Probability Theory

Ikhan Choi

November 23, 2021

Contents

Part I

Random variables

Chapter 1

Probability distributions

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

1.2 Joint probability

functions of random variables independent random variables

1.3 Conditional probablity

1.1 (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 $P(B|X,Y) = 0$. $\square$

1.4 Discrete probability distributions

1.5 Continuous probability distributions

Chapter 2

Independence

2.1 Monotone class lemma

2.1 (Dynkin's π-λ theorem)**.** 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 $\sigma(\mathcal{P}) \subset \mathcal{L}$.

monotone class

2.2 Independent σ-algebras

2.3 Zero-one laws

2.2 (The Kolmogorov zero-one law)**.** Let $X_n : \Omega \to S$ be independent random variables. The *tail σ-algebra* is the σ-algebra $\mathcal{T}$ defined by $\mathcal{T} := \limsup_n \mathcal{F}_n$.

2.3 (The Hewitt-Savage zero-one law)**.** Let $X_n : \Omega \to S$ be i.i.d. random variables.

Chapter 3

Statistical inference

Part II

Limit theorems

Chapter 4

Laws of large numbers

4.1 Weak laws of large numbers

4.1. Let $X_n : \Omega \to \mathbb{R}$ be uncorrelated random variables.

(a) If $E(X_n) = \mu$ and $E(X_n^2) \lesssim 1$, then $S_n/n \to \mu$ in probability.

(b) If $nP(|X_n| > b_n) \to 0$, $\frac{n}{b_n^2}E(|X|^2\mathbf{1}_{|X|\le b_n}) \to 0$, and $b_n \sim nE(X\mathbf{1}_{|X|\le b_n})$, then $S_n/b_n \to 1$ in probability.

4.2 (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 $|f(x)-f(y)| < \varepsilon$,

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

4.3 (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 $V(Y_n) \le 1$.

large deviation technique: Lp?

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

4.5 (An occupancy problem)**.**

4.6 (The St. Petersburg paradox)**.**

4.7 (Kolmogorov-Feller theorem). Suppose X_i satisfies the Feller condition

$$xP(|X_i| > x) \to 0$$

as $x \to \infty$.

(a)

4.2 Almost sure convergence

4.3 Strong laws of large numbers

Proof by Etemadi and proof by random series.

infinite monkey

Chapter 5

Weak convergence

5.1 Weak convergence in $\mathbb{R}$

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

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

(b) If $f_n \to f$ in L^1, then $f_n \to f$ in total variation.

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

5.2. (a) If $F_n \to F$ weakly, then there are random variables X_n and X with distributions F_n and F such that $X_n \to X$ almost surely.

5.3 (Portemanteau theorem). (a)

5.4 (Helly's selection theorem). (a)

(b) F_n has a weekly convergent subsequence F_{n_k}.

(c) If $\{F_n\}$ is tight, then

5.2 The space of probability measures

5.5. 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, if $\{\mu_n\}$ is tight.

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

Proof. (a) The bounded total variations of $\|\mu_n\| = 1$ is crucial. $\square$

5.6 (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.

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

Cb* weak topology is stronger than C0* vague topology

probability measures P subset Cb* subset C0*

positive linear functional on Cc infty is in Cc*, finite positive linear functional on Cc is in C0*, and also in Cb*

unitization C(X0) multiplier C(bX)=Cb(X)

Since X is not compact, C0(X) is not unital so that Prob(X)=S(C0(X)) is not compact.

5.3 Characteristic functions

5.8 (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 μ.

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

(b) If $\varphi \in L^1(\mathbb{R})$, then μ has density $f \in C_0(\mathbb{R}) \cap L^1(\mathbb{R})$.

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

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

5.10 (Lévy's continuity theorem). (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 $\mu_n \to \mu$ weakly.

5.11 (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.

5.4 Moments

moment problem

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

Chapter 6

Central limit theorems

Proof by continuity theorem (3.4.1)

6.1 (Classical CLT). Let $X_n : \Omega \to \mathbb{R}$ be i.i.d. random variables with $EX_i = \mu$ and $VX_i = \sigma^2$ for $0 < \sigma < \infty$. Then,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1)$$

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

6.2 (Lyapunov CLT). 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 $s_n^2 := VS_n$.

6.3 (Lindeberg CLT). Let $X_{i,n} : \Omega \to \mathbb{R}$ be independent random variables with $EX_{i,n} = \mu_{i,n}$ and $VX_{i,n} = \sigma_{i,n}^2$. If for every $\varepsilon > 0$ the *Lindeberg condition*

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

is satisfied, then

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

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

6.1 Berry-Esseen ineaulity

6.2 Poisson convergence

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

6.3 Stable laws

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

Part IV

Stochastic calculus