y; X \leq c, Y \leq c),$$

   Since $\mathbb{E}(X - Y; X > c, Y \leq c) \geq 0$, one has $\mathbb{E}(X - Y; X \leq c, Y \leq c) \leq 0$. Similarly, by exchanging X and Y and following the same steps, given that $\mathbb{E}(Y \mid X) = X$ almost surely, one has $\mathbb{E}(Y - X; X \leq c, Y \leq c) \leq 0$, and thus

   $$\mathbb{E}(Y - X; X \leq c, Y \leq c) = 0$$

Plugging into the previous equation, one has $\mathbb{E}(X-Y;X>c,Y\leq c)=0$. Since $X>c\geq Y$, this is the expectation of a nonnegative random variable s.t. $(X-Y)\mathbb{1}_{X>c\geq Y}=0$ almost surely, which can only happen if the event $[X>c\geq Y]$ has probability zero. Now, since

$$[X>Y]=\bigcup_{c\in\mathbb{Q}}[X>c\geq Y],$$

one has $\mathbb{P}(X>Y)=0$. By symmetry, $\mathbb{P}(Y>X)=0$. Hence, $X=Y$ a.s. □

2.4.3 Markov Chains

1. Let $X_1, X_2\dots$ be i.i.d. Bernoulli random variables ($\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=1/2$) and $S_n=\sum_{i=1}^n X_i$. Which of the following sequences are Markovian? If Markovian, give the transition matrix.

   (a) $(S_n^2-n)_{n\geq 0}$. (b) $(S_{2n})_{n\geq 0}$. (c) $(|S_n|)_{n\geq 0}$.

   Solution. (a) It is Markovian. Let $T_n=S_n^2-n, n\geq 0$. Then,

$$\begin{aligned}&\mathbb{P}(T_{n+1}=y\mid T_n=x_n,\dots,T_0=x_0)\\=&\mathbb{P}(S_{n+1}^2=y_{n+1}\mid S_n^2=y_n,\dots,S_0^2=y_0)\end{aligned}$$

Here we have three cases. Case (i): $y_1,\dots,y_n\neq 0$, implying that $S_1,\dots,S_n$ have the same sign. WLOG, let y_i be positive. Then,

$$\begin{aligned}&\mathbb{P}(S_{n+1}^2=(\sqrt{y_n}+1)^2\mid S_n^2=y_n,\dots,S_0^2=y_0)\\=&\mathbb{P}(X_{n+1}=1\mid X_n=\sqrt{y_n}-\sqrt{y_{n-1}},\dots,X_1=\sqrt{y_1}-\sqrt{y_0})\\=&\mathbb{P}(X_{n+1}=1)=\frac{1}{2}\end{aligned}$$

and similarly,

$$\mathbb{P}(S_{n+1}^2=(\sqrt{y_n}-1)^2\mid S_n^2=y_n,\dots,S_0^2=y_0)=\mathbb{P}(X_{n+1}=-1)=\frac{1}{2}$$

Case (ii): $y_n=0$: $\mathbb{P}(S_{n+1}^2=1\mid S_n^2=0,\dots)=1$. Case (iii): there exists some $k\leq n-1$ s.t. $y_k=0$. Let k to be the smallest number s.t. $S_k^2=0$. Then, one has

$$\begin{aligned}&\mathbb{P}(S_{n+1}^2=y_{n+1}\mid S_n^2=y_n,\dots,S_0^2=y_0)\\=&\mathbb{P}(S_{n+1}^2=y_{n+1}\mid S_n^2=y_n,\dots,S_k^2=0)\end{aligned}$$

Repeat this, we can assume k to be the largest number s.t. $S_k^2=0$. Then $S_{k+1}^2,\dots,S_n^2\neq 0$, and the case reduces to (i). Either case, one has $\mathbb{P}(T_{n+1}=y\mid T_n,\dots,T_0)=\mathbb{P}(T_{n+1}=y\mid T_n)$.

In the following, I use notation $P_{ij} := \mathbb{P}(X_{n+1} = j \mid X_n = i)$ where X_n are terms of the Markov process. Transition matrix:

$$\begin{aligned}\mathbf{P} := (P_{ij}), \quad & P_{i,i\pm 2\sqrt{i+n}} = \frac{1}{2}, \forall n > 0, i \in (-n, n^2 - n] \\ & P_{i,i} = 1, \forall i \leq 0, \text{ where } n = -i \text{ is fixed}\end{aligned}$$

and zero elsewhere.

(b) It is Markovian. The Markov property

$$\mathbb{P}(S_{2n} = x \mid S_0, S_2, \ldots, S_{2n-2}) = \mathbb{P}(S_{2n} = x \mid S_{2n-2})$$

will follow by the fact, since X_{2n}, X_{2n-1} are independent of $S_0, \ldots, S_{2n-2}$, that

$$\begin{aligned}\text{LHS} &= \mathbb{P}(X_{2n-1} + X_{2n} = x - S_{2n-2} \mid S_0, S_2, \ldots, S_{2n-2}) \\ &= \mathbb{P}(X_{2n-1} + X_{2n} = x - S_{2n-2} \mid S_{2n-2}) = \text{RHS}\end{aligned}$$

Note that $\mathbb{P}(S_{2n+2} = t + 2 \mid S_{2n} = t) = \mathbb{P}(S_{2n+2} = t - 2 \mid S_{2n} = t) = 1/4$ and $\mathbb{P}(S_{2n+2} = t \mid S_{2n} = t) = 1/2$. Transition matrix:

$$\mathbf{P} := (P_{ij}), \quad P_{i,i+2} = \frac{1}{4}, P_{i,i} = \frac{1}{2}, i \in \{-2n, -2n+2, \ldots, 2n-2, 2n\}, n \in \mathbb{N}$$

and zero elsewhere.

(c) It is Markovian. Consider $\mathbb{P}(S_n = y \mid |S_n| = y, |S_{n-1}| = y_{n-1}, \ldots, |S_1| = y_1)$. Similar to part (a), if $j = \max\{k : 0 \leq k \leq n : y_k = 0\}$, then one has

$$\begin{aligned}&\mathbb{P}(S_n = y \mid |S_n| = y, \ldots, |S_1| = y_1) \\ =&\, \mathbb{P}(S_n = y \mid |S_n| = y, \ldots, |S_j| = 0) \\ =&\, \frac{\mathbb{P}(S_n = y, \ldots, |S_j| = 0)}{\mathbb{P}(|S_n| = y, \ldots, |S_j| = 0)}\end{aligned}$$

where

$$\begin{aligned}&\mathbb{P}(|S_n| = y, \ldots, |S_j| = 0) \\ =&\, \mathbb{P}(S_n = y, \ldots, |S_j| = 0) + \mathbb{P}(S_n = -y, \ldots, |S_j| = 0) \\ =&\, (1/2)^{\frac{n-j}{2} + \frac{y}{2}} (1/2)^{\frac{n-j}{2} - \frac{y}{2}} + (1/2)^{\frac{n-j}{2} - \frac{y}{2}} (1/2)^{\frac{n-j}{2} + \frac{y}{2}}\end{aligned}$$

and so

$$\begin{aligned}&\frac{\mathbb{P}(S_n = y, \ldots, |S_j| = 0)}{\mathbb{P}(|S_n| = y, \ldots, |S_j| = 0)} \\ =&\, \frac{(1/2)^{\frac{n-j}{2} + \frac{y}{2}} (1/2)^{\frac{n-j}{2} - \frac{y}{2}}}{(1/2)^{\frac{n-j}{2} + \frac{y}{2}} (1/2)^{\frac{n-j}{2} - \frac{y}{2}} + (1/2)^{\frac{n-j}{2} - \frac{y}{2}} (1/2)^{\frac{n-j}{2} + \frac{y}{2}}} = \frac{1}{2}\end{aligned}$$

Then, one has

$$\begin{aligned}
&\mathbb{P}(|S_{n+1}| = y+1 \mid |S_n| = y, \ldots, |S_1| = y_1)\\
&= \frac{1}{2}\mathbb{P}(S_{n+1} = y+1 \mid S_n = y) + \frac{1}{2}\mathbb{P}(S_{n+1} = -y-1 \mid S_n = -y) = \frac{1}{2}
\end{aligned}$$

and similar for $|S_{n+1}| = y-1$ by symmetry.

Transition matrix:

$$\mathbf{P} := (P_{ij}), \quad P_{i,i+1} = \frac{1}{2}, \forall i > 0, \quad P_{0,1} = 1$$

and zero elsewhere.

□

2. Consider a Markov chain X with state space $\{0, 1, \ldots, n\}$ and transition matrix

$$\pi(0,k) = \frac{1}{2^{k+1}},\ 0 \le k \le n-1,\ \pi(0,n) = \frac{1}{2^n}$$
$$\pi(k,k-1) = 1,\ 1 \le k \le n-1,\ \pi(n,n) = \pi(n,n-1) = \frac{1}{2}.$$

(a) Prove that the chain has a unique invariant probability measure μ and calculate it.

Proof. Let $I = \{0, 1, \ldots, n\}$. Denote $\mu_k := \mu(k), k \in I$. It suffices to solve the system

$$\begin{cases} \sum_{i=0}^n \mu_i \pi_{ij} = \mu_j, & \forall j \in I \\ \mu_0 + \mu_1 + \ldots + \mu_n = 1 \\ \mu_0, \ldots, \mu_n \ge 0 \end{cases}$$

After calculation, the system has only one solution that

$$\mu = (\mu_0, \ldots, \mu_n) : \quad \mu_k = \frac{1}{2^{k+1}}, k \in [\![0, n-1]\!], \quad \mu_n = \frac{1}{2^n}$$

□

(b) Prove that for any $0 \le x_0 \le n-1$, $\pi^{(x_0+1)}(x_0, \cdot) = \mu$.

Proof. We prove by induction and use the *Chapman-Kolmogorov Equation*. For $x_0 = 0$, $\pi^{(1)}(x_0, \cdot)$ is by definition

$$\pi(0,k) = \frac{1}{2^{k+1}},\ 0 \le k \le n-1,\ \pi(0,n) = \frac{1}{2^n}$$

which corresponds with μ. Note that

$$\pi_{1j}^{(2)} = \sum_k \pi_{1k}\pi_{kj} = \pi_{0j} = \mu_j, \forall j \in I; \pi_{2j}^{(2)} = \sum_k \pi_{2k}\pi_{kj} = \pi_{1j}$$

$$\ldots \text{ assume } \pi_{(n-2),j}^{(n-1)} = \mu_j, \forall j \in I, \text{ and note that}$$

$$\pi_{(n-1),j}^{(n-1)} = \sum_k \pi_{(n-1),k}^{(n-2)}\pi_{kj} = \cdots = \pi_{1j},$$

$$\implies \pi_{(n-1),j}^{(n)} = \sum_k \pi_{(n-1),k}^{(n-1)}\pi_{kj} = \sum_k \pi_{1k}\pi_{kj} = \pi_{0j} = \mu_j, \forall j \in I$$

Therefore, by induction, one has for any $0 \le x_0 \le n-1$, $\pi^{(x_0+1)}(x_0, \cdot) = \mu$. □

(c) Prove that for any $0 \le x_0 \le n$, $\pi^{(n)}(x_0, \cdot) = \mu$.

Proof. When $0 \le x_0 \le n-1$, $\pi_{x_0,\cdot}^{(n)} = \pi(0, \cdot) = \mu$ since μ is invariant, which implies $\pi^{(k)}(0, \cdot) = \mu$.

When $x_0 = n$, one has

$$\pi_{x_0,0}^{(n)} = \mathbb{P}(n \to n-1 \text{ at first step}) = \frac{1}{2},$$

$$\pi_{x_0,1}^{(n)} = \mathbb{P}(n \to n \to n-1 \to \cdots \to 1) = \mathbb{P}(n \to n \to n-1) = \frac{1}{2^2}, \ldots$$

Inductively, one has $\pi_{x_0,k}^{(n)} = \frac{1}{2^{k+1}}$ for $0 \le k \le n-1$, and $\pi_{x_0,n}^{(n)} = \frac{1}{2^n}$. Hence, $\pi^{(n)}(n, \cdot) = \mu$ as well. □

(d) For any $t \ge 1$, calculate

$$d(t) := \frac{1}{2}\sum_{x=0}^{n} \left|\pi^{(t)}(n, x) - \mu(x)\right|,$$

and plot $t \mapsto d(t)$.

Solution. Note when $t \ge n$, one has $\pi_{n,x}^{(t)} = \mu_x$, since we have proved $\pi(n)_{n,x} = \mu_x$, and for any more steps, the distribution stays invariant. Hence, when $t \ge n$, $d(t) = 0$. When $t < n$, since n can at most go to $n - t$, one has

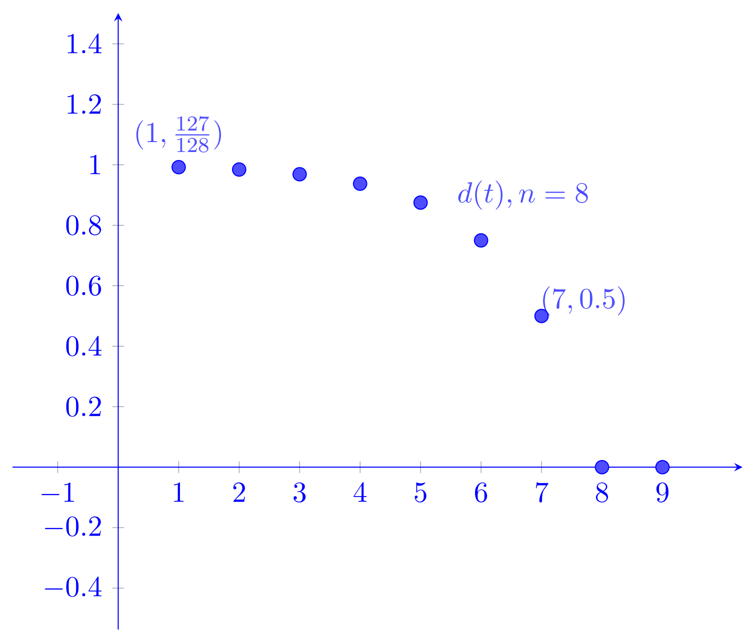
$$\begin{cases} \pi_{n,k}^{(t)} = 0, & k < n-t \\ \pi_{n,k}^{(t)} = 1/2^{k-(n-t)+1}, & n-t \le k \le n-1 \\ \pi_{n,n}^{(t)} = 1/2^t \end{cases}$$

Then, one has

$$
\begin{aligned}
d(1) &= \frac{1}{2}\left(\sum_{0}^{n-2}\mu_x + \left(\frac{1}{2} - \mu_{n-1}\right) + \left(\frac{1}{2} - \mu_n\right)\right)\\
&= 1 - \mu_{n-1} - \mu_n
\end{aligned}
$$

and, noting that $\pi_{n,x}^{(t)} - \mu_x > 0$ as $k + 1 - (n - t) < k + 1$,

$$
\begin{aligned}
d(t) &= \frac{1}{2}\left(\sum_{x=0}^{n-t-1}\mu_x + \sum_{x=n-t}^{n}\left|\pi_{n,x}^{(t)} - \mu_x\right|\right)\\
&= \frac{1}{2}\left(1 - \sum_{x=n-t}^{n}\mu_x + 1 - \sum_{x=n-t}^{n}\mu_x\right)\\
&= 1 - \sum_{x=n-t}^{n}\mu_x
\end{aligned}
$$



$\square$

3. For fixed $p, q \in [0, 1]$, consider a Markov chain X with two states $\{1, 2\}$, with transition matrix

$$
\pi = (\pi(i,j))_{1\le i,j\le 2} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}
$$

(a) For which p, q is the chain irreducible? Aperiodic (see Lemma 4.14 and Remark 4.10 in Varadhan)?

Solution. By definition, the chain is irreducible if $i \longleftrightarrow j, i \neq j$, so that we need $p > 0$ and $q > 0$, namely $p, q \in (0, 1]$.

By definition, the chain is aperiodic if all states are aperiodic: a state i is aperiodic if $d(i) := \gcd\{n \geq 1 : \pi_{ii}^{(n)} > 0\} = 1$. Hence, we have four cases:

- $p < 1, q < 1$: then the chain is aperiodic.
- $p = 1, q < 1$: then $q \neq 0$ (since otherwise, we have $d(1) = \infty$), and it is verified that $d(1) = 1$ when $q \in (0, 1)$ (both $n = 2$ and $n = 3$ are positive for $\pi_{ii}^{(n)}$).
- $p < 1, q = 1$: similarly, $p \neq 0$.
- $p = q = 1$: it is periodic, since $d(1) = d(2) = 2$.

Combining the cases above, one has the chain being aperiodic when $p, q \in [0, 1)$ or $p = 1, q \in (0, 1)$ or $q = 1, p \in (0, 1)$.

Remark. In Varadhan Remark 4.10, it says that an irreducible chain with period $d = 1$ is called aperiodic. However, in general, the definition of aperiodic does not rely on irreducibility, and in fact we have chains that are aperiodic but not irreducible, for example $\pi = I_{2\times 2}$. If one has to follow this remark, then just exclude the cases of $p = 0$ or $q = 0$.

□

(b) What are the invariant probability measures of X?

Solution. Let $\vec{\mu} = (\mu_1, \mu_2)$. By solving the equation

$$\vec{\mu}\pi = \vec{\mu},$$

one has (i) for $p = q = 0$, $\vec{\mu}$ can be any probability measure satisfying $\mu_1 + \mu_2 = 1$, $\mu_1, \mu_2 \geq 0$; (ii) otherwise, $\vec{\mu} = \left[\frac{q}{p+q}, \frac{p}{p+q}\right]$. □

(c) Compute $\pi^{(n)}, n \geq 1$.

Solution. By definition, $\pi^{(n)} = \pi^n$. For $q = 0$, it is easy to observe that

$$\pi^{(n)} = \begin{pmatrix} (1-p)^n & 1-(1-p)^n \\ 0 & 1 \end{pmatrix}, \quad n \in \mathbb{N}_+$$

For $q \neq 0$, we can diagonalize π by

$$\pi = \begin{pmatrix} -p/q & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -p-q+1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{-q}{p+q} & \frac{q}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$$

so that

$$\pi^{(n)} = \begin{pmatrix} -p/q & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} (-p-q+1)^n & 0 \\ 0 & 1^n \end{pmatrix} \begin{pmatrix} \frac{-q}{p+q} & \frac{q}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$$
$$= \frac{1}{p+q} \begin{pmatrix} q+p(1-p-q)^n & p-p(1-p-q)^n \\ q-q(1-p-q)^n & p+q(1-p-q)^n \end{pmatrix}$$

□

(d) When X is irreducible, for this invariant probability measure μ, calculate

$$d_1(n) := \frac{1}{2}\left(|\mathbb{P}_1(X_n=1)-\mu(1)| + |\mathbb{P}_1(X_n=2)-\mu(2)|\right)$$
$$d_2(n) := \frac{1}{2}\left(|\mathbb{P}_2(X_n=1)-\mu(1)| + |\mathbb{P}_2(X_n=2)-\mu(2)|\right)$$

where $\mathbb{P}_x$ means the chain starts at x.

Solution. (i) For $p,q \in (0,1]$, following (c), one has

$$d_1(n) = \frac{1}{2}\left|\frac{p}{p+q}(1-p-q)^n\right| + \frac{1}{2}\left|\frac{p}{p+q}(1-p-q)^n\right| = \frac{p}{p+q}\left|(1-p-q)^n\right|$$

(ii) if $p=0, q\neq 0$, then $d_1(n)=0$. (iii) if $p\neq 0, q=0$, then $d_1(n)=(1-p)^n$. (iv) for $p=q=0$, one has $P=I$ and $d_1(n)=\mu_2$.

Similarly, (i) if $p,q\in(0,1]$, then

$$d_2(n) = \frac{q}{p+q}\left|(1-p-q)^n\right|;$$

(ii) if $p=0, q\neq 0$, then $d_2(n)=(1-q)^n$; (iii) if $p\neq 0, q=0$, then $d_2(n)=0$; (iv) if $P=I$, then $d_2(n)=\mu_1$. □

4. Let T be a stopping time for a filtration $(\mathcal{F}_n)_{n\geq 1}$. Prove that $\mathcal{F}_T$ is a σ-field.

   Proof. Note that $\mathcal{F}_T = \{A\in\mathcal{F} : A\cap\{T\leq n\}\in\mathcal{F}_n \text{ for each } n\}$. We verify by definition: (i) $\varnothing\cap\{T\leq n\} = \varnothing\in\mathcal{F}_n$. (ii) If $A\in\mathcal{F} : A\cap\{T\leq n\}\in\mathcal{F}_n$ for each n, one has $A^\complement\cap\{T\leq n\} = (A\cap\{T\leq n\})^\complement\cap\{T\leq n\}\in\mathcal{F}_n$ as both sets are in $\mathcal{F}_n$. (iii) If $(A_k)_{k\geq 0}\in\mathcal{F}$, then

   $$\left(\bigcup_{k=1}^\infty A_k\right)\cap\{T\leq n\} = \bigcup_{k=1}^\infty (A_k\cap\{T\leq n\})\in\mathcal{F}_n$$

   since every $A_k\cap\{T\leq n\}$ is in $\mathcal{F}_n$. Hence, $\mathcal{F}_T$ is a σ-field. □

5. Let S and T be stopping times for a filtration $(\mathcal{F}_n)_{n\geq 1}$. Prove that $\max(S,T)$ and $\min(S,T)$ are stopping times.

Proof. Note that by definition, one has $\{\min(S,T) \le n\} = \{S \le n\} \cup \{T \le n\} \in \mathcal{F}_n$ and $\{\max(S,T) \le n\} = \{S \le n\} \cap \{T \le n\} \in \mathcal{F}_n$ for all n, hence max and $\min(S,T)$ are stopping times by definition. □

6. Let $S \le T$ be two stopping times and $A \in \mathcal{F}_S$. Define $U(\omega) = S(\omega)$ if $\omega \in A$, $U(\omega) = T(\omega)$ if $\omega \notin A$. Prove that U is a stopping time.

   Proof. We first show that if $S \le T$ are stopping times, then $\mathcal{F}_S \subset \mathcal{F}_T$: let $A \in \mathcal{F}_S$, then

   $$A \cap \{T \le t\} = (A \cap \{S \le t\}) \cap \{T \le t\} \in \mathcal{F}_T$$

   Now, since $A \in \mathcal{F}_S$, one has $A^\complement \in \mathcal{F}_S \subset \mathcal{F}_T$. Then, for each $t \ge 0$, one has $A \cap \{S \le t\} \in \mathcal{F}_t$ and $A^\complement \cap \{T \le t\} \in \mathcal{F}_t$. Hence, $\{U \le t\} = (A \cap \{S \le t\}) \cup \left(A^\complement \cap \{T \le t\}\right) \in \mathcal{F}_t$ is a stopping time by definition. □

7. An ant walks on a round clock, starting at 0. At each second, it walks either clockwise or counterclockwise, with probability $1/2$ to a neighbouring number, and through independent steps. Let X be the last number visited by the ant. Prove it is equidistributed on $\{1, 2, \dots, 11\}$.

   Proof. Note that we have 12 vertices on this circle. If $1, 2, \dots, 11$ are vertices assigned clockwise, we also assign vertices $-1, -2, \dots, -11$ counter-clockwise, meaning that number k and $k - 12$ represent the same vertex.

   Now, for $a, b \in \mathbb{N}$, consider the events for the ant

   $$\begin{aligned} E &= \{\text{visit every vertex in } [-a, b] \text{ before visiting vertex } -a-1 \text{ or } b+1\} \\ E_1 &= \{\text{visit } b \text{ from } 0 \text{ before visiting } -a, \text{ and then from } b \text{ to } -a \text{ before visiting } b+1\} \\ E_2 &= \{\text{visit } -a \text{ from } 0 \text{ before visiting } b, \text{ and then from } -a \text{ to } b \text{ before visiting } -a-1\} \end{aligned}$$

   Observe that $E_1 \cup E_2 = E, E_1 \cap E_2 = \varnothing$. Recall from *Gambler's ruin* that for a simple symmetric random walk starting from i, $\mathbb{P}(\{\text{visit } i-j \text{ before visiting } i+k\}) = \frac{k}{k+j}, \forall j, k \in \mathbb{N}$. Therefore, one has $\mathbb{P}(E_1) = \frac{a}{a+b}\frac{1}{a+b+1}, \mathbb{P}(E_2) = \frac{b}{a+b}\frac{1}{a+b+1}$ and thus $\mathbb{P}(E) = \frac{1}{a+b+1}$.

   Now, let $x \in [\![1, 11]\!]$ and let $b = x - 1, a = 11 - x$. Then, the event E is equivalent to the event E_x that the ant visits vertex x (or $x - 12$ which is the same) lastly, and one has $\mathbb{P}(E_x) = \frac{1}{(11-x)+(x-1)+1} = \frac{1}{11}, \forall x \in [\![1, 11]\!]$, i.e. in the original question one has $\mathbb{P}(X = x) = \frac{1}{11}, \forall x \in [\![1, 11]\!]$, i.e. X is equidistributed on $\{1, 2, \dots, 11\}$. □

8. Let $X_1, X_2, \dots$ be i.i.d., $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and $S_n = X_1 + \dots + X_n$. Prove that the following random variable converges in distribution as $n \to \infty$, and identify the limit:

   $$\left(\sum_{k=1}^{n} e^{S_k}\right)^{\frac{1}{\sqrt{n}}}$$

Proof. Let $M_n = \max\{S_1, \ldots, S_n\}, Z \sim \mathcal{N}(0,1)$. We first show that $M_n/\sqrt{n} \xrightarrow{\mathcal{D}} |Z|$. Note that by *reflection principle*, one has $\mathbb{P}(M_n \geq a) = 2\mathbb{P}(S_n \geq a)$. Note also that $\mathbb{E}[X_1] = 0, \mathrm{Var}[X_1] = 1$. Following the CLT, one has

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Therefore, $\mathbb{P}(M_n/\sqrt{n} \geq a/\sqrt{n}) = 2\mathbb{P}(S_n/\sqrt{n} \geq a/\sqrt{n}), \forall a \geq 0$ implies that

$$\frac{M_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} |Z|, \quad Z \sim \mathcal{N}(0,1)$$

Now, one has

$$e^{M_n} \leq \sum_{k=1}^{n} e^{S_k} \leq n e^{M_n}$$
$$\implies M_n \leq \log\left(\sum_{k=1}^{n} e^{S_k}\right) \leq M_n + \log n$$

Multiply $1/\sqrt{n}$ in all sides of the inequality and take the limit, and by noting that $\log n/\sqrt{n} \to 0$ as $n \to \infty$, one has

$$\lim_{n\to\infty} \frac{M_n}{\sqrt{n}} = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \log\left(\sum_{k=1}^{n} e^{S_k}\right)$$

which implies that

$$\frac{1}{\sqrt{n}} \log\left(\sum_{k=1}^{n} e^{S_k}\right) \xrightarrow{\mathcal{D}} |Z| \implies \left(\sum_{k=1}^{n} e^{S_k}\right)^{\frac{1}{\sqrt{n}}} \xrightarrow{\mathcal{D}} \boxed{e^{|Z|}}, \quad Z \sim \mathcal{N}(0,1)$$

□

9. (Birth-Death Chain) Consider a Markov chain X with state space $\mathbb{N}$ and transition matrix

$$\pi(0,0) = r_0,\ \pi(0,1) = p_0, \text{and } \forall i \geq 1,\ \pi(i,i-1) = q_i,\ \pi(i,i) = r_i,\ \pi(i,i+1) = p_i,$$

with $p_0, r_0 > 0$, $p_0 + r_0 = 1$ and for all $i \geq 1$, $p_i, q_i > 0$, $p_i + q_i + r_i = 1$. Prove that the chain is irreducible, aperiodic. Give a necessary and sufficient condition for the chain to have an invariant probability measure.

Proof. By definition, X is irreducible if all states communicate. Note that $p_0, q_1 > 0$, so $0 \longleftrightarrow 1$. For any $i \geq 1$, one has $p_i, q_{i+1} > 0$ by definition, so $i \longleftrightarrow i+1$. Since communication is transitive, all states $(0, 1, 2, \ldots)$ communicate.

By definition, the chain is aperiodic if all states are aperiodic: a state i is aperiodic if $d(i) := \gcd\{n \geq 1 : \pi_{ii}^{(n)} > 0\} = 1$. Note that $r_0 > 0$, so state 0 is aperiodic. For $i \geq 1$, $\pi_{ii}^{(2)} \geq p_i q_{i+1} + q_i p_{i-1} > 0$, (using $\geq$ as we do not yet know if r_i is strictly positive). On the other hand,

$$\pi_{ii}^{(2i+1)} \geq r_0 \cdot \prod_{k=1}^{i} (q_k p_{k-1}) > 0,$$

since all p, q and r_0 are by definition positive. Hence,

$$1 \leq \gcd\{n \geq 1 : \pi_{ii}^{(n)} > 0\} \leq \gcd(2, 2i+1) = 1,$$

implying that any state $i, i \geq 1$, is aperiodic. So is the chain.

As for the existence of invariant probability measure, note that X here is irreducible, and by the renewal theorem[22], it is sufficient to find a necessary and sufficient condition for X to be positive recurrent. We now show that X is positive recurrent if and only if

$$\sum_{k=1}^{\infty} \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} < \infty$$

Let π be the stationary distribution for the chain. Then we have

$$\begin{aligned} \pi_k &= \pi_{k-1} p_{k-1} + \pi_k (1 - p_k - q_k) \pi_{k+1} q_{k+1}, \quad k \geq 1 \\ \pi_0 &= \pi_0 (1 - p_0) + \pi_1 q_1 \end{aligned}$$

Rearranging the above equation, we have

$$\begin{aligned} q_{k+1}\pi_{k+1} - p_k \pi_k &= q_k \pi_k - p_{k-1}\pi_{k-1} \\ q_1 \pi_1 - p_0 \pi_0 &= 0 \end{aligned}$$

By induction, one has

$$\pi_{k+1} = \frac{p_k}{q_{k+1}} \pi_k \implies \pi_k = \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} \pi_0$$

If

$$\sum_{k=1}^{\infty} \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} < \infty$$

converges, let

$$\pi_0 = \left(\sum_{k=1}^{\infty} \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} \right)^{-1},$$

then we have $\pi_k > 0$. This implies the chain is positive recurrent. □

[22] An irreducible chain has a stationary distribution if and only if all states are positive recurrent.

10. Let $(G, \cdot)$ be a group, μ a probability measure on G and X the Markov chain such that $\pi(g, h \cdot g) = \mu(h)$. We call such a process X a random walk on G with jump kernel μ.

    (a) Explain why the usual random walk on $\mathbb{Z}^d$ is such process. Same question for the usual random walk on $(\mathbb{Z}/n\mathbb{Z})^d$, $n \geq 1$.

    Proof. A usual random walk in $\mathbb{Z}^d$ is a sequence of random variables S_n starting at $S_0 = 0$ with probability 1, and for each $n \in \mathbb{N}_+$ we define S_n to be $S_n = X_1 + \cdots + X_n$. Here the X_j are i.i.d. random variables such that, for each time index j and dimension index k,

    $$\mathbb{P}\{X_j = \mathbf{e}_k\} = \mathbb{P}\{X_j = -\mathbf{e}_k\} = \frac{1}{2d}$$

    with $\mathbf{e}_k$ denoting the k-th unit basis vector. Similarly, on $\mathbb{Z}/n\mathbb{Z}$, it is equivalent to a random walk on a n-cycle with vertices $0, 1, \dots, n-1$, same as the ant-clock problem in some previous problem set. □

    (b) Consider the following shuffling of a deck of $n \geq 2$ cards: pick two such distinct cards uniformly at random and exhange their positions in the deck. Show that this is also an example of a random walk on a group.

    Proof. It is equivalent to define on the symmetric group S_n (which can be viewed as the set of all bijective maps from $\{1, \dots, n\}$ to itself equipped with composition) that

    $$\pi(g, x \cdot g) = \mu(x) = \begin{cases} 1/n & \text{if } x = \mathrm{id} \\ 2/n^2 & \text{if } x \text{ is a transposition} \\ 0 & \text{otherwise} \end{cases}$$

    Note that the random transposition is well defined since $1/n + \binom{n}{2} 2/n^2 = 1$. □

    (c) Let $\mathcal{H} = \{h_1 \cdot h_2 \cdot \cdots \cdot h_n, \mu(h_i) > 0, 1 \leq i \leq n, n \in \mathbb{N}\}$. Discuss irreductibility of X depending on $\mathcal{H}$.

    Proof. X is irreducible if and only if $\mathcal{H} = G$. □

    (d) If X is irreducible on finite G, what are the invariant probability measures? What if G is not finite?

    (e) Make some search to define a reversible Markov chain. In the context of this exercise, show that X is reversible if and only if $\mu(h) = \mu(h^{-1})$ for any $h \in G$.

    Proof. A Markov chain that has invariant probabilities $\{\pi_i; i \geq 0\}$ is reversible if $P_{ij} = \pi_j P_{ji}/\pi_i$ for all i, j. □

    (f) Give an example of an irreducible random walk on a group which is not reversible.

2.5 Martingales

1. Let $(X_n)_{n\geq 1}$ be independent such that $\mathbb{E}(X_i) = m_i$, $\mathrm{Var}(X_i) = \sigma_i^2$, $i \geq 1$. Let $S_n = \sum_{i=1}^n X_i$ and $\mathcal{F}_n = \sigma(X_i, 1 \leq i \leq n)$.

   (a) Find sequences $(b_n)_{n\geq 1}$, $(c_n)_{n\geq 1}$ of real numbers such that $(S_n^2 + b_n S_n + c_n)_{n\geq 1}$ is a $(\mathcal{F}_n)_{n\geq 1}$-martingale.

   Solution. Note that for any $b_n, c_n, Y_n := S_n^2 + b_n S_n + c_n$ is integrable and $\mathcal{F}_n$-measurable, so it is sufficient to find b_n, c_n such that $\mathbb{E}\left[Y_{n+1} \mid \mathcal{F}_n\right] = Y_n$. Note that

   $$\begin{aligned} Y_{n+1} &= (S_n + X_{n+1})^2 + b_{n+1} S_n + b_{n+1} X_{n+1} + c_{n+1} \\ &= S_n^2 + b_{n+1} S_n + 2 S_n X_{n+1} + X_{n+1}^2 + b_{n+1} X_{n+1} + c_{n+1} \end{aligned}$$

   Taking conditional expectation, one has (following $\mathbb{E}\left[S_n X_{n+1} \mid \mathcal{F}_n\right] = S_n \mathbb{E}\left[X_{n+1}\right]$)

   $$\mathbb{E}\left[Y_{n+1} \mid \mathcal{F}_n\right] = S_n^2 + b_{n+1} S_n + 2 S_n m_{n+1} + (\sigma_{n+1}^2 + m_{n+1}^2) + b_{n+1} m_{n+1} + c_{n+1}$$

   Hence, b_n, c_n must satisfy

   $$b_{n+1} + 2 m_{n+1} = b_n, \quad \sigma_{n+1}^2 + m_{n+1}^2 + b_{n+1} m_{n+1} + c_{n+1} = c_n$$

   Solving by induction, one has

   $$b_n = b_1 - 2\sum_{i=2}^n m_i, \quad c_n = c_1 - \sum_{i=2}^n \left(\sigma_i^2 + m_i^2 + b_i m_i\right) \quad n \geq 2$$

   □

   (b) Assume moreover that there is a real number λ such that $e^{\lambda X_i} \in \mathrm{L}^1$ for any $i \geq 1$. Find a sequence $\left(a_n^{(\lambda)}\right)_{n\geq 1}$ such that $\left(e^{\lambda S_n - a_n^{(\lambda)}}\right)_{n\geq 1}$ is a $(\mathcal{F}_n)_{n\geq 1}$-martingale.

   Proof. For any $\lambda \in \mathbb{R}$, Y_n is integrable and $\mathcal{F}_n$-measurable, so it is sufficient to show $\mathbb{E}\left[Y_{n+1} \mid \mathcal{F}_n\right] = Y_n$. Note that

   $$\begin{aligned} \mathbb{E}\left[Y_{n+1} \mid \mathcal{F}_n\right] &= \mathbb{E}\left(e^{\lambda S_{n+1} - a_{n+1}} \middle| \mathcal{F}_n\right) \\ &= e^{\lambda S_n - a_{n+1}} \cdot \mathbb{E}\left(e^{\lambda X_{n+1}}\right) \\ &= e^{\lambda S_n - a_n} \\ \implies e^{-a_{n+1}} \cdot \mathbb{E}\left(e^{\lambda X_{n+1}}\right) &= e^{-a_n} \end{aligned}$$

   Since we do not know the distribution of X, just let $\mu_i^{(\lambda)} := \mathbb{E}\left(e^{\lambda X_i}\right)$. Then, one has

   $$a_n^{(\lambda)} = a_1 + \sum_{i=2}^n \log\left(\mu_i^{(\lambda)}\right)$$

   □

2. Let $(X_k)_{k\geq 0}$ be i.i.d. random variables, $\mathcal{F}_m = \sigma(X_1, \ldots, X_m)$ and $Y_m = \prod_{k=1}^m X_k$. Under which conditions is $(Y_m)_{m\geq 1}$ a $(\mathcal{F}_m)_{m\geq 1}$-submartingale, supermartingale, martingale?

Solution. We first need $\mathbb{E}[|X_k|] < \infty$ for either submartingale, supermartingale, or martingale. And since X_i are independent, Y_m as a product is integrable and $\mathcal{F}_m$-measurable. Now we only need to ensure for any n,

$$\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = Y_n \mathbb{E}(X_{n+1}|\mathcal{F}_n) = Y_n \mathbb{E}(X_{n+1}) \begin{cases} = Y_n & \mathrm{MG} \\ \geq Y_n & \mathrm{sub-MG} \\ \leq Y_n & \mathrm{super-MG} \end{cases}$$

(a) Martingale: either $\mathbb{E}(X_{n+1}) = 1$ or $X_1 = 0$ almost surely.

(b) Sub-martingale: $Y_n \geq 0$ (if and only if $X_1 \geq 0$ a.s.), $\mathbb{E}[X_{n+1}] \geq 1$.

(c) Super-martingale: $Y_n \geq 0$ (if and only if $X_1 \geq 0$ a.s.), $\mathbb{E}[X_{n+1}] \leq 1$.

□

3. Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration and $(X_n)_{n\geq 0}$ be a sequence of integrable random variables with $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = 0$, and assume X_n is $\mathcal{F}_n$-measurable for every n. Let $S_n = \sum_{k=0}^n X_k$. Show that $(S_n)_{n\geq 0}$ is a $(\mathcal{F}_n)_{n\geq 0}$-martingale.

Proof. Since X_n is integrable and $\mathcal{F}_n$-measurable, its partial sum S_n is also integrable and $\mathcal{F}_n$-measurable for any n. But one also has

$$\mathbb{E}(S_{n+1} \mid \mathcal{F}_n) = S_n + \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = S_n$$

given the assumption that $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = 0$. Hence, $(S_n)_{n\geq 0}$ is a $(\mathcal{F}_n)_{n\geq 0}$-martingale. □

4. Let $a > 0$ be fixed, $(X_i)_{i\geq 1}$ be iid, $\mathbb{R}^d$-valued random variables, uniformly distributed on the ball $\mathrm{B}(0, a)$. Set $S_n = x + \sum_{i=1}^n X_i$.

   (a) Let f be a superharmonic function. Show that $(f(S_n))_{n\geq 1}$ defines a supermartingale.

   Proof. Note that S_n, as the sum of uniform X_i, is integrable and $\mathcal{F}_n$-measurable (where $\mathcal{F}_n := \sigma\{X_1, \ldots, X_n\}$). By definition, $f(S_n)$ is also integrable. It is now sufficient to show $\mathbb{E}\{f(S_{n+1}) \mid \mathcal{F}_n\} \leq f(S_n)$.

   Let $S_0 = x$, then one has $f(S_{n+1}) = f(S_n + X_{n+1}), n \geq 0$. Let $\mathbb{E}(f(S_n + X_{n+1})|\mathcal{F}_n) = g(S_n)$ where $g(x) := \mathbb{E}[f(x, Y)] = \int f(x+y)\mathbb{P}(\mathrm{d}y)$ [23]. Since X_i is uniform on $\mathrm{B}(0, a)$, one has

$$\mathbb{P}(\mathrm{d}y) = \mathbb{1}_{B(0,a)} \frac{1}{|B(0,a)|} \mathrm{d}y,$$

[23] By the tower property, one has "if $g(x) = \mathbb{E}(f(x, Y))$, then $g(X) = \mathbb{E}(f(X, Y)|X)$."

which yields (since f is super-harmonic, by definition, its values are no smaller than the values of a harmonic function on the boundary of a ball)

$$g(S_n) = \int f(S_n + y)\mathbb{1}_{B(0,a)} \frac{1}{|B(0,a)|}\,\mathrm{d}y = \frac{1}{|B(0,a)|}\int_{B(S_n,a)} f(y)\,\mathrm{d}y \le f(S_n),$$

implying that $(f(S_n))_{n\ge1}$ defines a supermartingale. □

(b) Prove that if $d \le 2$ any nonnegative superharmonic function is constant. Does this result remain true when $d \ge 3$?

Proof. From part (a), we have $M_n := f(S_n)$ as a supermartingale. Since f nonnegative,by the martingale convergence theorem, M_n converges a.s. Suppose $M_n \to M$ a.s. Since superharmonic functions are lower semicontinuous, $f \le M$ at every point, but as a supermartingale, $\mathbb{E}f(S_n)$ does not increase with n. Therefore, $M \le f(x)$ where $x = S_0$ defined in part (a). Since x was arbitrary, it follows that $f(x) \equiv M$. But (S_n) is irreducible and recurrent, and so it visits every neighborhood of every point infinitely often. Thus with probability 1, $f(S_n)$ takes on every f value infinitely often. The process is then neighborhood recurrent for $d \le 2$. Thus f is constant, since $M_n = f(X_n)$ cannot take on distinct values infinitely often and still converge.

It is not in general true for $d > 2$. There exist positive superharmonic functions such as $\min\left(1, |x|^{2-d}\right)$. The process then is not neighborhood recurrent when $d > 2$, hence the result does not remain true. □

5. In the following exercise you can use the following fact: If a martingale is a.s. bounded by a deterministic constant, it converges almost surely.

Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of random variables, and assume (Y_n) converges in distribution to a limiting Y. Also, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the sequence of independent random variables $X := (X_n)_{n\in\mathbb{N}^*}$ is defined, and we assume that the sequence of partial sums $(S_n)_{n\in\mathbb{N}}$ (i.e. $S_0 = 0$ and $S_n := \sum_{j=1}^n X_j$) converges in distribution. Set $(\mathcal{F}_n)$ the natural filtration of X and $\Phi_n(t) = \mathbb{E}(\exp(\mathrm{i}tS_n))$ for $t \in \mathbb{R}$.

(a) Establish that $(\Phi_{Y_n}(\cdot))_{n\ge1}$ converges uniformly on every compact set, i.e. show that for any $a > 0$, $\max_{t\in[-a,a]} |\Phi_{Y_n}(t) - \Phi_Y(t)| \to 0$ as $n \to \infty$. Establish moreover that there exists $a > 0$ such that for any $n \ge 1$, $\min_{t\in[-a,a]} |\Phi_{Y_n}(t)| \ge 1/2$.

Proof. The pointwise convergence of the characteristic functions follows directly from the definition of weak convergence. Indeed, since $f(x) := e^{\imath x\xi}$ is for each $\xi \in \mathbb{R}$ continuous and bounded, we have

$$\phi_n(\xi) := \int e^{\imath x\xi} d\mu_n(x) \to \int e^{\imath x\xi} d\mu(x) =: \phi(\xi).$$

The uniform convergence on compact intervals is more delicate.

Step 1: The family of measure $\{\mu_n; n \in \mathbb{N}\}$ is tight, i.e. for any $\varepsilon > 0$ there exists a compact set K such that

$$\mu_n(K^c) \leq \varepsilon$$

Proof. Choose $r > 0$ such that $\mu(B[0,r]) < \varepsilon$ and set $K := B[0,2r]$. Pick a continuous function χ satisfying $1_{B[0,r]} \leq \chi \leq 1_K$. In particular, we have $1-\chi(x) = 1$ for any $x \in K^c$ and $1-\chi(x) = 0$ for any $x \in B[0,r]$. Hence,

$$\mu_n(K^c) \leq \int (1-\chi) d\mu_n \overset{n\to\infty}{\to} \int (1-\chi) d\mu \leq \mu(B[0,r]^c) < \varepsilon.$$

This shows that for $n \geq N$ sufficiently large, $\mu_n(K^c) \leq \varepsilon$. Enlarging K yields

$$\mu_n(K^c) \leq \varepsilon$$

for all $n \in \mathbb{N}$. □

Step 2: $(\varphi_n)_{n\in\mathbb{N}}$ is uniformly equicontinuous, i.e. for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any $|\xi - \eta| \leq \delta$ and $n \in \mathbb{N}$, we have

$$|\phi_n(\xi) - \phi_n(\eta)| \leq \varepsilon$$

Proof. Let $\varepsilon > 0$ and K as in step 1. Since the mapping $\xi \mapsto e^{\imath\xi}$ is continuous, we can pick $\delta > 0$ such that

$$\left|1 - e^{\imath x(\xi-\eta)}\right| \leq \varepsilon$$

for any $|\xi - \eta| \leq \delta$ and $x \in K$. Consequently,

$$\begin{aligned}|\phi_n(\eta) - \phi_n(\xi)| &\leq \int_K \underbrace{\left|e^{\imath\xi x} - e^{\imath\eta x}\right|}_{|1-e^{\imath x(\xi-\eta)}|\leq\varepsilon} d\mu_n(x) + \int_{K^c} \underbrace{\left|e^{\imath\xi x} - e^{\imath\eta x}\right|}_{\leq 2} d\mu_n(x) \\ &\leq \varepsilon + 2\mu_n(K^c) \leq 3\varepsilon.\end{aligned}$$

□

Step 3: Fix $\xi \in \mathbb{R}$ and $\varepsilon > 0$. Choose δ as in step 2. Since ϕ is continuous, we may assume that

$$|\phi(\xi) - \phi(\eta)| \leq \varepsilon$$

for any $\eta \in B(\xi, \delta)$. Hence,

$$|\phi_n(\eta) - \phi(\eta)| \leq \underbrace{|\phi_n(\eta) - \phi_n(\xi)}_{\leq\varepsilon} + \underbrace{|\phi_n(\xi) - \phi(\xi)}_{\substack{n\to\infty \\ \to 0}} + \underbrace{|\phi(\xi) - \phi(\eta)|}_{\leq\varepsilon}.$$

Letting $n \to \infty$ and $\varepsilon \to 0$ shows local uniform convergence. Since local uniform convergence is equivalent to uniform convergence on compact sets, this finishes the proof. □

(b) Show that there exists $t_0 > 0$ such that if $t \in [-t_0, t_0]$, then $(\exp(\mathrm{i}tS_n)/\Phi_n(t))_{n\geq 0}$ is a $(\mathcal{F}_n)-$martingale (i.e. both its real and imaginary parts are martingales).

(c) Prove that we can choose $t_0 > 0$ such that for any $t \in [-t_0, t_0]$, $\lim_{n\to\infty} \exp(\mathrm{i}tS_n)$ exists $\mathbb{P}$-a.s.

(d) Set

$$C = \Big\{(t, \omega) \in [-t_0, t_0] \times \Omega : \lim_{n\to\infty} \exp(\mathrm{i}tS_n(\omega)) \text{ exists}\Big\}$$

Prove that C is measurable, i.e. in the product of $\mathcal{B}([-t_0, t_0])$ with $\mathcal{F}$.

(e) Establish that $\int_{-t_0}^{t_0} \mathbb{1}_C(t, \omega)\mathbb{P}(\mathrm{d}\omega)\mathrm{d}t = 2t_0$.

(f) Prove that $\lim_{n\to\infty} S_n$ exists $\mathbb{P}$-a.s.

Appendix A

Practices for Final

A.1 Practice Final

1. (a) State the Borel-Cantelli lemma.

   (b) State the strong law of large numbers.

   (c) State the central limit theorem.

2. Prove that convergence in probability implies almost sure convergence along a subsequence.

3. On the same probability space, let X and Y be two bounded random variables, i.e. there exists $C > 0$ such that $|X(\omega)| + |Y(\omega)| < C$ for any $\omega \in \Omega$. Prove that X and Y are independent if and only if for any $k, \ell \in \mathbb{N}$, we have

$$\mathbb{E}\left[X^k Y^\ell\right] = \mathbb{E}\left[X^k\right] \cdot \mathbb{E}\left[Y^\ell\right]$$

4. Let $(p_n)_{n\geq 1}$ be a sequence of real numbers in $[0, 1]$ converging to $p \in (0, 1)$. Let Y_n be a random variable, binomial with parameters n and p_n : Y_n is equal in distribution to the sum of n independent Bernoulli random variables with parameter p_n. State and prove a central limit theorem for Y_n.

5. Let $n \geq 2$ be fixed and consider the Markov chain corresponding to the standard random walk on $\mathbb{Z}^2 \times (\mathbb{Z}/n\mathbb{Z})$:

$$\pi\left(((x, y), z), \left((x', y'), z'\right)\right) = \begin{cases} \frac{1}{8} & \text{if } |(x, y) - (x', y')|_2 = 1, z - z' = \pm 1 \bmod n \\ 0 & \text{otherwise} \end{cases}$$

   Is it transcient? Null recurrent? Positive recurrent?

6. Let $(U_n)_{n\geq 0}$ be a sequence of i.i.d random variables, with uniform distribution on $[0,1]$, on $(\Omega,\mathcal{F},\mathbb{P})$. Let $\mathcal{F}_0=\{\emptyset,\Omega\}$ and $\mathcal{F}_n=\sigma\left(U_1,\ldots,U_n\right)$. Define a sequence $(X_n)_{n\geq 0}$ through

$$X_0=p\in(0,1),\quad X_{n+1}=\theta X_n+(1-\theta)\mathbb{1}_{[0,X_n]}\left(U_{n+1}\right),$$

where $\theta\in(0,1)$ is given.

   (a) Prove that X is a $(\mathcal{F}_n)_{N\geq 0}$-martingale included in $[0,1]$.

   (b) Prove that X converges a.s. and in any L^p to a random variable denoted L.

   (c) What is the distribution of L?

A.2 Final

1. (a) What is a σ-field?

   (b) State the central limit theorem for i.i.d. random variables.

   (c) State the Radon-Nikodym theorem.

2. (a) Does convergence in distribution imply convergence in probability? If yes, prove it, it no, give a counterexample.

   (b) Does convergence in distribution to a constant imply convergence in probability? If yes, prove it, it no, give a counterexample.

3. Give an example of random variables X,Y,Z such that (X,Y), (X,Z), and (Y,Z) are Gaussian vectors, but not (X,Y,Z).

4. Consider the following Markov chain on $\mathbb{N}\cup\{0\}$:

$$p_{ij}=\begin{cases}1/3 & j=i+1\\ 2/3 & j=i-1, i>0\\ 2/3 & i=j=0\\ 0 & \text{otherwise}\end{cases}$$

   Is it transcient? Null recurrent? Positive recurrent?

5. Consider the random walk $S_n=\sum_k^n X_k$, the X_k 's being i.i.d., $\mathbb{P}\left(X_1=1\right)=\mathbb{P}\left(X_1=-1\right)=1/2, \mathcal{F}_n=\sigma\left(X_i, 0\leq i\leq n\right)$.

   (a) Prove that $\left(S_n^2-n, n\geq 0\right)$ is a $\left(\mathcal{F}_n\right)-$ martingale.

   (b) Let τ be a bounded stopping time. Prove that $\mathbb{E}\left(S_\tau^2\right)=\mathbb{E}(\tau)$.

   (c) Take now $\tau=\inf\{n\mid S_n\in\{-a,b\}\}$, where $a,b\in\mathbb{N}^*$. Prove that $\mathbb{E}\left(S_\tau\right)=0$ and $\mathbb{E}\left(S_r^2\right)=\mathbb{E}(\tau)$.

(d) What is $\mathbb{P}(S_\tau = -a)$? What is $\mathbb{E}(\tau)$?

6. We consider a collection $(X_n)_{n\geq 1}$ of random variables, not necessarily independent, such that the following holds. There exists $M > 0$ such that $\mathbb{E}\left[|X_n|^2\right] < M$ for all $n, \mathbb{E}[X_n] = 0$ and $\operatorname{Cov}(X_n, X_m) = 0$ for all $n \neq m$. Prove that $\frac{X_1+X_2+\cdots+X_n}{n}$ converges to 0 almost surely as $n \to \infty$.

Bibliography

[1] Billingsley, Patrick. *Probability and Measure, anniversary ed.* Wiley, 2012.

[2] Grimmett, Geoffrey R. Stirzaker, David R. *Probability and Random Processes, 3rd ed.* Oxford University Press, 2001.

[3] Varadhan, S. R. S. *Courant Lecture Notes: Probability Theory*. American Mathematical Society, 2001.

[4] Walsh, John B. *Knowing the Odds: An Introduction to Probability*. American Mathematical Society, 2012.