Probability Theory

Ikhan Choi

October 11, 2024

Contents

Part I

Probability distributions

Chapter 1

Random variables

1.1 Probability distributions

1.1 (Sample space). A *sample space* is a probability space, that is, a measure space $(\Omega, \mathcal{F}, P)$ 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 $X : \Omega \to S$ to a measurable space S, called the *state space*. The state space S is usally taken to be a Polish space together with its Borel σ-algebra. If $S = \mathbb{R}$ or $\mathbb{R}^d$, then we call the random element X as a *random variable* or *random vector* respectively.

Consider a statistical study of ages of people in the earth at a time. We conduct an experiment in which n people are randomly chosen with replacement in order to verify a hypothesis. We set the *population* $\mathcal{P}$ be the set of all people in the earth and the age function $a : \mathcal{P} \to \mathbb{Z}_{\geq 0}$. If we denote by X_i the age of ith person, then the reasonable choice for the domain of the random variables X_i is $\Omega = \mathcal{P}^n$, since the independence of X_i and X_j for $i \neq j$ can be easily realized by defining $X_i(p_1, p_2, \cdots) := a(p_i)$ by the product measure. In probability theory and statistics, we are interested in the distribution of age, that is, 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, and it is expected to be achieved approximately as n tends to infinity. Believing the determinism, an experiment is in fact recognized as an operation of revealing a pre-determined fate ω in the universal space Ω of possible world lines. The sample space Ω can be sufficiently enlarged when we require a finer domain of discourse such as the case $n \to \infty$, and we do not care of any concrete description of Ω except when discussing the mathematical existence issues.

1.2 (Probability distribution). Let $X : \Omega \to S$ be a random element, where S is a topological space. The (probability) *distribution* of X is the pushforward measure X_*P on $\mathbb{R}$. The right continuous non-decreasing 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* $p : \operatorname{supp}(X_*P) \to [0, 1]$. 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) Every single probability Borel measure on S 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)

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

Exercises

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

Chapter 2

Independence

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

2.2 (Monotone class lemma)**.**

2.3 (Kolmogorov extension theorem)**.** Let $\{(S_i, \mu_i)\}$ be a family of probability spaces. Let $S := \prod_i S_i$ be the product set with projections $\pi_i : S \to S_i$ and $\pi_J : S \to S_J := \prod_{j \in J} S_j$ for finite $J \subset I$ so that S_J is a probability space with the probability measure μ_J by the Fubini theorem. A *cylinder set* is a subset of the form $\pi_J^{-1}(A_J) \subset S$, where A_J is measurable in S_J for some J. Let $\mathcal{A}$ be the set of all cylinder sets in S.

Suppose the family $\{\mu_J\}$ satisfies the *consistency condition* for cylinder sets. Then, we can define a set function $\mu_0 : \mathcal{A} \to [0, \infty]$ by $\mu_0(A) := \mu_J(\pi_J(A))$ for $A \in \mathcal{A}$.

Proof. To apply the Carathéodory extension for outer measures to extend μ_0 to a measure, we need to check that μ_0 is monotonically countably additive and every cylinder set is Carathéodory measurable.

Let $C_i \in \mathcal{A}$ be a sequence of cylinder sets that covers a cylinder set $C \in \mathcal{A}$. Then,

$$\mu_0(C) \le \sum_i \mu_0(C_i).$$

□

(a) μ_0 is well-defined.

(b) μ_0 is finitely additive.

(c) μ_0 is countably additive if $\mu_0(B_n) \to 0$ for cylinders $B_n \downarrow \varnothing$ as $n \to \infty$.

(d) If $\mu_0(B_n) \ge \delta$, then we can find decreasing $D_n \subset B_n$ such that $\mu_0(D_n) \ge \frac{\delta}{2}$ and $D_n = D_n^* \times \mathbb{R}^{\mathbb{N}}$ for a compact rectangle D_n^*.

Proof. (d) Let $B_n = B_n^* \times \mathbb{R}^{\mathbb{N}}$ for a rectangle $B_n^* \subset \mathbb{R}^{r(n)}$. 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,

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

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

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 $\bigcap_{i=1}^\infty D_i \subset \bigcap_{i=1}^\infty B_n = \varnothing$, which is a contradiction that leads $\mu_0(B_n) \to 0$. □

2.1 Conditional probability

filtered probability space disintegration

Exercises

2.4 (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 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 distributions

3.1 Convergence in distribution

For a Polish space S, let $\mathrm{Prob}(S)$ be the space of probability Borel measures. Note that regularity automatically follows for finite Borel measures on a metrizable space.

3.1 (Portmanteau theorem)**.** Let S be a metrizable space. We say a net μ_i in $\mathrm{Prob}(S)$ *converges in distribution* or *weakly* to μ if

$$\int f\,d\mu_i \to \int f\,d\mu, \qquad f \in C_b(S).$$

The following statements are all equivalent.

(a) $\mu_i \to \mu$ in distribution.

(b) $\mu_i(g) \to \mu(g)$ for every uniformly continuous $g \in C_b(S)$.

(c) $\limsup_i \mu_i(F) \le \mu(F)$ for every closed $F \subset S$.

(d) $\liminf_i \mu_i(U) \ge \mu(U)$ for every open $U \subset S$.

(e) $\lim_i \mu_i(A) = \mu(A)$ for every Borel $A \subset S$ such that $\mu(\partial A) = 0$.

Proof. (a)⇒(b) Clear.

(b)⇒(c) Let U be an open set such that $F \subset U$. Since S is normal, there is $g \in C_b(S)$ such that $1_F \le g \le 1_U$. Therefore,

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

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

(c)⇔(d) Clear.

(c)+(d)⇒(e) It easily follows from

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

(e)⇒(a) Let $g \in C_b(S)$ and $\varepsilon > 0$. Since the pushforward measrue $g_*\mu$ 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 $f_\varepsilon := \sum_{i=0}^{n-1} t_i 1_{A_i}$ so that we have $\sup_{x\in S} |g_\varepsilon(x) - g(x)| \le \varepsilon$. From

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

we get

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

Since ε is arbitrary, we are done. □

3.2 (Lévy-Prokhorov metric)**.** Let S be a metric space. Define $\pi : \mathrm{Prob}(S) \times \mathrm{Prob}(S) \to [0, \infty)$ such that

$$\pi(\mu, \nu) := \inf\{\, r > 0 : \mu(A) \le \nu(B(A, r)) + r,\ \ \nu(A) \le \mu(B(A, r)) + r,\ A \in \mathcal{B}(S)\,\},$$

where $B(A, r) := \bigcup_{a\in A} B(a, r)$.

(a) π is a metric.

(b) If $\mu_n \to \mu$ in π, then $\mu_n \to \mu$ in distribution.

(c) If $\mu_i \to \mu$ in distribution, then $\mu_i \to \mu$ in π, when S is separable.

(d) (S, d) is separable if and only if $(\mathrm{Prob}(S), \pi)$ is separable.

(e) (S, d) is compact if and only if $(\mathrm{Prob}(S), \pi)$ is compact

(f) (S, d) is complete if and only if $(\mathrm{Prob}(S), \pi)$ is complete.

Proof. (c) □

3.3 (Prokhorov theorem)**.** Let S be a Polish space. Let $\mathrm{Prob}(S)$ be the space of probability Borel measures on S endowed with the topology of convergence in distribution. We say a set $M \subset \mathrm{Prob}(S)$ is *tight* if for each $\varepsilon > 0$ there is compact $K \subset S$ such that

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

(a) If M is relatively compact, then it is tight.

(b) If M is tight, then it is relatively compact.

Proof. (a) Fix $\varepsilon > 0$. We first claim as a lemma that for an open cover $\{B_i\}_{i\in I}$ of S we have

$$\sup_J \inf_{\mu\in M} \mu(B_J) = 1,$$

where $B_J := \bigcup_{j\in J} B_j$ and J runs through all finite subsets of I. Suppose the claim is false so that there are $\varepsilon > 0$ and a net (μ_J) in M such that $\mu_J(B_J) \le 1 - \varepsilon$. Because $\overline{M}$ is compact, we have a subnet μ_{J_α} of μ_J that converges to $\mu \in \overline{M}$ in distribution, then by the Portmanteau theorem we have for any finite $J \subset I$ that

$$\mu(B_J) \le \liminf_{\alpha} \mu_{J_\alpha}(B_J) \le \liminf_{\alpha} \mu_\alpha(B_{J_\alpha}) \le 1 - \varepsilon.$$

By limiting $J \uparrow I$, we lead a contradiction, so the claim is verified.

Now we use that S is Polish. Let $\{x_i\}_{i=1}^\infty$ be a dense set in S. Fix a metric d on S and consider the family of open covers of balls $\{B(x_i, m^{-1})\}$ parametrized by integers m. By the above claim, there is a finite $n_m > 0$ such that

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

Define

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

which compact since S is complete in d and it is closed and totally bounded. Moreover, we can verify

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

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

(b) We first prove that we have a natural embedding $i_* : \mathrm{Prob}(S) \to \mathrm{Prob}(\beta S)$ with respect to the topology of convergence in distribution, where βS is the Stone-Čech compactification and the map i_* is the pushforward of the natural embedding $i : S \to \beta S$ taken thanks to that S is completely regular. Be cautious that the space $\mathrm{Prob}(\beta S)$ is defined to be the space of probability regular Borel measures on βS because βS is no more metrizable. Let $\mu \in \mathrm{Prob}(S)$ and $\nu := i_*\mu$. Since ν is cleary a probability Borel measure on βS, so we prove it is regular. For any Borel $E \subset \beta S$ and any $\varepsilon > 0$, there is relatively closed $F \subset E \cap S$ in S such that $\mu(E \cap S) < \mu(F) + \varepsilon/2$ by the inner regularity of μ, and there is K that is compact in S such that $\mu(S \setminus K) < \varepsilon/2$ by the tightness of μ. Then, the inequality

$$\nu(E) = \mu(E \cap S) < \mu(F) + \frac{\varepsilon}{2} < \mu(F \cap K) + \varepsilon = \nu(F \cap K) + \varepsilon$$

proves that ν is regular since $F \cap K$ is closed in βS by compactness and satisfies $F \cap K \subset E$. Now we prove that for a net (μ_α) in $\mathrm{Prob}(S)$, if $\nu_\alpha := i_*\mu_\alpha \to \nu := i_*\mu$ in distribution, then $\mu_\alpha \to \mu$ in distribution. By assumption, we have

$$\int_{\beta S} f \, d\nu_\alpha \to \int_{\beta S} f \, d\nu, \qquad f \in C(\beta S).$$

Since $\nu_\alpha(\beta S \setminus S) = \nu(\beta S \setminus S) = 0$ (this expression is pretty bad because S may not be Borel in βS) and the restriction $C(\beta S) \to C_b(S)$ is an isomorphism due to the universal property of βS, we have

$$\int_S f \, d\mu_\alpha \to \int_S f \, d\mu, \qquad f \in C_b(S),$$

so $\mu_\alpha \to \mu$ in distribution. Hence, we have the embedding $i_* : \mathrm{Prob}(S) \to \mathrm{Prob}(\beta S)$.

Let M be a tight subset of $\mathrm{Prob}(S)$. Let (μ_α) be a net in M. Because the topology of convergence in distribution on $\mathrm{Prob}(\beta S)$ is compact by the Banach-Alaoglu theorem and the Riesz-Markov-Kakutani representation theorem, the net of regular Borel measures $\nu_\alpha := i_*\mu_\alpha$ has a subnet ν_β that converges to $\nu \in \mathrm{Prob}(\beta S)$ in distribution. By the tightness of $\{\mu_\beta\}$, for each $\varepsilon > 0$, there is 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. $\nu(S) = 1$, which means that ν is contained the image of $\mathrm{Prob}(S)$. By restriction ν on S we obtain μ, the limit of μ_β. □

3.4 (Skorokhod representation theorem).

3.5 (Continuous mapping theorem).

3.6 (Slutsky theorem).

3.2 Characteristic functions

3.7 (Characteristic functions). Let μ be a probability Borel measure on $\mathbb{R}$. Then, the *characteristic function* of μ is a function $\varphi : \mathbb{R} \to \mathbb{C}$ 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 $\mu \in \mathrm{Prob}(S) \subset \mathcal{S}'(\mathbb{R})$.

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

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

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

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

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

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

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

3.9 (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 $\mu_n \to \mu$ in distribution, then $\varphi_n \to \varphi$ pointwise.

(b) If $\varphi_n \to \varphi$ pointwise and φ is continuous at zero, then (μ_n) is tight and $\mu_n \to \mu$ in distribution.

Proof. (a) For each t,

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

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

(b)

□

3.10 (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.

3.3 Moments

moment problem

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

Exercises

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

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

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

(c) If $f_n \to f$ in total variation, then $f_n \to f$ in distribution.

3.12 (Convergence on real line).

(a) Portmanteau: $F_n(x) \to F(x)$ for every continuity point x of F.

(b) Easy proof of the Skorokhod representation

(c) Easy proof of continuous mapping theorem

(d) Easy proof of the Slutsky theorem

(e) Helly selection theorem, which uses S^1 instead of $\beta\mathbb{R}$.

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

(a) $S \to \mathrm{Prob}(S)$ is a topological embedding.

(b) $S \subset \mathrm{Prob}(S)$ is sequentially closed.

(c)

Proof. (a) It uses Urysohn.
(b) It uses (b)=>(c) of Portmanteau. □

3.14. 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.

3.15 (Convergence determining class)**.**

3.16 (Vauge convergence)**.** Let S be a locally compact Hausdorff space.

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

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

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

Proof. □

Part II

Stochastic processes

Chapter 4

Limit theorems

4.1 Laws of large numbers

Let X_i be a squence of random variables, and let $S_n := \sum_{i=1}^n X_i$.

Laws of large numbers are techniques to find increasing sequences $a_n \gtrsim b_n$ such that

$$S_n = a_n + o(b_n), \qquad n \to \infty$$

in probability or almost everywhere.

Central limit theorems consist of techniques to find $a_n \gtrsim b_n$ such that

$$S_n = a_n + b_n + o(b_n)$$

4.1 (Weak law of large numbers)**.** Let X_i be an uncorrelated sequence of random variables, that is, $E(X_iX_j) = EX_iEX_j$ for all i, j. Define

$$g(x) := \sup_i xP(|X_i| > x).$$

The boundedness of g is a necessary condition for $\sup_i E|X_i| < \infty$ and is a sufficient condition for $\sup_i E|X_i|^{1-\varepsilon} < \infty$ for any $\varepsilon > 0$. In particular, if we have $\lim_{x\to\infty} g(x) = 0$, then X_i is said to satisfy the *Kolmogorov-Feller condition*. Consider the truncation $Y_{n,i} := X_i 1_{|X_i|\le c_n}$ and $T_n := \sum_{i=1}^n Y_{n,i}$. Write

$$P\left(\left|\frac{S_n - ET_n}{n}\right| > \varepsilon\right) \le P(S_n \ne T_n) + P\left(\left|\frac{T_n - ET_n}{b_n}\right| > \varepsilon\right)$$

Let $a_n \sim ET_n$. We claim

$$\frac{S_n - a_n}{b_n} \to 0 \qquad \text{in probability.}$$

(a) If $(n/c_n)g(c_n) \to 0$, then

$$P(S_n \ne T_n) \to 0.$$

(b) If $(nc_n/b_n^2)\int_0^\infty g(c_nx)\,dx \to 0$, then

$$P\left(\left|\frac{T_n - ET_n}{b_n}\right| > \varepsilon\right) \to 0.$$

Proof. (a) It follows from

$$P(S_n \ne T_n) \le \sum_{i=1}^n P(|X_i| > c_n) \le \sum_{i=1}^n \frac{1}{c_n} g(c_n) = \frac{ng(c_n)}{c_n} \to 0.$$

If the Kolmogorov-Feller condition holds, then we may let $c_n \sim n$.

(b) We write

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

We are done. If the Kolmogorov-Feller condition holds, then we may let $nc_n \sim b_n^2$ by the bounded convergence theorem.

□

4.2 (Borel-Cantelli lemmas).

4.3 (Kolmogorov maximal inequality). If (X_i) is the sequence of independent random variables such that $EX_i = 0$ and $VX_i < \infty$, then

$$P(S_n^* > \varepsilon) \le \frac{1}{\varepsilon^2} VS_n,$$

where $S_n^* := \max_{i \le n} |S_i|$. We can prove it by construction of a linear martingale $S_{n\wedge\tau}$ with a stopping time to hit ε: independence and zero mean are necessary. This is a special case of the Doob maximal inequality for $S_{n\wedge\tau}^2$.

4.4 (Kolmogorov three series theorem). Let (X_i) be a sequence of independent random variables. Suppose for a constant $c > 0$ and $Y_i := X_i 1_{|X_i| \le c}$ that the following three series are convergent:

$$\sum_{i=1}^\infty P(|X_i| > c), \qquad \sum_{i=1}^\infty EY_i, \qquad \sum_{i=1}^\infty VY_i.$$

4.5 (Strong laws of large numbers). Let (X_i) be a sequence of independent random variables. The Kolmogorov condition:

$$\sum_{n=1}^\infty \frac{E|Y_n|^2}{b_n^2} < \infty.$$

It is satisfied when $E|X_i| < \infty$. Kronecker lemma

4.6 (Etemadi theorem). Extend the theorem for pairwise independent. But for pairwise uncorrelated, we need a lower bound. By extracting a exponentially fast but sparse subsequence, prove the a.s. convergence. And as we do in renewel theory, we may assume the sequence is non-decreasing and apply the squeeze.

4.2 Renewal theory

4.3 Central limit theorems

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

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

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

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

(a) We have

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

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

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

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

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

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

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

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

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

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

(c) Since

$$|e^{-x} - (1-x)| = \Big|\int_0^x (x-y)e^{-y}\,dy\Big| \le \frac{1}{2}x^2$$

for $x \ge 0$, we have

$$\Big|e^{-\frac{1}{2}E(tY)^2} - \big(1 - \frac{1}{2}E(tY)^2\big)\Big| \lesssim_t (EY^2)^2 \le EY^2(\varepsilon^2 + EY^2 1_{|Y|>\varepsilon}).$$

□

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

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

is satisfied, then

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

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

Berry-Esseen ineaulity

Exercises

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

$$|f_n(x) - f(x)| \le E(|f(S_n/n) - f(x)|) \le \varepsilon + 2\|f\|P(|S_n/n - x| \ge \delta) \to \varepsilon.$$

4.11 (High-dimensional cube is almost a sphere). Let $X_n \sim \mathrm{Unif}(-1,1)$ be independent and identically distributed random variables and $Y_n := X_n^2$. Then, $E(Y_n) = \frac{1}{3}$ and $V(Y_n) \le 1$.

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

4.13 (An occupancy problem).

4.14 (St. Peterburg paradox). For $P(X_n = 2^m) = 2^{-m}$, $g \le 1$ so that $(S_n - n\log_2 n)/n^{1+\varepsilon} \to 0$ in probability.

4.15 (Head runs).

4.16. 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 attibution)

Chapter 5

Discrete stochastic processes

5.1 Martingales

In this chapter we do not use the countability of the index set $\mathbb{N}$.

5.1. (a) If $EX_n = 0$, then S_n is a martingale.

(b) If $EX_n = 0$ and $VX_n = \sigma^2$, then $S_n^2 - n\sigma^2$ is a martingale.

(c) If $EX_n = 1$ and $X_n \ge 0$, then $M_n := \prod_{i=1}^n X_i$ is a martingale.

(d) If X_n is a martingel and φ is convex, then $\varphi(X_n)$ is a submartingale.

(e) If X_n is a submartingale and φ is non-decreasing convex, then $\varphi(X_n)$ is a submartingale.

(f) If $H_n \ge 0$ is predictable and X_n is a (super/sub)martingale, then the *(super/sub)martingale transform*

$$(H \cdot X)_n := H_1 X_1 + \sum_{i=2}^n H_i (X_i - X_{i-1})$$

is a (super/sub)martingale. For a martingale, the condition $H_n \ge 0$ is not required.

5.2 (Martingale convergence theorems)**.** Let (X_n) be a submartingale of random variables and let $a < b$. Let $\tau^0 < \tau_1 < \tau^1 < \tau_2 < \cdots$ be a sequence of hitting times inductively defined by $\tau^0 := 0$ and

$$\tau_k := \min\{n > \tau^{k-1} : X_n \le a\}, \qquad \tau^k := \min\{n > \tau_k : X_n \ge b\}, \qquad k \ge 1.$$

Let $u_n := \max\{k : \tau^k \le n\}$ be the number of upcrossing completed by time n.

(a) We have

$$(b-a)Eu_n \le E(X_n - a)^+, \qquad n \ge 1.$$

It is called the *upcrossing inequality* by Doob.

(b) If $\sup_n EX_n^+ < \infty$, then X_n converges a.s. to a random variable X such that $E|X| < \infty$.

Proof. (a) Let $Y_n := (X_n - a)^+$. Note that $\tau^{u_n} \le n < \tau^{u_n+1}$. Define a predictable sequence

$$H_n := \sum_{k=1}^\infty 1_{(\tau_k, \tau^k]}(n) = 1_{\{\tau^{u_n}\}}(n) + 1_{(\tau_{u_n+1}, \tau^{u_n+1})}(n).$$

Since $Y_{\tau_k} = 0$ for any $k \ge 1$, we have

$$(H \cdot Y)_n - (H \cdot Y)_{\tau^{u_n}} = \sum_{i=\tau^{u_n}+1}^n H_i (Y_i - Y_{i-1}) = 1_{(\tau_{u_n+1}, \tau^{u_n+1})}(n) \cdot (Y_n - Y_{\tau_{u_n+1}}) \ge 0,$$

so

$$(b-a)u_n = \sum_{k=1}^{u_n}(b-a) \le \sum_{k=1}^{u_n}(Y_{\tau^k} - Y_{\tau_k}) = (H\cdot Y)_{\tau_{u_n}} \le (H\cdot Y)_n.$$

Since (Y_n) is also a submartingale and $1-H_n \ge 0$, we have

$$E((1-H)\cdot Y)_n \ge E((1-H)\cdot Y)_1 = E((1-H_1)Y_1) \ge 0,$$

hence

$$(b-a)Eu_n \le E(H\cdot Y)_n \le E(1\cdot Y)_n = EY_n - EY_1 \le EY_n.$$

(b) The condition $\sup_n EX_n^+ < \infty$ implies that $\sup_n Eu_n < \infty$ by the upcrossing inequality, so the increasing sequence u_n converges a.s. It means that

$$P\Big(\bigcup_{a,b\in\mathbb{Q}}\{\liminf_n X_n < a < b < \limsup_n X_n\}\Big) = 0,$$

in other words, the limit $\lim_n X_n$ exists a.s. in $[-\infty,\infty]$. By the Fatou lemma,

$$E(\lim_n |X_n|) \le \liminf_n E|X_n| \le \liminf_n (2EX_n^+ - EX_1) < \infty$$

implies $\lim_n X_n \in (-\infty,\infty)$ a.s. □

5.3 (Doob inequality)**.** If (X_n) is a non-negative submartingale, then we have the following Doob's (maximal or submartingale) inequality

$$P(X_n^* > \varepsilon) \le \frac{1}{\varepsilon}EX_n.$$

For $p > 1$, if $\sup_n E|X_n|^p < \infty$, then X_n converges a.s. and in L^p.

5.4 (Uniform integrability)**.** We say a set of random variables $\{X_i\}$ is *uniformly integrable* if

$$\lim_{c\to\infty}\sup_i E(|X_i|1_{|X_i|>c}) = 0.$$

5.5 (Optional stopping theorem)**.** For a process X, the process *stopped at* a random time τ is the process X^τ defined by $X_n^\tau := X_{t\wedge\tau}$. If $H_n := 1_{n\le\tau}$, then $(H\cdot X)_n = X_{n\wedge\tau}$, so X^τ is a martingale if X is martingale. Wald equations

5.6 (Stopping time σ-algebra)**.** *** Let τ be a stopping time, that is, $\{\tau \le t\} \in \mathcal{F}_t$.

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \le t\} \in \mathcal{F}_t \ \forall t\}.$$

(a) $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.

(b) Début theorem: for a càdlàg process X, the hitting time of a Borel set is a stopping time.

(c) If X is a uniformly integrable martingale and τ is a stopping time, then so is the X^τ.

Proof. Since $\tau : \{\tau \le t\} \to \mathbb{R}_{\ge 0}$ is measurable with respect to $\mathcal{F}_t$, for all $x \in \mathbb{R}$ we have $\{X_\tau \le x\} \in \mathcal{F}_\tau$.

(b) Difficult.

(c) By the optional stopping, we have $X_t^\tau = E(X_\tau|\mathcal{F}_{t\wedge\tau})$, and in fact since $X_\tau 1_{\{\tau\le t\}}$ is $\mathcal{F}_t$-measurable, we further have

$$\begin{aligned} X_t^\tau &= E(X_\tau 1_{\{\tau\le t\}} + X_\tau 1_{\{\tau>t\}}|\mathcal{F}_{t\wedge\tau}) \\ &= X_\tau 1_{\{\tau\le t\}} + E(X_\tau|\mathcal{F}_{t\wedge\tau})1_{\{\tau>t\}} \\ &= X_\tau 1_{\{\tau\le t\}} + E(X_\tau|\mathcal{F}_t)1_{\{\tau>t\}} \\ &= E(X_\tau 1_{\{\tau\le t\}} + X_\tau 1_{\{\tau>t\}}|\mathcal{F}_t) \\ &= E(X_\tau|\mathcal{F}_t). \end{aligned}$$

So we are done. □

5.2 Markov chains

A Markov process on a discrete state space S is a weakly* continuous affine action of $\mathbb{R}_{\geq 0}$ on $\mathrm{Prob}(S)$.

Random walks

Poisson process

Ornstein-Uhlenbeck

5.3 Ergodic theory

Exercises

Chapter 6

Continuous stochastic processes

6.1

Kolmogorov extension Poisson process, Wiener process, Lévy process, Feller process, Markov process
Meyer's section theorems.

6.2 Semi-martingales

6.1 (Filtered proability space). A *filtered probability space* is a probability space $(\Omega, \mathcal{F}, P)$ together with a *filtration*, which is a non-decreasing family $(\mathcal{F}_t)$ of σ-subalgebras of $\mathcal{F}$ indexed by a totally ordered set $\mathbb{T}$. A filtered probability space is said to satisfy the *usual condition* if

(i) every subset of a negilible set belongs to $\mathcal{F}_t$ for every t, (completeness)

(ii) $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$. (right continuity)

From now on, we will always assume that a filtered probability space satisfies the usual condition.

The totally ordered set $\mathbb{T}$ of indices is endowed with the natural Borel σ-algebra induced from the order topology. A random variable τ that is valued in $\mathbb{T}$ is called a *stopping time* if $\{\tau \le t\} \in \mathcal{F}_t$ for all $t \in \mathbb{T}$.

(a) If $\{\tau < t\} \in \mathcal{F}_t$ for all $t \in \mathbb{T}$, then τ is a stopping time.

6.2 (Càdlàg modifications). A *stochastic process* or simply a *process* is a collection $X = (X_t)$ of random variables indexed by $\mathbb{T}$. We say a process X is *adapted* if $X_t \in \mathcal{F}_t$ for all $t \in \mathbb{T}$.

modification: $L^0_{\mathrm{loc}}(\Omega, \mathbb{R}^d)^{\mathbb{T}}$ indistinguishable: $L^0_{\mathrm{loc}}(\Omega, (\mathbb{R}^d)^{\mathbb{T}})$

Two processes X and Y are called *modifications* of each other if $X_t = Y_t$ almost surely for each $t \in \mathbb{T}$, and are *indistinguishable* if $\sup_{t\in\mathbb{T}} |X_t - Y_t| = 0$ almost surely.

jump process ΔX defined such that $\Delta X_t := X_t - X_{t-}$.

(a) If X and Y are right continuous, then they are indistinguishable if and only if they are modifications.

(b) For a submartingale X, X has a càdlàg modification if and only if the non-decreasing function $t \mapsto EX_t$ is right continuous. The modification is unique. In particular, we can always assume a martingale is càdlàg because we assume the usual condition.

doob inequality and optional stopping

6.3 (Local martingales). A process X is called a *local martingale* if there is a sequence of stopping times τ_n such that $\tau_n \to \infty$ a.s. and X^{τ_n} is a uniformly integrable martingale for each n. When is a local martingale actually a martingale? Kazamaki-Novikov criteria.

"P. A. Meyer (1973) showed that there are no local martingales in discrete time; they are a continuous time phenomenon."

6.4 (Semi-martingales).

6.5 (Doob-Meyer decomposition).

Girsanov theorem

6.3 Wiener spaces

Cameron-Martin centered Gaussian law Ornstein-Uhlenbeck process radonifying martingale representation theorem, malliavin calculus

Part III

Stochastic analysis

Chapter 7

Stochastic integral

7.1 Itô integral

Stieltjes integral for locally bounded variation processes square integrable martingale integral (by simple processes) local martingale integral Kunita-Watanabe inequality Ito formula and semi-martingale

7.2 Stratonovich integral

Chapter 8

Stochastic differential equations

Chapter 9

Part IV

Stochastic models

phase transition, percolation