Probability Theory

Ikhan Choi

May 9, 2022

Contents

Part I

Probability distributions

Chapter 1

Random variables

1.1 Sample spaces and distributions

sample space of an "experiment" random variables distributions expectation, moments, inequalities
equally likely outcomes coin toss dice roll ball drawing number permutation life time of a light bulb
joint distribution transformation of distributions distribution computations

1.2 Discrete probability distributions

1.3 Continuous probability distributions

1.4 Independence

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

1.2 (Monotone class lemma)**.**

Chapter 2

Conditional probablity

2.1 (Monty Hall problem)**.** Suppose you're on a game show, and you're given the choice of three doors A, B, and C. Behind one door is a car; behind the others, goats. 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?

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 challenger picked A, and Y 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, $P(A) = P(B) = P(C) = P(X) = 1/3$, and

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

Therefore,

$$\begin{aligned} P(C|X,Y) &= \frac{P(X \cap Y \cap C)}{P(X \cap Y)} \\ &= \frac{P(Y|X,C)P(X \cap C)}{P(Y|X,A)P(X \cap A) + P(Y|X,B)P(X \cap B) + P(Y|X,C)P(X \cap C)} \\ &= \frac{1 \cdot \frac{1}{9}}{\frac{1}{2} \cdot \frac{1}{9} + 0 \cdot \frac{1}{9} + 1 \cdot \frac{1}{9}} = \frac{2}{3}. \end{aligned}$$

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

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

(a) $F_n(x) \to F(x)$ 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 $F_n : \mathbb{R} \to [0,1]$ has a pointwise convergent subsequence.

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

3.2 Weak convergence in metric spaces

3.6. On metric spaces.

(a) Every single measure is regular if X is perfectly normal.

(b) Every single measure is tight if X is Polish.

3.7 (Portemanteau theorem). Let μ_n and μ be probability measures on a metric space S. We will define the *weak convergence* as follows: μ_n converges weakly to μ if

$$\int f\,d\mu_n \to \int f\,d\mu$$

for every $f \in C_b(S)$.

(a) $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ for all closed sets F.

(b) $\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$ for all open sets G.

3.8 (Skorokhod representation theorem).

3.9 (Continuous mapping theorem).

3.10 (Slutsky's theorem).

3.3 The space of probability measures

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

(a) If $f_n \to f$ a.s., 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 $f_n \to f$ weakly.

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

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

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

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

Proof. □

3.13 (Lévy-Prokhorov metric). Let S be a metric space, and $\mathrm{Prob}(S)$ be the set of probability Borel measures on S. Define $\pi : \mathrm{Prob}(S) \times \mathrm{Prob}(S) \to [0, \infty)$ 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) $\mu_n \to \mu$ in π implies $\mu_n \Rightarrow \mu$.

(c) $\mu_\alpha \Rightarrow \mu$ implies $\mu_\alpha \to \mu$ in π, if S is separable.

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

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

Proof. (c) □

3.14 (Prokhorov's theorem). Let S be a metrizable space. Let $\mathrm{Prob}(S)$ be the space of probability measures on S endowed with the topology of weak convergence.

(a) If S is Polish, then the relative compactness implies the tightness.

(b) The tightness implies the relative compactness.

3.4 Characteristic functions

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

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

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

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

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

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

3.17 (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 $\mu_n \to \mu$ weakly, then $\varphi_n \to \varphi$ pointwise.

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

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)

□

3.18 (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.

3.5 Moments

moment problem

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

Exercises

3.19. 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.

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 $(S_n - a_n)/b_n \to 0$ in probability or almost surely.

4.1 (Truncation method)**.** Let $X_{n,i} : \Omega \to \mathbb{R}$ be uncorrelated random variables (with respect to i for each n) and $S_n := X_{n,1}+\cdots+X_{n,n}$. For a positive sequence $(c_n)_{n=1}^\infty$, let $Y_{n,i} := X_{n,i}\mathbf{1}_{|X_{n,i}|\le c_n}$ be truncated random variables and $T_n := Y_{n,1} + \cdots + Y_{n,n}$. Suppose that the truncation level c_n satisfies the approximation condition

$$\lim_{n\to\infty}\sum_{i=1}^n P(|X_{n,i}| > c_n) = 0.$$

(a) If $(T_n - ET_n)/b_n \to 0$ in probability, then $(S_n - ET_n)/b_n \to 0$ in probability.

(b) If $(T_n - ET_n)/b_n \to Z$ in distribution, then $(S_n - ET_n)/b_n \to Z$ in distribution.

Proof. (a) Write

$$P\left(\left|\frac{S_n - ET_n}{b_n}\right| > \varepsilon\right) \le P(S_n \neq T_n) + P\left(\left|\frac{T_n - ET_n}{b_n}\right| > \varepsilon\right) \to 0$$

since

$$P(S_n \neq T_n) \le \sum_{i=1}^n P(X_{n,i} \neq Y_{n,i}) = \sum_{i=1}^n P(|X_{n,i}| > c_n) \to 0$$

as $n \to \infty$.

(b) By the Slutsky theorem. □

4.2 (Weak laws of large numbers)**.** Let $X_{n,i} : \Omega \to \mathbb{R}$ be uncorrelated random variables and $S_n := X_{n,1} + \cdots + X_{n,n}$. For a positive sequence $(c_n)_{n=1}^\infty$, let $Y_{n,i} := X_{n,i}\mathbf{1}_{|X_{n,i}|\le c_n}$ be truncated random variables and $T_n := Y_{n,1} + \cdots + Y_{n,n}$.

(a) If

$$b_n^2 \gg \sum_{i=1}^n E|Y_{n,i} - EY_{n,i}|^2,$$

then $(T_n - ET_n)/b_n \to 0$ in probability.

(b) Take slow c_n as possible such that

$$1 \gg \sum_{i=1}^n P(|X_{n,i}| > c_n).$$

Take slow b_n as possible such that

$$b_n^2 \gg \sum_{i=1}^n E|Y_{n,i}|^2.$$

then $(S_n - ET_n)/b_n \to 0$.

(c) If

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

then $(S_n - ET_n)/n \to 0$ in probability. This is called the *Kolmogorov-Feller condition*.

Proof. (a) Since X_n are uncorrelated, we have for any $\varepsilon > 0$ that

$$P\left(\left|\frac{S_n - ES_n}{c_n}\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2 c_n^2} VS_n \to 0$$

as $n \to \infty$.

(c) Write $g(x) := \sup_i xP(|X_i| > x)$. Then, the truncation condition for $b_n = n$ is satisfied as

$$\sum_{i=1}^n P(|X_i| > n) \le \sum_{i=1}^n \frac{1}{n} g(n) = g(n) \to 0$$

as $n \to \infty$.

On the other hand,

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n E|Y_i|^2 = \frac{1}{n^2}\sum_{i=1}^n \int_0^\infty 2xP(|Y_i| > x)\,dx = \frac{1}{n^2}\sum_{i=1}^n \int_0^n 2xP(|X_i| > x)\,dx \\ \le \frac{2}{n}\int_0^n g(x)\,dx = 2\int_0^1 g(nx)\,dx.\end{aligned}$$

Since $g(x) \le x$ and $g(x) \to 0$ as $x \to \infty$, g is bounded so that the bounded convergence theorem implies $\int_0^1 g(nx)\,dx \to 0$ as $n \to \infty$.

Therefore, $(T_n - ET_n)/n \to 0$ in probability. By the truncation □

4.3 (Borel-Cantelli lemmas).

4.4 (Strong laws of large numbers). Proof by Etemadi

Random series

4.2 Renewal theory

4.3 Central limit theorems

4.5 (Lyapunov central limit theorem). Let $X_n : \Omega \to \mathbb{R}$ be independent random variables with $EX_i = \mu_i$ and $VX_i = \sigma_i^2$. If there is $\delta > 0$ such that the *Lyapunov condition*

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

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 $s_n^2 := VS_n$.

4.6 (Lindeberg-Feller central limit theorem). Let $X_{i,n} : \Omega \to \mathbb{R}$ be independent random variables and $S_n := X_1 + \cdots + X_n$.

(a) If

$$s_n^2 \sim \sum_{i=1}^{n} E|X_i - EX_i|^2,$$

and if for every $\varepsilon > 0$ we have

$$s_n^2 \gg \sum_{i=1}^{n} E|X_i - EX_i|^2 \mathbf{1}_{|X_i - EX_i| > \varepsilon s_n},$$

then $(S_n - ES_n)/s_n \to N(0,1)$ in distribution. This is called the *Lindeberg condition*.

(b)

Berry-Esseen ineaulity

Exercises

4.7 (Bernstein polynomial). Let $X_n \sim \mathrm{Bern}(x)$ be i.i.d. 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 $|f(x)-f(y)| < \varepsilon$,

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

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

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

4.10 (An occupancy problem).

4.11 (The St. Petersburg paradox).

4.12. 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

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 Kolomogorov 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 $\varnothing$ 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 $\mu_0 : \mathcal{A} \to [0, \infty]$ by $\mu_0(A) = \mu_n(A^*)$ and $\mu_0(\varnothing) = 0$.

(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 $\mu_0(B_n) \geq \delta$, 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 $B_n^* \subset \mathbb{R}^{r(n)}$. 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,

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

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

Part IV

Stochastic calculus