

Probability Theory

Ikhan Choi

September 19, 2022

Contents

I	Probability distributions	2
1	Random variables	3
1.1	Probability distributions	3
1.2	Discrete distributions	4
1.3	Continuous distributions	4
1.4	Independence	4
2	Conditional probability	5
3	Convergence of probability measures	6
3.1	Weak convergence in \mathbb{R}	6
3.2	Weak topology in the space of probability measures	6
3.3	Characteristic functions	9
3.4	Moments	10
II	Discrete stochastic process	11
4	Limit theorems	12
4.1	Laws of large numbers	12
4.2	Renewal theory	14
4.3	Central limit theorems	14
5	Martingales	16
5.1	Submartingales	16
5.2	Martingale convergence theorem	16
5.3	Convergence in L^p and uniform integrability	16
5.4	Optional stopping theorem	16
6	Markov chains	17
III	Continuous stochastic processes	18
7	Brownian motion	19
7.1	Kolomogorov extension	19
IV	Stochastic calculus	20

Part I

Probability distributions

Chapter 1

Random variables

1.1 Probability distributions

1.1 (Sample space). Mathematically, a *sample space* is defined as a measure space (Ω, \mathcal{F}, P) with $P(\Omega) = 1$. Elements and measurable subsets of a sample space are called *outcomes* and *events*, respectively. Let Ω be a fixed sample space. Then, a *random element* is a measurable function $X : \Omega \rightarrow S$ to a measurable space S , called the *state space*. If the state space S is the set of real numbers \mathbb{R} together with the Borel σ -algebra, we call the random element X as a *random variable*.

Consider a statistical study of ages of people in the earth. For the study, we set the *population* $\mathcal{P} = \{ \text{people in the earth} \}$ and the age function $a : \mathcal{P} \rightarrow \mathbb{Z}_{\geq 0}$. In probability theory and statistics, we are interested in the estimation of the size of $a^{-1}(k)$ for each $k \in \mathbb{Z}_{\geq 0}$, not in the exact description of the age function a .

Let us say that we conducted an experiment in which n people are randomly chosen with replacement, to verify a hypothesis. If we denote by p_i the i th person, then

Then a reasonable choice for the domain of the functions X_i is $\Omega = \mathcal{P}^n$.

Believing the fatalism, an experiment can be seen as a process of revealing a pre-determined fate ω , which is what we call sample or outcome.

(a)

1.2 (Probability distribution). Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. The (probability) *distribution* of X is the pushforward measure X_*P on \mathbb{R} . The right continuous increasing function F corresponded to X_*P is called the (cumulative) *distribution function*.

If the distribution has discrete support, then we say X is *discrete*. Since a probability measure of discrete support is a countable convex combination of Dirac measures, we can define the (probability) *mass function* $p : \text{supp}(X_*P) \rightarrow [0, 1]$. If the distribution is absolutely continuous with respect to the Lebesgue measure, then we say X is *continuous*. By the Radon-Nikodym theorem, we can define the (probability) *density function* $f \in L^1(\mathbb{R})$. The mass and density functions are effective ways to describe distributions of random variables in most applications.

(a)

1.3 (Expectation and moments). Chebyshev's inequality

1.4 (Joint distribution).

1.5 (Distribution of functions). transformation, function

1.2 Discrete distributions

1.3 Continuous distributions

1.4 Independence

1.6 (Dynkin's π - λ lemma). Let \mathcal{P} be a π -system and \mathcal{L} a λ -system respectively. Denote by $\ell(\mathcal{P})$ the smallest λ -system containing \mathcal{P} .

- (a) If $A \in \ell(\mathcal{P})$, then $\mathcal{G}_A := \{B : A \cap B \in \ell(\mathcal{P})\}$ is a λ -system.
- (b) $\ell(\mathcal{P})$ is a π -system.
- (c) If a λ -system is a π -system, then it is a σ -algebra.
- (d) If $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

1.7 (Monotone class lemma).

Exercises

equally likely outcomes coin toss dice roll ball drawing number permutation life time of a light bulb

Chapter 2

Conditional probability

Exercises

2.1 (Monty Hall problem). Suppose you are on a game show, and given the choice of three doors A , B , and C . Behind one door is a car; behind the others, goats. You know that the probabilities a , b , and $c = 1 - a - b$. You pick a door, say A , and the host, who knows what's behind the doors, opens another door, say B , which has a goat. He then says to you, "Do you want to pick door C ?" Is it to your advantage to switch your choice?

- (a) Find the condition for a, b, c that the participant benefits when changed the choice.

Proof. Let A , B , and C be the events that a car is behind the doors A , B , and C , respectively. Let X be the event that the game host opened B . Note $\{A, B, C\}$ is a partition of the sample space Ω , and X is independent to A , B , and C . Then, $P(A) = P(B) = P(C) = 1/3$, and

$$P(X|A) = \frac{1}{2}, \quad P(X|B) = 0, \quad P(X|C) = 1.$$

Therefore,

$$\begin{aligned} P(C|X) &= \frac{P(X \cap C)}{P(X)} \\ &= \frac{P(X|C)P(C)}{P(X|A)P(A) + P(X|B)P(B) + P(X|C)P(C)} \\ &= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{2}{3}. \end{aligned}$$

Similarly, $P(A|X) = \frac{1}{3}$ and $P(B|X) = 0$. □

Chapter 3

Convergence of probability measures

3.1 Weak convergence in \mathbb{R}

3.1 (Portemanteau theorem). Let F_n and F be distribution functions $\mathbb{R} \rightarrow [0, 1]$. We will define the *weak convergence* as follows: F_n converges weakly to F if $F_n(x) \rightarrow F(x)$ for every continuity point x of $F(x)$.

(a) $F_n(x) \rightarrow F(x)$ for all continuity points x of F .

3.2 (Skorokhod representation theorem).

3.3 (Continuous mapping theorem).

3.4 (Slutsky's theorem).

3.5 (Helly's selection theorem). (a) Monotonically increasing functions $F_n : \mathbb{R} \rightarrow [0, 1]$ has a point-wise convergent subsequence.

(b) If $(F_n)_n$ is tight, then

3.6 (Properties of probability Borel measures). Let S be a topological space.

(a) Every single probability Borel measure is regular if S is perfectly normal. (inner approximation by closed sets)

(b) Every single probability Borel measure is tight if S is Polish. (inner approximation by compact sets)

3.2 Weak topology in the space of probability measures

3.7 (Local limit theorems). Suppose f_n and f are density functions.

(a) If $f_n \rightarrow f$ a.s., then $f_n \rightarrow f$ in L^1 . (Scheffé's theorem)

(b) $f_n \rightarrow f$ in L^1 if and only if in total variation.

(c) If $f_n \rightarrow f$ in total variation, then $f_n \rightarrow f$ weakly.

3.8 (Portmanteau theorem). Let S be a normal space and, μ_α be a net in $\text{Prob}(S)$. We define the *weak convergence* as follows: μ_α converges weakly to μ if

$$\int f d\mu_\alpha \rightarrow \int f d\mu$$

for every $f \in C_b(S)$. The following statements are all equivalent.

- (a) $\mu_\alpha \Rightarrow \mu$
- (b) $\mu_\alpha(g) \rightarrow \mu(g)$ for every uniformly continuous $g \in C_b(S)$.
- (c) $\limsup_\alpha \mu_\alpha(F) \leq \mu(F)$ for every closed F .
- (d) $\liminf_\alpha \mu_\alpha(U) \geq \mu(U)$ for every open U .
- (e) $\lim_\alpha \mu_\alpha(A) = \mu(A)$ for every Borel A such that $\mu(\partial A) = 0$.

Proof. (a) \Rightarrow (b) Clear.

(b) \Rightarrow (c) Let U be an open set such that $F \subset U$. There is uniformly continuous $g \in C_b(S)$ such that $\mathbf{1}_F \leq g \leq \mathbf{1}_U$. Therefore,

$$\limsup_\alpha \mu_\alpha(F) \leq \limsup_\alpha \mu_\alpha(g) = \mu(g) \leq \mu(U).$$

By the outer regularity of μ , we obtain $\limsup_\alpha \mu_\alpha(F) \leq \mu(F)$.

(c) \Leftrightarrow (d) Clear.

(c)+(d) \Rightarrow (e) It easily follows from

$$\limsup_\alpha \mu_\alpha(\bar{A}) \leq \mu(\bar{A}) = \mu(A) = \mu(A^\circ) \leq \liminf_\alpha \mu_\alpha(A^\circ).$$

(e) \Rightarrow (a) Let $g \in C_b(S)$ and $\varepsilon > 0$. Since the pushforward measure $g_*\mu$ has at most countably many mass points, there is a partition $(t_i)_{i=0}^n$ of an interval containing $[-\|g\|, \|g\|]$ such that $|t_{i+1} - t_i| < \varepsilon$ and $\mu(\{x : g(x) = t_i\}) = 0$ for each i . Let $(A_i)_{i=0}^{n-1}$ be a Borel decomposition of S given by $A_i := g^{-1}([t_i, t_{i+1}))$, and define $f_\varepsilon := \sum_{i=0}^{n-1} t_i \mathbf{1}_{A_i}$ so that we have $\sup_{x \in S} |g_\varepsilon(x) - g(x)| \leq \varepsilon$. From

$$\begin{aligned} |\mu_\alpha(g) - \mu(g)| &\leq |\mu_\alpha(g - g_\varepsilon)| + |\mu_\alpha(g_\varepsilon) - \mu(g_\varepsilon)| + |\mu(g_\varepsilon) - \mu(g)| \\ &\leq \varepsilon + \sum_{i=0}^{n-1} |t_i| |\mu_\alpha(A_i) - \mu(A_i)| + \varepsilon, \end{aligned}$$

we get

$$\limsup_\alpha |\mu_\alpha(g) - \mu(g)| < 2\varepsilon.$$

Since ε is arbitrary, we are done. □

3.9 (Embedding by Dirac measures). Let S be a normal space.

- (a) $S \rightarrow \text{Prob}(S)$ is an embedding.
- (b) $S \subset \text{Prob}(S)$ is sequentially closed.
- (c)

Proof. (a) It uses Urysohn.

(b) It uses (b) \Rightarrow (c) of Portmanteau. □

3.10 (Lévy-Prokhorov metric). Let S be a metric space, and $\text{Prob}(S)$ be the set of probability (regular) Borel measures on S . Define $\pi : \text{Prob}(S) \times \text{Prob}(S) \rightarrow [0, \infty)$ such that

$$\pi(\mu, \nu) := \inf\{\alpha > 0 : \mu(A) \leq \nu(A^\alpha) + \alpha, \nu(A) \leq \mu(A^\alpha) + \alpha, \forall A \in \mathcal{B}(S)\},$$

where A^α is the α -neighborhood of A .

- (a) π is a metric.
- (b) $\mu_n \rightarrow \mu$ in π implies $\mu_n \Rightarrow \mu$.
- (c) $\mu_\alpha \Rightarrow \mu$ implies $\mu_\alpha \rightarrow \mu$ in π , if S is separable.

- (d) (S, d) is separable if and only if $(\text{Prob}(S), \pi)$ is separable.
- (e) (S, d) is compact if and only if $(\text{Prob}(S), \pi)$ is compact
- (f) (S, d) is complete if and only if $(\text{Prob}(S), \pi)$ is complete.

Proof. (c) □

3.11 (Direct direction of Prokhorov's theorem). Let S be a Polish space. Let $\text{Prob}(S)$ be the space of probability measures on S endowed with the topology of weak convergence. Prokhorov's theorem states that a subset of $\text{Prob}(S)$ is relatively compact if and only if it is tight. We prove one direction, in which the construction of a sufficiently large compact set is a main issue.

Let $\mu \in \text{Prob}(S)$ and let M be a relatively compact subset of $\text{Prob}(S)$.

- (a) Every open cover $\{B_\alpha\}_\alpha$ of S has a finite subcollection $\{B_i\}_i$ for each $\varepsilon > 0$ such that

$$\mu\left(\bigcup_i B_i\right) > 1 - \varepsilon.$$

- (b) Every open cover $\{B_\alpha\}_\alpha$ of S has a finite subcollection $\{B_i\}_i$ for each $\varepsilon > 0$ such that

$$\inf_{\mu \in M} \mu\left(\bigcup_i B_i\right) > 1 - \varepsilon.$$

- (c) M is tight: there is a compact $K \subset S$ for each $\varepsilon > 0$ such that

$$\inf_{\mu \in M} \mu(K) > 1 - \varepsilon.$$

Proof. (a) Since a separable metric space is Lindelöf, we may assume $\{B_\alpha\}_\alpha = \{B_i\}_{i=1}^\infty$ is countable. Then, we can deduce the conclusion from the continuity from below and the fact $\mu_0(S) = 1$.

- (b) Suppose that the conclusion is not true so that there are $\varepsilon > 0$ and a sequence $\mu_n \in M$ such that

$$\mu_n\left(\bigcup_{i=1}^n B_i\right) \leq 1 - \varepsilon.$$

If we take a subsequence $(\mu_{n_k})_k$ that converges weakly to $\mu \in \overline{M}$ using the compactness of \overline{M} , then by the Portmanteau theorem we have for any n that

$$\mu\left(\bigcup_{i=1}^n B_i\right) \leq \liminf_{k \rightarrow \infty} \mu_{n_k}\left(\bigcup_{i=1}^n B_i\right) \leq \liminf_{k \rightarrow \infty} \mu_{n_k}\left(\bigcup_{i=1}^{n_k} B_i\right) \leq 1 - \varepsilon.$$

By taking n sufficiently large, we lead a contradiction to the part (a).

(c) Here we need metrization, which leads to the existence of countable fundamental system of uniformity for $\frac{\varepsilon}{2^m}$ argument. Also we need the completeness to change the total boundedness to compactness.

Let $\{x_i\}_{i=1}^\infty$ be a dense set in S . Then, since $\{B(x_i, \frac{1}{m})\}_{i=1}^\infty$ is a countable open cover of S for each integer $m > 0$, there is a finite $n_m > 0$ such that

$$\inf_{\mu \in M} \mu\left(\bigcup_{i=1}^{n_m} B(x_i, \frac{1}{m})\right) > 1 - \frac{\varepsilon}{2^m}.$$

Define

$$K := \bigcap_{m=1}^\infty \bigcup_{i=1}^{n_m} \overline{B(x_i, \frac{1}{m})}.$$

It is closed and totally bounded in a complete metric space S , so K is compact. Moreover, we can verify

$$1 - \mu(K) = \mu\left(\bigcup_{m=1}^\infty \bigcap_{i=1}^{n_m} \overline{B(x_i, \frac{1}{m})}^c\right) \leq \sum_{m=1}^\infty \left(1 - \mu\left(\bigcup_{i=1}^{n_m} B(x_i, \frac{1}{m})\right)\right) < \varepsilon$$

for every $\mu \in M$, so M is tight. □

3.12 (Converse direction of Prokhorov's theorem). The “converse” direction of Prokhorov's theorem is related to a construction of measure and considered to be more difficult. However, it holds in a general setting.

Let S be a normal space. Let $\text{Prob}(S)$ be the space of probability measures on S endowed with the topology of weak convergence. Let M be a tight subset of $\text{Prob}(S)$ and let $(\mu_\alpha)_\alpha \subset M$ be a net. We want to show that it has a convergent subnet in $\text{Prob}(S)$.

(a) M is relatively compact.

Proof. Let βS be the Stone-Ćech compactification of S . The inclusion $\iota : S \rightarrow \beta S$ is a topological embedding because S is completely regular. Pushforward the measures μ_α to make them probability Borel measures $\nu_\alpha := \iota_* \mu_\alpha$ on βS . We want to take a convergent subnet of $\nu_\alpha \in \text{Prob}(\beta S)$, and to show the limit is in fact contained in $\text{Prob}(S)$.

Our first claim is that the measure ν_α is regular for each α , that is, $\nu_\alpha \in \text{Prob}(\beta S)$. For any Borel $E \subset \beta S$ and any $\varepsilon > 0$, there is $F \subset E \cap S$ that is closed in S such that $\mu_\alpha(E \cap S) < \mu_\alpha(F) + \varepsilon/2$ by inner regularity, and there is K that is compact in S such that $\mu_\alpha(S \setminus K) < \varepsilon/2$ by tightness. Then, the inequality

$$\nu_\alpha(E) = \mu_\alpha(E \cap S) < \mu_\alpha(F) + \frac{\varepsilon}{2} < \mu_\alpha(F \cap K) + \varepsilon = \nu_\alpha(F \cap K) + \varepsilon$$

proves the regularity of ν_α since $F \cap K$ is compact in both S and βS with $F \cap K \subset E$. The space $\text{Prob}(\beta S)$ is compact by the Banach-Alaoglu theorem and the Riesz-Markov-Kakutani representation theorem. Therefore, ν_α has a subnet ν_β that converges to $\nu \in \text{Prob}(\beta S)$.

Recall that μ_β is tight. For each $\varepsilon > 0$, there is a compact $K \subset S$ such that $\nu_\beta(K) = \mu_\beta(K) \geq 1 - \varepsilon$ for all β . Then, by the Portmanteau theorem, we have

$$\nu(S) \geq \nu(K) \geq \limsup_{\beta} \nu_\beta(K) \geq 1 - \varepsilon.$$

Since ε is arbitrary, ν is concentrated on S , i.e. $\nu(S) = 1$. Now we restrict ν to S in order to obtain μ , which is a probability Borel measure on S .

From the definition of weak convergence we have

$$\int_{\beta S} f d\nu_\beta \rightarrow \int_{\beta S} f d\nu$$

for all $f \in C(\beta S)$. Since $\nu_\beta(\beta S \setminus S) = \nu(\beta S \setminus S) = 0$ and the restriction $C(\beta S) \rightarrow C_b(S)$ is an isomorphism due to the universal property of βS ,

$$\int_S f d\mu_\beta \rightarrow \int_S f d\mu$$

for all $f \in C_b(S)$, so μ_β converges weakly to $\mu \in \text{Prob}(S)$. □

3.3 Characteristic functions

3.13 (Characteristic functions). Let μ be a probability measure on \mathbb{R} . Then, the *characteristic function* of μ is defined by

$$\varphi(t) := Ee^{itX} = \int e^{itx} d\mu(x).$$

Note that $\varphi(t) = \hat{\mu}(-t)$ where $\hat{\mu}$ is the Fourier transform of $\mu \in \mathcal{S}'(\mathbb{R})$.

(a) $\varphi \in C_b(\mathbb{R})$.

3.14 (Inversion formula). Let μ be a probability measure on \mathbb{R} and φ its characteristic function.

(a) For $a < b$, we have

$$\mu((a, b)) + \frac{1}{2}\mu(\{a, b\}) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.$$

(b) For $a \in \mathbb{R}$, we have

$$\mu(\{a\}) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-ita} \varphi(t) dt$$

(c) If $\varphi \in L^1(\mathbb{R})$, then μ has density

$$f(x) = \frac{1}{2\pi} \int e^{-itx} \varphi(t) dt$$

in $C_0(\mathbb{R}) \cap L^1(\mathbb{R})$.

3.15 (Lévy's continuity theorem). The continuity theorem provides with a tool to verify the weak convergence in terms of characteristic functions. Let μ_n and μ be probability distributions on \mathbb{R} with characteristic functions φ_n and φ .

(a) If $\mu_n \rightarrow \mu$ weakly, then $\varphi_n \rightarrow \varphi$ pointwise.

(b) If $\varphi_n \rightarrow \varphi$ pointwise and φ is continuous at zero, then $(\mu_n)_n$ is tight and $\mu_n \rightarrow \mu$ weakly.

Proof. (a) For each t ,

$$\varphi_n(t) = \int e^{itx} d\mu_n(x) \rightarrow \int e^{itx} d\mu(x) = \varphi(t)$$

because $e^{itx} \in C_b(\mathbb{R})$.

(b)

□

3.16 (Criteria for characteristic functions). Bochner's theorem and Polya's criterion

There are two ways to represent a measure: A measure μ is absolutely continuous iff its distribution F is absolutely continuous iff its density f is integrable. So, the fourier transform of an absolutely continuous measure is just the fourier transform of L^1 functions.

3.4 Moments

moment problem

moment generating function defined on $|t| < \delta$

Exercises

3.17. Let φ_n be characteristic functions of probability measures μ_n on \mathbb{R} . If there is a continuous function φ such that $\varphi_n = \varphi$ on $n^{-1}\mathbb{Z}$, then μ_n converges weakly.

3.18 (Convergence determining class).

3.19 (Vague convergence). Let S be a locally compact Hausdorff space.

(a) $\mu_\alpha \rightarrow \mu$ vaguely if and only if $\int g d\mu_\alpha \rightarrow \int g d\mu$ for all $g \in C_c(S)$.

(b) $\mu_\alpha \rightarrow \mu$ weakly if and only if vaguely.

(c) $\delta_n \rightarrow 0$ vaguely but not weakly. (escaping to infinity)

Proof.

□

Part II

Discrete stochastic process

Chapter 4

Limit theorems

4.1 Laws of large numbers

Our purpose is to find appropriate a_n and slowly growing b_n such that $(S_n - a_n)/b_n \rightarrow 0$ in probability or almost surely.

4.1 (Kolmogorov-Feller theorem). Let X_i be an uncorrelated sequence of random variables such that

$$\lim_{x \rightarrow \infty} \sup_i xP(|X_i| > x) = 0.$$

This condition is called the *Kolmogorov-Feller* condition. Let $Y_{n,i} := X_i \mathbf{1}_{|X_i| \leq c_n}$.

(a) We have

$$\lim_{n \rightarrow \infty} P(S_n \neq T_n) = 0$$

if $n \lesssim c_n$.

(b) We have

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{T_n - ET_n}{b_n}\right| > \varepsilon\right) = 0$$

if $nc_n \lesssim b_n^2$.

(c) We have

$$\frac{S_n - ET_n}{n} \rightarrow 0$$

in probability.

Proof. Write $g(x) := \sup_i xP(|X_i| > x)$ so that $g(x) \rightarrow 0$ as $x \rightarrow \infty$.

(a) It follows from

$$P(S_n \neq T_n) \leq \sum_{i=1}^n P(|X_i| > c_n) \leq \sum_{i=1}^n \frac{1}{c_n} g(c_n) \lesssim g(c_n).$$

(b) We write

$$\begin{aligned}
P\left(\left|\frac{T_n - ET_n}{b_n}\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2 b_n^2} E|T_n - ET_n|^2 \\
&= \frac{1}{\varepsilon^2 b_n^2} \sum_{i=1}^n E|Y_{n,i} - EY_{n,i}|^2 \\
&\leq \frac{1}{\varepsilon^2 b_n^2} \sum_{i=1}^n E|X_i \mathbf{1}_{|X_i| \leq c_n}|^2 \\
&= \frac{1}{\varepsilon^2 b_n^2} \sum_{i=1}^n \int_0^{c_n} 2xP(|X_i| > x) dx \\
&\leq \frac{2n}{\varepsilon^2 b_n^2} \int_0^{c_n} g(x) dx \\
&= \frac{2nc_n}{\varepsilon^2 b_n^2} \int_0^1 g(c_n x) dx \\
&\lesssim \int_0^1 g(c_n x) dx.
\end{aligned}$$

Since $g(x) \leq x$ and $g(x) \rightarrow 0$ as $x \rightarrow \infty$, the function g is bounded. By the bounded convergence theorem, we get $\int_0^1 g(c_n x) dx \rightarrow 0$ as $n \rightarrow \infty$. \square

4.2 (St. Petersburg paradox). We want see the asymptotic behavior of the partial sums S_n of i.i.d. random variables X_i such that $E|X_i| = \infty$. Let

$$P(X_n = 2^m) = 2^{-m} \quad \text{for } m \geq 1.$$

Let $Y_{n,i} := X_i \mathbf{1}_{|X_i| \leq c_n}$.

(a) We have

$$\lim_{n \rightarrow \infty} P(S_n \neq T_n) = 0$$

if $n \ll c_n$.

(b) We have

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{T_n - ET_n}{b_n}\right| > \varepsilon\right) = 0$$

if $nc_n \ll b_n^2$.

(c) We have

$$\frac{S_n - n \log_2 n}{n^{1+\varepsilon}} \rightarrow 0$$

in probability for every $\varepsilon > 0$.

Proof. (a) It follows from

$$P(S_n \neq T_n) \leq \sum_{i=1}^n P(X_i \neq Y_{n,i}) = \sum_{i=1}^n P(|X_i| > c_n) \leq \sum_{i=1}^n \frac{2}{c_n} = \frac{2n}{c_n}.$$

(b) It follows from

$$\begin{aligned}
P\left(\left|\frac{T_n - ET_n}{b_n}\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2 b_n^2} E|T_n - ET_n|^2 \\
&= \frac{1}{\varepsilon^2 b_n^2} \sum_{i=1}^n E|Y_{n,i} - EY_{n,i}|^2 \\
&\leq \frac{1}{\varepsilon^2 b_n^2} \sum_{i=1}^n E|X_i \mathbf{1}_{|X_i| \leq c_n}|^2 \\
&\leq \frac{1}{\varepsilon^2 b_n^2} n \cdot 2c_n
\end{aligned}$$

□

4.3 (Borel-Cantelli lemmas).

4.4 (Head runs).

4.5 (Strong laws of large numbers for L^1). Proof by Ettemadi

Random series proof

4.2 Renewal theory

4.3 Central limit theorems

4.6 (Central limit theorem for L^3). Replacement method by Lindeman and Lyapunov

4.7 (Lindeberg-Feller theorem). Let X_i be independent random variables such that for every $\varepsilon > 0$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n E|X_i - EX_i|^2 \mathbf{1}_{|X_i - EX_i| > \varepsilon s_n} = 0.$$

This condition is called the *Lindeberg-Feller* condition. Let $Y_{n,i} := \frac{X_i - EX_i}{s_n}$.

(a) We have

$$|Ee^{it(S_n - ES_n)/s_n} - e^{-\frac{1}{2}t^2}| \leq \sum_{i=1}^n |Ee^{itY_{n,i}} - e^{-\frac{1}{2}E(tY_{n,i})^2}|.$$

(b) For any $\varepsilon > 0$, we have an estimate

$$\left|Ee^{itY} - \left(1 - \frac{1}{2}E(tY)^2\right)\right| \lesssim_t \varepsilon EY^2 + EY^2 \mathbf{1}_{|Y| > \varepsilon}$$

for all random variables Y such that $EY^2 < \infty$.

(c) For any $\varepsilon > 0$, we have an estimate

$$\left|e^{-\frac{1}{2}E(tY)^2} - \left(1 - \frac{1}{2}E(tY)^2\right)\right| \lesssim_t EY^2(\varepsilon^2 + EY^2 \mathbf{1}_{|Y| > \varepsilon}).$$

for all random variables Y such that $EY^2 < \infty$.

(d)

Proof. (a) Note

$$Ee^{it(S_n - ES_n)/s_n} = \prod_{i=1}^n Ee^{itY_{n,i}} \quad \text{and} \quad e^{-\frac{1}{2}t^2} = \prod_{i=1}^n e^{-\frac{1}{2}E(tY_{n,i})^2}.$$

(b) Since

$$\left| e^{ix} - \left(1 + ix - \frac{1}{2}x^2\right) \right| = \left| \frac{i^3}{2} \int_0^x (x-y)^2 e^{iy} dy \right| \leq \min\left\{\frac{1}{6}|x|^3, x^2\right\}$$

for $x \in \mathbb{R}$, we have

$$\begin{aligned} \left| Ee^{itY} - \left(1 - \frac{1}{2}E(tY)^2\right) \right| &\leq E \left| e^{itY} - \left(1 - \frac{1}{2}(tY)^2\right) \right| \\ &\lesssim_t E \min\{|Y|^3, Y^2\} \\ &\leq E|Y|^3 \mathbf{1}_{|Y| \leq \varepsilon} + EY^2 \mathbf{1}_{|Y| > \varepsilon} \\ &\leq \varepsilon EY^2 + EY^2 \mathbf{1}_{|Y| > \varepsilon}. \end{aligned}$$

(c) Since

$$|e^{-x} - (1-x)| = \left| \int_0^x (x-y)e^{-y} dy \right| \leq \frac{1}{2}x^2$$

for $x \geq 0$, we have

$$\left| e^{-\frac{1}{2}E(tY)^2} - \left(1 - \frac{1}{2}E(tY)^2\right) \right| \lesssim_t (EY^2)^2 \leq EY^2(\varepsilon^2 + EY^2 \mathbf{1}_{|Y| > \varepsilon}).$$

□

4.8. Let $X_n : \Omega \rightarrow \mathbb{R}$ be independent random variables. If there is $\delta > 0$ such that the *Lyapunov condition*

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E|X_i - EX_i|^{2+\delta} = 0$$

is satisfied, then

$$\frac{S_n - ES_n}{s_n} \rightarrow N(0, 1)$$

weakly, where $S_n := \sum_{i=1}^n X_i$ and $s_n^2 := VS_n$.

Berry-Esseen inequality

Exercises

4.9 (Bernstein polynomial). Let $X_n \sim \text{Bern}(x)$ be i.i.d. random variables. Since $S_n \sim \text{Binom}(n, x)$, $E(S_n/n) = x$, $V(S_n/n) = x(1-x)/n$. The L^2 law of large numbers implies $E(|S_n/n - x|^2) \rightarrow 0$. Define $f_n(x) := E(f(S_n/n))$. Then, by the uniform continuity $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$,

$$|f_n(x) - f(x)| \leq E(|f(S_n/n) - f(x)|) \leq \varepsilon + 2\|f\|P(|S_n/n - x| \geq \delta) \rightarrow \varepsilon.$$

4.10 (High-dimensional cube is almost a sphere). Let $X_n \sim \text{Unif}(-1, 1)$ be i.i.d. random variables and $Y_n := X_n^2$. Then, $E(Y_n) = \frac{1}{3}$ and $V(Y_n) \leq 1$.

4.11 (Coupon collector's problem). $T_n := \inf\{t : |\{X_i\}_i| = n\}$ Since $X_{n,k} \sim \text{Geo}(1 - \frac{k-1}{n})$, $E(X_{n,k}) = (1 - \frac{k-1}{n})^{-1}$, $V(X_{n,k}) \leq (1 - \frac{k-1}{n})^{-2}$. $E(T_n) \sim n \log n$

4.12 (An occupancy problem).

4.13. Find the probability that arbitrarily chosen positive integers are coprime.

Poisson convergence, law of rare events, or weak law of small numbers (a single sample makes a significant attribution)

Chapter 5

Martingales

5.1 Submartingales

5.2 Martingale convergence theorem

5.1 (Doob's upcrossing inequality). (a)

5.2 (Martingale convergence theorems). (a)

5.3. (a)

5.3 Convergence in L^p and uniform integrability

5.4 Optional stopping theorem

Chapter 6

Markov chains

Part III

Continuous stochastic processes

Chapter 7

Brownian motion

7.1 Kolmogorov extension

7.1 (Kolmogorov extension theorem). A *rectangle* is a finite product $\prod_{i=1}^n A_i \subset \mathbb{R}^n$ of measurable $A_i \subset \mathbb{R}$, and *cylinder* is a product $A^* \times \mathbb{R}^{\mathbb{N}}$ where A^* is a rectangle. Let \mathcal{A} be the semi-algebra containing \emptyset and all cylinders in $\mathbb{R}^{\mathbb{N}}$. Let $(\mu_n)_n$ be a sequence of probability measures on \mathbb{R}^n that satisfies *consistency condition*

$$\mu_{n+1}(A^* \times \mathbb{R}) = \mu_n(A^*)$$

for any rectangles $A^* \subset \mathbb{R}^n$, and define a set function $\mu_0 : \mathcal{A} \rightarrow [0, \infty]$ by $\mu_0(A) = \mu_n(A^*)$ and $\mu_0(\emptyset) = 0$.

- (a) μ_0 is well-defined.
- (b) μ_0 is finitely additive.
- (c) μ_0 is countably additive if $\mu_0(B_n) \rightarrow 0$ for cylinders $B_n \downarrow \emptyset$ as $n \rightarrow \infty$.
- (d) If $\mu_0(B_n) \geq \delta$, then we can find decreasing $D_n \subset B_n$ such that $\mu_0(D_n) \geq \frac{\delta}{2}$ and $D_n = D_n^* \times \mathbb{R}^{\mathbb{N}}$ for a compact rectangle D_n^* .
- (e) If $\mu_0(B_n) \geq \delta$, then $\bigcap_{i=1}^{\infty} B_i$ is non-empty.

Proof. (d) Let $B_n = B_n^* \times \mathbb{R}^{\mathbb{N}}$ for a rectangle $B_n^* \subset \mathbb{R}^n$. By the inner regularity of $\mu_{r(n)}$, there is a compact rectangle $C_n^* \subset B_n^*$ such that

$$\mu_0(B_n \setminus C_n) = \mu_{r(n)}(B_n^* \setminus C_n^*) < \frac{\delta}{2^{n+1}}.$$

Let $C_n := C_n^* \times \mathbb{R}^{\mathbb{N}}$ and define $D_n := \bigcap_{i=1}^n C_i = D_n^* \times \mathbb{R}^{\mathbb{N}}$. Then,

$$\mu_0(B_n \setminus D_n) \leq \mu_0\left(\bigcup_{i=1}^n B_n \setminus C_i\right) \leq \mu_0\left(\bigcup_{i=1}^n B_i \setminus C_i\right) < \frac{\delta}{2},$$

which implies $\mu_0(D_n) \geq \frac{\delta}{2}$.

(e) Take any sequence $(\omega_n)_n$ in $\mathbb{R}^{\mathbb{N}}$ such that $\omega_n \in D_n$. Since each $D_n^* \subset \mathbb{R}^{r(n)}$ is compact and non-empty, by diagonal argument, we have a subsequence $(\omega_k)_k$ such that ω_k is pointwise convergent, and its limit is contained in $\bigcap_{i=1}^{\infty} D_i \subset \bigcap_{i=1}^{\infty} B_i = \emptyset$, which is a contradiction that leads $\mu_0(B_n) \rightarrow 0$. \square

Part IV

Stochastic calculus