STA347: Probability Theory Lecture Notes

by

Blair Bilodeau

Department of Statistical Sciences
University of Toronto
August 2021

COPYRIGHT

I have compiled these notes to help me while teaching STA347 at the University of Toronto, and provided them to the students as a record of what we cover in lecture. As such, they are intended to supplement the lectures, and should not be treated as the sole reference to study from. All of the results in these notes, besides the mistakes, can be more-or-less found somewhere in four main texts.

The main material follows the structure of Rosenthal (2006), with the measure theoretic aspects sanitized away, while Durrett (2013) is also used as a reference for rigorous results (primarily his treatment of the Weak LLN). Meanwhile, both Ross (2007) (especially chapters 4 and 8) and Rice (2007) are used to provide motivating examples and allow the students to follow along in a more elementary text.

Durrett, R. (2013). Probability: Theory and Examples, Fourth edition. *Cambridge University Press.*

Rice, J. (2007). Mathematical statistics and data analysis, Third edition. *Cengage Learning.*

Rosenthal, J. (2006). A first look at rigorous probability theory, Second edition. *World Scientific.*

Ross, S. (2007). Introduction to probability models, Tenth edition. *Academic Press.*

TABLE OF CONTENTS

LECTURE 1

PROBABILITY BASICS

1.1 Set Theory Fundamentals

Definition 1.1.1 (Convergence of Real-Valued Sequences).
A sequence $(x_n)_{n=1}^{\infty} \subseteq \mathbb{R}$ converges to x if

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \qquad n > N \implies |x_n - x| < \varepsilon. \tag{1.1}$$

A sequence $(f_n)_{n=1}^{\infty} \subseteq \mathbb{R}^{\mathcal{X}}$ converges *pointwise* to f if

$$\forall \varepsilon > 0 \; \forall x \in \mathcal{X} \; \exists N_x \in \mathbb{N} \qquad n > N_x \implies |f_n(x) - f(x)| < \varepsilon. \tag{1.2}$$

A sequence $(f_n)_{n=1}^{\infty} \subseteq \mathbb{R}^{\mathcal{X}}$ converges *uniformly* to f if

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall x \in \mathcal{X} \qquad n > N \implies |f_n(x) - f(x)| < \varepsilon. \tag{1.3}$$

Definition 1.1.2 (Limiting Processes).
Let $A \subseteq \mathbb{R}$ be a set.

$$\begin{aligned} u = \sup A \quad &\iff \quad \forall \varepsilon > 0 \; \exists x \in A \quad u - \varepsilon < x \quad \text{and} \quad \forall x \in A \quad x \leq u. \\ v = \inf A \quad &\iff \quad \forall \varepsilon > 0 \; \exists x \in A \quad u + \varepsilon > x \quad \text{and} \quad \forall x \in A \quad x \geq u. \end{aligned} \tag{1.4}$$

Let $(x_n) \subseteq \mathbb{R}$ be a sequence.

$$\begin{aligned} \limsup_{n\to\infty} x_n &= \lim_{n\to\infty} \sup_{m \geq n} x_m = \inf_{n \geq 0} \sup_{m \geq n} x_m. \\ \liminf_{n\to\infty} x_n &= \lim_{n\to\infty} \inf_{m \geq n} x_m = \sup_{n \geq 0} \inf_{m \geq n} x_m. \end{aligned} \tag{1.5}$$

Definition 1.1.3 (Function Limits).
For any function $f : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned} f(x-) = \lim_{y \uparrow x} f(y) = L \quad &\iff \quad \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall y \in (x - \delta, x) \quad |f(y) - L| < \varepsilon. \\ f(x+) = \lim_{y \downarrow x} f(y) = L \quad &\iff \quad \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall y \in (x, x + \delta) \quad |f(y) - L| < \varepsilon. \\ \lim_{y \to x} f(y) = L \quad &\iff \quad f(x-) = f(x+) = L. \end{aligned} \tag{1.6}$$

1.2 Probability Measures

Definition 1.2.1. A *sample space* Ω is any non-empty set.

Example 1.2.2. The sample space contains "all the states of the world".
Coin flipping: $\Omega = \{H, T\}$.
Weather: $\{\text{hot, cold, mild}\} \times \{\text{wet, dry}\}$.
Extreme example: $\Omega = \{\text{all locations of every molecule on Earth}\}$.

Definition 1.2.3. For a sample space Ω, we say $\mathbb{P}$ is a *probability measure* if the following hold:

i) $\forall E \subseteq \Omega, 0 \leq \mathbb{P}(E) \leq 1$, that is, $\mathbb{P} : \mathscr{P}(\Omega) \to [0, 1]$;
ii) $\mathbb{P}(\Omega) = 1$;
iii) $\forall E_1, E_2, \ldots \subseteq \Omega$ such that $E_i \cap E_j = \emptyset \; \forall \; i \neq j$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i). \tag{1.7}$$

Remark. If this truly has to hold for all subsets of Ω, it may be impossible for such a $\mathbb{P}$ to exist. To define exactly which subsets we want $\mathbb{P}$ to satisfy these axioms for would require more advanced mathematics (the subject of measure theory). For this class, we will not worry about these technicalities, and assume $\mathbb{P}$ is defined for all subsets of Ω.

Example 1.2.4. Probabilities of coin flipping. $\mathbb{P}(H) = \mathbb{P}(T) = 1/2$. $\mathbb{P}(H \cup T) = 1/2 + 1/2 = 1$.

Example 1.2.5 (Uniform/Lebesgue measure). Let $\Omega = [L, U]$ for some $L < U \in \mathbb{R}$. Define $\mathbb{P}([a, b)) = \frac{b-a}{U-L}$ when $L \leq a \leq b \leq U$.

Proof. In this class we can assume this exists. □

Proposition 1.2.6. The following properties are satisfied by any probability measure $\mathbb{P}$.

1. $\forall E \subseteq \Omega, \mathbb{P}(E^c) = 1 - \mathbb{P}(E)$;
2. $\mathbb{P}(\emptyset) = 0$;
3. $\forall E, F \subseteq \Omega$ such that $E \subseteq F$, $\mathbb{P}(E) \leq \mathbb{P}(F)$;
4. $\forall E, F \subseteq \Omega$, $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$;
5. $\forall E, F \subseteq \Omega$, $\mathbb{P}(E \cap F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cup F)$;
6. $\forall E, F \subseteq \Omega$, $\mathbb{P}(F \setminus E) = \mathbb{P}(F) - \mathbb{P}(F \cap E)$.

Proof. Exercise (review from STA257). □

Lemma 1.2.7. Consider a sequence of sets E_i and set $E \subseteq \bigcup_{i=1}^{\infty} E_i$. Then,

$$\mathbb{P}(E) \leq \sum_{i=1}^{\infty} \mathbb{P}(E_i) \tag{1.8}$$

Proof. Let $F_i = E_i \cap E$. Observe that $E = \bigcup_{i=1}^{\infty} F_i$ and $F_i \subseteq E_i$ so $\mathbb{P}(F_i) \leq \mathbb{P}(E_i)$. Further, define $G_1 = F_1$, and $G_k = F_k \cap \left(\bigcup_{i=1}^{k-1} F_i\right)^c$ for $k \geq 2$. Observe that the set of G_i's are disjoint with $\bigcup_{i=1}^{\infty} G_i = \bigcup_{i=1}^{\infty} F_i = E$, and $G_i \subseteq F_i$. Then,

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} G_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(G_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(F_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(E_i). \tag{1.9}$$

□

Lemma 1.2.8. Let $\mathbb{P}$ be the uniform measure on $[0, 1]$. Then $\mathbb{P}(E) = 0$ for any countable set E.

Proof. Let $E = \{x_i\}_{i=1}^{\infty}$. Fix $\varepsilon > 0$ and define $E_i(\varepsilon) = [x_i - 2^{-i}\varepsilon, x_i + 2^{-i}\varepsilon)$. Observe that $E \subseteq \bigcup_{i=1}^{\infty} E_i(\varepsilon)$. Thus,

$$\mathbb{P}(E) \leq \sum_{i=1}^{\infty} \mathbb{P}(E_i(\varepsilon)) = \sum_{i=1}^{\infty}[x_i + 2^{-i}\varepsilon - x_i + 2^{-i}\varepsilon] = 2\varepsilon \sum_{i=1}^{\infty} 2^{-i} = 2\varepsilon. \tag{1.10}$$

Since ε was arbitrary, $\mathbb{P}(E) = 0$. $\square$

1.3 Independence

Definition 1.3.1. Events $E_1, \ldots, E_n \subseteq \Omega$ are independent with respect to a probability measure $\mathbb{P}$ if for each $\mathcal{I} \subseteq [n]$

$$\mathbb{P}\left(\bigcap_{i \in \mathcal{I}} E_i\right) = \prod_{i \in \mathcal{I}} \mathbb{P}(E_i). \tag{1.11}$$

Proposition 1.3.2. If $E_1, \ldots, E_n$ are independent, then $E_1^c, \ldots, E_n$ are independent.

Proof. Fix $\mathcal{I} \subseteq [n]$. If $1 \notin \mathcal{I}$, then clearly

$$\mathbb{P}\left(\bigcap_{i \in \mathcal{I}} E_i\right) = \prod_{i \in \mathcal{I}} \mathbb{P}(E_i). \tag{1.12}$$

If $1 \in \mathcal{I}$, then define $\mathcal{I}' = \mathcal{I} \setminus \{1\}$.

$$\begin{aligned}
\mathbb{P}\left(E_1^c \cap \bigcap_{i \in \mathcal{I}'} E_i\right) &= \mathbb{P}\left(\bigcap_{i \in \mathcal{I}'} E_i \setminus E_1\right) \\
&= \mathbb{P}\left(\bigcap_{i \in \mathcal{I}'} E_i\right) - \mathbb{P}\left(E_1 \cap \bigcap_{i \in \mathcal{I}'} E_i\right) \\
&= \prod_{i \in \mathcal{I}'} \mathbb{P}(E_i) - \prod_{i \in \mathcal{I}} \mathbb{P}(E_i) \\
&= [1 - \mathbb{P}(E_1)] \prod_{i \in \mathcal{I}'} \mathbb{P}(E_i) \\
&= \mathbb{P}(E_1^c) \prod_{i \in \mathcal{I}'} \mathbb{P}(E_i).
\end{aligned} \tag{1.13}$$

$\square$

Definition 1.3.3. An infinite collection of events $\{E_\alpha : \alpha \in \mathcal{I}\}$ are independent if for any finite subset $\mathcal{J} \subseteq \mathcal{I}$, the events $\{E_i : i \in \mathcal{J}\}$ are independent.

1.4 Random Variables

Definition 1.4.1. Given a sample space Ω, a function $X : \Omega \longrightarrow \mathbb{R}$ is a *random variable.*

Example 1.4.2. Coin flipping. $X(H) = 1, X(T) = 0$.

Definition 1.4.3. Random variables $X_1, \dots, X_n$ are *independent* with respect to a measure $\mathbb{P}$ if for all sets $A_1, \dots, A_n \subseteq \mathbb{R}$,

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{t=1}^{n} \mathbb{P}(X_i \in A_i). \tag{1.14}$$

Definition 1.4.4. An infinite collection of random variables $\{X_\alpha : \alpha \in \mathcal{I}\}$ are independent if for any finite subset $\mathcal{J} \subseteq \mathcal{I}$, the random variables $\{X_i : i \in \mathcal{J}\}$ are independent.

1.5 Distributions

Lemma 1.5.1. A random variable X and probability measure $\mathbb{P}$ induce a probability measure μ on $\mathbb{R}$ defined by

$$\mu(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \tag{1.15}$$

for any $A \subseteq \mathbb{R}$.

Proof. $\mu(B) \in [0,1]$ is trivial since $\mathbb{P}$ is a measure. Also, since $X(\omega) \in \mathbb{R}$ for all $\omega \in \Omega$, $\mu(\mathbb{R}) = \mathbb{P}(\Omega) = 1$. Finally, consider disjoint $A_1, A_2, \dots \subseteq \mathbb{R}$, and define $E_i = \{\omega : X(\omega) \in A_i\}$. Since X is a function, it is impossible for $X(\omega) = a$ and $X(\omega) = b$ when $a \neq b$, so the E_i's are also disjoint. Then,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(X(\omega) \in \bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \sum_{i=1}^{\infty} \mu(A_i). \tag{1.16}$$

□

Definition 1.5.2. A random variable X and measure $\mathbb{P}$ generate the *cumulative distribution function* defined by

$$F(x) = \mu((-\infty, x]) = \mathbb{P}(X \leq x). \tag{1.17}$$

Theorem 1.5.3. A distribution function F for a random variable X uniquely defines the measure μ.

Proof. Outside the scope of this class. □

Example 1.5.4. Uniform. $\mathbb{P}$ is uniform on $[0,1]$ and $X(\omega) = \omega$. For $x \in [0,1]$,

$$F(x) = \mu([0,x)) = \mathbb{P}(X(\omega) \in [0,x)) = \mathbb{P}(\omega \in [0,x)) = x. \tag{1.18}$$

Theorem 1.5.5. A distribution function satisfies the following properties:

i) For all $x \leq y \in \mathbb{R}$, $F(x) \leq F(y)$;
ii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$;
iii) F is right-continuous.

Proof. i) follows from monotonicity of measure, while ii) and iii) require a result from next lecture.

□

Definition 1.5.6. The *inverse CDF* of a random variable X is defined by

$$F^{-1}(y) = \sup\{x : F(x) < y\}. \tag{1.19}$$

Theorem 1.5.7. If F satisfies properties i) to iii), it is the CDF of some random variable.

Proof. Let $U \sim \text{Uniform}(0,1)$ and define the random variable $Y(\omega) = F^{-1}(U(\omega))$. Consider arbitrary $x, t \in [0,1]$.

First, suppose $F^{-1}(t) > x$. That is, $\sup\{y : F(y) < t\} > x$, so $F(x) < t$ since F is non-decreasing.

Next, suppose $F^{-1}(t) \leq x$. By the same logic, $F(x+\delta) \geq t$ for any $\delta > 0$. Since F is right continuous, this gives $F(x) \geq t$.

These combined have given that $\{t : F^{-1}(t) \leq x\} = \{t : t \leq F(x)\}$. Thus,

$$\mathbb{P}(Y \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x), \tag{1.20}$$

so Y is a random variable with the CDF F. □

Lemma 1.5.8. If F is a CDF, it can only have countably many discontinuities.

Proof. Let $\mathcal{D}$ be the set of discontinuities of F. If $x \in \mathcal{D}$, then $F(x-) < F(x+)$, so there exists a rational $q_x \in (F(x-), F(x+))$. Further, since F is non-decreasing, for $x \neq y \in \mathcal{D}$, $q_x \neq q_y$, so each $x \mapsto q_x$ is injective, and thus $|\mathcal{D}| \leq |\mathbb{Q}|$. It remains to observe that $\mathbb{Q}$ is countable. □

Definition 1.5.9. A random variable has a *density function* $f : \mathbb{R} \to \mathbb{R}_+$ if for all $x \in \mathbb{R}$,

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y. \tag{1.21}$$

Lemma 1.5.10. If X has a density function, $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$.

Proof.

$$\mathbb{P}(X = x) = \lim_{\delta \to 0} \mathbb{P}(x - \delta < X \leq x + \delta) = \lim_{\delta \to 0} \int_{x-\delta}^{x+\delta} f(y)\mathrm{d}y = 0. \tag{1.22}$$

□

Definition 1.5.11. We define the *joint distribution function* of a *random vector* $X = (X_1, \ldots, X_n)$ by

$$F(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n), \tag{1.23}$$

using the notation $\mathbb{P}(A, B) \stackrel{\text{def}}{=} \mathbb{P}(A \cap B)$.

Theorem 1.5.12. A joint distribution function $F : \mathbb{R}^n \to \mathbb{R}$ satisfies the following properties:

i) If $(x_1, \ldots, x_n)$ and $(x_1', \ldots, x_n')$ satisfy $x_i \leq x_i'$ for all $i \in [n]$, $F(x_1, \ldots, x_n) \leq F(x_1', \ldots, x_n')$;
ii) $\lim_{x_i \to -\infty} F(x_1, \ldots, x_n) = 0$ for all $i \in [n]$ and $\lim_{x_1 \to \infty, \ldots, x_n \to \infty} F(x_1, \ldots, x_n) = 1$;
iii) $\lim_{h \to 0^+} F(x_1 + h, \ldots, x_n) = \cdots = \lim_{h \to 0^+} F(x_1, \ldots, x_n + h) = F(x_1, \ldots, x_n)$.

Proof. Analogous to the proof of Theorem 1.5.5. □

Definition 1.5.13. A random vector has a *joint density function* $f : \mathbb{R}^n \to \mathbb{R}_+$ if $\forall (x_1, \ldots, x_n) \subseteq \mathbb{R}^n$

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(y_1, \ldots, y_n)\mathrm{d}y_1 \cdots \mathrm{d}y_n. \tag{1.24}$$

Theorem 1.5.14. Random variables $X_1, \ldots, X_n$ with CDFs $F_1, \ldots, F_n$ are independent if and only if

$$F(x_1, \ldots, x_n) = \prod_{t=1}^{n} F_i(x_i) \qquad \forall (x_1, \ldots, x_n) \subseteq \mathbb{R}^n. \tag{1.25}$$

Proof. The reverse direction is obvious by definition. The forward definition is outside the scope of this class. □

1.6 Exercises

Exercise 1.1. Prove Proposition 1.2.6.

Exercise 1.2. Find an example of an Ω, $\mathbb{P}$, and sets A, B, C such that

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

but A, B, C are not independent. *Hint: Ω does not need to have more than 4 elements.*

Exercise 1.3. Prove that if $\{A_\alpha\}_{\alpha\in\mathcal{I}}$ is independent then so is $\{A_\alpha^c\}_{\alpha\in\mathcal{I}}$.

Exercise 1.4. Prove that if X and Y are independent, then X and $f(Y)$ are independent for any function $f : \mathbb{R} \to \mathbb{R}$.

Exercise 1.5. Let $\mathbb{P}$ be the uniform measure on $[0, 1]$. Define $A = (a, b)$ and $B = (c, d)$, with $a < c$. State necessary and sufficient conditions for A and B to be independent.

Exercise 1.6. Review the exponential family of distributions (note this is not just the exponential distribution, but the exponential *family*).

LECTURE 2

CONVERGENCE

2.1 Tail Events

Proposition 2.1.1 (Continuity of Measure.). If $\{A_n\} \nearrow A$ or $\{A_n\} \searrow A$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Proof. Suppose $\{A_n\} \nearrow A$, and by convention let $A_0 = \emptyset$. Define $B_1 = A_1$ and $B_n = A_n \cap A_{n-1}^c$ for $n \geq 2$. Observe that

$$\begin{aligned}\bigcup_{m=1}^{n} B_m &= \bigcup_{m=1}^{n} (A_m \cap A_{m-1}^c) \\ &= \bigcup_{m=1}^{n} A_m \cap \bigcup_{m=1}^{n} A_{m-1}^c \\ &= A_n \cap A_0^c \\ &= A_n.\end{aligned} \tag{2.1}$$

Then,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(\bigcup_{m=1}^{\infty} B_m) \\ &= \sum_{m=1}^{\infty} \mathbb{P}(B_m) \\ &= \lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{P}(B_m) \\ &= \lim_{n\to\infty} \mathbb{P}(\bigcup_{m=1}^{n} B_m) \\ &= \lim_{n\to\infty} \mathbb{P}(A_n).\end{aligned} \tag{2.2}$$

If $\{A_n\} \searrow A$, then $\{A_n^c\} \nearrow A^c$, so

$$\begin{aligned}\mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\ &= 1 - \lim_{n\to\infty} \mathbb{P}(A_n^c) \\ &= 1 - \lim_{n\to\infty} [1 - \mathbb{P}(A_n)] \\ &= \lim_{n\to\infty} \mathbb{P}(A_n).\end{aligned} \tag{2.3}$$

□

Example 2.1.2. Uniform measure and $A_n = [0, 1 - 1/n]$.

Example 2.1.3.

$$A_n = \begin{cases} \Omega, & n \text{ odd} \\ \emptyset, & n \text{ even} \end{cases} . \tag{2.4}$$

Theorem 1.5.5. A distribution function satisfies the following properties:

i) For all $x \leq y \in \mathbb{R}$, $F(x) \leq F(y)$;
ii) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$;
iii) F is right-continuous.

Proof. It remains to prove ii) and iii). Suppose $x_n \uparrow \infty$ and $y_n \downarrow -\infty$, so that $\{\omega : X(\omega) < x_n\} \nearrow \{\omega : X(\omega) < \infty\}$ and $\{\omega : X(\omega) < y_n\} \searrow \{\omega : X(\omega) < -\infty\}$. By continuity of measure, $F(x_n) \uparrow \mathbb{P}(X(\omega) < \infty) = 1$ and $F(y_n) \downarrow \mathbb{P}(X(\omega) < -\infty) = 0$. Since these were arbitrary sequences, ii) holds.

Now, fix an arbitrary $x \in \mathbb{R}$ and suppose $x_n \downarrow x$. By the same logic, $\{\omega : X(\omega) < x_n\} \searrow \{\omega : X(\omega) < x\}$, so $F(x_n) \downarrow F(x)$, showing iii). $\square$

Definition 2.1.4. Consider a sequence $A_1, A_2, \ldots \subseteq \Omega$. Define the *tail events* by

$$\limsup_{n\longrightarrow\infty} A_n = \{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \tag{2.5}$$

and

$$\liminf_{n\longrightarrow\infty} A_n = \{A_n \text{ a.a.}\} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \tag{2.6}$$

Corollary 2.1.5. $\mathbb{P}(A_n \text{ i.o.}) = 1 - \mathbb{P}(A_n^c \text{ a.a.})$.

Proposition 2.1.6.

$$\mathbb{P}(A_n \text{ a.a.}) \leq \liminf_{n\longrightarrow\infty} \mathbb{P}(A_n) \leq \limsup_{n\longrightarrow\infty} \mathbb{P}(A_n) \leq \mathbb{P}(A_n \text{ i.o.}). \tag{2.7}$$

Proof. Observe that $\bigcap_{k=n}^{\infty} A_k \subseteq \bigcap_{k=n+1}^{\infty} A_k$ for all n. So,

$$\begin{aligned}
\mathbb{P}(A_n \text{ a.a.}) &= P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) \\
&= \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\
&= \liminf_{n\longrightarrow\infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\
&\leq \liminf_{n\longrightarrow\infty} P\left(A_n\right).
\end{aligned} \tag{2.8}$$

The second inequality is by definition. The third inequality is an exercise. $\square$

Theorem 2.1.7 (Borel-Cantelli Lemma).

i) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$;
ii) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and $\{A_n\}$ are independent, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. For i), observe that $\bigcup_{k=n+1}^{\infty} A_k \subseteq \bigcup_{k=n}^{\infty} A_k$ for all n. Thus,

$$\begin{aligned}\mathbb{P}(A_n \text{ i.o.}) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &\le \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) \\ &= 0.\end{aligned} \tag{2.9}$$

For ii), observe that $\bigcap_{k=n}^{\infty} A_k^c \subseteq \bigcap_{k=n+1}^{\infty} A_k^c$ for all n. Thus,

$$\begin{aligned}1 - \mathbb{P}(A_n \text{ i.o.}) &= \mathbb{P}((A_n \text{ i.o.})^c) \\ &= \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) \\ &= \lim_{n\to\infty} \prod_{k=n}^{\infty} [1 - \mathbb{P}(A_k)] \\ &\le \lim_{n\to\infty} \prod_{k=n}^{\infty} e^{-\mathbb{P}(A_k)} \\ &= \lim_{n\to\infty} e^{-\sum_{k=n}^{\infty}\mathbb{P}(A_k)} \\ &= \lim_{n\to\infty} 0 \\ &= 0.\end{aligned} \tag{2.10}$$

□

Example 2.1.8 (converse does not hold for i)). Uniform measure and $A_n = [0, 1/n]$. Then,

$$\begin{aligned}A_n \text{ i.o.} &= \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} [0, 1/k] \\ &= \{0\},\end{aligned} \tag{2.11}$$

but $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} 1/n = \infty$.

Example 2.1.9 (independence is needed for ii)). Define $c_1, c_2, \ldots$ such that c_i is a fair coin toss. Let $A_1, A_2, \ldots$ be such that $A_i = \{c_1 = 1\}$. Then, $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} 1/2 = \infty$ and $\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(A_n) = 1/2$.

2.2 Types of Convergence

Definition 2.2.1. A sequence of random variables X_n *converges almost surely* to X if

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1, \tag{2.12}$$

and is denoted by $X_n \longrightarrow X$ a.s.

Proposition 2.2.2. If for all $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0$, then $X_n \longrightarrow X$ a.s.

Proof. Consider that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ if and only if for all $\varepsilon > 0$, $|X_n(\omega) - X(\omega)| < \varepsilon$ for all but finitely many n. Thus,

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = \mathbb{P}(\forall\varepsilon > 0, |X_n(\omega) - X(\omega)| < \varepsilon \text{ a.a.}) = 1 - \mathbb{P}(\exists\varepsilon > 0, |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ i.o.}). \tag{2.13}$$

Next,

$$\begin{aligned}\mathbb{P}(\exists\varepsilon > 0, |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ i.o.}) &\leq \mathbb{P}(\exists\varepsilon \in \mathbb{Q}_+, |X_n(\omega) - X(\omega)| \geq \varepsilon \text{ i.o.}) \\ &\leq \sum_{\varepsilon\in\mathbb{Q}_+} \mathbb{P}(|X_n(\omega) - X(\omega)| \geq \varepsilon \text{ i.o.}) \\ &= 0.\end{aligned} \tag{2.14}$$

□

Corollary 2.2.3. If for all $\varepsilon > 0$, $\sum_{n=1}^{\infty} \mathbb{P}(|X_n(\omega) - X(\omega)| \geq \varepsilon) < \infty$, $X_n \longrightarrow X$ a.s.

Proof. Borel-Cantelli combined with the assumption implies the hypothesis of Proposition 2.2.2. □

Definition 2.2.4. A sequence of random variables X_n *converges in probability* to X if for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \leq \varepsilon) = 1, \tag{2.15}$$

and is denoted by $X_n \xrightarrow{P} X$.

Proposition 2.2.5. If $X_n \longrightarrow X$ a.s. then $X_n \xrightarrow{P} X$.

Proof. Fix $\varepsilon > 0$ and let $E_n = \{\omega : \exists m \geq n, |X_m(\omega) - X(\omega)| \geq \varepsilon\}$. Observe that $E_{n+1} \subseteq E_n$, and if $\omega \in \bigcap_{n=1}^{\infty} E_n$ then $X_n(\omega) \not\to X(\omega)$. Thus, using continuity of probability,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(|X_n(\omega) - X(\omega)| \geq \varepsilon) &\leq \lim_{n\to\infty} \mathbb{P}(E_n) \\ &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} E_n\right) \\ &\leq \mathbb{P}(X_n \not\to X) \\ &= 0.\end{aligned} \tag{2.16}$$

□

Example 2.2.6. X_n independent with $\mathbb{P}(X_n = 1) = 1/n$, $\mathbb{P}(X_n = 0) = 1 - 1/n$. For all $\varepsilon > 0$,

$$\mathbb{P}(X_n > \varepsilon) = 1/n \longrightarrow 0. \tag{2.17}$$

So, $X_n \xrightarrow{P} 0$. But, $P(X_n = 1 \text{ i.o.}) = 1$, so $P(X_n \longrightarrow 0) = 0$.

Theorem 2.2.7. If $X_n \xrightarrow{P} X$, there exists a subsequence such that $X_{n_k} \longrightarrow X$ a.s.

Proof. By definition, for each $k \in \mathbb{N}$, there exists n_k such that for $n \geq n_k$

$$\mathbb{P}(|X_n - X| > 2^{-k}) \leq 2^{-k}. \tag{2.18}$$

Further, choose these such that $n_{k+1} \geq n_k$, and define the sets

$$A_k = \{\omega : |X_{n_k}(\omega) - X(\omega)| > 2^{-k}\}. \tag{2.19}$$

Clearly,

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) \leq \sum_{k=1}^{\infty} 2^{-k} < \infty, \tag{2.20}$$

so by Borel Cantelli $\mathbb{P}(A_k \text{ i.o.}) = 0$. Finally, observe that $|X_{n_k}(\omega) - X(\omega)| > 2^{-k}$ only finitely many times implies $X_{n_k}(\omega) \longrightarrow X(\omega)$, so

$$1 = \mathbb{P}[(A_k \text{ i.o.})^c] \leq \mathbb{P}(X_{n_k} \longrightarrow X). \tag{2.21}$$

□

Theorem 2.2.8 (Continuous Mapping Theorem). If f is a continuous function then

i) $X_n \longrightarrow X$ a.s. implies that $f(X_n) \longrightarrow f(X)$ a.s.;
ii) $X_n \xrightarrow{P} X$ implies that $f(X_n) \xrightarrow{P} f(X)$;

Proof.

i) f continuous means that $X_n(\omega) \to X(\omega)$ implies $f(X_n(\omega)) \to f(X(\omega))$, so

$$1 = \mathbb{P}(X_n \to X) \leq \mathbb{P}(f(X_n) \to f(X)); \tag{2.22}$$

ii) For all $\varepsilon > 0$, there exists $\delta > 0$ such that $|X_n(\omega) - X(\omega)| \leq \delta$ implies $|f(X_n(\omega)) - f(X(\omega))| \leq \varepsilon$, but

$$1 = \lim_{n\to\infty} \mathbb{P}(|X_n - X| \leq \delta) \leq \lim_{n\to\infty} \mathbb{P}(|f(X_n) - f(X)| \leq \varepsilon); \tag{2.23}$$

□

2.3 Exercises

Exercise 2.1. Prove the third inequality of Proposition 2.1.6.

Exercise 2.2 (Rosenthal 3.6.7). Consider $\Omega = \{a, b, c\}$ with the measure $\mathbb{P}(a) = \mathbb{P}(b) = \mathbb{P}(c) = 1/3$. Find examples of $A_n \subseteq \Omega$ such that the inequalities in Proposition 2.1.6 are strict.

Exercise 2.3 (Rosenthal 3.6.9). Prove that for any collections $\{A_n\}$ and $\{B_n\}$,

$$\limsup(A_n \cap B_n) \subseteq \limsup A_n \cap \limsup B_n, \tag{2.24}$$

and find example where the inclusion is strict and where it is equality.

Exercise 2.4 (Rosenthal 3.6.12). Let X be a random variable such that $\mathbb{P}(X > 0) > 0$. Prove that there exists a $\delta > 0$ such that $\mathbb{P}(X \geq \delta) > 0$.

Exercise 2.5. Find an example of $\mathbb{P}$ and A_n such that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ but $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Exercise 2.6. Find a sequence X_n and X such that $X_n \xrightarrow{P} X$ but $X_n \not\to X$ a.s.

Exercise 2.7. Show that $X_n \longrightarrow X$ a.s. or $X_n \xrightarrow{P} X$ if and only if $(X_n - X) \longrightarrow 0$ a.s. or $(X_n - X) \xrightarrow{P} 0$ respectively.

Exercise 2.8. Show that if $X_n - a_n \xrightarrow{P} 0$ and $a_n \to a$, then $X_n \xrightarrow{P} a$.

LECTURE 3

EXPECTATION

3.1 Simple Random Variables

Definition 3.1.1 (Simple Expectation). If X only takes finitely many values $x_1, \dots, x_n$, and $A_i = \{\omega : X(\omega) = x_i\}$, define the *expectation* of X by

$$\mathbb{E}X = \sum_{i=1}^{n} x_i \mathbb{P}(A_i). \tag{3.1}$$

Example 3.1.2. For an arbitrary set $A \subset \Omega$, let $Y(\omega) = \mathbb{I}\{\omega \in A\}$. Then, $\mathbb{E}Y = \mathbb{P}(A)$.

Example 3.1.3. P is uniform measure on $\Omega = [0, 1]$, and

$$X(\omega) = \begin{cases} 5, & \omega > 1/3 \\ 3, & \omega \leq 1/3 \end{cases}. \tag{3.2}$$

Proposition 3.1.4. If X and Y are simple random variables, the following properties hold.

i) If $X \geq 0$ a.s., $\mathbb{E}X \geq 0$;
ii) For all $a \in \mathbb{R}$, $\mathbb{E}aX = a\mathbb{E}X$;
iii) $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$.

Proof. Both i) and ii) are trivial. For iii), let X take values $x_1, \dots, x_n$ on sets $A_1, \dots, A_n$ and Y take values $y_1, \dots, y_m$ on sets $B_1, \dots, B_m$. Observe that there are mn events $C_{ij} = \{X = x_i \cap Y = y_j\}$. Then,

$$\begin{aligned} \mathbb{E}(X+Y) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) \mathbb{P}(C_{ij}) \\ &= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} \mathbb{P}(X = x_i \cap Y = y_j) + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} \mathbb{P}(X = x_i \cap Y = y_j) \\ &= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) + \sum_{j=1}^{m} y_j \mathbb{P}(Y = y_j). \end{aligned} \tag{3.3}$$

□

Lemma 3.1.5. If properties i) to iii) hold, the following properties also hold.

iv) If $X \leq Y$ a.s. then $\mathbb{E}X \leq \mathbb{E}Y$;
v) If $X = Y$ a.s. then $\mathbb{E}X = \mathbb{E}Y$;
vi) $|\mathbb{E}X| \leq \mathbb{E}\,|X|$.

Proof.

iv) $X \leq Y$ a.s. $\implies Y - X \geq 0$ a.s. $\implies \mathbb{E}(Y - X) \geq 0 \implies \mathbb{E}Y \geq \mathbb{E}X$;
v) $X = Y$ a.s. $\implies X \leq Y$ a.s and $Y \leq X$ a.s. so $\mathbb{E}X \leq \mathbb{E}Y$ and $\mathbb{E}Y \leq \mathbb{E}X$;
vi) Trivially, $X \leq |X|$ and $-X \leq |X|$, so $\mathbb{E}X \leq \mathbb{E}|X|$ and $-\mathbb{E}X = \mathbb{E}(-X) \leq \mathbb{E}|X|$.

□

3.2 Bounded Random Variables

Definition 3.2.1. A random variable X is *bounded* if there exists $M < \infty$ such that $|X| \leq M$ a.s.

Proposition 3.2.2. If X is bounded,

$$\inf\{\mathbb{E}Z : Z \text{ simple}, Z \geq X \text{ a.s.}\} = \sup\{\mathbb{E}Y : Y \text{ simple}, Y \leq X \text{ a.s.}\}. \tag{3.4}$$

Proof. Define $L = \sup\{\mathbb{E}Y : Y \text{ simple}, Y \leq X \text{ a.s.}\}$ and $U = \inf\{\mathbb{E}Z : Z \text{ simple}, Z \geq X \text{ a.s.}\}$. Since $Y \leq X$ a.s. and $X \leq Z$ a.s. implies $\mathbb{E}Y \leq \mathbb{E}Z$, $L \leq U$.

Now, consider an arbitrary n, and for each k in $\{-n, \ldots, n\}$, define

$$E_k = \left\{\omega \in \Omega : \frac{kM}{n} \geq X(\omega) > \frac{(k-1)M}{n}\right\}. \tag{3.5}$$

Define the simple random variables

$$\psi_n(\omega) = \sum_{-n}^{n} \frac{kM}{n} \mathbb{I}\{\omega \in E_k\} \text{ and } \phi_n(\omega) = \sum_{-n}^{n} \frac{(k-1)M}{n} \mathbb{I}\{\omega \in E_k\}. \tag{3.6}$$

Observe that $\psi_n(\omega) - \phi_n(\omega) = M/n$, $\psi_n(\omega) \geq X(\omega)$, and $\phi_n(\omega) \leq X(\omega)$. Thus,

$$\begin{aligned} U &\leq \mathbb{E}\psi_n \\ &= \mathbb{E}\phi_n + M/n \\ &\leq L + M/n. \end{aligned} \tag{3.7}$$

Since this is true for all n, $U \leq L$. □

Definition 3.2.3 (Bounded Expectation). If X is bounded, define

$$\inf\{\mathbb{E}Z : Z \text{ simple}, Z \geq X \text{ a.s.}\} = \mathbb{E}X = \sup\{\mathbb{E}Y : Y \text{ simple}, Y \leq X \text{ a.s.}\}. \tag{3.8}$$

Proposition 3.2.4. If X is bounded, $\mathbb{E}X$ satisfies properties i) to iii).

Proof. For i), if $X \geq 0$ a.s., there exists simple Y such that $X \geq Y \geq 0$ a.s., so $\sup\{\mathbb{E}Y : Y \text{ simple}, Y \leq X \text{ a.s.}\} \geq \mathbb{E}Y \geq 0$. For ii), consider $a > 0$. Then,

$$\begin{aligned} \mathbb{E}X &= \sup\{\mathbb{E}Y : Y \text{ simple}, Y \leq X \text{ a.s.}\} \\ &= \frac{1}{a} \sup\{\mathbb{E}aY : Y \text{ simple}, aY \leq aX \text{ a.s.}\} \\ &= \frac{1}{a}\mathbb{E}aX. \end{aligned} \tag{3.9}$$

If $a < 0$ then use inf. For iii),

$$\begin{aligned}
\mathbb{E}X + \mathbb{E}Y &= \sup\{\mathbb{E}W : W \text{ simple}, W \leq X \text{ a.s.}\} + \sup\{\mathbb{E}V : V \text{ simple}, V \leq Y \text{ a.s.}\} \\
&= \sup\{\mathbb{E}W + \mathbb{E}V : W, V \text{ simple}, W \leq X \text{ a.s. and } V \leq Y \text{ a.s.}\} \\
&= \sup\{\mathbb{E}(W + V) : W, V \text{ simple}, W \leq X \text{ a.s. and } V \leq Y \text{ a.s.}\} \\
&\leq \sup\{\mathbb{E}(W + V) : W, V \text{ simple}, (W + V) \leq (X + Y) \text{ a.s.}\} \\
&= \mathbb{E}(X + Y).
\end{aligned} \tag{3.10}$$

Also,

$$\begin{aligned}
\mathbb{E}X + \mathbb{E}Y &= \inf\{\mathbb{E}W : W \text{ simple}, W \geq X \text{ a.s.}\} + \inf\{\mathbb{E}V : V \text{ simple}, V \geq Y \text{ a.s.}\} \\
&= \inf\{\mathbb{E}W + \mathbb{E}V : W, V \text{ simple}, W \geq X \text{ a.s. and } V \geq Y \text{ a.s.}\} \\
&= \inf\{\mathbb{E}(W + V) : W, V \text{ simple}, W \geq X \text{ a.s. and } V \geq Y \text{ a.s.}\} \\
&\geq \inf\{\mathbb{E}(W + V) : W, V \text{ simple}, (W + V) \geq (X + Y) \text{ a.s.}\} \\
&= \mathbb{E}(X + Y).
\end{aligned} \tag{3.11}$$

□

3.3 Non-Negative Random Variables

Definition 3.3.1 (Non-Negative Expectation). If $X \geq 0$ a.s., define

$$\mathbb{E}X = \sup\{\mathbb{E}Y : Y \text{ bounded}, 0 \leq Y \leq X \text{ a.s.}\}. \tag{3.12}$$

Lemma 3.3.2. Let $X \geq 0$ a.s., and recall notation $a \wedge b = \min\{a, b\}$. Then,

$$\lim_{n\to\infty} \mathbb{E}(X \wedge n) = \mathbb{E}X. \tag{3.13}$$

Proof. Define $X_n = X \wedge n$. Clearly $X_n \leq X$, so $\mathbb{E}X_n \leq \mathbb{E}X$. Also $\mathbb{E}X_n \leq \mathbb{E}X_{n+1}$ for all n, so the limit exists, and thus $\lim_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}X$. Consider a bounded Y such that $0 \leq Y \leq X$ a.s. Then, for large n, $\mathbb{E}X_n \geq \mathbb{E}Y$, so $\lim_{n\to\infty} \mathbb{E}X_n \geq \sup\{\mathbb{E}Y : Y \text{ bounded}, 0 \leq Y \leq X \text{ a.s.}\} = \mathbb{E}X$. □

Proposition 3.3.3. If $X \geq 0$ a.s., $\mathbb{E}X$ satisfies properties i) to iii).

Proof. For i) the proof is the same as for bounded random variables. For ii)

$$\begin{aligned}
\mathbb{E}X &= \sup\{\mathbb{E}Y : Y \text{ bounded}, 0 \leq Y \leq X \text{ a.s.}\} \\
&= \frac{1}{a} \sup\{\mathbb{E}aY : Y \text{ bounded}, 0 \leq aY \leq aX \text{ a.s.}\} \\
&= \frac{1}{a} \mathbb{E}aX.
\end{aligned} \tag{3.14}$$

For iii), by the same argument as for bounded random variables,

$$\mathbb{E}X + \mathbb{E}Y \leq \sup\{\mathbb{E}Z : Z \text{ bounded}, 0 \leq Z \leq X + Y \text{ a.s.}\} = \mathbb{E}(X + Y). \tag{3.15}$$

Also,

$$\begin{aligned}\mathbb{E}(X+Y) &= \lim_{n\to\infty}\mathbb{E}[(X+Y)\wedge n] \\ &\le \lim_{n\to\infty}\mathbb{E}[(X\wedge n)+(Y\wedge n)] \\ &= \lim_{n\to\infty}[\mathbb{E}(X\wedge n)+\mathbb{E}(Y\wedge n)] \\ &= \mathbb{E}X+\mathbb{E}Y.\end{aligned} \tag{3.16}$$

□

3.4 Integrable Random Variables

Example 3.4.1. Consider the uniform measure on [0,1], and let $X(\omega) = 2^k$ when $2^{-k} \le \omega < 2^{-(k-1)}$, $k \in \mathbb{N}$. Then, for any n, let $X_n(\omega) = 2^k$ when $2^{-k} \le \omega < 2^{-(k-1)}$, $k \in [n]$ and 0 otherwise. Clearly, $X_n \le X$, X_n simple, and $\mathbb{E}X_n = \sum_{k=1}^n 2^k(2^{-(k-1)} - 2^{-k}) = 2^k 2^{-k} = n$. Thus, for any n, $\sup\{\mathbb{E}Y : Y \text{ simple}, Y \le X \text{ a.s.}\} \ge n$, so $\mathbb{E}X = \infty$.

Definition 3.4.2. A random variable X is *integrable* if $\mathbb{E}\,|X| < \infty$.

Definition 3.4.3 (Integrable Expectation). For any random variable X, let $X^+(\omega) = \max\{X(\omega), 0\}$ and $X^-(\omega) = \max\{-X(\omega), 0\}$. Observe that both X^+ and X^- are non-negative, $X = X^+ - X^-$, and $|X| = X^+ + X^-$. Then, if X is integrable, define $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$.

Proposition 3.4.4. If X is integrable, properties i) to iii) hold.

Proof. For i), observe that $X^- = 0$. For ii), if $a > 0$,

$$\begin{aligned}\mathbb{E}aX &= \mathbb{E}(aX)^+ - \mathbb{E}(aX)^- \\ &= \mathbb{E}aX^+ - \mathbb{E}aX^- \\ &= a(\mathbb{E}X^+ - \mathbb{E}X^-) \\ &= a\mathbb{E}X.\end{aligned} \tag{3.17}$$

If $a < 0$,

$$\begin{aligned}\mathbb{E}aX &= \mathbb{E}(aX)^+ - \mathbb{E}(aX)^- \\ &= \mathbb{E}(-aX^-) - \mathbb{E}(-aX^+) \\ &= a(\mathbb{E}X^+ - \mathbb{E}X^-) \\ &= a\mathbb{E}X.\end{aligned} \tag{3.18}$$

For iii), since $\mathbb{E}(X+Y)^- \le \mathbb{E}\,|X+Y| \le \mathbb{E}\,|X| + \mathbb{E}\,|Y| < \infty$,

$$\begin{aligned}\mathbb{E}(X+Y) &= \mathbb{E}(X+Y)^+ - \mathbb{E}(X+Y)^- \\ &= \mathbb{E}(X^+ + Y^+) - \mathbb{E}(X^- + Y^-) \\ &= \mathbb{E}(X^+ - X^-) + \mathbb{E}(Y^+ - Y^-) \\ &= \mathbb{E}X + \mathbb{E}Y.\end{aligned} \tag{3.19}$$

□

3.5 Familiar Properties

Theorem 3.5.1. If X and Y are independent with $\mathbb{E}X, \mathbb{E}Y < \infty$, then $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Proof. First, suppose X and Y are simple, so that $X = \sum_{i=1}^n x_i \mathbb{I}\{A_i\}$ and $Y = \sum_{j=1}^m y_j \mathbb{I}\{B_j\}$. Then,

$$\begin{aligned}\mathbb{E}(XY) &= \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^m x_i y_j \mathbb{I}\{A_i \cap B_j\}\right] \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j \mathbb{P}(X \in A_i, Y \in B_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j \mathbb{P}(X \in A_i)\mathbb{P}(Y \in B_j) \\ &= \sum_{i=1}^n x_i \mathbb{P}(X \in A_i) \sum_{j=1}^m y_j \mathbb{P}(Y \in B_j) \\ &= (\mathbb{E}X)(\mathbb{E}Y).\end{aligned} \tag{3.20}$$

Next suppose that X and Y are bounded and non-negative. Then, since $Z_1 \geq X, Z_2 \geq Y$ implies $Z_1 Z_2 \geq XY$

$$\begin{aligned}(\mathbb{E}X)(\mathbb{E}Y) &= \inf\{\mathbb{E}Z_1 : Z_1 \text{ simple }, Z_1 \geq X \text{ a.s.}\} \inf\{\mathbb{E}Z_2 : Z_2 \text{ simple }, Z_2 \geq Y \text{ a.s.}\} \\ &= \inf\{(\mathbb{E}Z_1)(\mathbb{E}Z_2) : Z_1, Z_2 \text{ simple and independent }, Z_1 \geq X, Z_2 \geq Y \text{ a.s.}\} \\ &\geq \inf\{\mathbb{E}(Z_1 Z_2) : Z_1, Z_2 \text{ simple and independent }, Z_1 Z_2 \geq XY \text{ a.s.}\} \\ &\geq \inf\{\mathbb{E}(Z) : Z \text{ simple }, Z \geq XY \text{ a.s.}\} \\ &= \mathbb{E}(XY).\end{aligned} \tag{3.21}$$

Similarly,

$$\begin{aligned}(\mathbb{E}X)(\mathbb{E}Y) &= \sup\{\mathbb{E}Z_1 : Z_1 \text{ simple }, Z_1 \leq X \text{ a.s.}\} \sup\{\mathbb{E}Z_2 : Z_2 \text{ simple }, Z_2 \leq Y \text{ a.s.}\} \\ &= \sup\{(\mathbb{E}Z_1)(\mathbb{E}Z_2) : Z_1, Z_2 \text{ simple and independent }, Z_1 \leq X, Z_2 \leq Y \text{ a.s.}\} \\ &\leq \sup\{\mathbb{E}(Z_1 Z_2) : Z_1, Z_2 \text{ simple and independent }, 0 \leq Z_1 Z_2 \leq XY \text{ a.s.}\} \\ &\leq \sup\{\mathbb{E}(Z) : Z \text{ simple }, 0 \leq Z \leq XY \text{ a.s.}\} \\ &= \mathbb{E}(XY).\end{aligned} \tag{3.22}$$

That is $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$. Now, recalling that if X and Y are independent then all combinations of X^+, X^-, Y^+, and Y^- are independent, for arbitrary bounded and independent X, Y,

$$\begin{aligned}\mathbb{E}(XY) &= \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] \\ &= \mathbb{E}[X^+Y^+ - X^-Y^+ - X^+Y^- - X^-Y^-] \\ &= (\mathbb{E}X^+)(\mathbb{E}Y^+) - (\mathbb{E}X^-)(\mathbb{E}Y^+) - (\mathbb{E}X^+)(\mathbb{E}Y^-) - (\mathbb{E}X^-)(\mathbb{E}Y^-) \\ &= (\mathbb{E}X)(\mathbb{E}Y),\end{aligned} \tag{3.23}$$

where we applied the result for bounded and non-negative repeatedly. This argument can be repeated for the non-negative case and the integrable case.

□

Definition 3.5.2. The *variance* of a random variable X is defined as $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$. The *covariance* of random variables X and Y is $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$. The *correlation* between random variables X and Y is $\mathrm{Corr}(X,Y) = \mathrm{Cov}(X,Y)/\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$.

3.6 Exercises

Exercise 3.1. Show that the following properties hold:

i) $\mathrm{Var}(X) \geq 0$,
ii) For any $c \in \mathbb{R}$, $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$,
iii) If X and Y are independent, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Exercise 3.2. Prove that Definition 3.1.1 is well-defined. That is, if $\{A_i\}$ and $\{B_j\}$ are partitions of Ω with $\sum_{i=1}^n x_i \mathbb{I}_{A_i} = \sum_{j=1}^m y_j \mathbb{I}_{B_j}$, then $\sum_{i=1}^n x_i \mathbb{P}(A_i) = \sum_{j=1}^m y_j \mathbb{P}(B_j)$.

Exercise 3.3. Prove that if $X \sim \mathbb{P}$ and $Y \sim \mathbb{P}$, then $\mathbb{E}X = \mathbb{E}Y$. *Hint: start with simple random variables and work up.*

Exercise 3.4 (Rosenthal 4.5.2). For X such that $\mathbb{E}X < \infty$ and $a \in \mathbb{R}$, prove that $\mathbb{E}[\max\{X, a\}] \geq \max\{\mathbb{E}X, a\}$.

Exercise 3.5 (Rosenthal 4.5.3). Find random variables $X, Y : [0,1] \to \mathbb{R}$ such that under the uniform measure, $\mathbb{P}(X > Y) > 1/2$ but $\mathbb{E}X < \mathbb{E}Y$.

Exercise 3.6 (Rosenthal 4.5.10). Let X_1, X_2 be i.i.d. with $\mathbb{E}X_i = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$, and let N be integer valued with $\mathbb{E}N = m$ and $\mathrm{Var}(N) = v$ and independent from all X_i. Show that

$$\mathrm{Var}\left(\sum_{i=1}^N X_i\right) = \sigma^2 m + \mu^2 v. \tag{3.24}$$

Exercise 3.7 (Rosenthal 4.5.13). Find examples of $X : [0,1] \to \mathbb{R}$ with uniform measure $\mathbb{P}$ such that

a) $\mathbb{E}X^+ = \infty$ and $0 < \mathbb{E}X^- < \infty$,
b) $\mathbb{E}X^- = \infty$ and $0 < \mathbb{E}X^+ < \infty$,
c) $\mathbb{E}X^+ = \mathbb{E}X^- = \infty$,
d) $\mathbb{E}X < \infty$ but $\mathbb{E}X^2 = \infty$.

LECTURE 4

PROPERTIES OF EXPECTATION

4.1 Concentration of Measure

Theorem 4.1.1 (Markov's Inequality). If $X \geq 0$ a.s., then for all $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}. \tag{4.1}$$

Proof. Define $Z(\omega) = a\mathbb{I}\{X(\omega) \geq a\}$. Then, $Z \leq X$ a.s., and since Z is simple, $\mathbb{E}Z = aP(X \geq a)$. □

Corollary 4.1.2 (Chebyshev's Inequality). For all $t \geq 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}. \tag{4.2}$$

Proof. Apply Markov's to $Y = (X - \mathbb{E}X)^2$. □

Definition 4.1.3. The *moment generating function* of a random variable X is defined by

$$M_X(\lambda) = \mathbb{E}e^{\lambda X}. \tag{4.3}$$

Corollary 4.1.4 (Chernoff's Inequality). For all $t \geq 0$,

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \inf_{\lambda > 0} M_{X-\mathbb{E}X}(\lambda)e^{-\lambda t}. \tag{4.4}$$

Proof. Markov's. □

Definition 4.1.5. A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is *convex* if for all $\lambda \in (0, 1)$ and $x, y \in \mathbb{R}$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{4.5}$$

Lemma 4.1.6. If Y satisfies $\mathbb{E}Y = 0$ and $a \leq Y \leq b$ a.s., for all $\lambda > 0$

$$M_Y(\lambda) \leq e^{\lambda^2(b-a)^2/8}. \tag{4.6}$$

Proof. Since $Y \in [a, b]$, we can write it as a convex combination via $Y = \alpha a + (1 - \alpha)b$ for some $\alpha \in [0, 1]$. In particular, this holds for $\alpha = (b - Y)/(b - a)$. Since $e^{\lambda Y}$ is convex in Y,

$$e^{\lambda Y} = e^{\lambda[\alpha a + (1-\alpha)b]} \leq \alpha e^{\lambda a} + (1 - \alpha)e^{\lambda b} = \frac{b - Y}{b - a}e^{\lambda a} + \frac{Y - a}{b - a}e^{\lambda b}. \tag{4.7}$$

Taking expectation of both sides gives

$$\mathbb{E}e^{\lambda Y} \leq \frac{b}{b - a}e^{\lambda a} - \frac{a}{b - a}e^{\lambda b}. \tag{4.8}$$

Then, define $p = b/(b-a)$ and $u = (b-a)\lambda$. Also, consider the function

$$\varphi(u) = \log(pe^{\lambda a} + (1-p)e^{\lambda b}) = \lambda a + \log(p + (1-p)e^{\lambda(b-a)}) = (p-1)u + \log(p + (1-p)e^u). \quad (4.9)$$

Then, from Taylor expanding we see that there exists a $\xi \in \mathbb{R}$ such that

$$\varphi(u) = \varphi(0) + \varphi'(0)u + \frac{1}{2}\varphi''(\xi)u^2. \quad (4.10)$$

Observe that $\varphi(0) = 0$ and

$$\varphi'(x) = (p-1) + \frac{(1-p)e^x}{p + (1-p)e^x} = (p-1) + 1 - \frac{p}{p + (1-p)e^x}, \quad (4.11)$$

so $\varphi'(0) = 0$ as well. Finally,

$$\varphi''(x) = \frac{p(1-p)e^x}{[p + (1-p)e^x]^2}, \quad (4.12)$$

which you can check satisfies $\varphi''(x) \leq 1/4$ for all $x \in \mathbb{R}$. That is,

$$\mathbb{E}e^{\lambda Y} \leq e^{\varphi(u)} \leq e^{u^2/8} \leq e^{\lambda^2(b-a)^2/8}. \quad (4.13)$$

□

Theorem 4.1.7 (Hoeffding's Inequality). Suppose $X_1, X_2, \ldots$ are independent with $a_i \leq X_i \leq b_i$ a.s. for all i. Then, letting $S_n = \sum_{i=1}^n X_i$, for all $t > 0$

$$\mathbb{P}\left(|S_n - \mathbb{E}S_n| \geq t\right) \leq 2\exp\left\{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right\}. \quad (4.14)$$

Proof. First, we apply Chernoff's inequality to obtain

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq t) \leq \inf_{\lambda > 0} e^{-\lambda t} M_{S_n - \mathbb{E}S_n}(\lambda). \quad (4.15)$$

By independence and the above lemma,

$$M_{S_n - \mathbb{E}S_n}(\lambda) = \prod_{i=1}^n M_{X_i - \mathbb{E}X_i}(\lambda) \leq \prod_{i=1}^n e^{\lambda^2(b_i - a_i)^2/8} = \exp\left\{\frac{\lambda^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right\}. \quad (4.16)$$

Thus,

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq t) \leq \inf_{\lambda > 0} \exp\left\{-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right\}. \quad (4.17)$$

Taking $\lambda = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}$ gives one direction of the result. To get the other direction consider $-X_1, \ldots, -X_n$. □

4.2 Various Inequalities

Proposition 4.2.1 (Jensen's Inequality). If f is a convex function and X is a random variable such that $f(X)$ is integrable,

$$f(\mathbb{E}X) \leq \mathbb{E}f(X). \quad (4.18)$$

The inequality is flipped if f is concave.

Proof. Since f is convex, there exists a function g such that $g(x) = ax + b$ such that $g(x) \leq f(x)$ for all x and $g(\mathbb{E}X) = f(\mathbb{E}X)$. Then,

$$\mathbb{E}f(X) \geq \mathbb{E}g(X) = \mathbb{E}[aX + b] = a\mathbb{E}X + b = g(\mathbb{E}X) = f(\mathbb{E}X). \tag{4.19}$$

If concave then $-f$ is convex. □

Example 4.2.2. $(\mathbb{E}X)^2 \leq \mathbb{E}X^2$, $\mathbb{E}\log(X) \leq \log(\mathbb{E}X)$.

Lemma 4.2.3. If $0 < p < q$,

$$[\mathbb{E}\,|X|^p]^{1/p} \leq [\mathbb{E}\,|X|^q]^{1/q}. \tag{4.20}$$

Proof. Since $q/p > 1$, $f(x) = x^{q/p}$ is convex. Thus,

$$[\mathbb{E}X^p]^{q/p} \leq \mathbb{E}(X^p)^{q/p} = \mathbb{E}X^q. \tag{4.21}$$

□

Lemma 4.2.4. If $X \geq 0$ a.s. and $\mathbb{E}X = 0$, $X = 0$ a.s.

Proof. Define $A_n = \{\omega \in \Omega : X(\omega) > 1/n\}$ for $n \in \mathbb{N}$. By Markov, $P(A_n) = 0$ for all n. Also, $A_n \nearrow A = \{\omega \in \Omega : X(\omega) > 0\}$, so by continuity of probability $P(A) = \lim_{n\to\infty} P(A_n) = 0$. □

Proposition 4.2.5 (Holder's Inequality). If $p, q > 1$ such that $1/p + 1/q = 1$,

$$\mathbb{E}\,|XY| \leq [\mathbb{E}\,|X|^p]^{1/p}[\mathbb{E}\,|Y|^q]^{1/q}. \tag{4.22}$$

Proof. If $[\mathbb{E}\,|X|^p]^{1/p} = 0$, $|X|^p = 0$ a.s. so $X = 0$ a.s., which implies $|XY| = 0$ a.s. The same holds if $[\mathbb{E}\,|Y|^q]^{1/q} = 0$.

Otherwise, let

$$X^* = \frac{|X|}{[\mathbb{E}\,|X|^p]^{1/p}} \text{ and } Y^* = \frac{|Y|}{[\mathbb{E}\,|Y|^p]^{1/p}}. \tag{4.23}$$

Observe that $\mathbb{E}(X^*)^p = \mathbb{E}(Y^*)^q = 1$.

We now show $\frac{1}{p}x^p + \frac{1}{q}y^q \geq xy$ for all $x, y \geq 0$. To see this, let $h_y(x) = \frac{1}{p}x^p + \frac{1}{q}y^q - xy$. Then, $h_y'(x) = x^{p-1} - y$ and $h_y''(x) = (p-1)x^{p-2} \geq 0$, so the minimizer is $x^* = y^{1/(p-1)}$. Plugging this in, $h_y(x^*) = \frac{1}{p}y^{p/(p-1)} + \frac{1}{q}y^q - y^{1/(p-1)}y = y^q\left(\frac{1}{p} + \frac{1}{q}\right) - y^q = 0$.

Thus, $X^*Y^* \leq \frac{1}{p}(X^*)^p + \frac{1}{q}(Y^*)^q$, so

$$\frac{\mathbb{E}\,|XY|}{[\mathbb{E}\,|X|^p]^{1/p}[\mathbb{E}\,|Y|^p]^{1/p}} \leq \frac{1}{p} + \frac{1}{q} = 1. \tag{4.24}$$

□

Corollary 4.2.6 (Cauchy-Schwarz Inequality).

$$\mathbb{E}\,|XY| \leq \sqrt{\mathbb{E}X^2\mathbb{E}Y^2}. \tag{4.25}$$

Proposition 4.2.7 (Minkowski's Inequality). If $p > 1$,

$$[\mathbb{E}\,|X+Y|^p]^{1/p} \le [\mathbb{E}\,|X|^p]^{1/p} + [\mathbb{E}\,|Y|^p]^{1/p}. \tag{4.26}$$

Proof. Let $q = \frac{p}{p-1}$ so that $1/p + 1/q = 1$. Then, by Holder's,

$$\mathbb{E}\left[|X|\,|X+Y|^{p-1}\right] \le [\mathbb{E}\,|X|^p]^{1/p}\left[\mathbb{E}\,|X+Y|^{q(p-1)}\right]^{1/q} = [\mathbb{E}\,|X|^p]^{1/p}\,[\mathbb{E}\,|X+Y|^p]^{(p-1)/p}. \tag{4.27}$$

Similarly,

$$\mathbb{E}\left[|Y|\,|X+Y|^{p-1}\right] \le [\mathbb{E}\,|Y|^p]^{1/p}\,[\mathbb{E}\,|X+Y|^p]^{(p-1)/p}. \tag{4.28}$$

Thus,

$$\mathbb{E}\,|X+Y|^p \le \mathbb{E}\left[(|X|+|Y|)\,|X+Y|^{p-1}\right] \le \left([\mathbb{E}\,|X|^p]^{1/p} + [\mathbb{E}\,|Y|^p]^{1/p}\right)[\mathbb{E}\,|X+Y|^p]^{(p-1)/p}. \tag{4.29}$$

Rearrange to get the result. □

4.3 Limit Theorems

Theorem 4.3.1 (Bounded Convergence Theorem). Suppose that $|X_n| \le M$ a.s. and $X_n \xrightarrow{P} X$. Then, $\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_n$.

Proof. Fix $\varepsilon > 0$ and define $G_n = \{|X_n - X| > \varepsilon\}$. Then,

$$\begin{aligned} |\mathbb{E}X_n - \mathbb{E}X| &= |\mathbb{E}(X_n - X)| \\ &\le \mathbb{E}\,|X_n - X| \\ &= \mathbb{E}\left[|X_n - X|\,\mathbb{I}_{G_n}\right] + \mathbb{E}\left[|X_n - X|\,\mathbb{I}_{G_n^c}\right] \\ &\le 2M\mathbb{P}(G_n) + \varepsilon[1 - \mathbb{P}(G_n)] \\ &= \varepsilon + \mathbb{P}(G_n)[2M - \varepsilon]. \end{aligned} \tag{4.30}$$

By convergence in probability and arbitrary ε, $\lim_{n\to\infty} |\mathbb{E}X_n - \mathbb{E}X| = 0$. Real analysis fact that this implies the result. □

Theorem 4.3.2 (Fatou's Lemma). If $X_n \ge 0$ a.s. for all n, $\liminf_{n\longrightarrow\infty} \mathbb{E}X_n \ge \mathbb{E}(\liminf_{n\longrightarrow\infty} X_n)$.

Proof. For each n, define $Y_n = \inf_{m\ge n} X_m$. Clearly, $X_n \ge Y_n$ a.s. and $Y_n \uparrow \liminf_{n\longrightarrow\infty} X_n = Y$ a.s. Thus, $\liminf_{n\longrightarrow\infty} \mathbb{E}X_n \ge \liminf_{n\longrightarrow\infty} \mathbb{E}Y_n$. Fix an arbitrary M, and observe $(Y_n \wedge M) \longrightarrow (Y \wedge M)$ a.s., so by BCT we have $\liminf_{n\longrightarrow\infty} \mathbb{E}Y_n \ge \lim_{n\to\infty} \mathbb{E}(Y_n \wedge M) = \mathbb{E}(Y \wedge M)$. Taking the limit as $M \longrightarrow \infty$ and applying Lemma 3.3.2 gives the result. □

Theorem 4.3.3 (Monotone Convergence Theorem). If $X_n \ge 0$ a.s. and $X_n \uparrow X$ a.s., $\mathbb{E}X_n \uparrow \mathbb{E}X$.

Proof. Since $\mathbb{E}X_n \le \mathbb{E}X$, $\lim_{n\to\infty} \mathbb{E}X_n \le \mathbb{E}X$. But, by Fatou's,

$$\lim_{n\to\infty} \mathbb{E}X_n = \liminf_{n\longrightarrow\infty} \mathbb{E}X_n \ge \mathbb{E}(\liminf_{n\longrightarrow\infty} X_n) = \mathbb{E}X. \tag{4.31}$$

□

Theorem 4.3.4 (Dominated Convergence Theorem). If $X_n \longrightarrow X$ a.s. and $|X_n| \le Y$ a.s. for some integrable Y, $\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_n$.

Proof. Since $|X_n| \le Y$ a.s. for all n, $|X| \le Y$ a.s. and thus X is integrable. Since $X_n + Y \ge 0$ a.s. for all n, by Fatou's $\liminf_{n\longrightarrow\infty} \mathbb{E}(X_n + Y) \ge \mathbb{E}(\liminf_{n\longrightarrow\infty} X_n + Y) = \mathbb{E}(X + Y)$. Thus, $\liminf_{n\longrightarrow\infty} \mathbb{E}X_n \ge \mathbb{E}X$. It also holds that $Y - X_n \ge 0$, so $\liminf_{n\longrightarrow\infty} \mathbb{E}(Y - X_n) \ge \mathbb{E}(\liminf_{n\longrightarrow\infty} Y - X_n) = \mathbb{E}(Y - X)$. Rearranging (and using that the expectations are all finite) gives $\mathbb{E}X \ge \limsup_{n\longrightarrow\infty} \mathbb{E}X_n$, so the result holds. □

4.4 Computing Expected Value

Theorem 4.4.1. Suppose X is a random variable with density f. Then, for any function g,

$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(x)f(x)dx. \tag{4.32}$$

Proof. First, suppose $g(X)$ is a simple random variable taking values $g_1, \ldots, g_n$. Then, we can define $A_i = \{x \in \mathbb{R} : g(x) = g_i\}$ and $E_i = \{\omega \in \Omega : X(\omega) \in A_i\}$. Observe that the A_i partition the set $\mathrm{Img}(X) = \{x \in \mathbb{R} : \exists \omega \in \Omega \text{ s.t. } X(\omega) = x\}$ and consequently the E_i partition Ω. Then, $\mathbb{E}g(X) = \sum_{i=1}^n g_i \mathbb{P}(E_i)$, and since clearly $\mathrm{Img}(X) \supseteq \mathrm{Support}(X) = \{x \in \mathbb{R} : f(x) > 0\}$,

$$\begin{aligned}
\int_{\mathbb{R}} g(x)f(x)dx &= \int_{\mathrm{Img}(X)} g(x)f(x)dx \\
&= \int_{\mathrm{Img}(X)} \sum_{i=1}^n g_i \mathbb{I}\{x \in A_i\} f(x)dx \\
&= \sum_{i=1}^n g_i \int_{\mathrm{Img}(X)} \mathbb{I}\{x \in A_i\} f(x)dx \\
&= \sum_{i=1}^n g_i \mathbb{P}(X \in A_i) \\
&= \sum_{i=1}^n g_i \mathbb{P}(E_i).
\end{aligned} \tag{4.33}$$

Next, suppose $g(X) \ge 0$ a.s. For each $n \in \mathbb{N}$, define $g_n = \frac{\lfloor 2^n g(X) \rfloor}{2^n} \wedge n$. Observe that $0 \le g_n(X) \le g(X)$, $g_n \uparrow g$, and it takes finitely many values so it is a simple function. Thus, $\mathbb{E}g_n(X) = \int g_n(x)f(x)dx$ by the first part of the proof. By MCT, $\lim_{n\to\infty} \mathbb{E}g_n(X) = \mathbb{E}g(X)$ and since $g_n \ge 0$, $\lim_{n\to\infty} \int g_n(x)f(x)dx = \int g(x)f(x)dx$.

Finally, suppose $\mathbb{E}\,|g(X)| < \infty$. Then apply previous paragraph to $\mathbb{E}g^+(X)$ and $\mathbb{E}g^-(X)$.

□

4.5 Exercises

Exercise 4.1. Using a Taylor expansion, show that for a *Rademacher* random variable S (e.g., taking values 1 and -1 with probability $1/2$ each)

$$\mathbb{E}e^{\lambda S} \leq e^{\lambda^2/2} \ \forall \lambda \in \mathbb{R}. \tag{4.34}$$

Then, letting $Z = \sum_{i=1}^n S_i$ for i.i.d. Rademachers S_i, show that

$$\mathbb{P}(Z \geq t) \leq e^{-t^2/(2n)}. \tag{4.35}$$

Exercise 4.2 (Rosenthal 3.6.13). Suppose $\mathbb{E}X_n = 0$ and $\mathbb{E}(X_n^2) = 1$ for all n. Prove that $\mathbb{P}(X_n \geq n \text{ i.o.}) = 0$.

Exercise 4.3. Find a random variable X and $a > 0$ such that $\mathbb{P}(X > a) \geq \mathbb{E}X/a$, and identify what breaks in the proof of Markov's inequality for this example.

Exercise 4.4. For any X, Y, use Cauchy-Schwarz to show that $|\mathrm{Corr}(X, Y)| \leq 1$.

Exercise 4.5. Prove that $f(x) = \max\{x, a\}$ and $f(x) = (x - a)^2$ are convex for any $a \in \mathbb{R}$.

Exercise 4.6 (Rosenthal 5.5.9). Prove that if X is such that $\mathbb{E}X = m < \infty$ and $\mathrm{Var}(X) = v < \infty$, for all $a > 0$

$$\mathbb{P}(X - m \geq a) \leq \frac{v}{v + a^2}. \tag{4.36}$$

Exercise 4.7 (Rosenthal 5.5.10). Let $X_1, X_2, \dots$ satisfy $\mathbb{E}X_n = m < \infty$ and $\mathrm{Var}(X_n) = 1/\sqrt{n}$. Prove that $X_n \xrightarrow{P} m$.

Exercise 4.8 (Rosenthal 5.5.11). Give an example of $X_1, X_2, \dots$ such that $X_n/n \xrightarrow{P} 0$ and $X_n/n^2 \longrightarrow 0$ a.s., but $\mathbb{P}(X_n/n \longrightarrow 0) < 1$.

Exercise 4.9. Prove that if $X_n \longrightarrow X$ a.s., for all $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| \geq \varepsilon \text{ i.o.}) = 0. \tag{4.37}$$

Exercise 4.10. Show that if X only takes values in $\mathbb{N}$, $\mathbb{E}X = \sum_{k=1}^\infty \mathbb{P}(X \geq k)$.

Exercise 4.11. Show that if $X \geq 0$ and $p > 0$, $\mathbb{E}X^p = \int_0^\infty px^{p-1}\mathbb{P}(X \geq x)dx$.

Exercise 4.12 (Convolution Formula). Show that if X and Y are independent with densities f_X and f_Y, for all $z \in \mathbb{R}$,

$$\mathbb{P}(X + Y \leq z) = \int F_X(z - y) f_Y(y) dy. \tag{4.38}$$

LECTURE 5

LAWS OF LARGE NUMBERS

5.1 Familiar Results

Theorem 5.1.1 (Baby Weak LLN). Let $X_1, X_2, \dots$ be independent, and for all i, suppose $\mathbb{E}X_i = \mu$ and $\mathrm{Var}(X_i) \le \sigma^2 < \infty$. Define $S_n = \sum_{i=1}^n X_i$. Then,

$$\frac{1}{n} S_n \xrightarrow{P} \mu. \tag{5.1}$$

Proof. By linearity, $\mathbb{E}S_n/n = \mu$ and $\mathrm{Var}(S_n/n) \le \sigma^2/n$. Then, for any $\varepsilon > 0$, Chebyshev's gives that

$$\lim_{n\to\infty} \mathbb{P}(|S_n/n - \mu| > \varepsilon) \le \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon} = 0. \tag{5.2}$$

□

Theorem 5.1.2 (Baby Strong LLN). Let $X_1, X_2, \dots$ be independent, and for all i, suppose $\mathbb{E}X_i = \mu$ and $\mathbb{E}(X_i - \mu)^4 \le a < \infty$. Then,

$$\frac{1}{n} S_n \longrightarrow \mu \text{ a.s.} \tag{5.3}$$

Proof. First, observe that $\mathbb{E}(X_i - \mu)^2 \le \mathbb{E}(X_i - \mu)^4 + 1$, by considering the case when the variance is smaller and greater than 1. Without loss of generality, we can suppose $\mu = 0$. Then, we have that

$$\begin{aligned}
\mathbb{E}S_n^4 &= \mathbb{E}\left(\sum_{i=1}^n X_i\right)^4 \\
&= \mathbb{E}\Bigg(\sum_{i=1}^n X_i^4 + k_1 \sum_{i=1}^n \sum_{j\ne i} X_i^3 X_j + k_2 \sum_{i=1}^n \sum_{j\ne i} X_i^2 X_j^2 + k_3 \sum_{i=1}^n \sum_{j\ne i} \sum_{k\ne j,i} X_i^2 X_j X_k \\
&\qquad + k_4 \sum_{i=1}^n \sum_{j\ne i} \sum_{k\ne j,i} \sum_{\ell\ne j,i,k} X_i X_j X_k X_\ell\Bigg) \\
&= \mathbb{E}\sum_{i=1}^n X_i^4 + k_2 \mathbb{E}\sum_{i=1}^n \sum_{j\ne i} X_i^2 X_j^2 \\
&\le na + k_2 n(n-1)(a+1)^2 \\
&\le Kn^2.
\end{aligned}$$

Next, for any $\varepsilon > 0$, we can apply Markov's to get

$$\mathbb{P}\left(\left|\frac{1}{n} S_n\right| > \varepsilon\right) = \mathbb{P}(S_n^4 > n^4\varepsilon^4) \le \frac{\mathbb{E}S_n^4}{n^4\varepsilon^4} \le \frac{K}{n^2\varepsilon^4}. \tag{5.4}$$

Since $\sum_{n=1}^\infty \frac{K}{n^2\varepsilon^4} < \infty$, by Borel-Cantelli the result holds. □

5.2 Advanced Weak LLN

Definition 5.2.1. A sequence of random variables $(X_i)_{i\in\mathcal{I}}$ with $\mathbb{E}X_i^2 < \infty$ are uncorrelated if for all $i \neq j$, $\mathbb{E}X_iX_j = \mathbb{E}X_i\mathbb{E}X_j$.

Lemma 5.2.2. If $X_1, \ldots, X_n$ are uncorrelated, $\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i)$.

Proof. Exercise. □

Lemma 5.2.3. If $p > 0$ and $\mathbb{E}\,|X_n|^p \longrightarrow 0$, $X_n \xrightarrow{P} 0$.

Proof. Markov's. □

Theorem 5.2.4 (Uncorrelated Weak LLN). If $X_1, X_2, \ldots$ are uncorrelated with $\mathbb{E}X_i = \mu$ and $\mathrm{Var}(X_i) \leq \sigma^2 < \infty$,

$$\mathbb{E}\left(\frac{1}{n}S_n - \mu\right)^2 \to 0. \tag{5.5}$$

Proof.

$$\mathbb{E}\left(\frac{1}{n}S_n - \mu\right)^2 = \mathrm{Var}\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) \leq \frac{n\sigma^2}{n^2} \to 0. \tag{5.6}$$

□

Theorem 5.2.5 (Weak LLN). If $X_1, X_2, \ldots$ are i.i.d. with $\lim_{x\to\infty} x\mathbb{P}(|X_1| > x) = 0$, for $\mu_n = \mathbb{E}\left(X_1\mathbb{I}\left\{|X_1| \leq n\right\}\right)$,

$$\frac{1}{n}S_n - \mu_n \xrightarrow{P} 0. \tag{5.7}$$

Proof. Fix $\varepsilon > 0$. Let $\bar{X}_k^{(n)} = X_k\mathbb{I}\{|X_k| \leq n\}$ and $\bar{S}_n = \sum_{k=1}^n \bar{X}_k^{(n)}$. Then,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu_n\right| > \varepsilon\right) \leq \mathbb{P}(S_n \neq \bar{S}_n) + \mathbb{P}\left(\left|\frac{\bar{S}_n}{n} - \mu_n\right| > \frac{\varepsilon}{2}\right). \tag{5.8}$$

For the first term,

$$\begin{aligned}
\mathbb{P}(S_n \neq \bar{S}_n) &\leq \mathbb{P}\left(\bigcup_{k=1}^n \{\bar{X}_k^{(n)} \neq X_k\}\right) \\
&\leq \sum_{k=1}^n \mathbb{P}(\bar{X}_k^{(n)} \neq X_k) \\
&= \sum_{k=1}^n \mathbb{P}(|X_k| > n) \\
&= n\mathbb{P}(|X_1| > n) \\
&\longrightarrow 0.
\end{aligned} \tag{5.9}$$

For the second term, first observe that

$$\mathbb{E}\bar{S}_n = \mathbb{E}\sum_{k=1}^n \bar{X}_k^{(n)} = \sum_{k=1}^n \mathbb{E}[X_k\mathbb{I}\{|X_k| \leq n\}] = n\mu_n. \tag{5.10}$$

So, by Chebyshev's,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{\bar{S}_n}{n} - \mu_n\right| > \frac{\varepsilon}{2}\right) &\leq \frac{4}{n^2\varepsilon^2}\mathbb{E}\left(\bar{S}_n - n\mu_n\right)^2 \\ &= \frac{4}{n^2\varepsilon^2}\operatorname{Var}\left(\bar{S}_n\right) \\ &= \frac{4}{n^2\varepsilon^2}\sum_{k=1}^{n}\operatorname{Var}\left(\bar{X}_k^{(n)}\right) \\ &= \frac{4}{n\varepsilon^2}\operatorname{Var}\left(\bar{X}_1^{(n)}\right) \\ &\leq \frac{4}{n\varepsilon^2}\mathbb{E}(X_1\mathbb{I}\{|X_1| \leq n\})^2.\end{aligned} \tag{5.11}$$

Finally, recalling that $\mathbb{E}X^p = \int_0^\infty px^{p-1}\mathbb{P}(X \geq x)dx$,

$$\begin{aligned}\mathbb{E}(X_1\mathbb{I}\{|X_1| \leq n\})^2 &= \int_0^\infty 2x\mathbb{P}\left(\left|\bar{X}_k^{(n)}\right| \geq x\right)dx \\ &= \int_0^n 2x\mathbb{P}(|X_k| \geq x)dx \\ &= 2\int_0^n x\mathbb{P}(|X_1| \geq x)dx.\end{aligned} \tag{5.12}$$

Since $0 \leq x\mathbb{P}(|X_1| \geq x) \leq x$ for all x and goes to 0, $\sup_x x\mathbb{P}(|X_1| \geq x) < \infty$. Thus,

$$\lim_{n\to\infty}\frac{1}{n}\int_0^n x\mathbb{P}(|X_1| \geq x)dx = \lim_{n\to\infty}\int_0^1 ny\mathbb{P}(|X_1| > ny)dy = \int_0^1 \lim_{n\to\infty} ny\mathbb{P}(|X_1| > ny)dy = 0. \tag{5.13}$$

□

Corollary 5.2.6. If $X_1, X_2, \dots$ are i.i.d. with $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 = \mu < \infty$,

$$\frac{1}{n}S_n \xrightarrow{P} \mu. \tag{5.14}$$

Proof. Let $Y_n = |X_1|\,\mathbb{I}\{|X_1| > n\}$. For each $\omega \in \Omega$, $|X_1(\omega)| < \infty$, so $Y_n \longrightarrow 0$ a.s. Since $|Y_n| \leq |X_1|$ which is integrable, by the DCT we have $\mathbb{E}Y_n \longrightarrow 0$. Further, observe that $Y_n \geq n\mathbb{I}\{|X_1| > n\}$, so $\mathbb{E}Y_n \geq n\mathbb{P}(|X_1| > n)$. Thus, $\lim_{x\to\infty} x\mathbb{P}(|X_1| > x) = 0$.

Next, consider $Z_n = X_1\mathbb{I}\{|X_1| \leq n\}$. Again, since $|X_1(\omega)| < \infty$, $Z_n \longrightarrow X_1$ a.s. Thus, we can again apply DCT to obtain $\mu_n = \mathbb{E}Z_n \longrightarrow \mathbb{E}X_1 = \mu$. The result then follows by the Weak LLN. □

5.3 Advanced Strong LLN

Theorem 5.3.1 (Strong LLN). If $X_1, X_2, \dots$ are i.i.d. with $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 = \mu < \infty$,

$$\frac{1}{n}S_n \longrightarrow \mu \text{ a.s.} \tag{5.15}$$

Proof. Not required to be able to prove for this class. The main techniques are well summarized in the proof of the Weak LLN.

□

5.4 Applications

Proposition 5.4.1 (Polynomial Approximation). If $f : [0,1] \to \mathbb{R}$ is continuous, the Bernstein approximation

$$f_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} f(m/n) \tag{5.16}$$

satisfies

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| \longrightarrow 0. \tag{5.17}$$

Proof.

$$\begin{aligned}
&|f_n(x) - f(x)| \\
&= |\mathbb{E}f(S_n/n) - f(x)| \\
&\le \mathbb{E}\,|f(S_n/n) - f(x)| \\
&= \mathbb{E}\left[|f(S_n/n) - f(x)|\,\mathbb{I}\{|S_n/n - x| < \delta\}\right] + \mathbb{E}\left[|f(S_n/n) - f(x)|\,\mathbb{I}\{|S_n/n - x| \ge \delta\}\right] \\
&\le \varepsilon + 2M\mathbb{P}(|S_n/n - x| \ge \delta),
\end{aligned} \tag{5.18}$$

where $M = \sup_{x\in[0,1]} |f(x)| < \infty$ since f is continuous on a closed interval. Now, by Chebyshev's inequality,

$$\mathbb{P}(|S_n/n - x| \ge \delta) \le \frac{\mathbb{E}[(S_n/n - x)^2]}{\delta^2} = \frac{\mathrm{Var}(S_n/n)}{\delta^2} = \frac{\sum_{i=1}^n \mathrm{Var}(X_i)}{n^2\delta^2} = \frac{x(1-x)}{n\delta^2} \le \frac{1}{4n\delta^2}.$$

Since this was true for all $x \in [0,1]$, $\lim_{n\to\infty} \sup_{x\in[0,1]} |f_n(x) - f(x)| \le \varepsilon$, but ε was arbitrary so the proposition follows. □

Theorem 5.4.2 (Glivenko-Cantelli). Let $X_1, X_2, \ldots$ be i.i.d. with CDF F and define the empirical distribution function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_i \le x\}.$$

Then,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \longrightarrow 0 \qquad \text{a.s.} \tag{5.19}$$

Proof. Fix $x \in \mathbb{R}$ and define $Y_i = \mathbb{I}\{X_i \le x\}$. Since Y_i are i.i.d. and $\mathbb{E}Y_i = F(x) \le 1$, the Strong LLN says that $F_n(x) = \frac{1}{n}\sum_{i=1}^n Y_i \longrightarrow F(x)$ a.s. That is, pointwise convergence occurs almost surely. One might hope to automatically upgrade pointwise convergence of $F_n(x)$ to uniform convergence since F is bounded, but F may not be continuous, so we must show this upgrade can be done.

Still for a fixed x, let $Z_i = \mathbb{I}\{X_i < x\}$. For any function g, let $g(x-) = \lim_{y\uparrow x} g(y)$. Again, Z_i are i.i.d. with $\mathbb{E}Z_i = \mathbb{P}(X_i < x) = F(x-)$, so by the Strong LLN we have $\lim_{n\to\infty} F_n(x-) = F(x-)$ a.s. That is, we also have pointwise convergence of $F_n(x-)$ almost surely.

Now, fix an arbitrary $k \in \mathbb{N}$, and for $1 \le j \le k-1$, define $x_{j,k} = \inf\{x : F(x) \ge j/k\}$. Also, define $x_{0,k} = -\infty$ and $x_{k,k} = \infty$. By the pointwise convergence shown above, for each j and k, there exists $N_{j,k}$ such that if $n > N_{j,k}$ then

$$|F_n(x_{j,k}) - F(x_{j,k})| < \frac{1}{k} \text{ and } |F_n(x_{j,k}-) - F(x_{j,k}-)| < \frac{1}{k}. \tag{5.20}$$

Let $N_k = \sup_{0 \le j \le k} N_{j,k}$. Then, for any $x \in \mathbb{R}$, there exists $1 \le j \le k$ with $x \in (x_{j-1,k}, x_{j,k})$. If $n > N_k$, then by monotonicity of F_n and F, and $F(x_{j,k}-) - F(x_{j-1,k}) \le \frac{1}{k}$,

$$F_n(x) \le F_n(x_{j,k}-) \le F(x_{j,k}-) + \frac{1}{k} \le F(x_{j-1,k}) + \frac{2}{k} \le F(x) + \frac{2}{k}. \tag{5.21}$$

Similarly,

$$F_n(x) \ge F_n(x_{j-1,k}) \ge F(x_{j-1,k}) - \frac{1}{k} \ge F(x_{j,k}-) - \frac{2}{k} \ge F(x) - \frac{2}{k}. \tag{5.22}$$

That is, $|F_n(x) - F(x)| \le \frac{2}{k}$ for all $x \in \mathbb{R}$, since N_k does not depend on x. Since k was arbitrary, the result holds. $\square$

5.5 Exercises

Exercise 5.1. Prove Lemma 5.2.2.

Exercise 5.2 (Durrett 2.2.1). Suppose $X_1, X_2, \ldots$ are uncorrelated with $\mathbb{E}X_i = \mu_i$ and $\lim_{i\to\infty} \frac{\mathrm{Var}(X_i)}{i} = 0$. If $\nu_n = \frac{1}{n}\mathbb{E}S_n$, show that

$$\lim_{n\to\infty} \mathbb{E}\Big[\Big(\frac{S_n}{n} - \nu_n\Big)^2\Big] = 0.$$

Exercise 5.3 (Durrett 2.2.4). Let $X_1, X_2, \ldots$ be i.i.d. such that $\mathbb{P}(X_1 = (-1)^k k) = \frac{C}{k^2 \log(k)}$ for all integers $k \geq 2$, where C is a constant so that the probabilities sum to 1. Show that $\mathbb{E}\,|X_1| = \infty$ but there is a finite μ such that $\frac{S_n}{n} \xrightarrow{P} \mu$.

Hint #1: $\mu_n = \mathbb{E}[X_1 \mathbb{I}\{|X_1 \leq n|\}] = \sum_{k=2}^n (-1)^k k \frac{C}{k^2 \log(k)}$ *is an alternating sequence of real numbers that congerge to zero, so from calculus class* $\mu_n \longrightarrow \mu$ *for some real number* μ. *Thus, it suffices to show* $\frac{S_n}{n} - \mu_n \xrightarrow{P} 0$.
Hint #2: For any positive and decreasing function f, $\sum_{k=x+1}^\infty f(k) \leq \int_x^\infty f(y)dy \leq \sum_{k=x}^\infty f(k)$.

Exercise 5.4 (Durrett 2.2.5). Let $X_1, X_2, \ldots$ be i.i.d. such that $\mathbb{P}(X_1 > x) = \frac{e}{x \log(x)}$ for $x \geq e$. Show that $\mathbb{E}\,|X_1| = \infty$, but $\frac{S_n}{n} - \mu_n \xrightarrow{P} 0$.
Hint: Recall Exercise 4.11.

Exercise 5.5. For any sequence of random variables X_n and $\varepsilon > 0$, show there exist constants $c_n \to \infty$ such that $\mathbb{P}(|X_n| > \varepsilon c_n) < 2^{-n}$.

Exercise 5.6 (Durrett 2.3.10). For any sequence of random variables X_n, show there exist constants $c_n \to \infty$ such that

$$\frac{X_n}{c_n} \longrightarrow 0 \text{ a.s.}$$

Exercise 5.7 (Durrett 2.3.15). Suppose $X_1, X_2, \ldots$ are i.i.d. Show that $\mathbb{E}\,|X_1| < \infty$ if and only if

$$\frac{X_n}{n} \longrightarrow 0 \text{ a.s.}$$

Exercise 5.8 (Durrett 2.3.18). Suppose $X_1, X_2, \ldots$ are i.i.d. such that for all n, $\mathbb{P}(X_n > x) = e^{-x}$. Show that

$$\limsup_{n\to\infty} \frac{X_n}{\log(n)} = 1 \text{ a.s.}$$

LECTURE 6

CENTRAL LIMIT THEOREMS

6.1 Weak Convergence

Definition 6.1.1. A sequence of probability measures μ_n with corresponding distribution functions F_n *converge weakly* to a probability measure μ with distribution function F if for all continuity points x of F

$$\lim_{n\to\infty} F_n(x) = F(x), \tag{6.1}$$

and is denoted by $\mu_n \Longrightarrow \mu$, or $F_n \Longrightarrow F$, or $X_n \Longrightarrow X$ if $X_n \sim F_n$ and $X \sim F$.

Theorem 6.1.2 (Scheffé's Theorem). Let X_n each have density f_n and X have density f. Then, if $f_n(x) \longrightarrow f(x)$ a.e., $X_n \Longrightarrow X$.

Proof. Observe that the existence of a density implies F is continuous everywhere.

$$\begin{aligned} |F_n(y) - F(y)| &= \left|\int_{-\infty}^{y} [f_n(x) - f(x)]\, dx\right| \\ &\le \int_{-\infty}^{y} |f_n(x) - f(x)|\, dx \\ &\le \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx \\ &= \int_{\mathbb{R}} [f_n(x) - f(x)]^+\, dx + \int_{\mathbb{R}} [f_n(x) - f(x)]^-\, dx \\ &= 2\int_{\mathbb{R}} [f_n(x) - f(x)]^-\, dx \\ &= 2\int_{\mathbb{R}} [f(x) - f_n(x)]^+\, dx, \end{aligned} \tag{6.2}$$

using that $x^+ + x^- = |x|$, $x^+ - x^- = x$, and $(-x)^+ = x^-$. Then, observe that $[f(x) - f_n(x)]^+ \le [f(x)]^+ = f(x)$, so

$$\lim_{n\to\infty} |F_n(y) - F(y)| \le 2\int \lim_{n\to\infty} [f(x) - f_n(x)]^+\, dx = 0. \tag{6.3}$$

□

Theorem 6.1.3 (Skorokhod Representation on $\mathbb{R}$). If $F_n \Longrightarrow F$, there exist an Ω and $\mathbb{P}$ along with random variables $X, X_1, X_2, \ldots$ defined on Ω such that $X_n \sim F_n$, $X \sim F$ and $\mathbb{P}(X_n \longrightarrow X) = 1$.

Proof. Consider $x \in [0,1]$. First, suppose there is at most one $a \in \mathbb{R}$ such that $F(a) = x$. Then, by Exercise 6.6, $F_n^{-1}(x) \longrightarrow F^{-1}(x)$. Otherwise, suppose there are $a < b$ such that $F(a) = F(b) = x$. Then, there is a $q \in \mathbb{Q}$ such that $a < q < b$ and consequently $F(q) = x$, so there are only countably many x that can satisfy this property. Let $\Omega = [0,1]$ and $\mathbb{P}$ be the uniform distribution on $[0,1]$. Since $\mathbb{P}(A) = 0$ for any countable A, $\mathbb{P}(F_n^{-1}(\omega) \longrightarrow F^{-1}(\omega)) = 1$. So, taking $X_n(\omega) = F_n^{-1}(\omega)$ and $X(\omega) = F^{-1}(\omega)$ suffices. □

Theorem 6.1.4 (Portmanteau Lemma). $F_n \Longrightarrow F$ if and only if for all bounded and continuous $h : \mathbb{R} \to \mathbb{R}$, when $X_n \sim F_n$ and $X \sim F$ then

$$\mathbb{E}h(X_n) \longrightarrow \mathbb{E}h(X).$$

Proof. For the forward direction, choose $X_n \sim F_n$ and $X \sim F$ with $\mathbb{P}(X_n \longrightarrow X) = 1$ from the Skorokhod representation. Then, by the bounded convergence theorem, $\mathbb{E}h(X_n) \longrightarrow \mathbb{E}h(X)$.

For the reverse direction, fix $a \in \mathbb{R}$ and define $g_a(x) = \mathbb{I}\{x \le a\}$ and suppose $X_n \sim F_n$ and $X \sim F$. To approximate g_a with a continuous function, define

$$g_{a,\varepsilon}(x) = \begin{cases} 1 & x \le a \\ 1 - (x-a)/\varepsilon & a \le x \le a + \varepsilon \\ 0 & x \ge a + \varepsilon \end{cases}. \tag{6.4}$$

Observe that $F_n(a) = \mathbb{P}(X_n \le a) = \mathbb{E}g_a(X_n)$ and for all $\varepsilon > 0$, $g_{a-\varepsilon} \le g_{a-\varepsilon,\varepsilon} \le g_a \le g_{a,\varepsilon} \le g_{a+\varepsilon}$. Also, $g_{a,\varepsilon}$ is bounded and continuous, so $\lim_{n\to\infty} \mathbb{E}g_{a,\varepsilon}(X_n) = \mathbb{E}g_{a,\varepsilon}(X)$. Thus,

$$\limsup_{n\to\infty} F_n(a) \le \lim_{n\to\infty} \mathbb{E}g_{a,\varepsilon}(X_n) = \mathbb{E}g_{a,\varepsilon}(X) \le F(a+\varepsilon). \tag{6.5}$$

Thus, for all a, $\limsup_{n\to\infty} F_n(a) \le \lim_{\varepsilon\to 0} F(a+\varepsilon) = F(a)$. Similarly,

$$\liminf_{n\to\infty} F_n(a) \ge \lim_{n\to\infty} \mathbb{E}g_{a-\varepsilon,\varepsilon}(X_n) = \mathbb{E}g_{a-\varepsilon,\varepsilon}(X) \ge F(a-\varepsilon). \tag{6.6}$$

If a is a continuity point, then $\liminf_{n\to\infty} F_n(a) \ge \lim_{\varepsilon\to 0} F(a-\varepsilon) = \mathbb{P}(X < a) = \mathbb{P}(X \le a) = F(a)$.

□

Corollary 6.1.5. If $X_n \Longrightarrow X$, then $f(X_n) \Longrightarrow f(X)$ for any continuous f.

Proof. For any continuous and bounded g, $g \circ f$ is also continuous and bounded. □

Corollary 6.1.6. If $X_n \xrightarrow{P} X$ then $X_n \Longrightarrow X$.

Proof. Consider any subsequence X_{n_j}, and observe that $X_{n_j} \xrightarrow{P} X$ as well. Also, recall that there exists a further subsequence $X_{n_{j(k)}} \longrightarrow X$ a.s. Thus, for bounded and continuous h, the continuous mapping theorem and bounded convergence theorem give that $\mathbb{E}h(X_{n_{j(k)}}) \longrightarrow \mathbb{E}h(X)$. Since $\mathbb{E}h(X_n)$ is a sequence of real numbers, this implies $\mathbb{E}h(X_n) \longrightarrow \mathbb{E}h(X)$. □

Example 6.1.7. The reverse does not hold. Let $X \sim \text{Gaussian}(0,1)$ and $X_n = -X$ for $n \in \mathbb{N}$. Then, trivially $F_n = F$, so $X_n \Longrightarrow X$, but

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X| > \varepsilon/2) = c > 0. \tag{6.7}$$

Lemma 6.1.8. If $X_n \implies c$ where c is constant, $X_n \xrightarrow{P} c$.

Proof. First, let F be the CDF of the random variable $X \equiv c$, so that $F(y) = \mathbb{I}\{c \leq y\}$. Observe that F is continuous everywhere except $y = c$. Now, fix $\varepsilon > 0$.

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(|X_n - c| > \varepsilon) &= \lim_{n\to\infty} [\mathbb{P}(X_n < c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon)] \\
&\leq \lim_{n\to\infty} [\mathbb{P}(X_n \leq c - \varepsilon/2) + \mathbb{P}(X_n > c + \varepsilon)] \\
&= \lim_{n\to\infty} F_n(c - \varepsilon/2) + 1 - \lim_{n\to\infty} F_n(c + \varepsilon) \\
&= F(c - \varepsilon/2) + 1 - F(c + \varepsilon) \\
&= 0.
\end{aligned} \tag{6.8}$$

□

Theorem 6.1.9 (Slutsky's Theorem). If $X_n \implies X$ and $Y_n \implies c$ for a constant c,

- $X_n + Y_n \implies X + c$,
- $X_n Y_n \implies Xc$,
- $X_n/Y_n \implies X/c$ if $c \neq 0$.

Proof.

- Fix $\varepsilon > 0$ and z a continuity point of the CDF of $X + c$. Observe that $X_n \leq z - c - \varepsilon$ and $Y_n \leq c + \varepsilon$ implies $X_n + Y_n \leq z$. Thus,

$$\begin{aligned}
&\mathbb{P}(X_n + Y_n \leq z) \\
&\geq \mathbb{P}(X_n \leq z - c - \varepsilon \cap Y_n \leq c + \varepsilon) \\
&= \mathbb{P}(X_n \leq z - c - \varepsilon) + \mathbb{P}(Y_n \leq c + \varepsilon) - \mathbb{P}(X_n \leq z - c - \varepsilon \cup Y_n \leq c + \varepsilon) \\
&\geq \mathbb{P}(X_n \leq z - c - \varepsilon) + \mathbb{P}(Y_n \leq c + \varepsilon) - 1 \\
&= \mathbb{P}(X_n \leq z - c - \varepsilon) - \mathbb{P}(Y_n > c + \varepsilon).
\end{aligned} \tag{6.9}$$

Now, observe that the CDF of Y_n is continuous everywhere except at c. So,

$$\lim_{n\to\infty} \mathbb{P}(Y_n > c + \varepsilon) = \mathbb{P}(Y > c + \varepsilon) = 0. \tag{6.10}$$

For arbitrarily small ε we can take $z - c - \varepsilon$ to be a continuity point of the CDF of X without loss of generality, so

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq z - c - \varepsilon) = \mathbb{P}(X \leq z - c - \varepsilon). \tag{6.11}$$

That is,

$$\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \leq z) \geq \mathbb{P}(X + c \leq z - \varepsilon). \tag{6.12}$$

Since z is a continuity point of the CDF of $X + c$, taking $\varepsilon \to 0$ gives

$$\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \leq z) \geq \mathbb{P}(X + c \leq z). \tag{6.13}$$

Similarly, $X_n + Y_n \leq z$ and $Y_n \geq c - \varepsilon$ implies $X_n \leq z - c + \varepsilon$, so

$$\mathbb{P}(X_n \leq z - c + \varepsilon) \geq \mathbb{P}(X_n + Y_n \leq z) - \mathbb{P}(Y_n < c - \varepsilon). \tag{6.14}$$

Rearranging and taking limits in the same way gives

$$\limsup_{n\to\infty} \mathbb{P}(X_n + Y_n \leq z) \leq \mathbb{P}(X + c \leq z). \tag{6.15}$$

The remaining two are left as exercises. □

Remark. We have been somewhat implicitly working with random variables on $\mathbb{R}$, but of course the definition of weak convergence works just as well for random variables on $\mathbb{R}^d$. In fact, the Skorokhod Representation Theorem and Portmanteau Lemma can be equivalently shown for the multivariate case, with the latter implying a multivariate continuous mapping theorem for weak convergence. Consequently, if $(X_n, Y_n) \Longrightarrow (X, Y)$, then $X_n + Y_n \Longrightarrow X + Y$, $X_n Y_n \Longrightarrow XY$, and $X_n/Y_n \Longrightarrow X/Y$. As you'll see in Exercise 6.8, however, this is not the same as $X_n \Longrightarrow X$ and $Y_n \Longrightarrow Y$.

6.2 Characteristic Functions

Remark. This section requires the use of complex numbers. Recall i is defined as the solution to $x^2 = -1$, and a complex number $x \in \mathbb{C}$ is defined by real numbers $a, b \in \mathbb{R}$ such that $x = a + ib$. The *modulus* of x is $|x| = \sqrt{a^2 + b^2}$ and the *complex conjugate* is $\bar{x} = a - bi$. The *real* and *imaginary* parts of x are defined respectively by $\mathrm{Re}(x) = a$ and $\mathrm{Im}(x) = b$.

Lemma 6.2.1 (Euler's Formula). For any $x \in \mathbb{R}$,

$$e^{ix} = \cos(x) + i\sin(x) \tag{6.16}$$

Proof. See first year calculus book. □

Definition 6.2.2. For a random variable taking X taking values in $\mathbb{C}$, we define its expectation by

$$\mathbb{E}X = \mathbb{E}\mathrm{Re}(X) + i\mathbb{E}\mathrm{Im}(X). \tag{6.17}$$

Definition 6.2.3. A random variable X has a unique *characteristic function* defined by

$$\varphi(t) = \mathbb{E}e^{itX} = \mathbb{E}\cos(tX) + i\mathbb{E}\sin(tX). \tag{6.18}$$

Proposition 6.2.4. A characteristic function φ has the following properties:

i) $\varphi(0) = 1$.
ii) $\varphi(-t) = \overline{\varphi(t)}$.
iii) $|\varphi(t)| \leq 1$.
iv) $|\varphi(t+h) - \varphi(t)| \leq \mathbb{E}\left|e^{-ihX} - 1\right|$.
v) $\mathbb{E}e^{it(aX+b)} = e^{itb}\varphi(at)$.

Proof.

i) By definition.
ii) $\varphi(-t) = \mathbb{E}[\cos(-tX) + i\sin(-tX)] = \mathbb{E}[\cos(tX) - i\sin(tX)]$.
iii) $|\varphi(t)| = \left|\mathbb{E}e^{itX}\right| \leq \mathbb{E}\left|e^{itX}\right| = 1$, using that $\sqrt{x^2+y^2}$ is convex.
iv) $|\varphi(t+h) - \varphi(t)| = \left|\mathbb{E}e^{itX}e^{ihX} - \mathbb{E}e^{itX}\right| \leq \mathbb{E}\left[\left|e^{itX}\right|\left|e^{ihX} - 1\right|\right] = \mathbb{E}\left|e^{-ihX} - 1\right|$.
v) By definition.

□

Proposition 6.2.5. If X and Y are independent with characteristic functions φ_X and φ_Y, then $Z = X + Y$ has characteristic function

$$\varphi_Z(t) = \varphi_X(t)\varphi_Y(t). \tag{6.19}$$

Proof.

$$\varphi_Z(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}] = \mathbb{E}[e^{itX}]\mathbb{E}[e^{itY}] = \varphi_X(t)\varphi_Y(t). \tag{6.20}$$

□

Theorem 6.2.6. Common distributions have the following characteristic functions.

- $X \sim \text{Ber}(p)$: $\varphi(t) = 1 - p + pe^{it}$.
- $X \sim \text{Pois}(\lambda)$: $\varphi(t) = \exp\{\lambda(e^{it} - 1)\}$.
- $X \sim \text{Exp}(\lambda)$: $\varphi(t) = \frac{\lambda}{\lambda - it}$.
- $X \sim \text{Gaussian}(\mu, \sigma^2)$: $\varphi(t) = e^{i\mu t - \sigma^2 t^2/2}$.

Proof. Exercise.
(*Treat i as a constant – we won't worry about the technical details of complex analysis here.*) □

Theorem 6.2.7. Let X be a random variable and consider a function $f : [a, b] \times \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}f(t, X) < \infty$ and $\frac{\partial}{\partial t}f(t, X) = f'(t, X)$ exists for all $t \in (a, b)$. Further, suppose there is a random variable Y with $\mathbb{E}Y < \infty$ and $|f'(t, X)| \le Y$ a.s. for $t \in (a, b)$. Then, for all $t \in (a, b)$,

$$\frac{\partial}{\partial t}\mathbb{E}f(t, X) = \mathbb{E}f'(t, X). \tag{6.21}$$

Proof. First, observe that

$$f'(t, X) = \lim_{h \to 0} \frac{f(t + h, X) - f(t, X)}{h}, \tag{6.22}$$

so

$$\left| \frac{f(t + h, X) - f(t, X)}{h} \right| \le Y \tag{6.23}$$

for small h. Then, using the dominated convergence theorem,

$$\begin{aligned}
\frac{\partial}{\partial t}\mathbb{E}f(t, X) &= \lim_{h \to 0} \frac{\mathbb{E}f(t + h, X) - \mathbb{E}f(t, X)}{h} \\
&= \lim_{h \to 0} \mathbb{E}\frac{f(t + h, X) - f(t, X)}{h} \\
&= \mathbb{E}\lim_{h \to 0} \frac{f(t + h, X) - f(t, X)}{h} \\
&= \mathbb{E}f'(t, X).
\end{aligned} \tag{6.24}$$

□

Proposition 6.2.8. If X is a random variable with $\mathbb{E}|X|^k < \infty$, then for $0 \le j \le k$,

$$\varphi_X^{(j)}(t) = \mathbb{E}[(iX)^j e^{itX}]. \tag{6.25}$$

Proof. This is proved by induction. When $j = 0$, this is just the definition of φ. Suppose it holds for some j. Then,

$$\begin{aligned}\varphi_X^{(j+1)}(t) &= \frac{\partial}{\partial t}\varphi_X^{(j)}(t) \\ &= \frac{\partial}{\partial t}\mathbb{E}[(iX)^j e^{itX}] \\ &= \mathbb{E}\left[(iX)^j \frac{\partial}{\partial t} e^{itX}\right] \\ &= \mathbb{E}[(iX)^{j+1} e^{itX}],\end{aligned} \tag{6.26}$$

where we used the previous proposition to swap the order of expectation and differentiation. □

Theorem 6.2.9 (Continuity Theorem). Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures with characteristic functions $\varphi, \varphi_1, \varphi_2, \ldots$. Then, $\mu_n \implies \mu$ if and only if $\varphi_n(t) \longrightarrow \varphi(t)$ for all $t \in \mathbb{R}$.

Proof. Beyond the scope of this course. □

Lemma 6.2.10. If X_n, Y_n are independent for all n and X, Y are independent with $X_n \implies X$ and $Y_n \implies Y$, then

$$X_n + Y_n \implies X + Y. \tag{6.27}$$

Proof. Using characteristic functions,

$$\lim_{n\to\infty} \varphi_{X_n+Y_n}(t) = \lim_{n\to\infty} \varphi_{X_n}(t)\varphi_{Y_n}(t) = \varphi_X(t)\varphi_Y(t) = \varphi_{X+Y}(t). \tag{6.28}$$

Then apply the continuity theorem. □

6.3 Central Limit Theorem

Lemma 6.3.1. For any random variable with $\mathbb{E}\,|X|^k < \infty$ for $0 \le k \le m$,

$$\varphi_X(t) = \sum_{k=0}^{m} \frac{(it)^k}{k!}\mathbb{E}X^k + o(|t|^m). \tag{6.29}$$

Proof. Exercise. *Hint: use Taylor series error approximation.* □

Theorem 6.3.2. If $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}X_n = 0$ and $\mathbb{E}X_n^2 = \sigma^2 < \infty$,

$$Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \implies \text{Gaussian}(0, \sigma^2). \tag{6.30}$$

Proof. Using the previous lemma with $m = 2$,

$$\varphi_{X_i}(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2). \tag{6.31}$$

Thus,

$$\lim_{n\to\infty} \varphi_{Y_n}(t) = \lim_{n\to\infty} [\varphi_i(t/\sqrt{n})]^n = \lim_{n\to\infty} \left[1 - \frac{1}{2n}\sigma^2 t^2 + o(t^2/n)\right]^n = e^{-\sigma^2 t^2/2}. \tag{6.32}$$

This is the characteristic function of Gaussian$(0, \sigma^2)$, so the result follows from the continuity theorem. □

6.4 Exercises

Exercise 6.1. A Cauchy random variable has density defined by $f(x) = \frac{1}{\pi(1+x^2)}$ for all $x \in \mathbb{R}$. Show that for all $t \neq 0$, the MGF $M_X(t) = \mathbb{E}e^{tX} = \infty$.
Note: this is just a calculus question to motivate why we use characteristic functions. I won't ask you something like this on an exam.

Exercise 6.2 (Durrett 3.3.9). Suppose $X_n \sim \text{Gaussian}(0, \sigma_n^2)$ and $X_n \Longrightarrow X$ for some random variable X. Show that $\sigma_n \longrightarrow \sigma$ for some $\sigma \in [0, \infty)$.

Exercise 6.3. Prove Scheffé's Theorem for discrete pmfs instead of densities.

Exercise 6.4. Find an example of random variables X_n with densities f_n such that $X_n \Longrightarrow \text{Unif}(0,1)$ but $\{x : f_n(x) \longrightarrow 1\} = \emptyset$.

Exercise 6.5. Prove that if $\mathbb{P}_n \Longrightarrow \mathbb{P}$ and $\mathbb{P}_n \Longrightarrow \mathbb{P}'$, then $\mathbb{P} = \mathbb{P}'$. *Hint: use the Portmanteau lemma.*

Exercise 6.6. Prove that if $F_n \Longrightarrow F$ and x is such that there is at most one $a \in \mathbb{R}$ with $F(a) = x$, then $F_n^{-1}(x) \longrightarrow F^{-1}(x)$. *Hint: For any $\varepsilon > 0$, you can choose a y such that F is continuous at y (why?) and $F^{-1}(x) - \varepsilon < y < F^{-1}(x)$.*

Exercise 6.7. Prove the remainder of Theorem 6.1.9.

Exercise 6.8. Find an example such that $X_n \Longrightarrow X$ and $Y_n \Longrightarrow Y$ but $X_n + Y_n \not\Longrightarrow X + Y$.

Exercise 6.9. Prove Theorem 6.2.6.

Exercise 6.10. Prove Lemma 6.3.1.

Exercise 6.11. Let $X_1, X_2, \ldots$ be i.i.d. with characteristic function ψ. Show that if $\psi'(0) = ia$, then

$$\frac{S_n}{n} \xrightarrow{P} a.$$

LECTURE 7

CONDITIONAL EXPECTATION

7.1 Definition and Properties

Definition 7.1.1. For random variables $X \in \mathbb{R}^{d_1}$ and $Y \in \mathbb{R}^{d_2}$, we define a *conditional expectation of X given Y* by any random variable Z satisfying:

a) there exists $g : \mathbb{R}^{d_2} \to \mathbb{R}^{d_1}$ such that $Z = g(Y)$ and
b) $\mathbb{E}[Z\mathbb{I}\{Y \in A\}] = \mathbb{E}[X\mathbb{I}\{Y \in A\}]$ for all $A \subseteq \mathbb{R}^{d_2}$.

We denote such a Z by $\mathbb{E}[X \mid Y]$.

Example 7.1.2. $\Omega = [-1, 1]$ and $\mathbb{P}$ is uniform distribution. Define

$$\begin{aligned}
X(\omega) &= -1/2 + \mathbb{I}\{\omega \in [-1, -1/2] \cup [0, 1/2]\} + 2\mathbb{I}\{\omega \in [-1/2, 0]\} \\
Y(\omega) &= \mathbb{I}\{\omega \geq 0\} \\
Z(\omega) &= 1 - Y(\omega).
\end{aligned}$$

Observe that $\mathbb{E}[X \mid Y] = Z$ and $\mathbb{P}(X = Z) = 0$.

Proposition 7.1.3. Conditional expectation satisfies the following:

i) $X \geq 0$ a.s. implies $\mathbb{E}[X \mid Y] \geq 0$ a.s.;
ii) $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}X$;
iii) $\mathbb{E}[aX_1 + bX_2 \mid Y] = a\mathbb{E}[X_1 \mid Y] + b\mathbb{E}[X_2 \mid Y]$ for all $a, b \in \mathbb{R}$;
iv) X independent of Y implies $\mathbb{E}[X \mid Y] = \mathbb{E}X$;
v) $\mathbb{E}|X| < \infty$ implies $\mathbb{E}|\mathbb{E}[X \mid Y]| < \infty$.

Note we are considering $d_1 = d_2 = 1$ here, but analogous results hold for the multivariate case.

Proof.

i) Suppose $\mathbb{E}[X \mid Y] = g(Y)$ and consider $A = \{x \in \mathbb{R} : g(x) \leq 0\}$. Then, $\mathbb{E}[Z\mathbb{I}\{Y \in A\}] = \mathbb{E}[g(Y)\mathbb{I}\{g(Y) \leq 0\}] \leq 0$, and $\mathbb{E}[Z\mathbb{I}\{Y \in A\}] = \mathbb{E}[X\mathbb{I}\{Y \in A\}] \geq 0$, so $\mathbb{E}[Z\mathbb{I}\{Y \in A\}] = 0$. Thus, since $Z\mathbb{I}\{Z \leq 0\} \leq 0$, $Z\mathbb{I}\{Z \leq 0\} = 0$ a.s., so $Z \geq 0$ a.s.
ii) Take $A = \mathbb{R}$.
iii) Set $Z = a\mathbb{E}[X_1 \mid Y] + b\mathbb{E}[X_2 \mid Y]$. For any $A \subseteq \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[Z\mathbb{I}\{Y \in A\}] &= \mathbb{E}[(a\mathbb{E}[X_1 \mid Y] + b\mathbb{E}[X_2 \mid Y])\,\mathbb{I}\{Y \in A\}] \\
&= a\mathbb{E}\left[\mathbb{E}[X_1 \mid Y]\mathbb{I}\{Y \in A\}\right] + b\mathbb{E}\left[\mathbb{E}[X_2 \mid Y]\mathbb{I}\{Y \in A\}\right] \\
&= a\mathbb{E}[X_1\mathbb{I}\{Y \in A\}] + b\mathbb{E}[X_2\mathbb{I}\{Y \in A\}] \\
&= \mathbb{E}[(aX_1 + bX_2)\,\mathbb{I}\{Y \in A\}].
\end{aligned} \tag{7.1}$$

iv) Recall that X independent of $f(Y)$ for any f, so for any $A \subseteq \mathbb{R}$,

$$\mathbb{E}[X\mathbb{I}\{Y \in A\}] = \mathbb{E}[\mathbb{I}\{Y \in A\}]\mathbb{E}[X] = \mathbb{E}\left[(\mathbb{E}X)\mathbb{I}\{Y \in A\}\right].$$

v) Again let $\mathbb{E}[X \mid Y] = g(Y)$. Define $A = \{y \in \mathbb{R} : g(y) \geq 0\}$.

$$\begin{aligned} \mathbb{E}\left|\mathbb{E}[X \mid Y]\right| &= \mathbb{E}[\mathbb{E}[X \mid Y]^+] + \mathbb{E}[\mathbb{E}[X \mid Y]^-] \\ &= \mathbb{E}[g(Y)\mathbb{I}\{g(Y) \geq 0\}] - \mathbb{E}[g(Y)\mathbb{I}\{g(Y) < 0\}] \\ &= \mathbb{E}[g(Y)\mathbb{I}\{Y \in A\}] - \mathbb{E}[g(Y)\mathbb{I}\{Y \in A^c\}] \\ &= \mathbb{E}[X\mathbb{I}\{Y \in A\}] - \mathbb{E}[X\mathbb{I}\{Y \in A^c\}] \\ &\leq \mathbb{E}[|X|\,\mathbb{I}\{Y \in A\}] + \mathbb{E}[|X|\,\mathbb{I}\{Y \in A^c\}] \\ &= \mathbb{E}\,|X|\,. \end{aligned} \tag{7.2}$$

□

Theorem 7.1.4. Conditional expectation is unique. That is, for any X, Y, $Z_1 = \mathbb{E}[X \mid Y]$ and $Z_2 = \mathbb{E}[X \mid Y]$ implies $Z_1 = Z_2$ a.s.

Proof.

$$0 = \mathbb{E}[0 \mid Y] = \mathbb{E}[X - X \mid Y] = \mathbb{E}[X \mid Y] - \mathbb{E}[X \mid Y] = Z_1 - Z_2. \tag{7.3}$$

□

Lemma 7.1.5. For any $f : \mathbb{R} \to \mathbb{R}$, if $\mathbb{E}\,|X| < \infty$ and $\mathbb{E}\,|Xf(Y)| < \infty$, $\mathbb{E}[f(Y)X \mid Y] = f(Y)\mathbb{E}[X \mid Y]$.

Proof. Let $Z = f(Y)\mathbb{E}[X \mid Y]$. Consider any $A \subseteq \mathbb{R}$. Then,

$$\mathbb{E}[Z\mathbb{I}\{Y \in A\}] = \mathbb{E}[f(Y)\mathbb{E}[X \mid Y]\mathbb{I}\{Y \in A\}] = \mathbb{E}[h(Y)\mathbb{E}[X \mid Y]], \tag{7.4}$$

where $h(Y) = f(Y)\mathbb{I}\{Y \in A\}$.

First, suppose $f(Y) = \mathbb{I}\{Y \in B\}$ for some $B \subseteq \mathbb{R}$. Then,

$$\mathbb{E}[h(Y)\mathbb{E}[X \mid Y]] = \mathbb{E}[\mathbb{I}\{Y \in A \cap B\}\mathbb{E}[X \mid Y]] = \mathbb{E}[\mathbb{I}\{Y \in A \cap B\}X] = \mathbb{E}[f(Y)X\mathbb{I}\{Y \in A\}]. \tag{7.5}$$

Next, suppose $f(Y) = \sum_{i=1}^m b_i\mathbb{I}\{Y \in B_i\}$. By linearity of expectation and the result for indicator f, the result holds.

If $f(Y) \in [0, M]$ and $X \geq 0$ a.s., take $f_n \uparrow f$ where f_n are simple functions. Then, by MCT,

$$\mathbb{E}[f_n(Y)\mathbb{E}[X \mid Y]\mathbb{I}\{Y \in A\}] \longrightarrow \mathbb{E}[f(Y)\mathbb{E}[X \mid Y]\mathbb{I}\{Y \in A\}] \tag{7.6}$$

and

$$\mathbb{E}[f_n(Y)X\mathbb{I}\{Y \in A\}] \longrightarrow \mathbb{E}[f(Y)X\mathbb{I}\{Y \in A\}]. \tag{7.7}$$

Thus, since for all n,

$$\mathbb{E}[f_n(Y)\mathbb{E}[X \mid Y]\mathbb{I}\{Y \in A\}] = \mathbb{E}[f_n(Y)X\mathbb{I}\{Y \in A\}], \tag{7.8}$$

we have

$$\mathbb{E}[f(Y)\mathbb{E}[X \mid Y]\mathbb{I}\{Y \in A\}] = \mathbb{E}[f(Y)X\mathbb{I}\{Y \in A\}]. \tag{7.9}$$

Next, if $f(Y) \in [0, M]$ and $X \in \mathbb{R}$, repeat this using X^+ and X^-.

Finally, for integrable f, write $f = f^+ - f^-$, and use a sequence of bounded functions. □

Lemma 7.1.6. Consider integrable random variables X, Y, Z. Then,

$$\mathbb{E}[\mathbb{E}[X|Y] \mid Y, Z] = \mathbb{E}[\mathbb{E}[X|Y, Z] \mid Y] = \mathbb{E}[X \mid Y]. \tag{7.10}$$

Proof. Consider any $A \subseteq \mathbb{R}^2$. Then, $\mathbb{E}[X \mid Y] = g(Y)$ for some g, so we can define $h(Y, Z) = g(Y)$. Thus, trivially, $\mathbb{E}[\mathbb{E}[X|Y] \mid Y, Z] = \mathbb{E}[X \mid Y]$.

For the second equality, consider any $A \subseteq \mathbb{R}$. Then,

$$\begin{aligned}\mathbb{E}[\mathbb{I}\{Y \in A\}\mathbb{E}[X|Y]] &= \mathbb{E}[\mathbb{I}\{Y \in A\}X] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{I}\{Y \in A\}X \mid Y, Z]] \\ &= \mathbb{E}[\mathbb{I}\{Y \in A\}\mathbb{E}[X \mid Y, Z]].\end{aligned} \tag{7.11}$$

□

7.2 Computation

Definition 7.2.1. For random variables X, Y and sets A, B, if $\mathbb{P}(Y \in B) > 0$ then define the *conditional probability* (with respect to an event) by

$$\mathbb{P}(X \in A \mid Y \in B) = \frac{\mathbb{P}(X \in A, Y \in B)}{\mathbb{P}(Y \in B)}. \tag{7.12}$$

Definition 7.2.2. For random variables X, Y and a set A, define the *conditional probability* (with respect to a random variable) by

$$\mathbb{P}(X \in A \mid Y) = \mathbb{E}[\mathbb{I}\{X \in A\} \mid Y]. \tag{7.13}$$

Theorem 7.2.3. If X and Y have joint density $f_{X,Y}$ and marginal densities f_X and f_Y,

$$\mathbb{E}[X \mid Y] = \int_{\mathbb{R}} x \frac{f_{X,Y}(x, Y)}{f_Y(Y)} dx. \tag{7.14}$$

Proof. For any $A \subseteq \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\left[\left(\int_{\mathbb{R}} x \frac{f_{X,Y}(x, Y)}{f_Y(Y)} dx\right) \mathbb{I}\{Y \in A\}\right] &= \int_A \left(\int_{\mathbb{R}} x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx\right) f_Y(y) dy \\ &= \int_{\mathbb{R}^2} x \mathbb{I}\{y \in A\} f_{X,Y}(x, y) dx dy \\ &= \mathbb{E}[X \mathbb{I}\{Y \in A\}].\end{aligned} \tag{7.15}$$

□

Definition 7.2.4. For random variables X and Y, define the *conditional variance* by

$$\mathrm{Var}(X \mid Y) = \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2 \mid Y]. \tag{7.16}$$

Theorem 7.2.5. If X is a random variable with $\mathrm{Var}(X) < \infty$, then for all Y

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]). \tag{7.17}$$

Proof. Observe that

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X-\mathbb{E}[X])^2] \\
&= \mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X])^2 \mid Y]] \\
&= \mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X \mid Y]+\mathbb{E}[X \mid Y]-\mathbb{E}[X])^2 \mid Y]] \\
&= \mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X \mid Y])^2 \mid Y]] + \mathbb{E}[\mathbb{E}[(\mathbb{E}[X \mid Y]-\mathbb{E}[X])^2 \mid Y]] \\
&\quad + 2\mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y]-\mathbb{E}[X]) \mid Y]] \\
&= \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]) + 2\mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y]-\mathbb{E}[X]) \mid Y]].
\end{aligned} \tag{7.18}$$

Then, it remains to check that

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y]-\mathbb{E}[X]) \mid Y]] \\
&= \mathbb{E}[\mathbb{E}[X\mathbb{E}[X \mid Y] - (\mathbb{E}[X \mid Y])^2 - X\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[X \mid Y] \mid Y]] \\
&= \mathbb{E}[\mathbb{E}[X\mathbb{E}[X \mid Y] \mid Y] - \mathbb{E}[(\mathbb{E}[X \mid Y])^2 \mid Y] - \mathbb{E}[X\mathbb{E}[X] \mid Y] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[X \mid Y] \mid Y]] \\
&= \mathbb{E}[(\mathbb{E}[X \mid Y])^2 - (\mathbb{E}[X \mid Y])^2 - \mathbb{E}[X]\mathbb{E}[X \mid Y] + \mathbb{E}[X]\mathbb{E}[X \mid Y]] \\
&= 0.
\end{aligned} \tag{7.19}$$

□

7.3 Conditional Limit Theorems

Proposition 7.3.1 (Conditional Jensen's Inequality). If f is a convex function and X is a random variable such that $f(X)$ is integrable, for any Y

$$f(\mathbb{E}[X|Y]) \leq \mathbb{E}[f(X)|Y]. \tag{7.20}$$

The inequality is flipped if f is concave.

Proof. Since f is convex, we can write $f(x) = \sup\{\ell(x) : \ell \text{ is linear}, \ell(x) \leq f(x)\ \forall x\}$. Then, for all such ℓ,

$$\mathbb{E}[f(X) \mid Y] \geq \mathbb{E}[\ell(X) \mid Y] = \ell(\mathbb{E}[X \mid Y]). \tag{7.21}$$

Thus,

$$\mathbb{E}[f(X) \mid Y] \geq \sup_{\ell} \ell(\mathbb{E}[X \mid Y]) = f(\mathbb{E}[X \mid Y]). \tag{7.22}$$

If concave then $-f$ is convex. □

Theorem 7.3.2 (Conditional Monotone Convergence Theorem). If $X_n \geq 0$ a.s. and $X_n \uparrow X$ a.s., for any Y

$$\mathbb{E}[X_n \mid Y] \uparrow \mathbb{E}[X \mid Y]. \tag{7.23}$$

Proof. Since $X - X_n \geq 0$ a.s., $\mathbb{E}[X_n \mid Y] \leq \mathbb{E}[X \mid Y]$. Define $Z = \lim_{n\to\infty} \mathbb{E}[X_n \mid Y]$. Clearly, $Z = g(Y)$ for some g, and $\mathbb{E}[X_n \mid Y] \uparrow Z$. Finally, for any $A \subseteq \mathbb{R}$, $\mathbb{E}[\mathbb{E}[X_n \mid Y]\mathbb{I}\{Y \in A\}] = \mathbb{E}[X_n\mathbb{I}\{Y \in A\}]$. Thus,

$$\begin{aligned}
\mathbb{E}\left[\lim_{n\to\infty} \mathbb{E}[X_n \mid Y]\mathbb{I}\{Y \in A\}\right] &= \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X_n \mid Y]\mathbb{I}\{Y \in A\}] \quad \text{(by MCT)} \\
&= \lim_{n\to\infty} \mathbb{E}[X_n\mathbb{I}\{Y \in A\}] \\
&= \mathbb{E}[X\mathbb{I}\{Y \in A\}] \quad \text{(by MCT).}
\end{aligned} \tag{7.24}$$

□

Theorem 7.3.3 (Conditional Fatou's Lemma). If $X_n \geq 0$ a.s. for all n, for any Y

$$\liminf_{n \longrightarrow \infty} \mathbb{E}[X_n \mid Y] \geq \mathbb{E}[\liminf_{n \longrightarrow \infty} X_n \mid Y]. \tag{7.25}$$

Proof. For each n, define $Z_n = \inf_{m \geq n} X_m$. Clearly, $X_n \geq Z_n$ a.s. and $Z_n \uparrow \liminf_{n \longrightarrow \infty} X_n = Z$ a.s. Thus, by conditional MCT, $\mathbb{E}[Z_n \mid Y] \uparrow \mathbb{E}[Z \mid Y]$. That is,

$$\mathbb{E}[\liminf_{n \to \infty} X_n \mid Y] = \mathbb{E}[Z \mid Y] = \lim_{n \to \infty} \mathbb{E}[Z_n \mid Y] \leq \liminf_{n \to \infty} \mathbb{E}[X_n \mid Y]. \tag{7.26}$$

$\square$

Theorem 7.3.4 (Conditional Dominated Convergence Theorem). If $X_n \longrightarrow X$ a.s. and $|X_n| \leq Z$ a.s. for some integrable Z, for all Y

$$\mathbb{E}[X|Y] = \lim_{n \to \infty} \mathbb{E}[X_n \mid Y]. \tag{7.27}$$

Proof. Observe that $|\mathbb{E}[X_n \mid Y]| \leq \mathbb{E}[|X_n| \mid Y] \leq \mathbb{E}[Z \mid Y]$ a.s., and $\mathbb{E}[\mathbb{E}[Z \mid Y]] = \mathbb{E}Z < \infty$. So,

$$\begin{aligned}
\mathbb{E}\left[\lim_{n \to \infty} \mathbb{E}[X_n \mid Y]\mathbb{I}\{Y \in A\}\right] &= \lim_{n \to \infty} \mathbb{E}[\mathbb{E}[X_n \mid Y]\mathbb{I}\{Y \in A\}] \quad \text{(by DCT)} \\
&= \lim_{n \to \infty} \mathbb{E}[X_n \mathbb{I}\{Y \in A\}] \\
&= \mathbb{E}[X \mathbb{I}\{Y \in A\}] \quad \text{(by DCT)}.
\end{aligned} \tag{7.28}$$

$\square$

7.4 Exercises

Exercise 7.1 (Conditional Chebyshev's). Prove that if $a > 0$ then

$$\mathbb{P}(|X| \geq a \mid Y) \leq \frac{1}{a^2}\mathbb{E}(X^2 \mid Y).$$

Exercise 7.2 (Conditional Cauchy-Schwarz). Prove that for any random variables X, Y, Z,

$$\Big(\mathbb{E}[XY \mid Z]\Big)^2 \leq \mathbb{E}[X^2 \mid Z]\mathbb{E}[Y^2 \mid Z].$$

Exercise 7.3 (Rosenthal 13.4.6). Find an example of random variables X, Y, Z such that X and Y are independent but $\mathbb{E}[X \mid Z]$ and $\mathbb{E}[Y \mid Z]$ are not.

Exercise 7.4 (Rosenthal 13.4.10). Find an example of X and Y which are not independent but $\mathbb{E}[X \mid Y] = \mathbb{E}X$ a.s.

Exercise 7.5 (Rosenthal 13.4.11). Show that if $\mathbb{E}[X \mid Y] = \mathbb{E}X$ a.s. then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, but the reverse implication need not hold.

Exercise 7.6. Review how to compute conditional expectation. (e.g., Exercises from Section 3.5 of Evans and Rosenthal.)

LECTURE 8

MARKOV CHAINS

8.1 Random Walks

Definition 8.1.1. Consider $X_0 = 0$ and for each t, X_t is 1 with probability p and -1 with probability $1-p$. Then, $S_n = \sum_{t=0}^{n} X_t$ is called a *simple random walk.*

Proposition 8.1.2. Let S_n be a simple random walk. For any $a \in \mathbb{N}$, if $E_a = \mathbb{P}(\exists n \in \mathbb{N} \text{ s.t. } S_n = a)$,

$$E_a = \begin{cases} 1 & \text{if } p \geq 1/2 \\ (\frac{p}{1-p})^a & \text{if } p < 1/2 \end{cases}. \tag{8.1}$$

Proof. First, observe that $E_a = (E_1)^a$ for $a > 0$. Now, conditioning on X_1,

$$\begin{aligned} E_1 &= \mathbb{E}\Big[\mathbb{I}\{\exists n \in \mathbb{N} \text{ s.t. } S_n = 1\} \mid X_1 = 1\Big]\mathbb{P}(X_1 = 1) \\ &\quad + \mathbb{E}\Big[\mathbb{I}\{\exists n \in \mathbb{N} \text{ s.t. } S_n = 1\} \mid X_1 = -1\Big]\mathbb{P}(X_1 = -1) \\ &= 1 \cdot p + \mathbb{E}\Big[\mathbb{I}\{\exists n-1 \in \mathbb{N} \text{ s.t. } S_{n-1} - 1 = 1\} \mid X_1 = -1\Big](1-p) \\ &= p + E_2(1-p) \\ &= p + (1-p)(E_1)^2. \end{aligned} \tag{8.2}$$

Solving this quadratic gives

$$E_1 = \frac{1 \pm \sqrt{1 - 4(1-p)p}}{2(1-p)}. \tag{8.3}$$

Now, observe that $1 - 4(1-p)p = 1 - 4p + 4p^2 = (2p-1)^2$. Thus, possible solutions are

$$E_1 = \frac{1 \pm (2p-1)}{2(1-p)} = \begin{cases} \frac{p}{1-p} \\ 1 \end{cases}. \tag{8.4}$$

If $p \geq 1/2$, then $p/(1-p) \geq 1$, so we get $E_1 = 1$. That is, the random walk will hit every $a \in \mathbb{N}$ a.s. Now, consider $p < 1/2$. Then, both solutions of (8.4) are valid, so we need to do more work to determine which is correct. We define $E_a(n) = \mathbb{P}(\exists m \leq n \text{ s.t. } S_m = a)$. Observe that $E_1(1) = p \leq \frac{p}{1-p}$. We will now use induction to show $E_1(n) \leq \frac{p}{1-p}$ for all $n \in \mathbb{N}$. The induction hypothesis is $E_1(n) \leq \frac{p}{1-p}$. Then, by the same conditioning as above,

$$\begin{aligned} E_1(n+1) &= p + (1-p)E_2(n) \\ &\leq p + (1-p)E_1^2(n) \\ &\leq p + (1-p)\left(\frac{p}{1-p}\right)^2 \\ &= \frac{p(1-p) + p^2}{1-p} \\ &= \frac{p}{1-p}. \end{aligned} \tag{8.5}$$

This completes the induction. Finally, since $E_1(n)$ is increasing, by continuity of measure

$$E_1 = \lim_{n\to\infty} E_1(n) \le \frac{p}{1-p}. \tag{8.6}$$

Combined with (8.4), this gives the result. □

Proposition 8.1.3 (Gambler's Ruin). Consider $p \in [0,1]$ and and $N \in \mathbb{N}$. Let X_n be a process where for all $i \in \{1, \ldots, N-1\}$, $\mathbb{P}[X_{n+1} = i+1 \mid X_n = i] = p$ and $\mathbb{P}[X_{n+1} = i-1 \mid X_n = i] = 1-p = q$, but $\mathbb{P}[X_{n+1} = 0 \mid X_n = 0] = 1$ and $\mathbb{P}[X_{n+1} = N \mid X_n = N] = 1$. Then,

$$P_i = \mathbb{P}[\exists\, n \ge 1 \text{ s.t. } X_n = N \mid X_0 = i] = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)^N} & \text{if } p \neq 1/2 \\ \frac{i}{N} & \text{if } p = 1/2. \end{cases} \tag{8.7}$$

Proof. For $i \in \{1, \ldots, N-1\}$, $P_i = pP_{i+1} + qP_{i-1}$. Writing $P_i = pP_i + qP_i$ and combining these two equations gives $P_{i+1} - P_i = (q/p)(P_i - P_{i-1})$. Using that $P_0 = 0$,

$$\begin{aligned} P_2 - P_1 &= (q/p)(P_1 - P_0) = (q/p)P_1 \\ P_3 - P_2 &= (q/p)(P_2 - P_1) = (q/p)^2 P_1 \\ &\vdots \\ P_i - P_{i-1} &= (q/p)(P_{i-1} - P_{i-2}) = (q/p)^{i-1} P_1 \\ &\vdots \\ P_N - P_{N-1} &= (q/p)(P_{N-1} - P_{N-2}) = (q/p)^{N-1} P_1. \end{aligned} \tag{8.8}$$

Then,

$$P_i - P_1 = \sum_{k=2}^{i} P_k - P_{k-1} = \sum_{k=2}^{i} (q/p)^{k-1} P_1 = P_1 \sum_{k=1}^{i-1} (q/p)^k = \begin{cases} \frac{(q/p)^i - (q/p)}{(q/p)-1} P_1 & \text{if } q \neq p \\ (i-1)P_1 & \text{if } q = p. \end{cases} \tag{8.9}$$

That is,

$$P_i = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)} P_1 & \text{if } q \neq p \\ iP_1 & \text{if } q = p. \end{cases} \tag{8.10}$$

Finally, since $P_N = 1$,

$$P_1 = \begin{cases} \frac{1-(q/p)}{1-(q/p)^N} & \text{if } q \neq p \\ \frac{1}{N} & \text{if } q = p. \end{cases} \tag{8.11}$$

□

8.2 Chapman-Kolmogorov Equations

Definition 8.2.1. A stochastic process satisfies the *Markov property* if for all t, $\mathbb{E}[X_t \mid X_{t-1}, X_{t-2}, \ldots, X_0] = \mathbb{E}[X_t \mid X_{t-1}]$.

Definition 8.2.2. Consider a stochastic process X_n taking values in a countable state space $\mathcal{S}$ and satisfying the Markov property with the added requirement that the probability of moving from state i to state j is independent of the time n. Such a process is called a *Markov chain*, and is defined by the *transition matrix* $\mathbf{P}$ with entries $P_{ij} = \mathbb{P}[X_n = j \mid X_{n-1} = i]$.

Lemma 8.2.3. A simple random walk is a Markov chain.

Proof.
$$\mathbb{E}[S_n \mid S_{n-1}, \dots, S_0] = (S_{n-1}+1)(1/2) + (S_{n-1}-1)(1/2) = \mathbb{E}[S_n \mid S_{n-1}]. \tag{8.12}$$

Then, in the transition matrix, $P_{i,i+1} = p = 1 - P_{i,i-1}$ and $P_{ij} = 0$ otherwise. □

Theorem 8.2.4. Define $\mathbf{P}^n$ to be the n-step transition matrix, so that $P_{ij}^{(n)} = \mathbb{P}(X_{k+n} = j \mid P_k = i)$. Then, for any states $i, j \in \mathcal{S}$ and $n, m > 0$,

$$P_{ij}^{(n+m)} = \sum_{k \in \mathcal{S}} P_{ik}^{(n)} P_{kj}^{(m)}. \tag{8.13}$$

Proof.
$$\begin{aligned}
P_{ij}^{(n+m)} &= \mathbb{P}(X_{n+m} = j \mid X_0 = i) \\
&= \mathbb{P}\left(\cup_{k \in \mathcal{S}}[X_{n+m} = j, X_n = k] \mid X_0 = i\right) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}\left(X_{n+m} = j, X_n = k \mid X_0 = i\right) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}\left(X_{n+m} = j \mid X_n = k, X_0 = i\right) \mathbb{P}\left(X_n = k \mid X_0 = i\right) \qquad (8.14) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}\left(X_{n+m} = j \mid X_n = k\right) \mathbb{P}\left(X_n = k \mid X_0 = i\right) \\
&= \sum_{k \in \mathcal{S}} P_{jk}^{(m)} P_{ik}^{(n)}.
\end{aligned}$$

□

8.3 Classification Properties

Definition 8.3.1. We say state j is *accessible* from state i if for some $n \geq 0$, $P_{ij}^{(n)} > 0$. If i is also accessible from j, we say they *communicate*.

Definition 8.3.2. For any event A, use $\mathbb{P}_i(A) = \mathbb{P}(A \mid X_0 = i)$ to denote the probability conditional on the initial state. Also, let $f_{ij}^{(n)} = \mathbb{P}_i(X_n = j \cap X_m \neq j \ \forall 1 \leq m \leq n-1)$ and $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = \mathbb{P}_i(\exists n \geq 1 \text{ s.t. } X_n = j)$. Then, a state i is *recurrent* if it always returns back to itself, so $f_{ii} = 1$, and is *transient* if it is not recurrent, so $f_{ii} < 1$.

Theorem 8.3.3. A state i is transient if and only if $\mathbb{P}_i(X_n = i \text{ i.o.}) = 0$, which occurs if and only if $\sum_{n=1}^{\infty} P_{ii}^{(n)} < \infty$. A state i is recurrent if and only if $\mathbb{P}_i(X_n = i \text{ i.o.}) = 1$, which occurs if and only if $\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$.

Proof. First, by continuity of probabilities,

$$\mathbb{P}_i(X_n = i \text{ i.o.}) = \lim_{k\to\infty} \mathbb{P}_i(|\{n \geq 1 : X_n = i\}| \geq k) = \lim_{k\to\infty} (f_{ii})^k = \begin{cases} 0 & f_{ii} < 1 \\ 1 & f_{ii} = 1 \end{cases}. \tag{8.15}$$

Next, recalling $\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$ for non-negative discrete random variables,

$$\begin{aligned} \sum_{n=1}^{\infty} P_{ii}^{(n)} &= \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = i) \\ &= \sum_{n=1}^{\infty} \mathbb{E}_i \mathbb{I}\{X_n = i\} \\ &= \mathbb{E}_i \sum_{n=1}^{\infty} \mathbb{I}\{X_n = i\} \\ &= \mathbb{E}_i \left|\{n \geq 1 : X_n = i\}\right| \\ &= \sum_{k=1}^{\infty} \mathbb{P}_i(|\{n \geq 1 : X_n = i\}| \geq k) \\ &= \sum_{k=1}^{\infty} (f_{ii})^k \\ &= \begin{cases} \frac{f_{ii}}{1-f_{ii}} < \infty, & f_{ii} < 1 \\ \infty, & f_{ii} = 1 \end{cases}. \end{aligned} \tag{8.16}$$

□

Definition 8.3.4. A Markov chain is *irreducible* if for all states i and j, $f_{ij} > 0$.

Theorem 8.3.5. The following are equivalent for an irreducible Markov chain:

i) For all states i and j, $f_{ij} = 1$;
ii) There exists a state k such that $f_{kk} = 1$;
iii) There exist states k and ℓ such that $\sum_{n=1}^{\infty} P_{k\ell}^{(n)} = \infty$;
iv) For all states i and j, $\sum_{n=1}^{\infty} P_{ij}^{(n)} = \infty$.

Proof. i) implies ii) is obvious and ii) implies iii) follows from the previous theorem. Suppose iii) holds for some $k, \ell \in \mathcal{S}$ and fix $i, j \in \mathcal{S}$. By irreducibility, there exists $m, r > 0$ such that $P_{ik}^{(m)} > 0$ and $P_{\ell j}^{(r)} > 0$. Then, by Chapman-Kolmogorov, for any $n > 0$ we have

$$P_{ij}^{(m+n+r)} = \sum_{u \in \mathcal{S}} P_{iu}^{(m)} P_{uj}^{(n+r)} = \sum_{u,v \in \mathcal{S}} P_{iu}^{(m)} P_{uv}^{(n)} P_{vj}^{(r)} \geq P_{ik}^{(m)} P_{k\ell}^{(n)} P_{\ell j}^{(r)}. \tag{8.17}$$

Thus,

$$\sum_{n=1}^{\infty} P_{ij}^{(n)} = \sum_{n=1-(m+r)}^{\infty} P_{ij}^{(m+n+r)} \geq \sum_{n=1}^{\infty} P_{ik}^{(m)} P_{k\ell}^{(n)} P_{\ell j}^{(r)} = P_{ik}^{(m)} P_{\ell j}^{(r)} \sum_{n=1}^{\infty} P_{k\ell}^{(n)} = \infty. \tag{8.18}$$

Finally, suppose iv) holds and by way of contradiction suppose $f_{ij} < 1$ for some $i, j \in \mathcal{S}$. Then,

$$1 - f_{jj} = \mathbb{P}_j(\forall n \geq 1, X_n \neq j) \geq \mathbb{P}_j(X_n = i \text{ before } X_n = j)\mathbb{P}_i(\forall n \geq 1, X_n \neq j) > 0, \tag{8.19}$$

where the last inequality is by irreducibility and the contradiction assumption. However, this implies that $\sum_{n=1}^{\infty} P_{jj}^{(n)} < \infty$ by the previous theorem, which is a contradiction. □

Definition 8.3.6. If any of properties i) to iv) hold, the Markov chain itself is said to be *recurrent*.

8.4 Stationarity Convergence

Definition 8.4.1. A probability distribution $\boldsymbol{\pi}$ on the state space is the *stationary distribution* of a Markov chain X_n if $\sum_{i \in \mathcal{S}} \pi_i P_{ij} = \pi_j$ for all $j \in \mathcal{S}$. This can be written in matrix form as $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$.

Definition 8.4.2. The *period* of a state i, denoted by $d(i)$, is the greatest common divisor of $\{n \geq 1 : P_{ii}^{(n)} > 0\}$.

Example 8.4.3.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{bmatrix}. \tag{8.20}$$

Definition 8.4.4. A Markov chain is *aperiodic* if every state has period 1. Otherwise, it is *periodic*.

Lemma 8.4.5. If $f_{ij} > 0$ and $f_{ji} > 0$, states i and j have the same period.

Proof. There must exist some r, s such that $P_{ij}^{(r)} > 0$ and $P_{ji}^{(s)} > 0$. Then, $P_{ii}^{(r+s)} \geq P_{ij}^{(r)} P_{ji}^{(s)} > 0$, so $r + s$ is a multiple of $d(i)$. Now, for any $m \in \{n \geq 1 : P_{jj}^{(n)} > 0\}$, $P_{ii}^{(r+m+s)} > P_{ij}^{(r)} P_{jj}^{(m)} P_{ji}^{(s)} > 0$. Thus, $r + m + s$ is a multiple of $d(i)$, which implies m is a multiple of $d(i)$. That is, $d(j) \geq d(i)$. However, this argument was symmetric in i and j, so the reverse inequality can also be shown. □

Corollary 8.4.6. If a Markov chain is irreducible, all of its states have the same period.

Lemma 8.4.7. If a Markov chain is irreducible and has a stationary distribution $\boldsymbol{\pi}$, then it is recurrent.

Proof. Suppose by way of contradiction it is not recurrent. Then, for all $i, j \in \mathcal{S}$, $\sum_{n=1}^{\infty} P_{ij}^{(n)} < \infty$, so $\lim_{n\to\infty} P_{ij}^{(n)} = 0$. But, since $\pi_j = \sum_{i\in\mathcal{S}} \pi_i P_{ij}^{(n)}$ for all n, taking $n \to \infty$ shows $\pi_j = 0$ for all j. This contradicts $\sum_{j\in\mathcal{S}} \pi_j = 1$. □

Theorem 8.4.8. If a Markov chain is irreducible and aperiodic with stationary distribution $\boldsymbol{\pi}$, then for all states i and j

$$\lim_{n\to\infty} \mathbb{P}_i(X_n = j) = \pi_j. \tag{8.21}$$

Proof. Not too hard, but too long for this course. □

8.5 Stationarity Existence

Definition 8.5.1. The *mean recurrence time* of a state i is defined by

$$m_i = \mathbb{E}_i[\inf\{n \geq 1 : X_n = i\}]. \tag{8.22}$$

Definition 8.5.2. Observe that if i is transient, then $m_i = \infty$. If i is recurrent, then we say it is *null recurrent* if $m_i = \infty$ and *positive recurrent* if $m_i < \infty$.

Definition 8.5.3. A state i is *ergodic* if it is aperiodic and positive recurrent. If all states are ergodic, then we say the entire Markov chain is ergodic.

Theorem 8.5.4. If a Markov chain is irreducible and each state i is positive recurrent, then there exists a unique stationary distribution $\boldsymbol{\pi}$ with $\pi_i = 1/m_i$. Further, if each state is aperiodic (i.e., the Markov chain is ergodic), then $\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j = 1/m_j$.

Proof. Not too hard, but too long for this course. □

8.6 Exercises

Exercise 8.1. Solve for the stationary distribution of the Markov chain with

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}.$$

Exercise 8.2 (Ross 4.16). Show that if state i is recurrent and state i does not communicate with state j, then $P_{ij} = 0$.

Exercise 8.3 (Ross 4.20). A Markov chain is *doubly stochastic* if the sum over columns is also one; that is, $\sum_{i \in \mathcal{S}} P_{ij} = 1$ for all $j \in \mathcal{S}$. Show that if such a chain is irreducible and aperiodic with M states that the limiting probabilities are given by $\pi_j = 1/M$ for all $j \in \mathcal{S}$.

Exercise 8.4 (Ross 4.59). Recall the Gambler's Ruin setup, and define

$$M_i = \mathbb{E}_i[\inf\{n \geq 0 : X_n = N \text{ or } X_n = 0\}].$$

Show that

$$M_i = \begin{cases} \frac{i}{q-p} - \frac{N}{q-p} \frac{1-(q/p)^i}{1-(q/p)^N} & \text{if } p \neq 1/2 \\ i(N-1) & \text{if } p = 1/2. \end{cases}$$

Exercise 8.5 (Rosenthal 8.5.10). Show that for any Markov chain on a finite state space, at least one state must be recurrent. Also, give an example where exactly one state is recurrent. Finally, give an example where the state space is countably infinite and all states are transient.

Exercise 8.6 (Rosenthal 8.5.18). Prove that if $f_{ij} > 0$ and $f_{ji} = 0$ then i is transient.

Exercise 8.7 (Rosenthal 8.5.19). Prove that for a Markov chain with a finite state space, no states are null recurrent.

Exercise 8.8. Show that a simple random walk is recurrent but does not have a stationary distribution.

LECTURE 9

QUEUEING THEORY

9.1 Exponential Distribution

Definition 9.1.1. We say a nonnegative random variable $X \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$ if it has density $f(x) = \lambda e^{-\lambda x}$. Observe then it has CDF $F(x) = 1 - e^{-\lambda x}$, mean $\mathbb{E}X = 1/\lambda$, and MGF $M_X(t) = \frac{\lambda}{\lambda - t}$.

Lemma 9.1.2. $X \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$ if and only if it has the *memoryless property*:

$$\mathbb{P}[X > s + t \mid X > t] = \mathbb{P}[X > s]. \tag{9.1}$$

Proof. To show the first direction,

$$\mathbb{P}[X > s + t \mid X > t] = \frac{\mathbb{P}[X > s + t, X > t]}{\mathbb{P}[X > t]} = \frac{\mathbb{P}[X > s + t]}{\mathbb{P}[X > t]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda(t)}} = e^{-\lambda s} = \mathbb{P}[X > s]. \tag{9.2}$$

For the reverse direction, suppose X has the memory less property, and define $S(t) = \mathbb{P}[X > t]$. Then,

$$\begin{aligned}
\mathbb{P}[X > s + t \mid X > t] = \mathbb{P}[X > s] \quad &\iff \\
\frac{\mathbb{P}[X > s + t, X > t]}{\mathbb{P}[X > t]} = \mathbb{P}[X > s] \quad &\iff \\
S(s + t) = S(s)S(t).
\end{aligned} \tag{9.3}$$

This implies that for all $a \in \mathbb{Q}_+$, $S(a) = S(1)^a = e^{a \log(S(1))} = e^{-\lambda a}$, where $\lambda = -\log(S(1))$. To extend this to all $a > 0$, it remains to observe that $S(a)$ is monotonically decreasing. □

Definition 9.1.3. We say that $Y \sim \mathrm{Gamma}(\lambda, \alpha)$ for $\lambda, \alpha > 0$ if it has density $f(y) = \lambda e^{-\lambda y} \frac{(\lambda y)^{\alpha - 1}}{\Gamma(\alpha)}$, where $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha - 1} dx$ and $\Gamma(n) = (n-1)!$ for integer n.

Lemma 9.1.4. If $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$, then $Y = \sum_{i=1}^n X_i \sim \mathrm{Gamma}(\lambda, n)$.

Proof. We will prove this by induction. For $n = 1$, this is trivially true. Suppose this holds for $n \geq 2$. Then, by the convolution formula for densities (just differentiate the normal convolution formula),

$$\begin{aligned}
f_{X_1 + \cdots X_n + X_{n+1}}(t) &= \int_0^\infty f_{X_{n+1}}(t - s) f_{X_1 + \cdots X_n}(s) \mathbb{I}\{t - s \geq 0\} ds \\
&= \int_0^t \lambda e^{-\lambda(t-s)} \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} ds \\
&= \lambda e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \int_0^t s^{n-1} ds \\
&= \lambda e^{-\lambda t} \frac{(\lambda t)^n}{n!}.
\end{aligned} \tag{9.4}$$

□

9.2 Poisson Processes

Definition 9.2.1. Let $N(t)$ denote the number of times some event has occurred by time t. We call $\{N(t), t \geq 0\}$ a *Poisson process* if

a) $N(0) = 0$;
b) $N(t_1) - N(t_0)$ is independent from $N(s_1) - N(s_0)$ whenever $(t_0, t_1] \cap (s_0, s_1] = \emptyset$;
c) $\mathbb{P}[N(s+t) - N(s) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$ for all $s, t \geq 0$.

Proposition 9.2.2. $N(t)$ is a Poisson process if and only if

i) $N(0) = 0$;
ii) $N(t_1) - N(t_0)$ is independent from $N(s_1) - N(s_0)$ whenever $(t_0, t_1] \cap (s_0, s_1] = \emptyset$;
iii) For any s, $N(s+t) - N(s)$ has the same distribution for all $t \geq 0$;
iv) $\mathbb{P}[N(h) = 1] = \lambda h + o(h)$;
v) $\mathbb{P}[N(h) \geq 2] = o(h)$.

Proof. Beyond the scope of this course. □

Proposition 9.2.3. Let $E_0 = 0$ and E_n denote the time of the n^{th} event in a Poisson process with paramter $\lambda > 0$, and $T_n = E_n - E_{n-1}$. Then, for all n, $T_n \sim \text{Exp}(\lambda)$, and are i.i.d..

Proof. Observe that $\mathbb{P}[T_1 > t] = \mathbb{P}[N(t) = 0] = e^{-\lambda t}$. Then,

$$\mathbb{P}[T_2 > t] = \mathbb{E}\Big[\mathbb{P}[T_2 > t \mid T_1]\Big]. \tag{9.5}$$

Now, notice that if $T_1 = s$, then $T_2 > t$ if and only if $N(s+t) - N(s) = 0$, which occurs with probability $e^{-\lambda t}$ regardless of s. Thus,

$$\mathbb{E}\Big[\mathbb{P}[T_2 > t \mid T_1]\Big] = e^{-\lambda t}. \tag{9.6}$$

This argument can be repeated up to T_n. □

Corollary 9.2.4. $S_n = \sum_{i=1}^n T_i \sim \text{Gamma}(\lambda, n)$. That is, the amount of time until the n^{th} event has a gamma distribution.

9.3 Continuous-Time Markov Chains

Definition 9.3.1. A *continuous-time Markov chain* is a sequence of random variables $\{X(t) : t \geq 0\}$ taking values in $\mathcal{S}$ where for all $s, t \geq 0$ and deterministic functions $x : [0, \infty) \to \mathcal{S}$,

$$\mathbb{P}(X(t+s) = j \mid X(s) = i, X(u) = x(u)\ \forall u \in [0, s)) = \mathbb{P}(X(t+s) = j \mid X(s) = i). \tag{9.7}$$

We again suppose that the transition probabilities are homogeneous in time, so that

$$\mathbb{P}(X(t+s) = j \mid X(s) = i) = \mathbb{P}_i(X(t) = j) = P_{ij}(t). \tag{9.8}$$

Corollary 9.3.2. For each state $i \in \mathcal{S}$, let T_i be the time it takes for the process to move into a new state j. Then, T_i follows the exponential distribution with some rate ν_i.

Proof. Observe that T_i will have the memoryless property. □

Definition 9.3.3. We define $q_{ij} = \nu_i P_{ij}$ to be the *instantaneous transition rates*, where P_{ij} is the probability that the next state after state i is state j (regardless of how long it takes to change states).

Lemma 9.3.4. It holds that

$$\lim_{h\to 0} \frac{1 - P_{ii}(h)}{h} = \nu_i \quad \text{and} \quad \lim_{h\to 0} \frac{P_{ij}(h)}{h} = q_{ij}. \tag{9.9}$$

Proof. Since the length of transitions is exponential,

$$\mathbb{P}(T_i > h) = e^{-\nu_i h} = \sum_{t=0}^{\infty} (-1)^t \frac{(\nu_i h)^t}{t!} = 1 - \nu_i h + o(h). \tag{9.10}$$

Thus,

$$\lim_{h\to 0} \frac{1 - P_{ii}(h)}{h} = \lim_{h\to 0} \frac{\mathbb{P}(T_i \le h)}{h} = \lim_{h\to 0} \frac{\nu_i h + o(h)}{h} = \nu_i. \tag{9.11}$$

Similarly,

$$\lim_{h\to 0} \frac{P_{ij}(h)}{h} = \lim_{h\to 0} \frac{\mathbb{P}(T_i \le h) P_{ij}}{h} = \lim_{h\to 0} \frac{\nu_i h P_{ij} + o(h) P_{ij}}{h} = q_{ij}. \tag{9.12}$$

□

Lemma 9.3.5 (Continous Chapman-Kolmogorov Equations). For all $s, t \ge 0$ and $i, j \in \mathcal{S}$,

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s). \tag{9.13}$$

Proof. Exercise. □

Theorem 9.3.6 (Kolmogorov's Backward and Forward Equations). For all $i, j \in \mathcal{S}$ and $t \ge 0$,

$$P'_{ij}(t) = \sum_{k \ne i} q_{ik} P_{kj}(t) - \nu_i P_{ij}(t), \tag{9.14}$$

and

$$P'_{ij}(t) = \sum_{k \ne j} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t). \tag{9.15}$$

Proof. Beyond the scope of this course. □

Corollary 9.3.7. Suppose under some "nice" (analogous to discrete-time) conditions, we have limiting probabilities defined by $\pi_j = \lim_{t\to\infty} P_{ij}(t)$. Then, they satisfy that for all states j,

$$\nu_j \pi_j = \sum_{k \ne j} q_{kj} \pi_k. \tag{9.16}$$

Proof. Using the forward equation, we have

$$\lim_{t\to\infty} P'_{ij}(t) = \lim_{t\to\infty} \sum_{k \ne j} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t) = \sum_{k \ne j} q_{kj} \pi_k - \nu_j \pi_j, \tag{9.17}$$

where we have supposed we can swap the order of limit and summation. It remains to observe that since $P_{ij}(t) \in [0, 1]$ for all t, if $P'_{ij}(t)$ converges then it must converge to zero. □

9.4 Birth and Death Processes

Definition 9.4.1. Consider a continuous-time Markov chain where $X(t)$ represents the number of people in system at time t. Further, suppose that when there are n people in system, people arrive at exponential rate λ_n and depart at exponential rate μ_n. This is called a *birth and death process.*

Lemma 9.4.2. In a birth and death process, the transition rate between states is $\nu_n = \lambda_n + \mu_n$ and the transition probabilities are

$$P_{01} = 1, P_{n,n+1} = \frac{\lambda_n}{\lambda_n + \mu_n}, \text{ and } P_{n,n-1} = \frac{\mu_n}{\lambda_n + \mu_n}. \tag{9.18}$$

Proof. Suppose $X(t) = n$ and define random variables $A_n \sim \text{Exp}(\lambda_n)$ and $B_n \sim \text{Exp}(\mu_n)$. Then, the next state is $E_n = X(t + T_n) = (n+1)\mathbb{I}\{A_n < B_n\} + (n-1)\mathbb{I}\{A_n \geq B_n\}$. Observe that $\mathbb{P}[A_n > B_n] = \mathbb{E}\Big[\mathbb{P}[A_n > B_n \mid B_n]\Big] = \mathbb{E}[g(B_n)]$. Then, notice that $g(y) = \mathbb{P}[A_n > y] = e^{-\lambda_n y}$. Thus,

$$\mathbb{E}[g(B_n)] = \int_0^\infty e^{-\lambda_n y} \mu_n e^{-\mu_n y} dy = \mu_n \int_0^\infty e^{-(\lambda_n + \mu_n)y} dy = \frac{\mu_n}{\lambda_n + \mu_n}. \tag{9.19}$$

The other probability can be found analogously. $\square$

Theorem 9.4.3. Let $\boldsymbol{\pi}$ be the stationary distribution of the state taken by a birth and death process. Then, if $C_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^{n} \mu_i}$,

$$\boldsymbol{\pi}_0 = \frac{1}{1 + \sum_{n=1}^\infty C_n} \quad \text{and} \quad \boldsymbol{\pi}_n = \frac{C_n}{1 + \sum_{n=1}^\infty C_n}. \tag{9.20}$$

Proof. Recall that for all $j \in \mathcal{S}$, $\nu_j \pi_j = \sum_{k \neq j} q_{kj} \pi_k$. That is,

$$\begin{aligned} \lambda_0 \pi_0 &= \mu_1 \pi_1 \\ (\lambda_1 + \mu_1)\pi_1 &= \mu_2 \pi_2 + \lambda_0 \pi_0 \\ &\vdots \\ (\lambda_n + \mu_n)\pi_n &= \mu_{n+1}\pi_{n+1} + \lambda_{n-1}\pi_{n-1}. \end{aligned} \tag{9.21}$$

Combining each equation to the next one gives $\lambda_n \pi_n = \mu_{n+1}\pi_{n+1}$ for all $n \geq 0$. Solving this gives

$$\begin{aligned} \pi_1 &= \frac{\lambda_0}{\mu_1}\pi_0 \\ \pi_2 &= \frac{\lambda_1}{\mu_2}\pi_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1}\pi_0 \\ &\vdots \\ \pi_n &= \frac{\lambda_{n-1}}{\mu_n}\pi_{n-1} = \frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1}\pi_0. \end{aligned} \tag{9.22}$$

Finally, since $\sum_{n=0}^\infty \pi_n = 1$, we get

$$1 = \pi_0 + \pi_0 \sum_{n=1}^\infty \frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1}. \tag{9.23}$$

$\square$

9.5 M/M/s Queue

Definition 9.5.1. A customer of interest arrives to the queue to find $\mathcal{L}$ customers in system with $\mathcal{L}_Q$ of them waiting. They then will wait for $\mathcal{W}_Q$ units of time before having a service that lasts $\mathcal{S}$ units of time, meaning they spent a total of $\mathcal{W} = \mathcal{W}_Q + \mathcal{S}$ units of time in the system. The averages of these values are denoted by L, L_Q, W_Q, $1/\mu$, and W respectively.

Definition 9.5.2. For the M/M/s queue, arrivals occur according to a Poisson process at rate λ and departures occur exponentially at rate $n\mu$ if there are n customers in service, unless $n > s$, and then at rate $s\mu$.

Theorem 9.5.3 (PASTA). For a queue with Poisson arrivals,

$$\mathbb{P}[\text{a new customer finds } n \text{ in system}] = \pi_n. \tag{9.24}$$

Proof (Heuristic). Clearly,

$$\mathbb{P}[\text{new customer finds } n \text{ in system} \mid \text{customer arrived at time } t] = \mathbb{P}[n \text{ in system at time } t] = \pi_n. \tag{9.25}$$

However, since a Poisson process has independent increments, knowing the time the customer arrived at does not affect how the past proceeded, which means it does not affect how many are in system. □

Lemma 9.5.4. For an M/M/s queue, if $\rho = \frac{\lambda}{s\mu}$ then the probability a new customer finds the queue empty is

$$\pi_0 = \left[\sum_{n=0}^{s-1} \frac{\lambda^n}{n!\,\mu^n} + \frac{\lambda^s}{s!\,\mu^s(1-\rho)}\right]^{-1}. \tag{9.26}$$

Proof. Observe that this is a birth and death process with

$$C_n = \begin{cases} \frac{\lambda^n}{n!\mu^n} & \text{if } n \leq s \\ \frac{\lambda^n}{s^{n-s}s!\mu^n} & \text{if } n > s. \end{cases} \tag{9.27}$$

□

Theorem 9.5.5 (Little's Law). For any stable queue that is non-preemptive with $\lambda = \lim_{t\to\infty} \frac{N(t)}{t}$,

$$L = \lambda W \quad \text{and} \quad L_Q = \lambda W_Q. \tag{9.28}$$

Proof (M/M/1 Case). Since $C_n = \rho^n$, $\pi_n = \rho^n(1-\rho)$ for $n \geq 0$. Then,

$$L = \sum_{n=0}^{\infty} n\pi_n = (1-\rho)\sum_{n=0}^{\infty} n\rho^n = (1-\rho)\frac{\rho}{(1-\rho)^2} = \frac{\rho}{1-\rho}. \tag{9.29}$$

Now, if you arrive to find $\mathcal{L}$ in system, your total time in system will be $\mathcal{L}+1$ exponential service times. Then, using that the sum of exponentials is gamma, $\mathbb{E}[\mathcal{W} \mid \mathcal{L}] = \frac{\mathcal{L}+1}{\mu}$, so

$$W = \frac{L+1}{\mu} = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda} = \frac{L}{\lambda}. \tag{9.30}$$

Similarly, $W_Q = W - 1/\mu = \frac{\rho}{\mu(1-\rho)}$, and $L_Q = L - \rho = \frac{\rho^2}{1-\rho} = \lambda\frac{\rho}{\mu(1-\rho)} = \lambda W_Q$. □

9.6 Exercises

Exercise 9.1. Prove Lemma 9.3.5.

Exercise 9.2 (Ross 8.1). For the M/M/1 queue, compute the expected number of arrivals during a service period and the probability no customers arrive during a service period.

Exercise 9.3 (Ross 8.4). Show that for an M/M/1 queue, conditional on a customer waiting $x > 0$ units of time before entering service, $\mathcal{L} - 1 \sim \text{Pois}(\lambda)$. (That is, compute the conditional pmf.)

Exercise 9.4 (Ross 8.11). Consider a M/M/1 queue with the following variation: Whenever a service is completed a departure occurs only with probability α. With probability 1 - α the customer, instead of leaving, joins the end of the queue. Note that a customer may be serviced more than once.

a) Find $\boldsymbol{\pi}$.
b) Find the expected waiting time of a customer from the time they arrive until they enter service for the first time.
c) What is the probability that a customer enters service exactly n times, n = 1, 2, ...?
d) What is the expected amount of time that a customer spends in service (which does not include the time they spend waiting in line)?
e) What is the distribution of the total length of time a customer spends being served?

Exercise 9.5 (Ross 8.13). Two customers move about among three servers. Upon completion of service at server i, the customer leaves that server and enters service at whichever of the other two servers is free. (Therefore, there are always two busy servers.) If the service times at server i are exponential with rate μ_i, what proportion of time is server i idle?

Exercise 9.6 (Ross 8.16). Customers arrive at a two-server system according to a Poisson process having rate $\lambda = 5$. An arrival finding server 1 free will begin service with that server. An arrival finding server 1 busy and server 2 free will enter service with server 2. An arrival finding both servers busy goes away. Once a customer is served by either server, he departs the system. The service times at server i are exponential with rates μ_i, where $\mu_1 = 4$ and $\mu_2 = 2$. (What is the average time an entering customer spends in the system, and what proportion of time is server 2 busy?

Exercise 9.7 (Ross 8.28). Let D denote the time between successive departures in a stationary M/M/1 queue with $\lambda < \mu$. Show, by conditioning on whether or not a departure has left the system empty, that D is exponential with rate λ.