Probability Theory

Ikhan Choi

December 20, 2021

Contents

I Probability distributions 3

1 Random variables 4

1.1 Sample spaces and distributions 4
1.2 Discrete probability distributions 4
1.3 Continuous probability distributions 4
1.4 Independence 4

2 Conditional probablity 5

3 Convergence of probability measures 6

3.1 Polish spaces 6
3.2 Weak convergence on Polish spaces 6
3.3 The space of probability measures 6
3.4 Characteristic functions 7
3.5 Moments 8

II Limit theorems 9

4 Laws of large numbers 10

4.1 Weak and strong laws of large numbers 10
4.2 Random series 11
4.3 Renewal theory 11

5 Central limit theorems 13

5.1 Central limit theorems 13
5.2 Berry-Esseen ineaulity 14
5.3 Poisson convergence 14
5.4 Stable laws 14

6 **15**

III Stochastic processes 16

7 Martingales **17**

8 Markov chains **18**

9 Brownian motion **19**
9.1 Kolomogorov extension 19

IV Stochastic calculus 21

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 $P(B|X,Y) = 0$. □

Chapter 3

Convergence of probability measures

3.1 Polish spaces

3.2 Weak convergence on Polish spaces

3.1 (Portemanteau theorem)**.** Let $F_n : \mathbb{R} \to [0, 1]$ be distribution functions.

(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

3.3 The space of probability measures

3.6. Suppose f_n and f are density functions on $\mathbb{R}$.

(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.

3.7 (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.

Proof. $\square$

3.8 (Lévy-Prokhorov metric). (a) If S is a separable metrizable space, π generates the topology of weak convergence.

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

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

3.9 (Prokhorov's theorem). Let S be a separable metrizable space. Let $\mathrm{Prob}(S)$ be the space of probability measures on S. Let $\mathcal{F} \subset \mathrm{Prob}(S)$.

(a)

(b) $\mathcal{F}$ is weakly precompact if and only if it is tight.

3.4 Characteristic functions

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

3.11 (Inversion formula). (a) For $a < b$,

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

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

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

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

3.12 (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)

$\square$

3.13 (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 $|t| < \delta$

Part II

Limit theorems

Chapter 4

Laws of large numbers

4.1 Weak and strong laws of large numbers

4.1 (Truncation method)**.** Let $X_n : \Omega \to \mathbb{R}$ be uncorrelated random variables and $S_n := X_1 + \cdots + X_n$. For a positive sequence $(b_n)_{n=1}^\infty$, let $Y_n := X_n \mathbf{1}_{|X_n| \le b_n}$ be truncated random variables and $T_n := Y_1 + \cdots + Y_n$. Suppose that the truncation level b_n satisfies the approximation condition

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

(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

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

(b) By the Slutsky theorem. □

4.2 (Weak laws of large numbers)**.** Let $X_n : \Omega \to \mathbb{R}$ be uncorrelated random variables and $S_n := X_1 + \cdots + X_n$. For a positive sequence $(b_n)_{n=1}^\infty$, let $Y_n := X_n \mathbf{1}_{|X_n| \le b_n}$ be truncated random variables and $T_n := Y_1 + \cdots + Y_n$.

(a) If

$$\lim_{n\to\infty} \frac{1}{c_n^2} \sum_{i=1}^n E|X_i|^2 = 0,$$

then $(S_n - ES_n)/c_n \to 0$ in probability.

(b) If

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

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

Proof. (a) Since X_n are uncorrelated, we have

$$P\left(\left|\frac{S_n-ES_n}{b_n}\right|>\varepsilon\right)\le\frac{1}{\varepsilon^2 b_n^2}VS_n\le\frac{1}{\varepsilon^2 b_n^2}\sum_{i=1}^n E|X_i|^2\lesssim\frac{1}{n}\to 0$$

as $n\to\infty$.

(b) Write $g(x):=\sup_i xP(|X_i|>x)$. Then,

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

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 $n\to\infty$. $\square$

4.3 (Borel-Cantelli lemmas).

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

4.2 Random series

4.3 Renewal theory

Exercises

4.5 (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 $E(|S_n/n-x|^2)\to 0$. Define $f_n(x):=E(f(S_n/n))$. Then, by the uniform continuity $|x-y|<\delta$ implies $|f(x)-f(y)|<\varepsilon$,

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

4.6 (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 $V(Y_n) \leq 1$.

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

4.8 (An occupancy problem).

4.9 (The St. Petersburg paradox).

Chapter 5

Central limit theorems

5.1 Central limit theorems

5.1 (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 $s_n^2 := VS_n$.

5.2 (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

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

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

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

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

5.2 Berry-Esseen ineaulity

5.3 Poisson convergence

Law of rare events, or weak law of small numbers (a single sample makes a significant attibution)

5.4 Stable laws

Chapter 6

Part III

Stochastic processes

Chapter 7

Martingales

Chapter 8

Markov chains

Chapter 9

Brownian motion

9.1 Kolomogorov extension

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

Part IV

Stochastic calculus