Real Analysis II: Study Note

Chapter 9: Applications to Probability Theory

Tianqi Zhang

Department of Economics

University of Southern California

Spring 2026

Contents

1 Probability as a Measure Space 3

2 Distributions and Pushforward 4

3 Independence 4

4 Borel–Cantelli Lemmas 5

5 Kolmogorov Extension Theorem 5

6 Laws of Large Numbers 6

7 Central Limit Theorem 6

8 Characteristic Functions 6

9 Weak Convergence 7

10 Conditional Expectation 8

1 Probability as a Measure Space

Measure theory is the language of modern probability. A probability space is simply a measure space with total mass 1, but the vocabulary shifts: sets become events, measures become probabilities, and measurable functions become random variables. Every result from Chapters 1–8 has a direct probabilistic interpretation, and the culmination of this course — the Law of Large Numbers and Central Limit Theorem — are consequences of the measure-theoretic machinery built throughout.

Roadmap: probability space & r.v. → distributions & pushforward → independence & Borel–Cantelli → LLN → CLT → characteristic functions → weak convergence → conditional expectation.

Definition 1.1: Probability Space

A **probability space** is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}(\Omega) = 1$.

- Ω: **sample space** (all possible outcomes).
- $\mathcal{F}$: **σ-algebra of events**.
- $\mathbb{P}$: **probability measure**, $\mathbb{P}(A)$ = probability of event $A \in \mathcal{F}$.
- A **random variable** is a measurable function $X : \Omega \to \mathbb{R}$.
- $\mathbb{E}[X] := \int_\Omega X \,\mathrm{d}\mathbb{P}$ (if it exists, $X \in L^1$) is the **expected value**.
- $\mathrm{Var}(X) := \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (if exists, $X \in L^2$).

Canonical examples.

(1) **Binomial:** $\Omega = \{0, \ldots, n\}$, $\mathbb{P} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k$ (weighted counting / point measure).

(2) **Poisson:** $\Omega = \{0, 1, 2, \ldots\}$, $\mathbb{P} = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \delta_k$ (weighted counting measure).

(3) **Gaussian (Normal):** $\Omega = \mathbb{R}$, $\mathrm{d}\mathbb{P} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \mathrm{d}x$ (weighted Lebesgue measure), giving $X \sim N(\mu, \sigma^2)$.

Remark 1.1. *If $\mathbb{P} \ll \mathbb{Q}$ (Chapter 7), then $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$ exists by Radon–Nikodym, and integrals w.r.t. $\mathbb{P}$ can be converted to integrals w.r.t. $\mathbb{Q}$ via $\int f \,\mathrm{d}\mathbb{P} = \int f \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \,\mathrm{d}\mathbb{Q}$. Note that $L^p(\mathrm{d}\mathbb{P}) \subset L^p(\mathrm{d}\mathbb{Q})$ does* not *follow in general from $\mathbb{P} \ll \mathbb{Q}$ alone.*

2 Distributions and Pushforward

Definition 2.1: Distribution of a Random Variable

The **distribution** of $X : \Omega \to \mathbb{R}$ is the **pushforward measure** (Chapter 8, Example 3):

$$\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega : X(\omega) \in A\}), \qquad A \in \mathcal{B}(\mathbb{R}).$$

$\mathbb{P}_X$ is a Borel measure on $\mathbb{R}$.

Proposition 2.2: Change of Variables Formula

For $\phi \in L^1(\mathbb{R}, \mathrm{d}\mathbb{P}_X)$:

$$\int_{\mathbb{R}} \phi \, \mathrm{d}\mathbb{P}_X = \int_{\Omega} \phi(X) \, \mathrm{d}\mathbb{P}.$$

In particular, $\mathbb{E}[X] = \int_\Omega X \, \mathrm{d}\mathbb{P} = \int_{\mathbb{R}} t \, \mathrm{d}\mathbb{P}_X(t)$.

Proof: Check for simple functions and extend by standard approximation (simple $\to$ positive $\to$ integrable).

3 Independence

Definition 3.1: Independence

Events $A, B \in \mathcal{F}$ are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
A collection $\{A_i\}$ is independent if for every finite subcollection $\{A_{i_1}, \dots, A_{i_k}\}$:

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

Random variables $X_1, X_2, \dots$ are independent if $\{X_j^{-1}(B_j)\}$ form independent events for all Borel B_j.

Example. On $\Omega = [0,1]^d$ with Lebesgue measure, the coordinate projections $X_j(\omega) = \omega_j$ for $j = 1, \dots, d$ are independent: $\mathbb{P}(X_1 \in B_1, \dots, X_d \in B_d) = m(B_1) \cdots m(B_d) = \prod_j \mathbb{P}(X_j \in B_j)$, since the Lebesgue measure on $[0,1]^d$ is the product of d copies of Lebesgue measure on $[0,1]$.

Theorem 3.2: Joint Distribution and Independence

Random variables $X_1, \dots, X_d$ are independent if and only if the joint distribution factors:

$$\mathbb{P}_{X_1,\dots,X_d} = \mathrm{d}\mathbb{P}_{X_1} \otimes \cdots \otimes \mathrm{d}\mathbb{P}_{X_d} \quad \text{(product measure)}.$$

In particular, if $\{X_j\}$ are independent: $\mathbb{E}[X_{j_1} \cdots X_{j_m}] = \mathbb{E}[X_{j_1}] \cdots \mathbb{E}[X_{j_m}]$.

4 Borel–Cantelli Lemmas

Remark 4.1. *Recall from Chapter 1:* $\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$ *is the event that infinitely many* A_n *occur.* $\mathbb{P}(\limsup_n A_n) = \mathbb{P}(A_n \text{ i.o.})$ *(i.o. = infinitely often).*

Lemma 4.1: Borel–Cantelli I

If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$. (Independence is *not* required.)

Lemma 4.2: Borel–Cantelli II

If $\{A_n\}$ are **independent** and $\sum_n \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Example ("cat" appears infinitely often). Let Ω be an infinite sequence of random letters (each position uniform over 26 letters, positions independent). Let $A_n = \{\text{"cat" starts at position } n\}$. Then $\mathbb{P}(A_n) = 1/26^3$ for all n, so $\sum \mathbb{P}(A_n) = \infty$. However, adjacent events A_n and A_{n+1} are *not* independent (they share letters at positions $n+1$ and $n+2$). The correct argument uses the fact that $\{A_n, A_{n+3}, A_{n+6}, \ldots\}$ are independent (non-overlapping windows), and decomposes the full sequence into three such independent subsequences. Applying BC II to each subsequence gives $\mathbb{P}(\text{"cat" appears i.o.}) = 1$.

5 Kolmogorov Extension Theorem

To define an infinite sequence of i.i.d. random variables, we need a probability space supporting all of them simultaneously. The Kolmogorov Extension Theorem guarantees this.

Definition 5.1: i.i.d.

A sequence $X_1, X_2, \ldots$ is **independent and identically distributed (i.i.d.)** if the X_j are mutually independent and $\mathbb{P}_{X_j}$ is the same distribution $\mathbb{P}_0$ for all j.

Theorem 5.2: Kolmogorov Extension Theorem

Given Borel probability measures $\{\mathbb{P}_j\}_{j \in \mathbb{N}}$ on $\mathbb{R}$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and independent random variables X_j with $\mathbb{P}_{X_j} = \mathbb{P}_j$.

Construction. Set $\Omega = \mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \ldots) : \omega_j \in \mathbb{R}\}$ (sequence space, no topology needed). Let $\mathcal{F}$ be generated by **cylinder sets** $\{(\omega_1, \ldots) \in \Omega : (\omega_1, \ldots, \omega_n) \in B\}$ for $B \in \mathcal{B}(\mathbb{R}^n)$. The finite-dimensional distributions $\mathbb{P}^{(n)}(B) = \mathbb{P}_1 \times \cdots \times \mathbb{P}_n(B)$ are **consistent**: $\mathbb{P}^{(n)}(B \times \mathbb{R}) = \mathbb{P}^{(n-1)}(B)$. Kolmogorov's theorem states that consistent families extend to a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$.

6 Laws of Large Numbers

Theorem 6.1: Law of Large Numbers

Let $\{X_j\}$ be i.i.d. with $\mathbb{E}[X_j]$ defined, and $A_n := \frac{1}{n}\sum_{j=1}^{n} X_j$.

- **Weak LLN** (Bernoulli 1713 for Bernoulli r.v.s; Chebyshev's inequality in general): If $X_j \in L^2$, then $A_n \xrightarrow{p} \mu$ (convergence in probability): $\mathbb{P}(|A_n - \mu| > \varepsilon) \to 0$ for all $\varepsilon > 0$. *Proof:* $\mathrm{Var}(A_n) = \sigma^2/n \to 0$; apply Chebyshev (Markov for $p = 2$).

- **Strong LLN** (Kolmogorov 1933): If $X_j \in L^1$ with $\mathbb{E}[|X_j|] < \infty$ and $\mathbb{E}[X_j] = \mu$, then $A_n \to \mu$ **almost surely**: $\mathbb{P}(\{\omega : A_n(\omega) \to \mu\}) = 1$.

7 Central Limit Theorem

Theorem 7.1: Central Limit Theorem (CLT)

Let $\{X_k\}$ be i.i.d. with $X_k \in L^2$, $\mathbb{E}[X_k] = \mu$, $\mathrm{Var}(X_k) = \sigma^2$. Define

$$Z_n := \frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}}.$$

Then $Z_n \Rightarrow N(0,1)$ in some notion: (convergence in distribution):

$$\lim_{n\to\infty} \mathbb{P}(a \le Z_n \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,\mathrm{d}t.$$

Historical note. Special case for Binomial (normal approximation): de Moivre (1733) and Laplace (1812). General version via moment conditions: Lyapunov (1901). The characteristic function proof (Lévy, 1920s): show $\varphi_{Z_n}(t) \to e^{-t^2/2}$ pointwise, then invoke Lévy's continuity theorem to conclude $Z_n \Rightarrow N(0,1)$.

8 Characteristic Functions

Definition 8.1: Characteristic Function

Given X on $(\Omega, \mathcal{F}, \mathbb{P})$, its **characteristic function** is the Fourier transform of $\mathbb{P}_X$:

$$\varphi_X(t) := \mathbb{E}[e^{itX}] = \int_\Omega e^{itX}\,\mathrm{d}\mathbb{P} = \int_{\mathbb{R}} e^{itx}\,\mathrm{d}\mathbb{P}_X(x), \qquad t \in \mathbb{R}.$$

Proposition 8.2: Properties of Characteristic Functions

(i) **Smoothness:** If $X \in L^k$ (i.e. $\mathbb{E}[|X|^k] < \infty$), then $\varphi_X \in C^k(\mathbb{R})$ with

$$\varphi_X^{(k)}(t) = \int_{\mathbb{R}} (ix)^k e^{itx} \, \mathrm{d}\mathbb{P}_X(x).$$

Differentiating under the integral is justified by DCT, since $|(ix)^k e^{itx}| = |x|^k \in L^1(\mathrm{d}\mathbb{P}_X)$.

(ii) **Moment generation:** $\varphi_X(0) = 1$, $\varphi_X'(0) = i\mu$, $\varphi_X''(0) = -\mathbb{E}[X^2]$.

(iii) **Scaling:** for $c \in \mathbb{R}$, $\varphi_{cX}(t) = \varphi_X(ct)$.

(iv) **Independence:** if $X \perp Y$, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

Example: $X \sim N(0,1)$: $\varphi_X(t) = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, \mathrm{d}x = e^{-t^2/2}$. This follows by completing the square: $itx - x^2/2 = -\frac{1}{2}(x - it)^2 - t^2/2$, then shifting the contour of integration.

9 Weak Convergence

Definition 9.1: Weak Convergence

Let X_n, X be random variables with distributions $\mathbb{P}_{X_n}$, $\mathbb{P}_X$, characteristic functions φ_n, φ, and CDFs $F_n(t) = \mathbb{P}(X_n \le t)$, $F(t) = \mathbb{P}(X \le t)$. The following are equivalent (**TFAE**, with (i)⇔(iii) by definition, and (i)⇔(ii) by Lévy's Continuity Theorem):

(i) $\displaystyle\int h \, \mathrm{d}\mathbb{P}_{X_n} \to \int h \, \mathrm{d}\mathbb{P}_X$ for all bounded continuous $h : \mathbb{R} \to \mathbb{R}$.

(ii) $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$ (pointwise), and the limit φ is continuous at 0.

(iii) $F_n(t) \to F(t)$ at all continuity points t of F.

This is called **weak convergence** (or **convergence in distribution**): $X_n \Rightarrow X$.

Remark 9.1. *F is increasing and bounded, so by the structure of monotone functions it has at most countably many discontinuities. The condition $F_n(t) \to F(t)$ at all continuity points of F (not necessarily all t) is the correct notion.*

10 Conditional Expectation

Definition 10.1: Conditional Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X \in L^1(\Omega, \mathbb{P})$, and $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra. The **conditional expectation** $\mathbb{E}[X \mid \mathcal{G}]$ is the unique (up to $\mathbb{P}$-a.s. equivalence) random variable Y satisfying:

- Y is $\mathcal{G}$-measurable.
- $\displaystyle\int_E X \, \mathrm{d}\mathbb{P} = \int_E Y \, \mathrm{d}\mathbb{P}$ for all $E \in \mathcal{G}$.

Existence and uniqueness: the map $E \mapsto \int_E X \, \mathrm{d}\mathbb{P}$ defines a signed measure on $(\Omega, \mathcal{G})$ that is absolutely continuous w.r.t. $\mathbb{P}|_{\mathcal{G}}$. By the Radon–Nikodym theorem, its density w.r.t. $\mathbb{P}|_{\mathcal{G}}$ is the unique such Y.

Example: discrete conditioning. Partition Ω into countable disjoint non-trivial pieces $\{\Omega_j\}$ with $\mathbb{P}(\Omega_j) > 0$. Let $\mathcal{F} = \sigma(\{\Omega_j\})$ (the σ-algebra generated by the partition — "no more information within each Ω_j"). If $Y = \mathbb{E}[X|\mathcal{F}]$, then Y must be constant on each Ω_j:

$$Y = \sum_j c_j \chi_{\Omega_j}, \qquad c_j = \frac{\int_{\Omega_j} X \, \mathrm{d}\mathbb{P}}{\mathbb{P}(\Omega_j)} = \mathbb{E}[X|\Omega_j].$$

Interpretation: conditioning on $\mathcal{F}$ replaces X by its average over each atom Ω_j of the partition — exactly what conditioning means intuitively.