Lecture Notes on Probability Theory

Jason Swanson

August 5, 2024

Preface

These notes are for a sequence of graduate courses on probability and stochastic analysis. The notes are not intended to stand alone, but rather to use with the textbooks listed in the references. The notes do not contain a proof for every result. We have included proofs with techniques that are especially important to beginning students. The student should refer to the corresponding textbook for proofs that we have omitted.

To get the most out of these notes, the student should have had prior exposure to measure theory. In this case, Part I serves only as a review of measure and integration. Most of the proofs in this part are omitted. The student may regard Part I as a reference of results to recall as needed.

It is possible, though, to use these notes without any prior exposure to measure theory. For such a student, more proofs have been added at the end of Part I. In this case, the student should at least be proficient in undergraduate mathematical analysis. This includes, but is not limited to, the following topics:

1. set theory, including relations, equivalence relations, functions, arbitrary Cartesian products, and projection maps,
2. partial orderings and the well-ordering principle,
3. cardinality, countability, and the Schröder-Bernstein theorem,
4. the extended real number system, supremum, infimum, limit superior, and limit inferior, and
5. metric spaces, separability, continuity, compactness, total boundedness, the Bolzano-Weierstrass theorem, and the Heine-Borel theorem.

For a more complete list, see [5, Chapter 0].

In addition to this background, students should also have taken an undergraduate course in probability. For a review of undergraduate probability, see [12].

Contents

Part I

Review of Measure and Integration

Chapter 1

Measures

1.1 σ-algebras

If X is a set, we write 2^X for the power set of X.

Let X be a nonempty set and $\mathcal{M} \subset 2^X$ a nonempty collection of subsets of X. If

- $E_1, \ldots, E_n \in \mathcal{M}$ implies $\bigcup_{j=1}^n E_j \in \mathcal{M}$, and
- $E \in \mathcal{M}$ implies $E^c \in \mathcal{M}$,

then $\mathcal{M}$ is an **algebra** (or **field**) on X. If

- $E_1, E_2, \ldots \in \mathcal{M}$ implies $\bigcup_{n=1}^\infty E_n \in \mathcal{M}$, and
- $E \in \mathcal{M}$ implies $E^c \in \mathcal{M}$,

then $\mathcal{M}$ is a **σ-algebra** (or **σ-field**) on X.

If $\mathcal{M}$ is a σ-algebra, and $E_1, E_2, \ldots \in \mathcal{M}$, then $\bigcap_{n=1}^\infty E_n = (\bigcup_{n=1}^\infty E_n^c)^c \in \mathcal{M}$. The smallest σ-algebra on X is $\{\varnothing, X\}$, which is called the **trivial σ-algebra**. The largest σ-algebra on X is 2^X.

If $E \in \mathcal{M}$, define

$$\mathcal{M}|_E = \{A \cap E : A \in \mathcal{M}\} = \{A \in \mathcal{M} : A \subset E\}.$$

Then $\mathcal{M}|_E$ is a σ-algebra on E (check).

Proposition 1.1. *If $\mathscr{C} = \{\mathcal{M}_\alpha : \alpha \in A\}$ is a nonempty collection of σ-algebras on X, then*

$$\bigcap_{\alpha \in A} \mathcal{M}_\alpha = \{E \subset X : E \in \mathcal{M}_\alpha \text{ for all } \alpha \in A\}$$

is a σ-algebra.

Proof. Exercise 1.1. □

Let $\mathcal{E} \subset 2^X$. Let $\mathscr{C} = \{\mathcal{M}_\alpha : \alpha \in A\}$ denote the collection of all σ-algebras $\mathcal{M}_\alpha$ on X such that $\mathcal{E} \subset \mathcal{M}_\alpha$. Note that $\mathscr{C}$ is nonempty, since $2^X \in \mathscr{C}$. It therefore follows that $\sigma(\mathcal{E}) := \bigcap_{\alpha \in A} \mathcal{M}_\alpha$ is a σ-algebra. We call $\sigma(\mathcal{E})$ the σ-algebra generated by $\mathcal{E}$. It is usually described as the smallest σ-algebra containing $\mathcal{E}$. This description is justified by the following proposition.

Proposition 1.2. *If $\mathcal{E} \subset 2^X$, then $\sigma(\mathcal{E})$ is the unique σ-algebra such that*

(a) $\mathcal{E} \subset \sigma(\mathcal{E})$, and

(b) if $\mathcal{G}$ is a σ-algebra on X such that $\mathcal{E} \subset \mathcal{G}$, then $\sigma(\mathcal{E}) \subset \mathcal{G}$.

Proof. Exercise 1.4. □

1.1.1 Borel σ-algebras

Let X be a set and $\mathcal{T} \subset 2^X$. If

- $X \in \mathcal{T}$, $\varnothing \in \mathcal{T}$,
- $U_1, \ldots, U_n \in \mathcal{T}$ implies $\bigcap_{j=1}^n U_j \in \mathcal{T}$, and
- $\{U_\alpha : \alpha \in A\} \subset \mathcal{T}$ implies $\bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}$,

then $\mathcal{T}$ is a **topology** on X and $(X, \mathcal{T})$ is a **topological space**. A set $U \in \mathcal{T}$ is called an **open set**. A **neighborhood of** $x \in X$ is any $U \in \mathcal{T}$ such that $x \in U$. If $\{x_n\}$ is a sequence in X and $x \in X$, then we say x_n **converges to** x, written $x_n \to x$, if, for any neighborhood U of x, there exists $N \in \mathbb{N}$ such that $x_n \in U$ whenever $n \geqslant N$. If $(Y, \mathcal{U})$ is another topological space, then $f : X \to Y$ is **continuous** if $f^{-1}(U) \in \mathcal{T}$ for all $U \in \mathcal{U}$. We say that f is **continuous at** $x \in X$ if, for any neighborhood V of $f(x)$, there exists a neighborhood U of x such that $f(U) \subset V$. A function $f : X \to Y$ is continuous if and only if it is continuous at x for all $x \in X$ (see [10, Theorem 18.1]). If f is continuous at x, then $f(x_n) \to f(x)$ whenever $x_n \to x$ (see [10, Theorem 21.3]). Unlike in metric spaces, the converse is not true in general. A partial converse, however, is provided by Lemma 1.10.

If (X, ρ) is a metric space, then the collection of subsets of X which are open (in the sense of a metric space) forms a topology on X called the **metric topology**. Moreover, the definitions of convergence and continuous functions on a metric space are equivalent to the topological definitions applied to the metric topology (see [10, Section 20]). A topological space $(X, \mathcal{T})$ is **metrizable** if there exists a metric ρ on X such that $\mathcal{T}$ is the metric topology for (X, ρ).

If $(X, \mathcal{T})$ is a topological space and $X \neq \varnothing$, then $\mathcal{B}_X := \sigma(\mathcal{T})$ is called the **Borel σ-algebra on** X. A set $E \in \mathcal{B}_X$ is called a **Borel set**.

The following is in [5, Proposition 1.2].

Proposition 1.3. *Consider the following subsets of $2^{\mathbb{R}}$:*

(a) the open intervals: $\mathcal{E}_1 = \{(a, b) : a < b\}$,

(b) the closed intervals: $\mathcal{E}_2 = \{[a,b] : a < b\}$,

(c) the half-open intervals: $\mathcal{E}_3 = \{(a,b] : a < b\}$ *or* $\mathcal{E}_4 = \{[a,b) : a < b\}$,

(d) the open rays: $\mathcal{E}_5 = \{(a,\infty) : a \in \mathbb{R}\}$ *or* $\mathcal{E}_6 = \{(-\infty,a) : a \in \mathbb{R}\}$,

(e) the closed rays: $\mathcal{E}_7 = \{[a,\infty) : a \in \mathbb{R}\}$ *or* $\mathcal{E}_8 = \{(-\infty,a] : a \in \mathbb{R}\}$.

Then $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{E}_i)$ *for any* $i \in \{1,\dots,8\}$.

1.1.2 Product σ-algebras

Let $\{X_\alpha : \alpha \in A\}$ be an indexed collection of nonempty sets. Let $X = \prod_{\alpha\in A} X_\alpha$ be the Cartesian product of this collection. A typical element of X has the form $x = (x_\alpha)_{\alpha\in A}$, where $x_\alpha \in X_\alpha$ for all α. Let $\pi_\alpha : X \to X_\alpha$ denote the projection onto the α-th component. That is, $\pi_\alpha(x) = x_\alpha$.

For each $\alpha \in A$, let $\mathcal{M}_\alpha$ be a σ-algebra on X_α. We define

$$\bigotimes_{\alpha\in A} \mathcal{M}_\alpha = \sigma(\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{M}_\alpha, \alpha \in A\}), \tag{1.1}$$

which we call the **product σ-algebra on** X.

The following propositions are from [5, Section 1.2].

Proposition 1.4. *If* A *is countable, then*

$$\bigotimes_{\alpha\in A} \mathcal{M}_\alpha = \sigma\left(\left\{ \prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{M}_\alpha \right\}\right).$$

Proposition 1.5. *If, for each* $\alpha \in A$, *we have* $\mathcal{M}_\alpha = \sigma(\mathcal{E}_\alpha)$, *then*

$$\bigotimes_{\alpha\in A} \mathcal{M}_\alpha = \sigma(\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{E}_\alpha, \alpha \in A\}),$$

Proposition 1.6. *If* A *is countable and, for each* $\alpha \in A$, *we have* $\mathcal{M}_\alpha = \sigma(\mathcal{E}_\alpha)$, *then*

$$\bigotimes_{\alpha\in A} \mathcal{M}_\alpha = \sigma\left(\left\{ \prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{E}_\alpha \right\}\right).$$

Proposition 1.7. *Let* $X_1,\dots,X_n$ *be metric spaces and let* $X = \prod_{j=1}^n X_j$ *be equipped with the product metric. Then* $\bigotimes_{j=1}^n \mathcal{B}_{X_j} \subset \mathcal{B}_X$. *If the* X_j*'s are separable, then* $\bigotimes_{j=1}^n \mathcal{B}_{X_j} = \mathcal{B}_X$. *In particular,* $\bigotimes_{j=1}^n \mathcal{B}_{\mathbb{R}} = \mathcal{B}_{\mathbb{R}^n}$.

Remark 1.8. We will adopt the notation $\mathcal{R} = \mathcal{B}_{\mathbb{R}}$ and $\mathcal{R}^n = \mathcal{B}_{\mathbb{R}^n}$.

1.1.3 The topology of pointwise convergence

Let X be a set and $\mathcal{B} \subset 2^X$. If $\bigcup_{B\in\mathcal{B}} B = X$ and, whenever $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \cap B_2$, there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$, then $\mathcal{B}$ is a **basis** for a topology on X.

If $\mathcal{B}$ is a basis for a topology on X and

$$\mathcal{T} = \left\{ \bigcup_{A\in\mathcal{A}} A : \mathcal{A} \subset \mathcal{B} \right\},$$

then $\mathcal{T}$ is a topology on X. Moreover, $U \in \mathcal{T}$ if and only if for all $x \in U$, there exists $B \in \mathcal{B}$ such that $x \in B \subset U$. In particular, this implies that for each $U \in \mathcal{T}$, there exists $\mathcal{A} \subset \mathcal{B}$ such that $U = \bigcup_{A\in\mathcal{A}} A$. We call $\mathcal{T}$ the **topology generated by** $\mathcal{B}$ and say that $\mathcal{B}$ **is a basis for** $\mathcal{T}$ (see [10, Section 13]).

Let $\{(X_\alpha, \mathcal{T}_\alpha) : \alpha \in A\}$ be an indexed collection of nonempty topological spaces and let $X = \prod_{\alpha\in A} X_\alpha$. A set of the form

$$\pi_{\alpha_1}^{-1}(U_{\alpha_1}) \cap \cdots \cap \pi_{\alpha_n}^{-1}(U_{\alpha_n}) \subset X,$$

where $U_{\alpha_j} \in \mathcal{T}_{\alpha_j}$, is called a **cylinder set**. If $\mathcal{B} \subset 2^X$ is the collection of cylinder sets, then $\mathcal{B}$ is a basis for a topology on X. The topology generated by $\mathcal{B}$ is called the **product topology** (see [10, Section 19]). The product topology is the unique topology on X that satisfies the following property: if $(Y, \mathcal{U})$ is a topological space and $f : Y \to X$, then f is continuous if and only if $\pi_\alpha \circ f : Y \to X_\alpha$ is continuous for all $\alpha \in A$ (see [10, Theorem 19.6] and [9, Theorem 3.37]).

In the product topology, $x_n \to x$ in X if and only if $\pi_\alpha(x_n) \to \pi_\alpha(x)$ in X_α for all $\alpha \in A$ (see [10, Exercise 19.6]).

If A and B are sets, we write A^B for the set of all functions from B to A. Let X be a topological space and T a set. The set X^T can be identified with $\prod_{t\in T} X$ and endowed with the product topology. Since $f(t) = \pi_t(f)$, it follows that $f_n \to f$ in the product topology if and only if $f_n(t) \to f(t)$ for all $t \in T$. It is for this reason that the product topology is also called the **topology of pointwise convergence**.

We now wish to compare the Borel σ-algebra generated by the product topology with the product σ-algebra. We will present a theorem giving sufficient conditions for them to be equal. Before presenting this theorem, we need some final preliminary definitions.

Let $(X, \mathcal{T})$ be a topological space. Suppose that for all $x \in X$, there exists a countable set of neighborhoods of x, denoted $\mathcal{B}_x$, such that for any neighborhood U of x, there exists $A \in \mathcal{B}_x$ with $A \subset U$. Then we say that A is **first-countable**. If there exists $\mathcal{B} \subset \mathcal{T}$ such that $\mathcal{B}$ is countable and $\mathcal{B}$ is a basis for $\mathcal{T}$, then $(X, \mathcal{T})$ is said to be **second-countable**. Second-countability implies first-countability (see Exercise 1.7). A metric space is second-countable if and only if it is separable (see [10, Theorem 30.3(b) and Exercise 30.5(a)]). Moreover, a countable product of second-countable spaces is second-countable (see [10, Theorem 30.2]). A sometimes useful result in connection with this is that a countable product of metric spaces is metrizable (see [10, Exercise 21.3(b)]).

The following is a generalization of Proposition 1.7.

Theorem 1.9. *Let $\{(X_\alpha, \mathcal{T}_\alpha) : \alpha \in A\}$ be an indexed collection of topological spaces and let $X = \prod_{\alpha \in A} X_\alpha$, endowed with the product topology. Let $\mathcal{B}_X$ be the Borel σ-algebra on X and $\mathcal{B}_\alpha$ the Borel σ-algebra on X_α. Then $\bigotimes_{\alpha \in A} \mathcal{B}_\alpha \subset \mathcal{B}_X$. If A is countable and $(X_\alpha, \mathcal{T}_\alpha)$ is second-countable for all $\alpha \in A$, then $\bigotimes_{\alpha \in A} \mathcal{B}_\alpha = \mathcal{B}_X$.*

Proof. By Proposition 1.5,

$$\bigotimes_{\alpha \in A} \mathcal{B}_\alpha = \sigma(\{\pi_\alpha^{-1}(U_\alpha) : U_\alpha \in \mathcal{T}_\alpha, \alpha \in A\}).$$

Since each $\pi_\alpha^{-1}(U_\alpha)$ is a cylinder set and cylinder sets are open, it follows that $\bigotimes_{\alpha \in A} \mathcal{B}_\alpha \subset \mathcal{B}_X$.

Now suppose A is countable and $(X_\alpha, \mathcal{T}_\alpha)$ is second-countable for all $\alpha \in A$. It will suffice to show that for every open $U \subset X$, we have $U \in \bigotimes_{\alpha \in A} \mathcal{B}_\alpha$. Let $\widehat{\mathcal{B}}_\alpha$ be a countable basis for X_α and let $\widehat{\mathcal{B}}$ be the collection of cylinder sets of the form

$$\pi_{\alpha_1}^{-1}(U_{\alpha_1}) \cap \cdots \cap \pi_{\alpha_n}^{-1}(U_{\alpha_n}),$$

where $U_{\alpha_j} \in \widehat{\mathcal{B}}_{\alpha_j}$. Then $\widehat{\mathcal{B}}$ is countable and, by the proof of [10, Theorem 30.2], $\widehat{\mathcal{B}}$ is a basis for the product topology. Note that $\widehat{\mathcal{B}} \subset \bigotimes_{\alpha \in A} \mathcal{B}_\alpha$. Let $U \subset X$ be open. Then there exists $\mathcal{A} \subset \widehat{\mathcal{B}}$ such that $U = \bigcup_{A \in \mathcal{A}} A$. Since this is a countable union and each $A \in \mathcal{A}$ satisfies $A \in \bigotimes_{\alpha \in A} \mathcal{B}_\alpha$, it follows that $U \in \bigotimes_{\alpha \in A} \mathcal{B}_\alpha$. □

We now provide a useful application of the above theorem. The lemma below is [10, Theorem 30.1(b)].

Lemma 1.10. *Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces and $f : X \to Y$. Suppose X is first-countable and that $f(x_n) \to f(x)$ whenever $x_n \to x$. Then f is continuous at x.*

Theorem 1.11. *Let $(X, \mathcal{T})$ be a second-countable topological space (this is the case, for example, if X is a separable metric space), and let T be a countable set. Let $\mathcal{M} = \mathcal{B}_X$ and $\mathcal{M}^T = \bigotimes_{t \in T} \mathcal{B}_X$ be the product σ-algebra on X^T. Let $(Y, \mathcal{U})$ be a topological space and $G : X^T \to Y$. If $G(f_n) \to G(f)$ whenever $f_n(t) \to f(t)$ for all $t \in T$, then $G^{-1}(U) \in \mathcal{M}^T$ whenever $U \in \mathcal{U}$.*

Proof. Let us endow X^T with the product topology (or the topology of pointwise convergence). Suppose $G(f_n) \to G(f)$ whenever $f_n(t) \to f(t)$ for all $t \in T$. In other words, $G(f_n) \to G(f)$ whenever $f_n \to f$. Since a countable product of second-countable spaces is second-countable, and second-countability implies first-countability, Lemma 1.10 implies that G is continuous.

Now let $U \in \mathcal{U}$. By the definition of continuity, $G^{-1}(U)$ is open and therefore, $G^{-1}(U) \in \mathcal{B}_{X^T}$. But by Theorem 1.9, $\mathcal{B}_{X^T} = \mathcal{M}^T$. □

Remark 1.12. As we will see in Section 2.1 (specifically Proposition 2.2), under the hypotheses of Theorem 1.11, we may conclude that G is "$(\mathcal{M}^T, \mathcal{B}_Y)$-measurable".

Exercises

1.1. Prove Proposition 1.1.

1.2. Provide an example of a nonempty family of σ-algebras, $\mathscr{C} = \{\mathcal{M}_\alpha : \alpha \in A\}$, such that

$$\bigcup_{\alpha \in A} \mathcal{M}_\alpha = \{E \subset X : E \in \mathcal{M}_\alpha \text{ for some } \alpha \in A\}$$

is not a σ-algebra.

1.3. Prove that if $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots$ are σ-algebras, then $\bigcup_{n=1}^\infty \mathcal{M}_n$ is an algebra.

1.4. Prove Proposition 1.2.

1.5. A σ-algebra $\mathcal{M}$ is said to be **countably generated** if there exists a countable set $\mathcal{C} \subset 2^X$ such that $\mathcal{M} = \sigma(\mathcal{C})$. Prove that $\mathcal{R}$ is countably generated.

1.6. Show that every infinite σ-algebra is uncountable.

1.7. Prove that second-countability implies first-countability.

1.2 Measures

Let X be a nonempty set and $\mathcal{M}$ a σ-algebra on X. Then $(X, \mathcal{M})$ is called a **measurable space** and the sets $E \in \mathcal{M}$ are called **measurable sets**. A **measure** on $(X, \mathcal{M})$ is a function $\mu : \mathcal{M} \to [0, \infty]$ such that

(i) $\mu(\varnothing) = 0$, and

(ii) if $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ are disjoint, then

$$\mu\left(\biguplus_{n=1}^\infty E_n \right) = \sum_{n=1}^\infty \mu(E_n).$$

Note that an indexed collection of sets $\{E_n\}$ is disjoint if $E_n \cap E_m = \varnothing$ whenever $n \neq m$. Also note that $\biguplus$ means the same thing as $\bigcup$, except it indicates that the sets in the union are disjoint.

Property (ii) is called **countable additivity**. It implies **finite additivity**:

(ii′) if $\{E_j\}_{j=1}^n \subset \mathcal{M}$ are disjoint, then

$$\mu\left(\biguplus_{j=1}^n E_j \right) = \sum_{j=1}^n \mu(E_j),$$

since one can take $E_j = \varnothing$ for $j > n$. A function $\mu : \mathcal{M} \to [0, \infty]$ that satisfies (i) and (ii′), but not necessarily (ii), is called a **finitely additive measure** on $(X, \mathcal{M})$.

If μ is a measure on $(X, \mathcal{M})$, then $(X, \mathcal{M}, \mu)$ is called a **measure space**. If $\mu(X) < \infty$, then μ is a **finite** measure. If there exists $\{E_n\}_{n=1}^{\infty} \subset \mathcal{M}$ such that $X = \bigcup_{n=1}^{\infty} E_n$ and $\mu(E_n) < \infty$ for all n, then μ is a **σ-finite** measure. If $\mu(X) = 1$, then μ is called a **probability measure**. If $\mu(X) = 0$, then we say μ is **trivial**. If $\mu(X) > 0$, then μ is **nontrivial**.

Example 1.13. Let $(X, \mathcal{M})$ be any measurable space. For $E \in \mathcal{M}$, let $\mu(E) = |E|$. That is, $\mu(E)$ is the number of elements in E. Then μ is a measure on $(X, \mathcal{M})$ and is called **counting measure**.

Example 1.14. Let $(X, \mathcal{M})$ be any measurable space and fix $x_0 \in X$. For $E \in \mathcal{M}$, define $\mu(E) = 1$ if $x_0 \in E$ and $\mu(E) = 0$ otherwise. Then μ is a measure on $(X, \mathcal{M})$ and is called the **point mass measure** (or **Dirac measure**) at x_0.

The following is in [5, Theorem 1.8].

Theorem 1.15. *Let $(X, \mathcal{M}, \mu)$ be a measure space.*

(a) (**Monotonicity**) *If $E, F \in \mathcal{M}$ and $E \subset F$, then $\mu(E) \leqslant \mu(F)$.*

(b) (**Subadditivity**) *If $\{E_n\}_{n=1}^{\infty} \subset \mathcal{M}$, then $\mu(\bigcup_{n=1}^{\infty} E_n) \leqslant \sum_{n=1}^{\infty} \mu(E_n)$.*

(c) (**Continuity from below**) *If $\{E_n\}_{n=1}^{\infty} \subset \mathcal{M}$ and $E_1 \subset E_2 \subset \cdots$, then $\mu(\bigcup_{n=1}^{\infty} E_n) = \lim_{n\to\infty} \mu(E_n)$.*

(d) (**Continuity from above**) *If $\{E_n\}_{n=1}^{\infty} \subset \mathcal{M}$ and $E_1 \supset E_2 \supset \cdots$, and $\mu(E_1) < \infty$, then $\mu(\bigcap_{n=1}^{\infty} E_n) = \lim_{n\to\infty} \mu(E_n)$.*

Proof. Recall that for any two sets A and B, we have $A \backslash B = A \cap B^c$.

For (a), let $E, F \in \mathcal{M}$ with $E \subset F$. Since $F = E \uplus (F \backslash E)$, we have

$$\mu(F) = \mu(E) + \mu(F \backslash E).$$

Since $\mu(F \backslash E) \geqslant 0$, it follows that $\mu(F) \geqslant \mu(E)$.

For (b), let $\{E_n\}_{n=1}^{\infty} \subset \mathcal{M}$. Define $F_1 = E_1$ and, for $n \geqslant 2$, let

$$F_n = E_n \backslash \left(\bigcup_{j=1}^{n-1} E_j \right).$$

Then $\{F_n\}_{n=1}^{\infty}$ are disjoint, $\biguplus_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} E_n$, and $F_n \subset E_n$. Thus,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n \right) = \mu\left(\biguplus_{n=1}^{\infty} F_n \right) = \sum_{n=1}^{\infty} \mu(F_n) \leqslant \sum_{n=1}^{\infty} \mu(E_n),$$

where the final inequality follows from (a).

For (c), let $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ with $E_1 \subset E_2 \subset \cdots$. Define $\{F_n\}_{n=1}^\infty$ as above. Then

$$\mu\left(\bigcup_{n=1}^\infty E_n\right) = \mu\left(\biguplus_{n=1}^\infty F_n\right) = \sum_{n=1}^\infty \mu(F_n)$$
$$= \lim_{n\to\infty} \sum_{j=1}^n \mu(F_n) = \lim_{n\to\infty} \mu\left(\biguplus_{j=1}^n F_j\right) = \lim_{n\to\infty} \mu(E_n),$$

where we have used the fact that $\biguplus_{j=1}^n F_j = \bigcup_{j=1}^n E_j = E_n$.

Finally, for (d), let $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ with $E_1 \supset E_2 \supset \cdots$. Define $E_n' = E_1 \backslash E_n$. Then $E_1' \subset E_2' \subset \cdots$. Let $E = \bigcap_{n=1}^\infty E_n$ and $E' = \bigcup_{n=1}^\infty E_n'$. Then, by (c), we have $\mu(E') = \lim_{n\to\infty} \mu(E_n')$. Also note that

$$E' = \bigcup_{n=1}^\infty (E_1 \cap E_n^c) = E_1 \cap \left(\bigcup_{n=1}^\infty E_n^c\right) = E_1 \cap \left(\bigcap_{n=1}^\infty E_n\right)^c = E_1 \backslash E.$$

Thus, $\mu(E_1 \backslash E) = \lim_{n\to\infty} \mu(E_1 \backslash E_n)$. Next, since $E_1 = E \uplus (E_1 \backslash E)$, we have $\mu(E_1) = \mu(E) + \mu(E_1 \backslash E)$. By (a), we have $\mu(E) \leqslant \mu(E_1) < \infty$. Therefore, we may subtract it from both sides, giving $\mu(E_1 \backslash E) = \mu(E_1) - \mu(E)$. Similarly, we have $\mu(E_1 \backslash E_n) = \mu(E_1) - \mu(E_n)$. It follows that

$$\mu(E_1) - \mu(E) = \mu(E_1) - \lim_{n\to\infty} \mu(E_n).$$

Lastly, since $\mu(E_1) < \infty$, this implies the conclusion of (d). □

Let $(X, \mathcal{M}, \mu)$ be a measure space. If $N \in \mathcal{M}$ and $\mu(N) = 0$, then N is called a **null set**. If something is true for all $x \in X$, except for x in some null set, then we say the property is true μ-**almost everywhere**, abbreviated μ-a.e. For example, if $f : X \to \mathbb{R}$, then $f = 0$ μ-a.e. means there exists a null set N such that $f(x) = 0$ for all $x \in N^c$. When the measure is understood, we drop the μ and simply write $f = 0$ a.e.

A set is called **negligible** if it is a subset of a null set. A measure space $(X, \mathcal{M}, \mu)$ is **complete** if $\mathcal{M}$ contains all negligible sets. That is, a complete measure space has the property that if $N \in \mathcal{M}$, $\mu(N) = 0$, and $F \subset N$, then $F \in \mathcal{M}$. Note that by monotonicity, we necessarily have $\mu(F) = 0$ in this case.

The following is [5, Theorem 1.9].

Theorem 1.16. *Let $(X, \mathcal{M}, \mu)$ be a measure space, and let $\mathcal{N}$ be the collection of null sets. That is, $\mathcal{N} = \{N \in \mathcal{M} : \mu(N) = 0\}$. Let*

$$\overline{\mathcal{M}} = \{E \cup F : E \in \mathcal{M} \text{ and } F \subset N \text{ for some } N \in \mathcal{N}\}.$$

Then $\overline{\mathcal{M}}$ is a σ-algebra on X and there exists a unique measure $\overline{\mu}$ on $(X, \overline{\mathcal{M}})$ such that

(a) $\overline{\mu}(E) = \mu(E)$ for all $E \in \mathcal{M}$, and

(b) $(X, \overline{\mathcal{M}}, \overline{\mu})$ is a complete measure space.

*The measure $\overline{\mu}$ is called the **completion** of μ, and $\overline{\mathcal{M}}$ is the **completion** of $\mathcal{M}$ with respect to μ.*

Exercises

1.8. Prove that counting measure, defined in Example 1.13, is a measure.

1.9. Prove that the point mass measure, defined in Example 1.14, is a measure.

1.10. Let $X = \mathbb{R}$ and

$$\mathcal{M} = \{E \subset \mathbb{R} : E \text{ or } E^c \text{ is countable}\}.$$

Define $\mu : \mathcal{M} \to [0, \infty]$ by

$$\mu(E) = \begin{cases} 0 & \text{if } E \text{ is countable}, \\ 1 & \text{if } E^c \text{ is countable}. \end{cases}$$

Prove that $(\mathbb{R}, \mathcal{M}, \mu)$ is a complete measure space.

1.3 Premeasures and outer measures

Let X be a nonempty set and $\mathcal{A}$ an algebra on X. A **premeasure** on $(X, \mathcal{A})$ is a function $\mu_0 : \mathcal{A} \to [0, \infty]$ such that

(i) $\mu_0(\varnothing) = 0$, and

(ii) if $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ are disjoint and $\biguplus_{n=1}^\infty A_n \in \mathcal{A}$, then

$$\mu_0\left(\biguplus_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu_0(A_n).$$

Example 1.17. Let $X = \mathbb{R}$ and let $\mathcal{A}$ be the collection of sets of the form

$$A = \biguplus_{j=1}^n (a_j, b_j],$$

where $-\infty \leqslant a_j \leqslant b_j \leqslant \infty$. (If $a_j = b_j$, we interpret $(a_j, b_j] = \varnothing$, and if $b_j = \infty$, we interpret $(a_j, b_j] = (a_j, \infty)$.) For any such A, let us define

$$\mu_0(A) = \sum_{j=1}^n (b_j - a_j).$$

Then $\mathcal{A}$ is an algebra on $\mathbb{R}$ and μ_0 is a premeasure on $(\mathbb{R}, \mathcal{A})$. For a proof of this, see, for example, [5, Proposition 1.15].

Let X be a nonempty set. An **outer measure** on X is a function $\mu^* : 2^X \to [0, \infty]$ such that

(i) $\mu^*(\varnothing) = 0$,

(ii) if $A \subset B$, then $\mu^*(A) \leqslant \mu^*(B)$, and

(iii) $\mu^*(\bigcup_{n=1}^\infty A_n) \leqslant \sum_{n=1}^\infty \mu^*(A_n)$.

A set $A \subset X$ is called **μ^*-measurable** if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c),$$

for all $E \subset X$.

The following is [5, Proposition 1.10 and Proposition 1.13a].

Proposition 1.18. *Let $\mathcal{A}$ be an algebra on a nonempty set X, and let μ_0 be a premeasure on $(X, \mathcal{A})$. For any $E \subset X$, define*

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^\infty \mu_0(A_n) : A_n \in \mathcal{A}, E \subset \bigcup_{n=1}^\infty A_n \right\}. \tag{1.2}$$

Then μ^* is an outer measure on X with $\mu^*|_{\mathcal{A}} = \mu_0$, and every $A \in \mathcal{A}$ is μ^*-measurable.

The following theorem is the foundation for the creation of measures such as Lebesgue measure on the real line. It can be found, for example, in [5, Theorem 1.11].

Theorem 1.19 (Carathéodory's extension theorem)**.** *Let μ^* be an outer measure on a nonempty set X, let $\mathcal{M}^*$ be the collection of μ^*-measurable sets, and let $\overline{\mu} = \mu^*|_{\mathcal{M}^*}$. Then $(X, \mathcal{M}^*, \overline{\mu})$ is a complete measure space.*

Remark 1.20. If μ^* is created from a premeasure according to Proposition 1.18, then $\sigma(\mathcal{A}) \subset \mathcal{M}^*$, so we may define $\mu = \mu^*|_{\sigma(\mathcal{A})}$. Then $(X, \sigma(\mathcal{A}), \mu)$ is a measure space, but it is not necessarily complete. This gives us a way to take a premeasure μ_0 on $(X, \mathcal{A})$, and extend it to a measure μ on $(X, \sigma(\mathcal{A}))$.

There may, however, be other ways to extend μ_0. Suppose ν is another measure on $(X, \sigma(\mathcal{A}))$ such that $\nu|_{\mathcal{A}} = \mu_0$. We cannot necessarily conclude that $\mu = \nu$. However, we can say two things:

(i) For any $E \in \sigma(\mathcal{A})$, we have $\nu(E) \leqslant \mu(E)$, with equality when $\mu(E) < \infty$.

(ii) If there exists $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty A_n$ and $\mu_0(A_n) < \infty$ for all n, then $\mu = \nu$.

If the conditions of (ii) hold, then we say μ_0 is a σ-finite premeasure on $(X, \mathcal{A})$. If μ_0 is σ-finite, then according to (ii), the measure μ is the unique extension of μ_0 from $\mathcal{A}$ to $\sigma(\mathcal{A})$. Moreover, if μ_0 is σ-finite, then $(X, \mathcal{M}^*, \overline{\mu})$ is the completion of $(X, \sigma(\mathcal{A}), \mu)$.

Example 1.21. Continuing Example 1.17, recall the premeasure μ_0 we defined on $(\mathbb{R}, \mathcal{A})$. Let us define the outer measure μ^* on $\mathbb{R}$ by (1.2). It can then be shown that for any $E \subset \mathbb{R}$, we have

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^\infty (b_j - a_j) : E \subset \bigcup_{n=1}^\infty (a_j, b_j] \right\}.$$

Let $\mathcal{L}$ be the collection of μ^*-measurable subsets of $\mathbb{R}$, and let $\overline{\lambda} = \mu^*|_{\mathcal{L}}$. According to Carathéodory's extension theorem, $(\mathbb{R}, \mathcal{L}, \overline{\lambda})$ is a complete measure space with $\overline{\lambda}(A) = \mu_0(A)$ for all $A \in \mathcal{A}$. In particular, $\overline{\lambda}((a, b]) = b - a$. The measure $\overline{\lambda}$ is called **Lebesgue measure**, the σ-algebra $\mathcal{L}$ is called the **Lebesgue σ-algebra**, and the sets $E \in \mathcal{L}$ are called **Lebesgue measurable** sets.

It can be shown that $\sigma(\mathcal{A}) = \mathcal{R}$, the Borel σ-algebra on $\mathbb{R}$. Let $\lambda = \overline{\lambda}|_{\mathcal{R}}$. Then $(\mathbb{R}, \mathcal{R}, \lambda)$ is a measure space, although it is not complete. By an abuse of terminology, the measure λ is also called Lebesgue measure, even though it is only defined for Borel sets. According to Remark 1.20, Lebesgue measure λ is the unique measure on $(\mathbb{R}, \mathcal{R})$ that satisfies $\lambda((a, b]) = b - a$ for all half-open intervals, and $(\mathbb{R}, \mathcal{L}, \overline{\lambda})$ is the completion of $(\mathbb{R}, \mathcal{R}, \lambda)$.

As one further abuse of notation, we will typically omit the bar, writing $(\mathbb{R}, \mathcal{L}, \lambda)$ instead of $(\mathbb{R}, \mathcal{L}, \overline{\lambda})$. The reader will often need to rely on context to determine whether the domain of λ is meant to be $\mathcal{L}$ or $\mathcal{R}$.

Exercises

1.11. Let μ^* be an outer measure on a nonempty set X, and let $\mathcal{M}^*$ be the collection of μ^*-measurable sets. Suppose $\{A_n\}_{n=1}^\infty \subset \mathcal{M}^*$ are disjoint. Prove that

$$\mu^*\left(E \cap \left(\biguplus_{n=1}^\infty A_n\right)\right) = \sum_{n=1}^\infty \mu^*(E \cap A_n),$$

for any $E \subset X$.

1.4 Borel measures on $\mathbb{R}$

Let X be a topological space. A Borel measure on X is a measure μ on $(X, \mathcal{B}_X)$. The following is [5, Theorem 1.16]. In these notes, "increasing" is synonymous with "nondecreasing".

Theorem 1.22. *If $F : \mathbb{R} \to \mathbb{R}$ is any increasing, right-continuous function, then there is a unique Borel measure μ_F on $\mathbb{R}$ such that*

$$\mu_F((a, b]) = F(b) - F(a),$$

for all $a, b \in \mathbb{R}$ with $a < b$. If G is another such function, then $\mu_F = \mu_G$ if and only if $F - G$ is constant.

Conversely, if μ is a Borel measure on $\mathbb{R}$ that is finite on all bounded sets, then the function

$$F(x) = \begin{cases} \mu((0, x]) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -\mu((x, 0]) & \text{if } x < 0, \end{cases}$$

is increasing, right-continuous, and $\mu = \mu_F$.

Remark 1.23. The measure μ_F is constructed as in Example 1.21, using the outer measure

$$\mu_F^*(E) = \inf\left\{\sum_{j=1}^{\infty}(F(b_j) - F(a_j)) : E \subset \bigcup_{j=1}^{\infty}(a_j, b_j]\right\}, \tag{1.3}$$

for all $E \subset \mathbb{R}$. We then have $\mu_F = \mu_F^*|_{\mathcal{R}}$.

If $\mathcal{M}_F$ is the collection of all μ_F^*-measurable sets and $\overline{\mu}_F = \mu_F^*|_{\mathcal{M}_F}$, then $(\mathbb{R}, \mathcal{M}_F, \overline{\mu}_F)$ is the completion of $(\mathbb{R}, \mathcal{R}, \mu_F)$. As with Lebesgue measure, we will typically omit the bar and write $(\mathbb{R}, \mathcal{M}_F, \mu_F)$ instead of $(\mathbb{R}, \mathcal{M}_F, \overline{\mu}_F)$. The reader must depend on context to determine whether the domain of μ_F is $\mathcal{R}$ or $\mathcal{M}_F$. The measure μ_F, in either case, is called the **Lebesgue-Stieltjes measure** associated to F.

When $F(x) = x$, we have $\mathcal{M}_F = \mathcal{L}$, which is the Lebesgue σ-algebra, and $\mu_F = \lambda$, which is Lebesgue measure.

Equation (1.3) is useful for doing calculations, and it is often helpful to recognize that we can use open intervals, instead of half-open intervals. This is often useful when combined with the fact that any open subset of $\mathbb{R}$ can be written as a countable union of disjoint open intervals. The following is [5, Lemma 1.17].

Lemma 1.24. *Let μ_F be a Lebesgue-Stieltjes measure. Then*

$$\mu_F(E) = \inf\left\{\sum_{j=1}^{\infty}(F(b_j) - F(a_j)) : E \subset \bigcup_{j=1}^{\infty}(a_j, b_j)\right\},$$

for all $E \in \mathcal{M}_F$.

Another helpful tool in calculating Lebesgue-Stieltjes measures is the following, which is [5, Theorem 1.18].

Theorem 1.25. *Let μ_F be a Lebesgue-Stieltjes measure. Then*

$$\mu_F(E) = \inf\{\mu_F(U) : E \subset U \text{ and } U \text{ is open}\}, \text{ and}$$
$$\mu_F(E) = \sup\{\mu_F(K) : K \subset E \text{ and } K \text{ is compact}\},$$

for all $E \in \mathcal{M}_F$.

Recall that $\triangle$ denotes the symmetric difference, $E \triangle A = (E\backslash A) \cup (A\backslash E)$.

The following proposition is an example of the so-called Littlewood's first principle of real analysis: Every measurable set is nearly a finite union of intervals. This proposition can be found in [5, Proposition 1.20].

Proposition 1.26. *Let μ_F be a Lebesgue-Stieltjes measure. Let $E \in \mathcal{M}_F$ with $\mu_F(E) < \infty$. Then for every $\varepsilon > 0$, there is a set of the form $A = \bigcup_{j=1}^{n}(a_j, b_j)$ such that $\mu_F(E \triangle A) < \varepsilon$.*

At this point, it is possible to prove that Lebesgue measure behaves as you would anticipate with respect to translations and dilations. The following theorem is [5, Theorem 1.21].

Theorem 1.27. *If $E \in \mathcal{L}$, then*

$$E + s := \{x + s : x \in E\} \in \mathcal{L}, \text{ and}$$
$$rE := \{rx : x \in E\} \in \mathcal{L}.$$

Moreover, $\lambda(E + s) = \lambda(E)$ and $\lambda(rE) = |r|\lambda(E)$.

The previous two results show that Lebesgue measure fits with our intuition in some rather important ways. However, there are still many counterintuitive facts about Lebesgue measure that can catch us off guard. An important source of counterexamples is the Cantor set and the associated Cantor function.

The **Cantor set**, C, is the set obtained iteratively from $[0,1]$ by successively removing the middle third from each remaining subinterval. Informally, we generate a sequence of sets that begins with

$$E_0 = [0,1],$$
$$E_1 = [0,1/3] \cup [2/3,1],$$
$$E_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1].$$

This pattern continues, and we define $C = \bigcap_{n=0}^{\infty} E_n$. For a rigorous definition of the Cantor set using ternary expansions, see [5, Section 1.5].

The Cantor set is compact. It is also nowhere dense, which means its closure has empty interior. The Cantor set is totally disconnected, meaning the only connected subsets of C are single points. Moreover, the Cantor set contains no isolated points. It can be shown that the Cantor set is uncountable, and also that $\lambda(C) = 0$.

A variant on the Cantor set is something called the **generalized Cantor set**. Instead of removing the middle thirds at the n-th stage, we remove the middle α_n-ths, where each $\alpha_n \in (0,1)$. The resulting generalized Cantor set, K, is compact, nowhere dense, totally disconnected, and uncountable. However, if $\{\alpha_n\}$ is chosen so that $\alpha_n \to 0$ sufficiently fast, then $\lambda(K) > 0$. Generalized Cantor sets can provide examples of nowhere dense sets with positive Lebesgue measure. For details, see [5, Section 1.5].

The complement of the Cantor set, relative to $[0,1]$, is $C^c = \bigcup_{n=0}^{\infty} E_n^c$. Note that

$$E_0^c = \varnothing,$$
$$E_1^c = (1/3,2/3),$$
$$E_2^c = (1/9,2/9) \cup (1/3,2/3) \cup (7/9,8/9),$$

and, in general, $E_n^c = \biguplus_{j=1}^{2^n-1} I_{j,n}$, where $I_{1,n}, \ldots, I_{2^n-1,n}$ are open intervals, ordered so that the endpoints of $I_{j,n}$ are less than the endpoints of $I_{j+1,n}$. With this notation, the **Cantor function** (or **Cantor-Lebesgue function**) is the unique function $f : [0,1] \to [0,1]$ such that

- $f(x) = j2^{-n}$ for all $x \in I_{j,n}$, and
- f is continuous.

For a rigorous definition of the Cantor function using ternary expansions, see [5, Section 1.5]. The Cantor function is an increasing function. Since it is also continuous, it is a surjection from $[0, 1]$ onto itself. Note that f is constant on every open subinterval of C^c. Thus, for every $x \in C^c$, we have that $f'(x)$ exists and $f'(x) = 0$. Since $\lambda(C) = 0$, we have $f' = 0$ a.e. on $[0, 1]$. Thus, the Cantor function is an example of a continuous, nonconstant function whose derivative is zero almost everywhere.

The Cantor function can be used to construct sets which are Lebesgue measurable, but not Borel measurable. For an example, see [5, Exercise 2.9].

Exercises

1.12. Let μ_F be a Lebesgue-Stieltjes measure. For $x \in \mathbb{R}$, define

$$F(x-) = \lim_{y \to x^-} F(y),$$

which exists, since F is increasing. Prove that

(a) $\mu_F(\{a\}) = F(a) - F(a-)$,

(b) $\mu_F([a, b)) = F(b-) - F(a-)$,

(c) $\mu_F([a, b]) = F(b) - F(a-)$, and

(d) $\mu_F((a, b)) = F(b-) - F(a)$.

1.13. Let $E \in \mathcal{L}$ with $\lambda(E) > 0$. Prove that for any $\alpha < 1$, there is an open interval I such that $\lambda(E \cap I) > \alpha\lambda(I)$.

Chapter 2

Integration

2.1 Measurable functions

Let X and Y be sets, and let $f : X \to Y$. Recall that if $E \subset Y$, then

$$f^{-1}(E) = \{x \in X : f(x) \in E\}.$$

Also recall that

$$f^{-1}\left(\bigcup_{\alpha \in A} E_\alpha\right) = \bigcup_{\alpha \in A} f^{-1}(E_\alpha),$$
$$f^{-1}\left(\bigcap_{\alpha \in A} E_\alpha\right) = \bigcap_{\alpha \in A} f^{-1}(E_\alpha),$$

and $f^{-1}(E^c) = (f^{-1}(E))^c$.

Now let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. A function $f : X \to Y$ is said to be $(\mathcal{M}, \mathcal{N})$**-measurable** if $f^{-1}(E) \in \mathcal{M}$ whenever $E \in \mathcal{N}$. The following propositions are from [5, Section 2.1].

Proposition 2.1. *Let $(X, \mathcal{M})$, $(Y, \mathcal{N})$, and $(Z, \mathcal{O})$ be measurable spaces. If $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and $g : Y \to Z$ is $(\mathcal{N}, \mathcal{O})$-measurable, then $g \circ f : X \to Z$ is $(\mathcal{M}, \mathcal{O})$-measurable.*

Proof. Exercise 2.1. □

Proposition 2.2. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. Suppose $\mathcal{N} = \sigma(\mathcal{E})$ for some $\mathcal{E} \subset 2^Y$, and let $f : X \to Y$. Then f is $(\mathcal{M}, \mathcal{N})$-measurable if and only if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{E}$.*

Consequently, if X and Y are topological spaces, then every continuous function $f : X \to Y$ is $(\mathcal{B}_X, \mathcal{B}_Y)$-measurable.

Proof. We will just prove the "if" part of the first claim. The remainder of the proof is Exercise 2.2.

Suppose $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{E}$. We want to show that $f^{-1}(E) \in \mathcal{M}$ for all $E \in \sigma(\mathcal{E})$. Unfortunately, there is no nice way of taking a set $E \in \sigma(\mathcal{E})$ and writing it in terms of sets in $\mathcal{E}$. We therefore employ the following common proof technique.

Let

$$\mathcal{L} = \{E \subset Y : f^{-1}(E) \in \mathcal{M}\}.$$

Suppose $\{E_n\}_{n=1}^\infty \subset \mathcal{L}$ and let $E = \bigcup_{n=1}^\infty E_n$. Then

$$f^{-1}(E) = f^{-1}\left(\bigcup_{n=1}^\infty E_n\right) = \bigcup_{n=1}^\infty f^{-1}(E_n).$$

Since each $f^{-1}(E_n) \in \mathcal{M}$ and $\mathcal{M}$ is a σ-algebra, it follows that $f^{-1}(E) \in \mathcal{M}$. Thus, $E \in \mathcal{L}$, and we have shown that $\mathcal{L}$ is closed under countable unions.

Next, suppose $E \in \mathcal{L}$, and note that $f^{-1}(E^c) = (f^{-1}(E))^c$. Since $f^{-1}(E) \in \mathcal{M}$ and $\mathcal{M}$ is a σ-algebra, it follows that $f^{-1}(E^c) \in \mathcal{M}$. Thus, $E^c \in \mathcal{L}$, and we have shown that $\mathcal{L}$ is closed under complements.

Since $\mathcal{L}$ is closed under countable unions and complements, it follows that $\mathcal{L}$ is a σ-algebra. Moreover, by hypothesis, $\mathcal{E} \subset \mathcal{L}$. Since $\mathcal{L}$ is a σ-algebra, it follows that $\sigma(\mathcal{E}) \subset \mathcal{L}$. In other words, if $E \in \sigma(\mathcal{E})$, then $E \in \mathcal{L}$, which implies $f^{-1}(E) \in \mathcal{M}$. □

The technique in the above proof is very common. In general, there is no good way to represent a generic set $E \in \sigma(\mathcal{E})$ in terms of the generating sets $A \in \mathcal{E}$. (It is possible using something called transfinite induction, but few people would consider that a "good" way.) So the typical approach is to define the collection of sets, $\mathcal{L}$, that satisfy the property we wish to prove, and then show that this collection is a σ-algebra that contains $\mathcal{E}$.

Sometimes, however, it can be difficult to prove that $\mathcal{L}$ is a σ-algebra. In this case, we can use something called the π-λ theorem, which is often used in probability theory.

To state the theorem, we first need two pieces of terminology. Let X be a set. Then $\mathcal{E} \subset 2^X$ is a **π-system** if it is closed under intersections, meaning that $A \cap B \in \mathcal{E}$ whenever $A, B \in \mathcal{E}$. Also, $\mathcal{L} \subset 2^X$ is a **λ-system** if the following three properties hold.

(a) $X \in \mathcal{L}$.

(b) If $A, B \in \mathcal{L}$ and $A \subset B$, then $B \backslash A \in \mathcal{L}$.

(c) If $\{A_n\}_{n=1}^\infty \subset \mathcal{L}$ and $A_n \subset A_{n+1}$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^\infty A_n \in \mathcal{L}$.

The following is [2, Theorem 2.1.2].

Theorem 2.3 (The π-λ theorem)**.** *If $\mathcal{E}$ is a π-system, $\mathcal{L}$ is a λ-system, and $\mathcal{E} \subset \mathcal{L}$, then $\sigma(\mathcal{E}) \subset \mathcal{L}$.*

For an example of the π-λ theorem in action, see the proof of Theorem 6.11.

Let $(X, \mathcal{M})$ be a measurable space and Y a topological space. If $f : X \to Y$ is $(\mathcal{M}, \mathcal{B}_Y)$-measurable, then we will just say f is **$\mathcal{M}$-measurable**, or we shorten it further to simply say f is **measurable**. In other words, we always take the σ-algebra on the range space to be the Borel σ-algebra, unless otherwise specified.

For example, if we say that $f : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable, then it means that f is $(\mathcal{L}, \mathcal{R})$-measurable. But if we say that $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then it means that f is $(\mathcal{R}, \mathcal{R})$-measurable.

Also note that, especially in probability, we will often use the abuse of notation, $f \in \mathcal{M}$, to mean that f is $\mathcal{M}$-measurable.

Remark 2.4. Proposition 2.1 shows that the composition of Borel measurable functions is Borel measurable. But note that the composition of Lebesgue measurable functions is not necessarily Lebesgue measurable.

Proposition 2.5. *If $(X, \mathcal{M})$ is a measurable space and $f : X \to \mathbb{R}$, then the following are equivalent:*

(a) f is $\mathcal{M}$-measurable.

(b) $f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.

(c) $f^{-1}([a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.

(d) $f^{-1}((-\infty, a)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.

(e) $f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.

Proof. Exercise 2.4. □

Let X be a set and $(Y, \mathcal{N})$ a measurable space. Let $f : X \to Y$ and define

$$\sigma(f) := \{f^{-1}(E) : E \in \mathcal{N}\}.$$

Then $\sigma(f)$ is a σ-algebra, and f is $(\sigma(f), \mathcal{N})$-measurable. Moreover, if $\mathcal{M}$ is another σ-algebra on X such that f is $(\mathcal{M}, \mathcal{N})$-measurable, then $\sigma(f) \subset \mathcal{M}$. In other words, $\sigma(f)$ is the smallest σ-algebra on X that makes f a measurable function. We call $\sigma(f)$ the σ-algebra **generated** by f.

More generally, if $\{(Y_\alpha, \mathcal{N}_\alpha)\}_{\alpha \in A}$ is a family of measurable spaces, and for each $\alpha \in A$, we have a function $f_\alpha : X \to Y_\alpha$, then

$$\sigma(\{f_\alpha : \alpha \in A\}) = \sigma(\{f_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{N}_\alpha, \alpha \in A\}).$$

This is the σ-algebra **generated** by the family $\{f_\alpha\}$, and it is the smallest σ-algebra on X that makes all of the f_α's measurable.

As an example, take $X = \prod_{\alpha \in A} Y_\alpha$ and let $f_\alpha = \pi_\alpha$ be the coordinate projections. By (1.1), we have $\sigma(\{\pi_\alpha : \alpha \in A\}) = \bigotimes_{\alpha \in A} \mathcal{N}_\alpha$. In other words, the product σ-algebra is just the σ-algebra generated by the coordinate projections.

The following propositions are from [5, Section 2.1].

Proposition 2.6. *Let* $(X, \mathcal{M})$ *and* $\{(Y_\alpha, \mathcal{N}_\alpha)\}_{\alpha \in A}$ *be measurable spaces. Let* $Y = \prod_{\alpha \in A} Y_\alpha$ *and* $\mathcal{N} = \bigotimes_{\alpha \in A} \mathcal{N}_\alpha$. *Let* $\pi_\alpha : Y \to Y_\alpha$ *be the coordinate projections. Then* $f : X \to Y$ *is* $(\mathcal{M}, \mathcal{N})$*-measurable if and only if* $f_\alpha := \pi_\alpha \circ f$ *is* $(\mathcal{M}, \mathcal{N}_\alpha)$*-measurable for all* $\alpha \in A$.

Corollary 2.7. *Let* $(X, \mathcal{M})$ *be a measurable space and* $f : X \to \mathbb{R}^d$. *Let* $f_1, \ldots, f_d$ *be the components of* f, *so that* $f(x) = (f_1(x), \ldots, f_d(x))$. *Then* f *is* $\mathcal{M}$*-measurable if and only if* $f_1, \ldots, f_d$ *are all* $\mathcal{M}$*-measurable.*

Corollary 2.8. *Let* $(X, \mathcal{M})$ *be a measurable space and* $f : X \to \mathbb{C}$. *Then* f *is* $\mathcal{M}$*-measurable if and only if* $\mathrm{Re} f$ *and* $\mathrm{Im} f$ *are both* $\mathcal{M}$*-measurable.*

Proposition 2.9. *Let* $(X, \mathcal{M})$ *be a measurable space and* $f, g : X \to \mathbb{C}$. *If* f *and* g *are both* $\mathcal{M}$*-measurable, then so are* $f + g$ *and* fg.

We frequently need to allow our functions to take on the values ∞ and $-\infty$. For this reason, let us introduce the extended real line, $\mathbb{R}^* = [-\infty, \infty]$. We equip $\mathbb{R}^*$ with the metric $\rho(x, y) = |A(x) - A(y)|$, where $A(x) = \tan^{-1}(x)$. This generates a topology on $\mathbb{R}^*$, allowing us to define the Borel σ-algebra, $\mathcal{B}_{\mathbb{R}^*}$. It can be shown that $\mathcal{B}_{\mathbb{R}^*} = \{E \subset \mathbb{R}^* : E \cap \mathbb{R} \in \mathcal{R}\}$. We will adopt the notation $\mathcal{R}^* = \mathcal{B}_{\mathbb{R}^*}$.

When working in $\mathbb{R}^*$, keep in mind that $\infty - \infty$ is undefined. However, we will adopt the convention that $0 \cdot \infty = 0$.

Proposition 2.10. *Let* $(X, \mathcal{M})$ *be a measurable space and let* $f, g : X \to \mathbb{R}^*$ *be* $\mathcal{M}$*-measurable.*

(i) The function fg *is* $\mathcal{M}$*-measurable.*

(ii) Fix $a \in \mathbb{R}^*$ *and define*

$$h(x) = \begin{cases} a & \text{if } f(x) + g(x) \text{ is undefined,} \\ f(x) + g(x) & \text{otherwise.} \end{cases}$$

Then h *is* $\mathcal{M}$*-measurable.*

Proof. Exercise 2.6. ☐

If $a, b \in \mathbb{R}^*$, then $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. The following propositions are from [5, Section 2.1].

Proposition 2.11. *If* $\{f_n\}$ *is a sequence of* $\mathbb{R}^*$*-valued measurable functions on* $(X, \mathcal{M})$, *then the functions* $\sup_n f_n$, $\limsup_{n\to\infty} f_n$, $\inf_n f_n$, *and* $\liminf_{n\to\infty} f_n$ *are all measurable. If* $f(x) = \lim_{n\to\infty} f_n(x)$ *exists for all* $x \in X$, *then* f *is measurable.*

Corollary 2.12. *If* $f, g : X \to \mathbb{R}^*$ *are measurable, then so are* $f \vee g$ *and* $f \wedge g$.

Corollary 2.13. *If* $\{f_n\}$ *is a sequence of* $\mathbb{C}$*-valued measurable functions on* $(X, \mathcal{M})$, *and* $f(x) = \lim_{n\to\infty} f_n(x)$ *exists for all* $x \in X$, *then* f *is measurable.*

It is often useful to combine these propositions with the following.

Proposition 2.14. *If $\{f_n\}$ is a sequence of $\mathbb{R}^*$-valued measurable functions on $(X, \mathcal{M})$, then $\{x : \lim_{n\to\infty} f_n(x)$ exists$\}$ is a measurable set.*

Proof. Exercise 2.7. □

If $f : X \to \mathbb{R}^*$, then the **positive part** and **negative part** of f are $f^+ = f \vee 0$ and $f^- = (-f) \vee 0$, respectively. Note that if f is measurable, then so are f^+ and f^-.

If $E \subset X$, then the **indicator function** of E is defined by

$$1_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases}$$

It is easy to verify that 1_E is $\mathcal{M}$-measurable if and only if $E \in \mathcal{M}$.

A **simple function** on $(X, \mathcal{M})$ is an $\mathcal{M}$-measurable function $\varphi : X \to \mathbb{C}$ such that the range of φ is a finite subset of $\mathbb{C}$. Let φ be a simple function with range $\{a_1, \ldots, a_n\}$ and let $E_j = \varphi^{-1}(a_j)$. Then the a_j's are distinct, the collection $\{E_j\}$ is a partition of X, and we can write $\varphi = \sum_{j=1}^n a_j 1_{E_j}$. This expression is called the **standard representation** of φ. Note that there may be a j such that $a_j = 0$, but this term is still included in the standard representation.

We say that $f_n \to f$ **pointwise** if $f_n(x) \to f(x)$ for all $x \in X$. An essential part of the theory of integration is that a function is measurable if and only if it is a pointwise limit of simple functions. This is formally stated in the following theorem, which is [5, Theorem 2.10].

Theorem 2.15. *Let $(X, \mathcal{M})$ be a measurable space.*

(a) If $f : X \to [0, \infty]$ is measurable, then there exists a sequence $\{\varphi_n\}$ of simple functions with $0 \leqslant \varphi_1 \leqslant \varphi_2 \leqslant \cdots \leqslant f$ and satisfying $\varphi_n \to f$ pointwise and $\varphi_n \to f$ uniformly on any set on which f is bounded.

(b) If $f : X \to \mathbb{C}$ is measurable, then there exists a sequence $\{\varphi_n\}$ of simple functions with $0 \leqslant |\varphi_1| \leqslant |\varphi_2| \leqslant \cdots \leqslant |f|$ and satisfying $\varphi_n \to f$ pointwise and $\varphi_n \to f$ uniformly on any set on which f is bounded.

With this result in hand, we can present a result which is closely related to the π-λ theorem.

Theorem 2.16 (monotone class theorem)**.** *Let X be a set and $\mathcal{P} \subset 2^X$ a π-system such that $X \in \mathcal{P}$. Let $\mathcal{H}$ be a collection of functions from X to $\mathbb{R}$ satisfying:*

(i) If $A \in \mathcal{P}$, then $1_A \in \mathcal{H}$.

(ii) If $f, g \in \mathcal{H}$ and $c \in \mathbb{R}$, then $f + g \in \mathcal{H}$ and $cf \in \mathcal{H}$.

(iii) If $f_n \in \mathcal{H}$ are nonnegative and there exists a bounded f such that $f_n \uparrow f$, then $f \in \mathcal{H}$.

Then $\mathcal{H}$ contains all bounded functions that are $\sigma(\mathcal{P})$-measurable.

Proof. Let $\mathcal{L} = \{E : 1_E \in \mathcal{H}\}$. Since $X \in \mathcal{P}$, (ii) and (iii) imply that $\mathcal{L}$ is a λ-system. Condition (i) implies that $\mathcal{P} \subset \mathcal{L}$. By the π-λ theorem, $\sigma(\mathcal{P}) \subset \mathcal{L}$. Therefore, by (ii), $\mathcal{H}$ contains all simple, $\sigma(\mathcal{P})$-measurable functions.

Let f be bounded and $\sigma(\mathcal{P})$-measurable. By considering f^+ and f^-, we may assume without loss of generality that f is nonnegative. Choose simple, $\sigma(\mathcal{P})$-measurable functions φ_n such that $0 \leqslant \varphi_n \uparrow f$. As above, each $\varphi_n \in \mathcal{H}$. And so by (iii), $f \in \mathcal{H}$. □

The last two results of this section deal with complete measure spaces. The proof of the first is an exercise. The second is [5, Proposition 2.12].

Proposition 2.17. *The following implications are valid if and only if the measure space $(X, \mathcal{M}, \mu)$ is complete.*

(a) If f is measurable and $f = g$ μ-a.e., then g is measurable.

(b) If f_n is measurable for each $n \in \mathbb{N}$ and $f_n \to f$ μ-a.e., then f is measurable.

Proof. Exercise 2.8. □

Proposition 2.18. *Let $(X, \mathcal{M}, \mu)$ be a measure space and $(X, \overline{\mathcal{M}}, \overline{\mu})$ its completion. If f is an $\overline{\mathcal{M}}$-measurable function, then there exists an $\mathcal{M}$-measurable function g such that $f = g$ $\overline{\mu}$-a.e.*

Proposition 2.11 and Theorem 2.15 together show that a real-valued function is measurable if and only if it can be written as the pointwise limit of simple functions. The restriction that these functions be real (or extended real, or complex) is not necessary. This is in fact true for any separable metric space.

If $(X, \mathcal{M})$ is a measurable space and (M, ρ) is a metric space, then an **M-valued simple function** is a measurable function $f : X \to M$ with a finite range.

Theorem 2.19. *If M is totally bounded and $f : X \to M$ is measurable, then there exists a sequence of M-valued simple functions φ_n such that $\varphi_n \to f$ uniformly.*

Proof. Let $n \in \mathbb{N}$. Choose $y_1, \ldots, y_{m(n)} \in M$ such that $\bigcup_{j=1}^{m(n)} B_{1/n}(y_j) \supset M$. Let

$$A_{j,n} = B_{1/n}(y_j) \setminus \bigcup_{k=1}^{j-1} B_{1/n}(y_k)$$

and $E_j = f^{-1}(A_{j,n}) \in \mathcal{M}$. Define $\varphi_n = \sum_{j=1}^{m(n)} y_j 1_{E_j}$.

Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $1/N < \varepsilon$. Let $n \geqslant \mathbb{N}$ and let $x \in X$. Choose $j \in \{1, \ldots, m(n)\}$ such that $f(x) \in A_{j,n}$. Then

$$\rho(f(x), y_j) = \rho(f(x), \varphi_n(x)) < 1/n < \varepsilon,$$

which shows that $\varphi_n \to f$ uniformly. □

Corollary 2.20. *If M is separable and $f : X \to M$ is measurable, then there exists a sequence of M-valued simple functions φ_n such that $\varphi_n \to f$ pointwise.*

Proof. By [13, p. 182], a metric space is separable if and only if it is homeomorphic to a totally bounded metric space. Let $(\widetilde{M}, \widetilde{\rho})$ be totally bounded and let $\psi : M \to \widetilde{M}$ be a homeomorphism. Choose $\widetilde{M}$-valued simple functions $\widetilde{\varphi}_n$ such that $\widetilde{\varphi}_n \to \psi \circ f$ uniformly. Let $\varphi_n = \psi^{-1} \circ \widetilde{\varphi}_n$. Then φ_n are M-valued simple functions with $\varphi_n \to f$ pointwise. □

Theorem 2.21. *Let M be separable and $f_n : X \to M$ a sequence of measurable functions. If $f_n \to f$ pointwise, then f is measurable.*

Proof. Let $\{q_k\}$ be a countable dense set in M. Let $U \subset M$ be open. Let $x \in X$. Suppose $f(x) \in U$. Choose $r \in \mathbb{Q} \cap (0, \infty)$ such that $B_{2r}(f(x)) \subset U$. Choose n such that $q_k \in B_{r/2}(f(x))$. Note that $B_r(q_k) \subset B_{3r/2}(f(x)) \subset U$. Also note that $f(x) \in B_{r/2}(q_k)$, so there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, we have $f_n(x) \in B_{r/2}(q_k)$.

Let $S = \{(r, k) \in \mathbb{Q} \times \mathbb{N} : r > 0, B_r(q_k) \subset U\}$. We have shown here that if $f(x) \in U$, then there exists $(r, k) \in S$ and $N \in \mathbb{N}$ such that for all $n \geqslant N$, we have $f_n(x) \in B_{r/2}(q_k)$. In other words,

$$f^{-1}(U) \subset \bigcup_{(r,k)\in S} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} f_n^{-1}(B_{r/2}(q_k)).$$

Conversely, if there exists $(r, k) \in S$ and $N \in \mathbb{N}$ such that for all $n \geqslant N$, we have $f_n(x) \in B_{r/2}(q_k)$, then $\rho(f(x), q_k) \leqslant r/2$, so $f(x) \in B_r(q_k) \subset U$. Hence,

$$f^{-1}(U) = \bigcup_{(r,k)\in S} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} f_n^{-1}(B_{r/2}(q_k)).$$

Since S is countable and each f_n is measurable, it follows that $f^{-1}(U) \in \mathcal{M}$, and so f is measurable. □

More generally, this last result is true when M is a second countable, regular topological space. See `https://math.stackexchange.com/q/2587155`.

It can sometimes be helpful to combine Theorem 2.21 with the following.

Theorem 2.22. *Let M be complete and separable, and $f_n : X \to M$ a sequence of measurable functions. Then $E = \{x \in X : \lim_{n\to\infty} f_n(x)$ exists$\}$ is measurable.*

Proof. If M is complete, then $x \in E$ if and only if $\{f_n(x)\}_{n=1}^{\infty}$ is Cauchy. In other words, $x \in E$ if and only if, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n, m \geqslant N$, we have $\rho(f_n(x), f_m(x)) < \varepsilon$. Since it suffices that this holds for ε of the form $1/k$, this means that

$$E = \bigcap_{k=1}^{\infty} \bigcup_{N\in\mathbb{N}} \bigcap_{n,m\geqslant N} \{x \in X : \rho(f_n(x), f_m(x)) < 1/k\}.$$

Define $g_{n,m} : X \to M \times M$ by $g_{n,m}(x) = (f_n(x), f_m(x))$, so that the above may be rewritten as

$$E = \bigcap_{k=1}^{\infty} \bigcup_{N \in \mathbb{N}} \bigcap_{n,m \geqslant N} (\rho \circ g_{n,m})^{-1}([0, 1/k)).$$

By Proposition 2.6, the functions $g_{n,m}$ are $(\mathcal{M}, \mathcal{B}_M \otimes \mathcal{B}_M)$-measurable. By Theorem 1.9, we have $\mathcal{B}_M \otimes \mathcal{B}_M = \mathcal{B}_{M \times M}$. In general, a metric is a continuous function, so it follows that $\rho : M \times M \to [0, \infty)$ is $\mathcal{B}_{M\times M}$-measurable. Thus, $\rho \circ g_{n,m}$ is measurable, which implies E is measurable. □

Exercises

2.1. Prove Proposition 2.1.

2.2. Complete the proof of Proposition 2.2.

2.3. Let f and g be measurable functions from a measurable space $(X, \mathcal{M})$ to $\mathbb{R}$. Let $E \in \mathcal{M}$ and define $h : X \to \mathbb{R}$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \in E, \\ g(x) & \text{if } x \in E^c. \end{cases}$$

Prove that h is measurable.

2.4. Prove Proposition 2.5.

2.5. Let X be a set and $(Y, \mathcal{N})$ a measurable space, with $\mathcal{N} = \sigma(\mathcal{E})$ for some $\mathcal{E} \subset 2^Y$. Let $f : X \to Y$ and show that $\sigma(f) = \sigma(\{f^{-1}(A) : A \in \mathcal{E}\})$.

2.6. Prove Proposition 2.10.

2.7. Prove Proposition 2.14.

2.8. Prove Proposition 2.17.

2.2 Integration of nonnegative functions

Let $(X, \mathcal{M}, \mu)$ be a measure space and let φ be a nonnegative simple function with standard representation $\varphi = \sum_{j=1}^{n} a_j 1_{E_j}$. Then we define the **integral** of φ with respect to μ to be

$$\int \varphi \, d\mu = \sum_{j=1}^{n} a_j \mu(E_j).$$

Note that $\int \varphi \, d\mu \in [0, \infty]$. When it helps to explicitly display the argument of φ, we will write

$$\int \varphi(x) \, \mu(dx) = \int \varphi \, d\mu.$$

Other notation for the integral includes $\int \varphi$ and $\int \varphi(x)\, d\mu(x)$.

If $A \in \mathcal{M}$, then $\varphi 1_A$ is also a nonnegative simple function, and we define

$$\int_A \varphi\, d\mu = \int \varphi 1_A\, d\mu.$$

Note, then, that $\int_X \varphi\, d\mu = \int \varphi\, d\mu$.

All of these notational conventions will also apply to integrals of more general functions, which we will be defining shortly.

If $f : X \to [0, \infty]$ is measurable, then we define

$$\int f\, d\mu = \sup \left\{ \int \varphi\, d\mu : \varphi \text{ is simple and } 0 \leqslant \varphi \leqslant f \right\}.$$

It can be shown that this definition agrees with the previous definition when f is simple.

The following results are all contained in [5, Section 2.2].

Proposition 2.23. *Let $f, g : X \to [0, \infty]$ be measurable.*

(a) If $c \geqslant 0$, then $\int cf\, d\mu = c \int f\, d\mu$.

(b) $\int (f + g)\, d\mu = \int f\, d\mu + \int g\, d\mu$.

(c) If $f \leqslant g$ a.e., then $\int f\, d\mu \leqslant \int g\, d\mu$.

(d) If $f = g$ a.e., then $\int f\, d\mu = \int g\, d\mu$.

(e) $\int f\, d\mu = 0$ if and only if $f = 0$ a.e.

(f) If $\int f\, d\mu < \infty$, then $f < \infty$ a.e.

Theorem 2.24 (Monotone convergence theorem)**.** *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions from X to $[0, \infty]$. Suppose that for each $n \in \mathbb{N}$, we have $f_n \leqslant f_{n+1}$ a.e. Also suppose $f_n \to f$ a.e. for some measurable function f. Then*

$$\lim_{n\to\infty} \int f_n\, d\mu = \int \lim_{n\to\infty} f_n\, d\mu = \int f\, d\mu.$$

Theorem 2.25 (Fatou's lemma)**.** *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions from X to $[0, \infty]$. Then*

$$\int \liminf_{n\to\infty} f_n\, d\mu \leqslant \liminf_{n\to\infty} \int f_n\, d\mu.$$

Remark 2.26. To remember the direction of the inequality in Fatou's lemma, keep in mind the following example. Define $f_n : \mathbb{R} \to [0, \infty]$ by $f_n = n 1_{(0,1/n)}$. Recall that λ denotes Lebesgue measure. Then $\int f_n\, d\lambda = 1$ for all n, but $\int \lim f_n\, d\lambda = 0$, since $f_n \to 0$ pointwise.

Exercises

2.9. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to [0, \infty]$ be measurable. Define $\nu : \mathcal{M} \to [0, \infty]$ by $\nu(E) = \int_E f \, d\mu$. Prove that ν is a measure on $(X, \mathcal{M})$ and that

$$\int g \, d\nu = \int gf \, d\mu, \tag{2.1}$$

for all measurable $g : X \to [0, \infty]$. (Hint: Use Theorem 2.15.)

Remark: We often use the shorthand $d\nu = f \, d\mu$ to indicate that ν is defined by $\nu(E) = \int_E f \, d\mu$. This shorthand also reminds of the above formula for transforming integrals.

2.10. Let $\{f_n\}$ be a sequence of $\mathcal{M}$-measurable functions from X to $[0, \infty]$. Let $f = \sum_{n=1}^{\infty} f_n$. Show that f is well-defined and measurable, and that

$$\sum_{n=1}^{\infty} \int f_n \, d\mu = \int \sum_{n=1}^{\infty} f_n \, d\mu = \int f \, d\mu.$$

2.11. Let $\{f_n\}$ be a sequence of $\mathcal{M}$-measurable functions from X to $[0, \infty]$. Assume for each $n \in \mathbb{N}$, we have $f_n \geqslant f_{n+1}$ a.e. Also assume $\int f_1 \, d\mu < \infty$. Prove that

$$\lim_{n\to\infty} \int f_n \, d\mu = \int \lim_{n\to\infty} f_n \, d\mu.$$

2.12. Let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable with $\int f \, d\mu < \infty$. Prove that for each $\varepsilon > 0$, there exists $E \in \mathcal{M}$ such that $\mu(E) < \infty$ and

$$\int_E f \, d\mu > \int f \, d\mu - \varepsilon.$$

2.3 General integration

Let $(X, \mathcal{M}, \mu)$ be a measure space. If $f : X \to \mathbb{R}^*$ is measurable, then f^+ and f^- are both measurable and nonnegative, and we define

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu,$$

provided at least one of the integrals on the right-hand is finite. We say that f is **integrable** if $\int |f| \, d\mu < \infty$. Since $|f| = f^+ + f^-$, we have that f is integrable if and only if $\int f \, d\mu$ exists and is real.

If $g : X \to \mathbb{C}$ is measurable, then we say g is integrable if $\int |g| \, d\mu < \infty$. Since $|g| \leqslant |\mathrm{Re} g| + |\mathrm{Im} g| \leqslant 2|g|$, it follows that g is integrable if and only if both $\mathrm{Re} g$ and $\mathrm{Im} g$ are integrable. In this case, we define

$$\int g \, d\mu = \int (\mathrm{Re} g) \, d\mu + i \int (\mathrm{Im} g) \, d\mu.$$

Note that if $f : X \to \mathbb{R}^*$ is integrable, then $|f| < \infty$ a.e., so there exists $g : X \to \mathbb{R}$ such that $f = g$ a.e. Since $\mathbb{R} \subset \mathbb{C}$, we can regard g as a map from X to $\mathbb{C}$. Therefore, when talking about integrable functions, we will generally assume they are complex-valued, unless otherwise specified.

More generally, if $E \in \mathcal{M}$, we say that f is **integrable on** E if $\int_E |f| \, d\mu < \infty$.

The proof of the following proposition can be found in [5, Section 2.3].

Proposition 2.27. *Let $f, g : X \to \mathbb{C}$ be integrable, and let $a, b \in \mathbb{C}$.*

(a) $af + bg$ is integrable and $\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu$.

(b) $|\int f \, d\mu| \leqslant \int |f| \, d\mu$.

(c) The following are equivalent:

(i) $\int_E f \, d\mu = \int_E g \, d\mu$ for all $E \in \mathcal{M}$.

(ii) $\int |f - g| \, d\mu = 0$.

(iii) $f = g$ a.e.

We can define an equivalence relation on the set of integrable functions $f : X \to \mathbb{C}$, where f and g are equivalent if $f = g$ a.e. We then define $L^1(X, \mathcal{M}, \mu)$ to be the set of equivalence classes under this relation. More specifically, if $f : X \to \mathbb{C}$ is integrable, then $[f] \in L^1(X, \mathcal{M}, \mu)$, where $[f] = \{g : f = g \text{ a.e.}\}$. By the previous proposition, the value of $\int_E g \, d\mu$ is the same for all $g \in [f]$. In other words, we can change a function on a null set, and this will not affect the integral of this function over any measurable set.

Instead of writing $L^1(X, \mathcal{M}, \mu)$, we will frequently abuse notation and drop one or more of X, $\mathcal{M}$, and μ. We will also abuse notation and write $f \in L^1(X, \mathcal{M}, \mu)$ when what is meant is $[f] \in L^1(X, \mathcal{M}, \mu)$.

It is easy to check that L^1 is a normed vector space over $\mathbb{C}$ with norm $\|[f]\|_1 = \int |f| \, d\mu$. (The aforementioned equivalence relation is needed to ensure that $\|[f]\|_1 = 0$ implies $[f] = [0]$.) Again, we usually abuse notation by writing $\|f\|_1 = \int |f| \, d\mu$. The norm on L^1 induces a metric, so that the distance between f and g in L^1 is $\|f - g\|_1$. Convergence in this metric is called **convergence in** L^1. That is, $f_n \to f$ in L^1 means that $\|f_n - f\|_1 \to 0$ as $n \to \infty$.

The following two theorems can be found in [5, Section 2.3].

Theorem 2.28 (The dominated convergence theorem)**.** *Let $\{f_n\}_{n=1}^\infty$ be a sequence in $L^1(X, \mathcal{M}, \mu)$. Suppose there exists a measurable function f such that $f_n \to f$ a.e. Also suppose there exists $g \in L^1$ such that for each $n \in \mathbb{N}$, we have $|f_n| \leqslant g$ a.e. Then $f \in L^1$ and*

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Theorem 2.29 (A generalized dominated convergence theorem)**.** *Let $\{f_n\}_{n=1}^\infty$ be a sequence in $L^1(X, \mathcal{M}, \mu)$. Suppose there exists $f \in L^1$ such that $f_n \to f$*

a.e. Also suppose there exist $g_n, g \in L^1$ such that $g_n \to g$ a.e., $\int g_n \, d\mu \to \int g \, d\mu$, and for each $n \in \mathbb{N}$, we have $|f_n| \leqslant g_n$ a.e. Then

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

The following result is part of [5, Theorem 2.26]. It is an example of the so-called Littlewood's second principle of real analysis: Every integrable function is nearly continuous.

Theorem 2.30. *Let μ_F be a Lebesgue-Stieltjes measure on $(\mathbb{R}, \mathcal{M}_F)$. Let $f \in L^1(\mu_F)$ and let $\varepsilon > 0$. Then there exists a continuous $g : \mathbb{R} \to \mathbb{C}$ that vanishes outside a bounded interval such that $\|f - g\|_1 < \varepsilon$.*

The next theorem is [5, Theorem 2.27]. It gives criteria for differentiation under the integral. The proof uses the dominated convergence theorem. Variations on this theorem can also be found in [2, Section A.5].

Theorem 2.31. *Let $(X, \mathcal{M}, \mu)$ be a measure space and $-\infty < a < b < \infty$. Let $f : X \times [a, b] \to \mathbb{C}$ and suppose that $f(\cdot, t)$ is integrable for each $t \in [a, b]$. Let*

$$F(t) = \int_X f(x, t) \, \mu(dx).$$

(a) *Suppose there exists $g \in L^1(\mu)$ such that $|f(x, t)| \leqslant g(x)$ for all x and t. If, for every $x \in X$, we have $f(x, t) \to f(x, t_0)$ as $t \to t_0$, then $F(t) \to F(t_0)$ as $t \to t_0$. In particular, if $f(x, \cdot)$ is continuous for each $x \in X$, then F is continuous.*

(b) *Suppose the partial derivative $\partial_t f(x, t)$ exists for all x and t. If there exists $g \in L^1(\mu)$ such that $|\partial_t f(x, t)| \leqslant g(x)$ for all x and t, then F is differentiable and*

$$F'(t) = \frac{d}{dt} \int_X f(x, t) \, \mu(dx) = \int_X \partial_t f(x, t) \, \mu(dx).$$

A special case of the following theorem can be found in [2, Theorem 1.6.9].

Theorem 2.32. *Let $(X, \mathcal{M}, \mu)$ be a measure space, and $(S, \mathcal{S})$ a measurable space. Let $h : X \to S$ be $(\mathcal{M}, \mathcal{S})$-measurable and let $f : S \to \mathbb{C}$ be $\mathcal{S}$-measurable. If either $f \circ h$ is nonnegative or $f \circ h \in L^1(\mu)$, then*

$$\int_X f \circ h \, d\mu = \int_S f \, d(\mu \circ h^{-1}).$$

Remark 2.33. Since $h : X \to S$ is $(\mathcal{M}, \mathcal{S})$-measurable, it follows that h^{-1} is a function from $\mathcal{S}$ to $\mathcal{M}$. Thus $\mu \circ h^{-1}$ is a function from $\mathcal{S}$ to $[0, \infty]$, and it can be shown that this function is a measure on $(S, \mathcal{S})$.

Proof of Theorem 2.32. Let us first assume that $f = 1_B$, where $B \in \mathcal{S}$. Then

$$\begin{aligned}\int_X f \circ h \, d\mu = \int_X 1_B(h(x)) \, \mu(dx) &= \int_X 1_{\{y : h(y) \in B\}}(x) \, \mu(dx) \\ &= \int_X 1_{h^{-1}(B)} \, d\mu = \mu(h^{-1}(B)).\end{aligned}$$

On the other hand,

$$\int_S f \, d(\mu \circ h^{-1}) = \int_S 1_B \, d(\mu \circ h^{-1}) = (\mu \circ h^{-1})(B) = \mu(h^{-1}(B)),$$

and so the result holds. By linearity, the result also holds whenever f is a simple function.

Now suppose $f \circ h$ is nonnegative. In this case, $f \circ h = f^+ \circ h$, so without loss of generality, we may assume f is nonnegative. Choose simple φ_n such that $0 \leqslant \varphi_n \uparrow f$ pointwise. Then $0 \leqslant \varphi_n \circ h \uparrow f \circ h$ pointwise, and so by monotone convergence,

$$\int_X f \circ h \, d\mu = \lim_{n \to \infty} \int_X \varphi_n \circ h \, d\mu = \lim_{n \to \infty} \int_S \varphi_n \, d(\mu \circ h^{-1}) = \int_S f \, d(\mu \circ h^{-1}),$$

and the result holds for all $f \geqslant 0$.

Finally, assume $f \circ h \in L^1(\mu)$. Since $\operatorname{Re}(f \circ h) = (\operatorname{Re} f) \circ h$ and $\operatorname{Im}(f \circ h) = (\operatorname{Im} f) \circ h$, we may assume without loss of generality that f is real-valued. In this case, we have that

$$\int_X f \circ h \, d\mu = \int_X (f \circ h)^+ \, d\mu - \int_X (f \circ h)^- \, d\mu,$$

and both integrals are finite. Since $(f \circ h)^+ = f^+ \circ h$ and $(f \circ h)^- = f^- \circ h$, it follows from the result for nonnegative functions that

$$\int_X f \circ h \, d\mu = \int_S f^+ \, d(\mu \circ h^{-1}) - \int_S f^- \, d(\mu \circ h^{-1}),$$

and both integrals are finite. Thus, $f \in L^1(\mu \circ h^{-1})$, and

$$\int_X f \circ h \, d\mu = \int_S f \, d(\mu \circ h^{-1}),$$

which finishes the proof. □

Remark 2.34. The technique used in the above proof is extremely common in the theory of measure and integration. The result is proved in four stages. First, it is proved for indicator functions, then for simple functions by linearity, then for nonnegative functions by monotone convergence, and finally for integrable functions by considering the positive and negative parts. This proof technique is a reflection of the manner in which the integral is defined.

If $f \in L^1(\mathbb{R}, \mathcal{L}, \lambda)$, then $\int f \, d\lambda$ is called the **Lebesgue integral** of f. The Lebesgue integral is a generalization of the Riemann integral. If a function f is Riemann integrable on a bounded interval $[a, b]$, then $f1_{[a,b]} \in L^1(\lambda)$ and $\int_{[a,b]} f \, d\lambda$ agrees with the value of the Riemann integral of f over $[a, b]$. Since $\lambda(\{a\}) = 0$, it does not matter whether we integrate over $[a, b]$ or $(a, b]$. We will henceforth adopt the notation that

$$\int_a^b f(x) \, dx = \int_{(a,b]} f \, d\lambda.$$

While it is true that the Lebesgue integral can handle a larger class of integrands than the Riemann integral, the real power of the Lebesgue integral comes from its associated convergence theorems: monotone convergence, Fatou's lemma, and dominated convergence. For a detailed discussion of the connection between the Riemann and Lebesgue integrals, see [5, Section 2.3].

Remark 2.35. Suppose μ is counting measure on $(\mathbb{N}, 2^{\mathbb{N}})$ and $f \in L^1(\mu)$. Then $f : \mathbb{N} \to \mathbb{C}$ and $\int |f| \, d\mu < \infty$. If we write f_n instead of $f(n)$, then we have $\int |f| \, d\mu = \sum_{n=1}^\infty |f_n|$ and $\int f \, d\mu = \sum_{n=1}^\infty f_n$. In other words, $L^1(\mu)$ is the space of absolutely convergent series, and the integral is just the sum. In this way, we can apply monotone convergence, Fatou's lemma, and dominated convergence to infinite series.

Remark 2.36. Let $\{x_\alpha\}_{\alpha \in A}$ be a (possibly uncountable) collection of extended real numbers. Suppose $x_\alpha \geqslant 0$ for all α. Then we define

$$\sum_{\alpha \in A} x_\alpha = \sup \left\{ \sum_{\alpha \in F} x_\alpha : F \subset A \text{ and } F \text{ is finite} \right\}.$$

It can be shown that if $S = \{\alpha \in A : x_\alpha > 0\}$ is uncountable, then $\sum_{\alpha \in A} x_\alpha = \infty$. Of course, if S if finite, then $\sum_{\alpha \in A} x_\alpha = \sum_{\alpha \in S} x_\alpha$. On the other hand, if S is countably infinite, and $g : \mathbb{N} \to S$ is any bijection, then

$$\sum_{\alpha \in A} x_\alpha = \lim_{n \to \infty} \sum_{j=1}^n x_{g(j)}.$$

In other words, this definition of summation agrees with the usual definition of an infinite series.

Let us now drop the assumption that each x_α is nonnegative, and assume instead that $\sum_{\alpha \in A} |x_\alpha| < \infty$. Then $S = \{\alpha \in A : x_\alpha \neq 0\}$ is countable. If S is finite, then we define $\sum_{\alpha \in A} x_\alpha = \sum_{\alpha \in S} x_\alpha$. Assume S is countably infinite and let $g : \mathbb{N} \to S$ be a bijection. Then the series $\sum_{j=1}^\infty x_{g(j)}$ is absolutely convergent, and its sum does not depend on g. We can therefore define $\sum_{\alpha \in A} x_\alpha = \sum_{j=1}^\infty x_{g(j)}$.

Let X be any set, let μ be counting measure on $(X, 2^X)$, and let $f : X \to \mathbb{C}$. Then f is μ-integrable if and only if $\sum_{x \in X} |f(x)| < \infty$, and in this case, we have $\int f \, d\mu = \sum_{x \in X} f(x)$.

Remark 2.37. There are a number of wonderful exercises at the end of [5, Section 2.3]. Some involve proving general, abstract results, and some involve calculations with specific integrals and series. These exercises are an excellent resource for reviewing.

2.4 Modes of convergence

Let $(X, \mathcal{M}, \mu)$ be a measure space, $\{f_n\}$ a sequence of $\mathbb{C}$-valued functions on X, and $f : X \to \mathbb{C}$.

As usual, we say $f_n \to f$ uniformly if $\sup_{x \in X} |f_n(x) - f(x)| \to 0$ as $n \to \infty$. As defined earlier, $f_n \to f$ pointwise if, for all $x \in X$, we have $f_n(x) \to f(x)$ as $n \to \infty$. We also saw earlier that $f_n \to f$ a.e. if there exists a null set N such that, for all $x \in N^c$, we have $f_n(x) \to f(x)$.

Uniform convergence implies pointwise convergence, and pointwise convergence implies a.e. convergence.

We also saw that $f_n \to f$ in L^1 if $\int |f_n - f| \, d\mu \to 0$ as $n \to \infty$. Another important mode of convergence is convergence in measure. We say that $f_n \to f$ **in measure** as $n \to \infty$ if, for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \mu(\{x : |f_n(x) - f(x)| \geqslant \varepsilon\}) = 0.$$

The following results can be found in [5, Section 2.4].

Proposition 2.38. *If $f_n \to f$ in L^1, then $f_n \to f$ in measure.*

Proposition 2.39. *If $f_n \to f$ in measure, then there exists a subsequence $\{f_{n_j}\}$ such that $f_{n_j} \to f$ a.e.*

Proposition 2.40. *If $\mu(X) < \infty$ and $f_n \to f$ a.e., then $f_n \to f$ in measure.*

Exercises

2.13. Let $(X, \mathcal{M}, \mu)$ be a finite measure space and let $L^0(X, \mathcal{M}, \mu)$ denote the space of measurable $f : X \to \mathbb{C}$. (As with L^1, this space actually consists of equivalence classes, where two functions are equivalent if they are equal a.e.) For $f, g \in L^0$, define

$$\rho(f, g) = \int \frac{|f - g|}{1 + |f - g|} \, d\mu.$$

Prove that ρ is a metric on L^0 and $f_n \to f$ in this metric if and only if $f_n \to f$ in measure.

2.14. Suppose $f_n \to f$ in measure and $g_n \to g$ in measure.

(a) Prove that $f_n + g_n \to f + g$ in measure.

(b) Prove that if $\mu(X) < \infty$, then $f_n g_n \to fg$ in measure.

(c) Give an example to show that the conclusion of (b) can be false when $\mu(X) = \infty$.

2.15. Let $f \in L^1(X, \mathcal{M}, \mu)$. Prove that for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\int_E |f|\, d\mu < \varepsilon,$$

whenever $\mu(E) < \delta$.

2.5 Useful inequalities

Let $a, b \in \mathbb{R}^*$ with $a < b$. A function $\varphi : (a, b) \to \mathbb{R}$ is **convex** if

$$\varphi(\lambda x + (1-\lambda)y) \leqslant \lambda\varphi(x) + (1-\lambda)\varphi(y),$$

for all $\lambda \in (0,1)$ and $x, y \in (a, b)$.

The following is [2, Theorem 1.5.1], as well as [5, Exercise 3.42(d)].

Theorem 2.41 (Jensen's inequality)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) = 1$. Suppose $f : X \to (a, b)$ is integrable, and $\varphi : (a, b) \to \mathbb{R}$ is convex. Then $\int \varphi \circ f\, d\mu$ exists, and*

$$\varphi\left(\int f\, d\mu\right) \leqslant \int \varphi \circ f\, d\mu.$$

Corollary 2.42. *Let $(X, \mathcal{M}, \mu)$ be a finite, nontrivial measure space. Suppose $f : X \to (a, b)$ is integrable, and $\varphi : (a, b) \to \mathbb{R}$ is convex. Let $c = \mu(X) \in (0, \infty)$. Then*

$$\varphi\left(\int f\, d\mu\right) \leqslant \frac{1}{c}\int \varphi(cf(x))\, \mu(dx),$$

where the integral on the right-hand side exists.

Proof. Let $\nu = c^{-1}\mu$ and $\psi(x) = \varphi(cx)$. Then $\nu(X) = 1$ and ψ is convex, so by Jensen's inequality,

$$\varphi\left(\int f\, d\mu\right) = \psi\left(\int f\, d\nu\right) \leqslant \int \psi(f(x))\, \nu(dx) = \frac{1}{c}\int \varphi(cf(x))\, \mu(dx),$$

where the integral on the right-hand side exists. □

Let $(X, \mathcal{M}, \mu)$ be a measure space. If $f : X \to \mathbb{C}$ is measurable and $p \in (0, \infty)$, then we define

$$\|f\|_p = \left(\int |f|^p\, d\mu\right)^{1/p}.$$

We let $L^p(X, \mathcal{M}, \mu)$ denote the set of all functions f such that $\|f\|_p < \infty$. As with L^1, we identify any two functions that are equal almost everywhere.

When dealing with L^p spaces, it is often helpful to remember the inequality,

$$|a+b|^p \leqslant (2(|a| \vee |b|))^p = 2^p(|a|^p \vee |b|^p) \leqslant 2^p(|a|^p + |b|^p).$$

The following is [2, Exercise 1.5.3], as well as [5, Theorem 6.5].

Theorem 2.43 (Minkowski's inequality)**.** *If $p \in [1,\infty)$ and $f,g \in L^p$, then*

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p.$$

Using Minkowski's inequality, it is easy to check that $\|\cdot\|_p$ is a norm on L^p when $p \geqslant 1$. The norm on L^p induces a metric, so that the distance between f and g in L^p is $\|f-g\|_p$. Convergence in this metric is called **convergence in** L^p. That is, $f_n \to f$ in L^p means that $\|f_n - f\|_p \to 0$ as $n \to \infty$. By [5, Exercise 6.9], if $f_n \to f$ in L^p for some $p \in [1,\infty)$, then $f_n \to f$ in measure.

By induction, Minkowski's inequality extends to finite sums, so that

$$\left\| \sum_{j=1}^n f_j \right\|_p \leqslant \sum_{j=1}^n \|f_j\|_p,$$

whenever $p \in [1,\infty)$ and $f_j \in L^p$. It is frequently useful to note, however, that it also extends to integrals.

The following is [5, Theorem 6.19].

Theorem 2.44 (Minkowski's inequality for integrals)**.** *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ-finite measure spaces.*

(a) If $f : X \times Y \to [0,\infty]$ is $\mathcal{M}\otimes\mathcal{N}$-measurable and $p \in [1,\infty)$, then

$$\left\| \int f(\cdot,y)\,\nu(dy) \right\|_p \leqslant \int \|f(\cdot,y)\|_p\,\nu(dy). \tag{2.2}$$

(b) Let $f : X \times Y \to \mathbb{C}$ be $\mathcal{M}\otimes\mathcal{N}$-measurable and $p \in [1,\infty)$. Assume that $f(\cdot,y) \in L^p(\mu)$ for ν-a.e. y, and that $y \mapsto \|f(\cdot,y)\|_p$ is in $L^1(\nu)$. Then $f(x,\cdot) \in L^p(\nu)$ for μ-a.e. x, $\int f(\cdot,y)\nu(dy) \in L^p(\mu)$, and (2.2) holds.

Remark 2.45. Note that

$$\left(\int \left| \int f(x,y)\,\nu(dy) \right|^p \mu(dx) \right)^{1/p} \leqslant \int \left(\int |f(x,y)|^p\,\mu(dx) \right)^{1/p} \nu(dy)$$

is equivalent to (2.2).

If $f : X \to \mathbb{C}$ is measurable, then we define the **essential supremum** of f to be

$$\|f\|_\infty = \inf\{M \geqslant 0 : |f| \leqslant M \text{ a.e.}\}.$$

It can be shown that the infimum is actually obtained, that is, one can show that $|f| \leqslant \|f\|_\infty$ a.e. We let $L^\infty(X,\mathcal{M},\mu)$ denote the set of all functions f such that $\|f\|_\infty < \infty$. As with L^1, we identify any two functions that are equal almost everywhere.

A function $f \in L^\infty$ need not be bounded. But if $f \in L^\infty$, then there exists a bounded, measurable g such that $f = g$ a.e. For example, we can take $g = 1_E f$, where $E = \{x : |f(x)| \leqslant \|f\|_\infty\}$.

By [5, Theorem 6.8], $\|\cdot\|_\infty$ is a norm on L^∞. The norm on L^∞ induces a metric, so that the distance between f and g in L^∞ is $\|f-g\|_\infty$. Convergence in this metric is called **convergence in** L^∞. That is, $f_n \to f$ in L^∞ means that $\|f_n - f\|_\infty \to 0$ as $n \to \infty$.

Also by [5, Theorem 6.8], we have $f_n \to f$ in L^∞ if and only if there exists $N \in \mathcal{M}$ such that $\mu(N) = 0$ and $f_n \to f$ uniformly on N^c.

If $p, q \in (1, \infty)$ and $1/p + 1/q = 1$, then p and q are **conjugate exponents**. In addition, we say that 1 and ∞ are conjugate exponents. Note that each $p \in [1, \infty]$ has a unique conjugate exponent. Also note that $p = 2$ is its own conjugate exponent.

The following is a slight generalization of [2, Theorem 1.5.2], as well as a combination of [5, Theorems 6.2 and 6.8].

Theorem 2.46 (Hölder's inequality)**.** *Let $p \in [1, \infty]$ and let q be its conjugate exponent. If $f, g : X \to \mathbb{C}$ are measurable, then $\|fg\|_1 \leqslant \|f\|_p \|g\|_q$.*

Remark 2.47. The case $p = 2$ is called the **Cauchy-Schwarz inequality**.

Exercises

2.16. Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) = 1$, and let $f : X \to \mathbb{C}$ be measurable. Show that if $0 < p < q \leqslant \infty$, then $\|f\|_p \leqslant \|f\|_q$.

2.17. Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) = 1$, and let $f : X \to \mathbb{C}$ be measurable. Show that $\|f\|_p \to \|f\|_\infty$ as $p \to \infty$.

2.18. Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) = 1$, and suppose $\|f\|_p < \infty$ for some $p > 0$. Prove the following.

(a) $\int \log|f|\, d\mu \leqslant \log \|f\|_q$ for all $q \in (0, p)$.

(b) $\log \|f\|_q \leqslant q^{-1}(\int |f|^q\, d\mu - 1)$ for all $q \in (0, p)$.

(c) $q^{-1}(\int |f|^q\, d\mu - 1) \to \int \log|f|\, d\mu$ as $q \to 0$.

(d) $\|f\|_q \to \exp(\int \log|f|\, d\mu)$ as $q \to 0$.

2.6 Product measures

Let $\{(X_j, \mathcal{M}_j, \mu_j)\}_{j=1}^n$ be σ-finite measure spaces. Let $X = \prod_{j=1}^n X_j$ and $\mathcal{M} = \bigotimes_{j=1}^n \mathcal{M}_j$. Then there exists a unique measure $\mu = \mu_1 \times \cdots \times \mu_n$ on $(X, \mathcal{M})$ such that

$$\mu(A_1 \times \cdots \times A_n) = \prod_{j=1}^n \mu_j(A_j),$$

for all $A_j \in \mathcal{M}_j$. Moreover, the measure space $(X, \mathcal{M}, \mu)$ is σ-finite. The measure μ is called the **product measure**.

The main results concerning integration with respect to product measures are the theorems of Tonelli and Fubini. The construction of the product measure and the proof of Fubini-Tonelli can be found in [5, Section 2.5], which was the primary source for this section of the notes. This material is also found in [2, Section 1.7].

Theorem 2.48. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces.*

(a) (Tonelli) *Suppose $f : X \times Y \to [0, \infty]$ is $\mathcal{M} \otimes \mathcal{N}$-measurable. Then $f(x, \cdot)$ is $\mathcal{N}$-measurable for all $x \in X$ and the function $x \mapsto g(x) = \int_Y f(x, \cdot)\, d\nu$ is nonnegative and $\mathcal{M}$-measurable. Similarly, $f(\cdot, y)$ is $\mathcal{M}$-measurable for all $y \in Y$ and the function $y \mapsto h(y) = \int_X f(\cdot, y)\, d\mu$ is nonnegative and $\mathcal{N}$-measurable. Finally,*

$$\int_{X\times Y} f\, d(\mu \times \nu) = \int_X g\, d\mu = \int_Y h\, d\nu.$$

(b) (Fubini) *Suppose $f \in L^1(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$. Then $f(x, \cdot) \in L^1(Y, \nu)$ for μ-a.e. $x \in X$ and the a.e.-defined function $x \mapsto g(x) = \int_Y f(x, \cdot)\, d\nu$ is in $L^1(X, \mu)$. Similarly, $f(\cdot, y) \in L^1(X, \mu)$ for ν-a.e. $y \in Y$ and the a.e.-defined function $y \mapsto h(y) = \int_X f(\cdot, y)\, d\mu$ is in $L^1(Y, \nu)$. Finally,*

$$\int_{X\times Y} f\, d(\mu \times \nu) = \int_X g\, d\mu = \int_Y h\, d\nu.$$

Remark 2.49. Notice that

$$\begin{aligned}\int_X g\, d\mu &= \int_X g(x)\, \mu(dx)\\ &= \int_X \left[\int_Y f(x, \cdot)\, d\nu\right] \mu(dx) = \int_X \left[\int_Y f(x, y)\, \nu(dy)\right] \mu(dx).\end{aligned}$$

Similarly,

$$\int_Y h\, d\nu = \int_Y \left[\int_X f(x, y)\, \mu(dx)\right] \nu(dy).$$

Hence, the theorems of Tonelli and Fubini are saying that the integral over the product space can be computed as an iterated integral in either order.

Remark 2.50. The theorems of Tonelli and Fubini are typically used in tandem. For example, before one can use Fubini's theorem, one must know that $\int_{X\times Y} |f|\, d(\mu \times \nu) < \infty$. But $|f|$ is nonnegative, so one can use Tonelli's theorem to verify this.

The measure space $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$ is typically not complete. It is therefore often desirable to work with its completion, which we shall denote by $(X \times Y, \mathcal{O}, \varpi)$. In that case, however, one usually encounters functions $f(x, y)$ that are only $\mathcal{O}$-measurable, and not $\mathcal{M}\otimes\mathcal{N}$-measurable. Therefore, the version of Fubini-Tonelli given above will not apply. Instead, one can use the following, which is [5, Theorem 2.39].

Theorem 2.51 (Fubini-Tonelli for complete measures)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be complete, σ-finite measure spaces, and let $(X \times Y, \mathcal{O}, \varpi)$ be the completion of $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$.*

(a) (Tonelli) *Suppose $f : X \times Y \to [0, \infty]$ is $\mathcal{O}$-measurable. Then $f(x, \cdot)$ is $\mathcal{N}$-measurable for μ-a.e. $x \in X$ and the a.e.-defined function $x \mapsto g(x) = \int_Y f(x, \cdot)\, d\nu$ is nonnegative and $\mathcal{M}$-measurable. Similarly, $f(\cdot, y)$ is $\mathcal{M}$-measurable for ν-a.e. $y \in Y$ and the a.e.-defined function $y \mapsto h(y) = \int_X f(\cdot, y)\, d\mu$ is nonnegative and $\mathcal{N}$-measurable. Finally,*

$$\int_{X \times Y} f\, d\varpi = \int_X g\, d\mu = \int_Y h\, d\nu.$$

(b) (Fubini) *Suppose $f \in L^1(X \times Y, \mathcal{O}, \varpi)$. Then $f(x, \cdot) \in L^1(Y, \nu)$ for μ-a.e. $x \in X$ and the a.e.-defined function $x \mapsto g(x) = \int_Y f(x, \cdot)\, d\nu$ is in $L^1(X, \mu)$. Similarly, $f(\cdot, y) \in L^1(X, \mu)$ for ν-a.e. $y \in Y$ and the a.e.-defined function $y \mapsto h(y) = \int_X f(\cdot, y)\, d\mu$ is in $L^1(Y, \nu)$. Finally,*

$$\int_{X \times Y} f\, d\varpi = \int_X g\, d\mu = \int_Y h\, d\nu.$$

The above treatment of product measures, which comes from [5], concerns only product measures on finite product spaces. This is insufficient for probability theory, so we consider one final topic in this section: Kolmogorov's extension theorem.

Let $\mathbb{R}^\infty = \prod_{j=1}^\infty \mathbb{R}$ and $\mathcal{R}^\infty = \bigotimes_{j=1}^\infty \mathcal{R}$. Kolmogorov's extension theorem is concerned with the existence of probability measures on $(\mathbb{R}^\infty, \mathcal{R}^\infty)$. Before we can state the theorem, we need a piece of terminology.

For each $n \in \mathbb{N}$, let μ_n be a probability measure on $(\mathbb{R}^n, \mathcal{R}^n)$. We say that the measures $\{\mu_n\}_{n=1}^\infty$ are **consistent** if, for all $n \in \mathbb{N}$,

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]),$$

whenever $-\infty \leqslant a_j < b_j \leqslant \infty$.

The following is [2, Theorem A.3.1].

Theorem 2.52 (Kolmogorov's extension theorem)**.** *For each $n \in \mathbb{N}$, let μ_n be a probability measure on $(\mathbb{R}^n, \mathcal{R}^n)$. If $\{\mu_n\}_{n=1}^\infty$ are consistent, then there exists a unique probability measure μ on $(\mathbb{R}^\infty, \mathcal{R}^\infty)$ such that*

$$\mu(\{\omega : \omega_j \in (a_j, b_j] \text{ for } 1 \leqslant j \leqslant n\}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]),$$

for all $n \in \mathbb{N}$ and all $-\infty \leqslant a_j < b_j \leqslant \infty$.

Exercises

2.19. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and let $f : X \to [0, \infty)$ be measurable. The region under the graph of f is

$$G_f = \{(x, y) \in X \times [0, \infty) : y < f(x)\}.$$

(a) Show that $G_f \in \mathcal{M} \otimes \mathcal{R}$.

Hint: Let $g : X \times [0, \infty) \to [0, \infty)^2$ be given by $g(x, y) = (f(x), y)$ and let $h : [0, \infty)^2 \to \mathbb{R}$ be given by $h(u, y) = u - y$. Then $(h \circ g)(x, y) = f(x) - y$ and $G_f = (h \circ g)^{-1}((0, \infty))$.

(b) Show that the integral of f is the area under its graph. That is, show that

$$\int_X f\, d\mu = (\mu \times \lambda)(G_f).$$

(c) Show that

$$(\mu \times \lambda)(G_f) = \int_0^\infty \mu(\{x : f(x) > y\})\, dy.$$

2.7 Lebesgue integration on $\mathbb{R}^n$

Recall that $(\mathbb{R}, \mathcal{L}, \overline{\lambda})$ is the Lebesgue measure space. It is the completion of $(\mathbb{R}, \mathcal{R}, \lambda)$, where λ is the unique measure on $(\mathbb{R}, \mathcal{R})$ such that $\lambda((a, b]) = b - a$.

The Lebesgue measure space on $\mathbb{R}^n$ is the completion of the product of $(\mathbb{R}, \mathcal{R}, \lambda)$ with itself n-times. More specifically, consider the following. Note that $\mathbb{R}^n = \prod_{j=1}^n \mathbb{R}$. Also recall that $\mathcal{R}^n = \bigotimes_{j=1}^n \mathcal{B}_\mathbb{R} = \mathcal{B}_{\mathbb{R}^n}$. Using the results from the previous section, we can now define the product measure, $\lambda^n = \prod_{j=1}^n \lambda$, giving us the product measure space, $(\mathbb{R}^n, \mathcal{R}^n, \lambda^n)$. Let us denote its completion by $(\mathbb{R}^n, \mathcal{L}_{\mathbb{R}^n}, \overline{\lambda^n})$. The measure $\overline{\lambda^n}$ is **Lebesgue measure** on $\mathbb{R}^n$, the σ-algebra $\mathcal{L}_{\mathbb{R}^n}$ is the **Lebesgue σ-algebra** on $\mathbb{R}^n$, and sets $E \in \mathcal{L}_{\mathbb{R}^n}$ are **Lebesgue measurable** subset of $\mathbb{R}^n$. It can be shown that $(\mathbb{R}^n, \mathcal{L}_{\mathbb{R}^n}, \overline{\lambda^n})$ is also the completion of the product measure space $(\mathbb{R}^n, \bigotimes_{j=1}^n \mathcal{L}, \overline{\lambda}^n)$.

As usual, we will typically omit the bar, and use λ^n for both Lebesgue measure on $\mathcal{L}_{\mathbb{R}^n}$ and Lebesgue measure on $\mathcal{R}^n$. Also, when there is little chance of confusion, we will often write λ instead of λ^n and $\mathcal{L}$ instead of $\mathcal{L}_{\mathbb{R}^n}$. We will also write $\int f(x)\, dx$ to mean $\int f\, d\lambda^n$, although in the case $n > 1$, one should remember that x is a vector.

The following is [5, Proposition 2.40]. The first part expresses a regularity property of Lebesgue measure, and is an extension of Theorem 1.25. The second part is an instantiation of Littlewood's first principle, and is an extension of Theorem 1.26.

Proposition 2.53. *Suppose $E \in \mathcal{L}_{\mathbb{R}^n}$. Then:*

(a) $\lambda(E) = \inf\{\lambda(U) : E \subset U, U \text{ open}\} = \sup\{\lambda(K) : K \subset E, K \text{ compact}\}$.

(b) If $\lambda(E) < \infty$, then for any $\varepsilon > 0$, there exist disjoint sets $\{R_j\}_{j=1}^N$ with $R_j = \prod_{i=1}^n (a_{i,j}, b_{i,j}]$ and $\lambda(E \,\triangle\, (\bigcup_{j=1}^N R_j)) < \varepsilon$.

Now let $U \subset \mathbb{R}^n$ be open. For $j \in \{1, \dots, n\}$, let $g_j : U \to \mathbb{R}$ and let $G : U \to \mathbb{R}^n$ be given by $G(x) = (g_1(x), \dots, g_n(x))$. Suppose $G \in C^1(U)$, meaning that $\partial_j g_i$ exists and is continuous for each i and j. For $x \in U$, let

$D_xG : \mathbb{R}^n \to \mathbb{R}^n$ be the linear map whose matrix representation in the standard basis is $[D_xG] = [\partial_j g_i(x)]$. Note that if G is linear, then $D_xG = G$ for all $x \in U$.

The function G is called a C^1 **diffeomorphism** if G is injective and D_xG is invertible for all $x \in U$. If G is a C^1 diffeomorphism, then it can be shown that $G^{-1} : G(U) \to U$ is a C^1 diffeomorphism and $D_x(G^{-1}) = (D_{G^{-1}(x)}G)^{-1}$ for all $x \in G(U)$.

The following is [5, Theorem 2.47]. It states that the usual change of variable formula for integrals on $\mathbb{R}^n$ extends to Lebesgue measure.

Theorem 2.54. *Suppose $U \subset \mathbb{R}^n$ is open and $G : U \to \mathbb{R}^n$ is a C^1 diffeomorphism. If f is a function on $G(U)$ that is Lebesgue measurable, then $f \circ G$ is Lebesgue measurable. If f is $[0,\infty]$-valued or $f \in L^1(G(U), \lambda)$, then*

$$\int_{G(U)} f(x)\,dx = \int_U (f \circ G)(x) |\det[D_xG]|\,dx.$$

Example 2.55. Suppose $U = \mathbb{R}^n$ and G is an affine linear function. Then $G(x) = v + Ax$ for some linear function A and some $v \in \mathbb{R}^n$. Suppose further that A is invertible. Then $G(\mathbb{R}^n) = \mathbb{R}^n$ and $[D_xG] = [A]$ for all $x \in \mathbb{R}^n$. Thus,

$$\int_{\mathbb{R}^n} f(x)\,dx = |\det[A]| \int_{\mathbb{R}^n} f(v + Ax)\,dx,$$

whenever f is nonnegative or integrable on $\mathbb{R}^n$. In particular, if $f = 1_E$ for some $E \in \mathcal{L}_{\mathbb{R}^n}$, then

$$\lambda(E) = |\det A| \lambda(\{x : v + Ax \in E\}).$$

Here are three examples of this:

(i) If $A = I$, then the result simply expresses the fact that Lebesgue measure is translation invariant.

(ii) If $v = 0$ and $A = r^{-1}I$ for some $r \neq 0$, then we obtain $\lambda(rE) = |r|^n \lambda(E)$.

(iii) If $v = 0$ and A is a rotation, then the result states that Lebesgue measure is rotationally invariant.

Exercises

2.20. Show that $e^{-xy} \sin x$ is integrable with respect to λ^2 on the strip $0 < x < a$, $0 < y$. Use Fubini's theorem to show that

$$\int_0^a \frac{\sin x}{x}\,dx = \frac{\pi}{2} - (\cos a) \int_0^\infty \frac{e^{-ay}}{1 + y^2}\,dy - (\sin a) \int_0^\infty \frac{ye^{-ay}}{1 + y^2}\,dy,$$

and replace $1 + y^2$ by 1 to conclude $|\int_0^a (\sin x)/x\,dx - \pi/2| \leqslant 2/a$ for $a \geqslant 1$.

2.21. Let $E = [0,1] \times [0,1]$. Investigate the existence and equality of $\int_E f\,d\lambda^2$, $\int_0^1 \int_0^1 f(x,y)\,dx\,dy$, and $\int_0^1 \int_0^1 f(x,y)\,dy\,dx$ for the following f:

(a) $f(x,y) = (x^2 - y^2)(x^2 + y^2)^{-1}$.

(b) $f(x,y) = (1 - xy)^{-a}$, where $a > 0$.

(c) $f(x,y) = (x - 1/2)^{-3} 1_{\{(x,y):0<y<|x-1/2|\}}$.

2.22. Let $a > 0$ and suppose f is Lebesgue integrable on $(0,a)$. For $x \in (0,a)$, define $g(x) = \int_x^a t^{-1} f(t)\,dt$. Show that g is integrable on $(0,a)$ and that $\int_0^a g(x)\,dx = \int_0^a f(x)\,dx$.

Chapter 3

Signed Measures and Integration

3.1 Signed measures

The primary source for the material in this section, as well as Section 3.2, is [5]. However, much of this material is also covered in [2, Section A.4].

Let $(X, \mathcal{M})$ be a measurable space and $\nu : \mathcal{M} \to \mathbb{R}^*$. Then ν is a **signed measure** on $(X, \mathcal{M})$ if:

(i) $\nu(\varnothing) = 0$,

(ii) $\nu(\mathcal{M}) \subset (-\infty, \infty]$ or $\nu(\mathcal{M}) \subset [-\infty, \infty)$, and

(iii) if $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$ are disjoint, then $\nu(\biguplus_j E_j) = \sum_j \nu(E_j)$, where this sum converges absolutely whenever $\nu(\biguplus_j E_j) < \infty$.

Note that every measure is a signed measure. When helpful for clarity, we will sometimes refer to a measure as a **positive measure**.

Example 3.1. Suppose μ_1 and μ_2 are measures on a measurable space $(X, \mathcal{M})$. If at least one of these is a finite measure, then $\nu = \mu_1 - \mu_2$ is a signed measure on $(X, \mathcal{M})$.

Example 3.2. Let $(X, \mathcal{M}, \mu)$ be a measure space, and let $f : X \to \mathbb{R}^*$ be measurable. Suppose $\int f \, d\mu$ exists. Then one can check that $\nu(E) = \int_E f \, d\mu$ defines a signed measure on $(X, \mathcal{M})$. As in Exercise 2.9, we often use the shorthand notation $d\nu = f \, d\mu$ to indicate that ν is defined in this fashion.

Let ν be a signed measure on $(X, \mathcal{M})$. A set $E \in \mathcal{M}$ is **positive** for ν if $\nu(F) \geqslant 0$ for all $F \subset E$ such that $F \in \mathcal{M}$. A set $E \in \mathcal{M}$ is **negative** for ν if $\nu(F) \leqslant 0$ for all $F \subset E$ such that $F \in \mathcal{M}$. As set $E \in \mathcal{M}$ is **null** for ν if $\nu(F) = 0$ for all $F \subset E$ such that $F \in \mathcal{M}$.

Example 3.3. Let ν be as in Example 3.2. Then $E \in \mathcal{M}$ is positive for ν if and only if $f \geqslant 0$ μ-a.e. on E. Similarly, $E \in \mathcal{M}$ is negative for ν if and only if $f \leqslant 0$ μ-a.e. on E. And $E \in \mathcal{M}$ is null for ν if and only if $f = 0$ μ-a.e. on E.

The following is [5, Theorem 3.3].

Theorem 3.4 (Hahn decomposition)**.** *Let ν be a signed measure on $(X, \mathcal{M})$. Then there exists P, a positive set for ν, and N, a negative set for ν, such that $P \cap N = \varnothing$ and $P \cup N = X$. If P', N' are another such pair, then $P \triangle P' = N \triangle N'$ and this set is null for ν.*

The decomposition $X = P \cup N$ is called a **Hahn decomposition** for ν.

Let ν_1 and ν_2 be signed measures on $(X, \mathcal{M})$. Suppose there exists E_1, a null set for ν_1, and E_2, a null set for ν_2, such that $E_1 \cap E_2 = \varnothing$ and $E_1 \cup E_2 = X$. Then we say ν_1 and ν_2 are **mutually singular**, and we denote this by $\nu_1 \perp \nu_2$. We may also describe this by saying that ν_1 is **singular with respect to** ν_2, or vice versa.

The following is [5, Theorem 3.4]

Theorem 3.5 (Jordan decomposition)**.** *If ν is a signed measure on $(X, \mathcal{M})$, then there exist unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.*

Remark 3.6. Let ν be a signed measure on $(X, \mathcal{M})$ and suppose $X = P \cup N$ is a Hahn decomposition for ν. Then ν^+ and ν^- are given by $\nu^+(E) = \nu(E \cap P)$ and $\nu^-(E) = -\nu(E \cap N)$.

The measures ν^+ and ν^- are called the **positive** and **negative** parts of ν, respectively. The decomposition $\nu = \nu^+ - \nu^-$ is called the **Jordan decomposition** of ν. The **total variation** of ν is the measure $|\nu| = \nu^+ + \nu^-$. We say ν is a **finite** (or **σ-finite**) signed measure if $|\nu|$ is a finite or (σ-finite) measure.

Proposition 3.7. *Let ν be a signed measure on $(X, \mathcal{M})$. Then $|\nu|$ is a finite measure if and only if $\nu(\mathcal{M}) \subset \mathbb{R}$, and in this case, $\nu(\mathcal{M}) \subset [-K, K]$ for some $K \in \mathbb{R}$.*

Proof. Suppose $|\nu|$ is a finite measure and let $E \in \mathcal{M}$. Then

$$|\nu(E)| = |\nu^+(E) - \nu^-(E)| \leqslant \nu^+(E) + \nu^-(E) = |\nu|(E) \leqslant |\nu|(X).$$

Since $K := |\nu|(X) < \infty$, it follows that $\nu(E) \in [-K, K]$ for all $E \in \mathcal{M}$.

It remains only to show that $\nu(\mathcal{M}) \subset \mathbb{R}$ implies that $|\nu|$ is a finite measure. Suppose $\nu(\mathcal{M}) \subset \mathbb{R}$. Let $X = P \cup N$ be a Hahn decomposition for ν. Since $\nu(\mathcal{M}) \subset [-\infty, \infty)$, it follows that $\nu^+(X) = \nu(X \cap P) < \infty$. Hence, ν^+ is a finite measure. Similarly, since $\nu(\mathcal{M}) \subset (-\infty, \infty]$, it follows that $\nu^-(X) = -\nu(X \cap N) < \infty$. Hence, ν^- is a finite measure. Therefore, $|\nu| = \nu^+ + \nu^-$ is a finite measure. ∎

Proposition 3.8. *Let $(X, \mathcal{M})$ be a measurable space and $\nu : \mathcal{M} \to \mathbb{R}^*$. Then ν is a signed measure if and only if there exists a measure μ on $(X, \mathcal{M})$ and a measurable function $f : X \to \mathbb{R}$ such that $d\nu = f\, d\mu$. In this case, $d\nu^+ = f^+\, d\mu$, $d\nu^- = f^-\, d\mu$, and $d|\nu| = |f|\, d\mu$.*

Proof. The "if" part is covered by Example 3.2. For the "only if" part, assume ν is a signed measure on $(X, \mathcal{M})$. Let $\mu = |\nu|$ and $f = 1_P - 1_N$, where $X = P \cup N$ is a Hahn decomposition for ν. Then

$$\int_E f\,d\mu = \int_E (1_P - 1_N)\,d|\nu| = |\nu|(P \cap E) - |\nu|(N \cap E).$$

Since P is null for ν^-, it follows that for any $A \subset P$, we have

$$|\nu|(A) = \nu^+(A) + \nu^-(A) = \nu^+(A) = \nu(A \cap P) = \nu(A).$$

Similarly, for any $A \subset N$, we have $|\nu|(A) = -\nu(A)$. Thus,

$$\int_E f\,d\mu = \nu(P \cap E) + \nu(N \cap E) = \nu(E),$$

since $X = P \cup N$.

Finally, if $d\nu = f\,d\mu$, then $P = f^{-1}([0, \infty))$ and $N = P^c$ form a Hahn decomposition for ν. Thus,

$$\begin{aligned}\nu^+(E) &= \nu(E \cap P) = \int_E f^+\,d\mu,\\ \nu^-(E) &= -\nu(E \cap N) = \int_E f^-\,d\mu, \text{ and}\\ |\nu|(E) &= \nu^+(E) + \nu^-(E) = \int_E |f|\,d\mu,\end{aligned}$$

for every $E \in \mathcal{M}$. □

Let ν be a signed measure on $(X, \mathcal{M})$. We define $L^1(\nu) = L^1(\nu^+) \cap L^1(\nu^-)$ and

$$\int f\,d\nu = \int f\,d\nu^+ - \int f\,d\nu^-,$$

for all $f \in L^1(\nu)$.

Exercises

3.1. Let μ, ν be a signed measures on $(X, \mathcal{M})$ and $E \in \mathcal{M}$. Prove that:

(a) E is null for ν if and only if $|\nu|(E) = 0$.

(b) The following are equivalent:

(i) $\nu \perp \mu$,

(ii) $|\nu| \perp \mu$,

(iii) $\nu^+ \perp \mu$ and $\nu^- \perp \mu$.

3.2. Let ν be a signed measure on $(X, \mathcal{M})$. Show that:

(a) $L^1(\nu) = L^1(|\nu|)$.

(b) If $f \in L^1(\nu)$, then $|\int f \, d\nu| \leqslant \int |f| \, d|\nu|$.

(c) If $E \in \mathcal{M}$, then $|\nu|(E) = \sup\{|\int_E f \, d\nu| : |f| \leqslant 1\}$.

3.3. Let ν be a signed measure on $(X, \mathcal{M})$ and μ_1, μ_2 measures on $(X, \mathcal{M})$. Assume that $\nu = \mu_1 - \mu_2$. Prove that $\mu_1 \geqslant \nu^+$ and $\mu_2 \geqslant \nu^-$.

3.4. Let ν_1, ν_2 be signed measures on $(X, \mathcal{M})$ with $\nu_j(\mathcal{M}) \subset (-\infty, \infty]$ for $j \in \{1, 2\}$. Use Exercise 3.3 to prove that $|\nu_1 + \nu_2| \leqslant |\nu_1| + |\nu_2|$.

3.2 The Radon-Nikodym derivative

Let $(X, \mathcal{M}, \mu)$ be a measure space, and ν a signed measure on $(X, \mathcal{M})$. We say that ν is **absolutely continuous** with respect to μ if $\mu(E) = 0$ implies $\nu(E) = 0$. We denote this by $\nu \ll \mu$.

The following is [5, Theorem 3.8].

Theorem 3.9 (The Lebesgue-Radon-Nikodym theorem)**.** *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and ν a σ-finite signed measure on $(X, \mathcal{M})$. Then there exist unique σ-finite signed measures η, ρ on $(X, \mathcal{M})$ such that $\eta \perp \mu$, $\rho \ll \mu$, and $\nu = \eta + \rho$. Moreover, there is a measurable function $f : X \to \mathbb{R}$ such that $d\rho = f \, d\mu$, and any two such functions are equal μ-a.e.*

The decomposition $\nu = \eta + \rho$ is called the **Lebesgue decomposition** of ν with respect to μ.

If $\nu \ll \mu$, then $\eta = 0$, and we have $d\nu = f \, d\mu$, for some measurable function f. This function f is called the **Radon-Nikodym derivative** of ν with respect to μ. Since the function f is only unique μ-a.e., the Radon-Nikodym derivative is only defined up to a null set. We sometimes say that f is a version of the Radon-Nikodym derivative, and that any two versions are equal μ-a.e.

The Radon-Nikodym derivative is typically denoted by $d\nu/d\mu$, so that

$$\nu(E) = \int_E \frac{d\nu}{d\mu} \, d\mu,$$

for all $E \in \mathcal{M}$. Or, in shorthand,

$$d\nu = \frac{d\nu}{d\mu} \, d\mu.$$

Note that if ν is a positive measure, then $d\nu/d\mu \geqslant 0$ μ-a.e.

The results in the following proposition can be found in [5, Section 3.2].

Proposition 3.10. *Let $(X, \mathcal{M})$ be a measurable space, ν, ν_1, ν_2 σ-finite signed measures on $(X, \mathcal{M})$ and μ, η σ-finite measures on $(X, \mathcal{M})$. Suppose $\nu \ll \mu$, $\nu_1 \ll \mu$, $\nu_2 \ll \mu$, and $\mu \ll \eta$. Then:*

(a) $\nu_1 + \nu_2 \ll \mu$ *and*

$$\frac{d(\nu_1 + \nu_2)}{d\mu} = \frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu}.$$

(b) If $g \in L^1(\nu)$, *then* $g(d\nu/d\mu) \in L^1(\mu)$ *and*

$$\int g\, d\nu = \int g \frac{d\nu}{d\mu}\, d\mu.$$

(c) $\nu \ll \eta$ *and*

$$\frac{d\nu}{d\eta} = \frac{d\nu}{d\mu}\frac{d\mu}{d\eta}, \ \eta\text{-a.e.}$$

(d) If we also have $\eta \ll \mu$, *then*

$$\frac{d\eta}{d\mu}\frac{d\mu}{d\eta} = 1, \ \eta\text{-a.e. and } \mu\text{-a.e.}$$

Example 3.11. Let μ_F be the Lebesgue-Stieltjes measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ associated with $F(x) = 1_{[0,\infty)}(x)e^x$, and let λ denote Lebesgue measure. Then we can write $\mu_F = \delta_0 + \mu_G$, where δ_0 is the point mass measure at 0 and $G(x) = 1_{[0,\infty)}(x)(e^x - 1)$.

Since $\{0\}$ is null for λ and $\mathbb{R}\backslash\{0\}$ is null for δ_0, it follows that $\delta_0 \perp \lambda$. Let us defined the measure ν by $d\nu = F d\lambda$. It is easily verified that $\nu((a,b]) = G(b) - G(a)$ for all $a < b$. By Theorem 1.22, it follows that $\nu = \mu_G$, and hence, $d\mu_G = F d\lambda$. In particular, this shows that $\mu_G \ll \lambda$, so that $\mu_F = \delta_0 + \mu_G$ is the Lebesgue decomposition of μ_F with respect to λ. Moreover, this shows that F is a version of the Radon-Nikodym derivative, $d\mu_G/d\lambda$. Any other version of this Radon-Nikodym derivative will be equal to F λ-a.e.

Exercises

3.5. Let $(X, \mathcal{M}, \mu)$ be a measure space and ν a signed measure on $(X, \mathcal{M})$. Prove that the following are equivalent:

(a) $\nu \ll \mu$.

(b) $|\nu| \ll \mu$.

(c) $\nu^+ \ll \mu$ and $\nu^- \ll \mu$.

3.6. Let μ be counting measure on $([0,1], \mathcal{B}_{[0,1]})$ and let λ be Lebesgue measure. Show that $\lambda \ll \mu$, but there does not exist any function f such that $d\lambda = f\, d\mu$. Why does this not contradict the Lebesgue-Radon-Nikodym theorem?

3.7. Suppose that μ, ν are σ-finite measures on $(X, \mathcal{M})$ with $\nu \ll \mu$. Let $\eta = \mu + \nu$ and let f be a version of $d\nu/d\eta$. Prove that $0 \leqslant f < 1$ μ-a.e. and that $f/(1-f)$ is a version of $d\nu/d\mu$.

3.3 Complex measures

The primary source for the material in this section, as well as Section 3.4, is [5]. This material does not seem to be present in [2]. Even though it is material that may not be covered in a typical course on measure and integration, I am nonetheless including it in these notes for completeness. Section 3.4 is particularly important, as it includes the fundamental theorem of calculus for Lebesgue integrals, as well as concepts that are needed in the theory of stochastic integration.

Let $(X, \mathcal{M})$ be a measurable space and $\alpha : \mathcal{M} \to \mathbb{C}$. Then α is a **complex measure** on $(X, \mathcal{M})$ if:

(i) $\alpha(\varnothing) = 0$, and

(ii) if $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$ are disjoint, then $\alpha(\biguplus_j E_j) = \sum_j \alpha(E_j)$, where this sum converges absolutely.

Note that every finite signed measure is a complex measure.

Every complex measure α can be written as $\alpha = \alpha_r + i\alpha_i$, where α_r, α_i are finite signed measures. By Proposition 3.7, it follows that $\alpha(\mathcal{M})$ is a bounded subset of $\mathbb{C}$.

We define $L^1(\alpha) = L^1(\alpha_r) \cap L^1(\alpha_i)$ and, for $f \in L^1(\alpha)$, we define

$$\int f\, d\alpha = \int f\, d\alpha_r + i \int f\, d\alpha_i.$$

If α is a complex measure and ν is a signed measure, then we say $\alpha \perp \nu$ if $\alpha_r \perp \nu$ and $\alpha_i \perp \nu$. If β is another complex measure, then we say $\alpha \perp \beta$ if $\alpha \perp \beta_r$ and $\alpha \perp \beta_i$.

If μ is a positive measure, then we say $\alpha \ll \mu$ if $\alpha_r \ll \mu$ and $\alpha_i \ll \mu$.

By applying Theorem 3.9 to the real and imaginary parts of a complex measure, we obtain the following, which is [5, Theorem 3.12].

Theorem 3.12 (The Lebesgue-Radon-Nikodym theorem)**.** *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and α a complex measure on $(X, \mathcal{M})$. Then there exists a unique complex measure η and a μ-a.e. unique $f \in L^1(\mu)$ such that $\eta \perp \mu$ and $d\alpha = d\eta + f\, d\mu$.*

As for signed measures, if $\alpha \ll \mu$, then the function f is called the Radon-Nikodym derivative of α with respect to μ and denoted by $d\alpha/d\mu$.

Proposition 3.13. *Let α be a complex measure on a measurable space $(X, \mathcal{M})$. There exists a measure μ on $(X, \mathcal{M})$ and an $f \in L^1(\mu)$ such that $d\alpha = f\, d\mu$.*

If ν is another measure on $(X, \mathcal{M})$ and $g \in L^1(\nu)$ with $d\alpha = g\, d\nu$, then $|f|\, d\mu = |g| d\nu$.

Proof. Let $\mu = |\alpha_r| + |\alpha_i|$. Then $\alpha \ll \mu$, so by Theorem 3.12, we have $d\alpha = f\, d\mu$, where $f = d\alpha/d\mu$.

Suppose we also have $d\alpha = g\,d\nu$. Let $\rho = \mu + \nu$. Then $\mu \ll \rho$ and $\nu \ll \rho$, so by Proposition 3.10(b), we have

$$\alpha(E) = \int_E f\,d\mu = \int_E f \frac{d\mu}{d\rho}\,d\rho,$$

for all $E \in \mathcal{M}$. Similarly,

$$\alpha(E) = \int_E g\,d\nu = \int_E g \frac{d\nu}{d\rho}\,d\rho,$$

for all $E \in \mathcal{M}$. By Proposition 2.27(c), this implies

$$f \frac{d\mu}{d\rho} = g \frac{d\nu}{d\rho} \ \rho\text{-a.e.}$$

Since $d\mu/d\rho$ and $d\nu/d\rho$ are nonnegative functions, this implies

$$|f| \frac{d\mu}{d\rho} = |g| \frac{d\nu}{d\rho} \ \rho\text{-a.e.}$$

Hence,

$$\int_E |f|\,d\mu = \int_E |f| \frac{d\mu}{d\rho}\,d\rho = \int_E |g| \frac{d\nu}{d\rho}\,d\rho = \int_E |g|\,d\nu,$$

for all $E \in \mathcal{M}$. □

Let α be a complex measure on $(X, \mathcal{M})$. By Proposition 3.13, we can write $d\alpha = f\,d\mu$ for some measure μ and some $f \in L^1(\mu)$. Also by Proposition 3.13, we may unambiguously define the **total variation** of α to be the measure $|\alpha|$ on $(X, \mathcal{M})$ given by $d|\alpha| = |f|\,d\mu$. By Proposition 3.8, this definition agrees with our previous definition when α is a signed measure.

Proposition 3.14. *If α, β are complex measures on $(X, \mathcal{M})$, then $|\alpha + \beta| \leqslant |\alpha| + |\beta|$.*

Proof. Write $d\alpha = f\,d\mu$ and $d\beta = g\,d\nu$, where μ, ν are measures on $(X, \mathcal{M})$, $f \in L^1(\mu)$ and $g \in L^1(\nu)$. Let $\rho = \mu + \nu$. We then have

$$(\alpha + \beta)(E) = \int_E f\,d\mu + \int_E g\,d\nu = \int_E \left(f \frac{d\mu}{d\rho} + g \frac{d\nu}{d\rho} \right) d\rho,$$

for all $E \in \mathcal{M}$. Thus,

$$\begin{aligned} |\alpha + \beta|(E) = \int_E \left| f \frac{d\mu}{d\rho} + g \frac{d\nu}{d\rho} \right| d\rho &\leqslant \int_E |f| \frac{d\mu}{d\rho}\,d\rho + \int_E |g| \frac{d\nu}{d\rho}\,d\rho \\ &= \int_E |f|\,d\mu + \int_E |g|\,d\nu = |\alpha|(E) + |\beta|(E), \end{aligned}$$

for all $E \in \mathcal{M}$. □

Let α be a complex Borel measure on $\mathbb{R}^n$. Then α is **discrete** if there exists a countable set $\{x_j\}_{j=1}^{\infty} \subset \mathbb{R}^n$ and complex numbers c_j such that $\sum_{j=1}^{\infty} |c_j| < \infty$ and $\alpha = \sum_{j=1}^{\infty} c_j \delta_{x_j}$, where δ_x is the point mass measure at x. On the other hand, α is **continuous** if $\alpha(\{x\}) = 0$ for all $x \in \mathbb{R}^n$. Note that if $\alpha \ll \lambda$, then α is continuous. Also note that if α is discrete, then $\alpha \perp \lambda$.

Proposition 3.15. *Every complex Borel measure α on $\mathbb{R}^n$ can be written uniquely as*

$$\alpha = \alpha_d + \alpha_{ac} + \alpha_{sc},$$

where α_d is discrete, $\alpha_{ac} \ll \lambda$, and α_{sc} is continuous with $\alpha_{sc} \perp \lambda$.

Proof. Let $E = \{x \in \mathbb{R}^n : \alpha(\{x\}) \neq 0\}$ and let $E_k = \{x \in \mathbb{R}^n : |\alpha(\{x\})| > 1/k\}$, so that $E = \bigcup_{k=1}^{\infty} E_k$.

Suppose $x_1, \ldots, x_N$ are N distinct elements in $\mathbb{R}^n$ such that $\{x_1, \ldots, x_N\} \subset E_k$. Then

$$\sum_{j=1}^{N} |\alpha|(\{x_j\}) = |\alpha|(\{x_1, \ldots, x_N\}) \leqslant |\alpha|(E_k) \leqslant M,$$

where $M = |\alpha|(\mathbb{R}^n) < \infty$. Let us write $d\alpha = f\,d\nu$, where ν is a measure on $\mathbb{R}^n$ and $f \in L^1(\nu)$. Then

$$\alpha(\{x_j\}) = \int_{\{x_j\}} f\,d\nu = f(x_j)\nu(\{x_j\}), \text{ and}$$
$$|\alpha|(\{x_j\}) = \int_{\{x_j\}} |f|\,d\nu = |f(x_j)|\nu(\{x_j\}).$$

Since ν is a positive measure, it follows that $|\alpha|(\{x_j\}) = |\alpha(\{x_j\})|$. Thus,

$$M \geqslant \sum_{j=1}^{N} |\alpha(\{x_j\})| > N/k,$$

since each $x_j \in E_k$. It follows that $N < Mk$, so that E_k has fewer than Mk elements. In particular, E_k is finite, so E is countable. Therefore, if we define $\alpha_d(A) = \alpha(A \cap E)$ and $\alpha_c(A) = \alpha(A \backslash E)$, then $\alpha = \alpha_d + \alpha_c$, where α_d is discrete and α_c is continuous. It is easy to see that this is the unique way to decompose α into a discrete and continuous part.

Using Theorem 3.12, let $\alpha_c = \alpha_{sc} + \alpha_{ac}$ be the Lebesgue decomposition of α_c with respect to λ, where $\alpha_{sc} \perp \lambda$ and $\alpha_{ac} \ll \lambda$, and this gives us the desired decomposition of α. □

Remark 3.16. The "*sc*" in α_{sc} stands for "singularly continuous", to remind us that α_{sc} is singular with respect to Lebesgue measure, but also continuous, since $\alpha_{sc}(\{x\}) = 0$ for all $x \in \mathbb{R}^n$.

Exercises

3.8. Let $(X, \mathcal{M}, \mu)$ be a measure space and α, β complex measures on $(X, \mathcal{M})$. Prove that:

(a) $\alpha \perp \beta$ if and only if $|\alpha| \perp |\beta|$.

(b) $\alpha \ll \mu$ if and only if $|\alpha| \ll \mu$.

3.9. Let α be a complex measure on $(X, \mathcal{M})$ with $\alpha(X) = |\alpha|(X)$. Prove that $\alpha = |\alpha|$.

3.4 Functions of bounded variation

For a function $F : \mathbb{R} \to \mathbb{C}$, we define

$$\begin{aligned} F(x+) &= \lim_{y \downarrow x} F(y), \\ F(x-) &= \lim_{y \uparrow x} F(y), \\ F(\infty) &= \lim_{x \to \infty} F(x), \text{ and} \\ F(-\infty) &= \lim_{x \to -\infty} F(x), \end{aligned}$$

provided these limits exist. Also, in this section, the phrase "almost everywhere" will always be with respect to Lebesgue measure.

The following is [5, Theorem 3.23].

Theorem 3.17. *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and define $G(x) = F(x+)$. Then:*

(a) The set of points at which F is discontinuous is countable.

(b) The function G is increasing, right-continuous, and $G = F$ a.e.

(c) F and G are differentiable a.e., and $F' = G'$ a.e.

If $a, b \in \mathbb{R}$, $a < b$, and $F : [a, b] \to \mathbb{C}$, then the **total variation** of F on $[a, b]$ is

$$\sup \left\{ \sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, a = x_0 < \cdots < x_n = b \right\}.$$

If $F : \mathbb{R} \to \mathbb{C}$, then we define the **total variation function** of F to be

$$T_F(x) = \sup \left\{ \sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, -\infty < x_0 < \cdots < x_n = x \right\}.$$

It can be verified that in this case, if $a, b \in \mathbb{R}$ and $a < b$, then the total variation of F on $[a, b]$ is $T_F(b) - T_F(a)$, provided this latter quantity is well-defined. Thus, $T_F : \mathbb{R} \to [0, \infty]$ is an increasing function.

The set of functions $F : [a, b] \to \mathbb{C}$ that have finite total variation on $[a, b]$ is denoted by $BV([a, b])$. The set of functions $F : \mathbb{R} \to \mathbb{C}$ that have $T_F(\infty) < \infty$ is denoted by BV.

If $F \in BV$, then $F|_{[a,b]} \in BV([a, b])$. Conversely, suppose $F \in BV([a, b])$. For $x < a$, define $F(x) = F(a)$, and for $x > b$, define $F(x) = F(b)$. Then F extended this way satisfies $F \in BV$. Consequently, we will lose no generality by focusing our attention on BV, rather than $BV([a, b])$.

Proposition 3.18.

(a) If $F : \mathbb{R} \to \mathbb{R}$ is bounded and increasing, then $F \in BV$ and $T_F(x) = F(x) - F(-\infty)$ for all $x \in \mathbb{R}$.

(b) If $F, G \in BV$ and $a, b \in \mathbb{C}$, then $aF + bG \in BV$.

(c) If F is differentiable on $\mathbb{R}$ and F' is bounded and $a, b \in \mathbb{R}$ with $a < b$, then $F|_{[a,b]} \in BV([a, b])$.

Proof. Exercise 3.10. □

The following is [5, Theorem 3.27].

Theorem 3.19.

(a) $F \in BV$ if and only if $\mathrm{Re}F \in BV$ and $\mathrm{Im}F \in BV$.

(b) If $F : \mathbb{R} \to \mathbb{R}$, then $F \in BV$ if and only if F is the difference of two increasing functions. For $F \in BV$, those functions may be taken to be $(T_F + F)/2$ and $(T_F - F)/2$.

(c) If $F \in BV$ and $x \in \mathbb{R}$, then $F(x+)$, $F(x-)$, $F(\infty)$, and $F(-\infty)$ exist.

(d) If $F \in BV$ and $G(x) = F(x+)$, then F' and G' exist a.e. and $F' = G'$ a.e.

Let $F \in BV$. Define $P_F = (T_F + F)/2$ and $N_F = (T_F - F)/2$, so that P_F and N_F are increasing functions with $F = P_F - N_F$.

The functions P_F and N_F are called the **positive** and **negative variations** of F, respectively. Since $x^+ := x \vee 0 = (|x| + x)/2$ and $x^- := -(x \wedge 0) = (|x| - x)/2$, it follows that

$$P_F(x) = \frac{1}{2}F(-\infty) + \sup\left\{ \sum_{j=1}^{\infty} (F(x_j) - F(x_{j-1}))^+ : n \in \mathbb{N}, -\infty < x_0 < \cdots < x_n = x \right\},$$

$$N_F(x) = -\frac{1}{2}F(-\infty) + \sup\left\{ \sum_{j=1}^{\infty} (F(x_j) - F(x_{j-1}))^- : n \in \mathbb{N}, -\infty < x_0 < \cdots < x_n = x \right\}.$$

The space of normalized bounded variation functions is defined by

$$NBV = \{F \in BV : F \text{ is right continuous and } F(-\infty) = 0\}.$$

It can be shown that if $F \in BV$ and $G(x) = F(x+) - F(-\infty)$, then $G \in NBV$ and $G - F$ is constant almost everywhere. (See Exercise 3.11.)

The following theorem is a combination of [5, Theorem 3.29] and [5, Exercise 3.29].

Theorem 3.20. *If α is a complex Borel measure on $\mathbb{R}$ and $F(x) = \alpha((-\infty, x])$, then $F \in NBV$. Conversely, if $F \in NBV$, then there is a unique complex Borel measure α_F such that $F(x) = \alpha_F((-\infty, x])$, and in this case, $|\alpha_F| = \alpha_{T_F}$. Moreover, if F is real-valued, then $\alpha_F^+ = \alpha_{P_F}$ and $\alpha_F^- = \alpha_{N_F}$.*

A function $F : \mathbb{R} \to \mathbb{C}$ is **absolutely continuous** if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for any finite set of disjoint intervals, $(a_1, b_1), \ldots, (a_n, b_n)$, we have

$$\sum_{j=1}^{n} (b_j - a_j) < \delta \text{ implies } \sum_{j=1}^{n} |F(b_j) - F(a_j)| < \varepsilon.$$

More generally, F **is absolutely continuous on** $[a, b]$ if this condition is satisfied whenever the intervals (a_j, b_j) all lie in $[a, b]$.

By taking $N = 1$, we see that absolute continuity implies uniform continuity. On the other hand, using the mean value theorem, we see that if $F'(x)$ exists for all x and F' is bounded, then F is absolutely continuous.

The following result is contained in [5, Propositions 3.30 and 3.32].

Proposition 3.21. *Let $F \in NBV$. Then $F' \in L^1(\lambda)$. Also, $\alpha_F \perp \lambda$ if and only if $F' = 0$ a.e. And $\alpha_F \ll \lambda$ if and only if F is absolutely continuous.*

If $F \in NBV$, then we will adopt the notation

$$\int_a^b g(x)\, dF(x) = \int_{(a,b]} g\, d\alpha_F. \tag{3.1}$$

We may also use variations on this notation, such as $\int_{(a,b]} g\, dF$. Integrals such as these are called **Lebesgue-Stieltjes integrals**.

The following integration by parts formula is [5, Theorem 3.36].

Theorem 3.22. *If $F, G \in NBV$ and at least one of them is continuous, then*

$$\int_{(a,b]} F\, dG = F(b)G(b) - F(a)G(a) - \int_{(a,b]} G\, dF.$$

Finally, we have the following, which is [5, Theorem 3.35].

Theorem 3.23 (The fundamental theorem of calculus for Lebesgue integrals)**.** *Let $a, b \in \mathbb{R}$ with $a < b$ and $F : [a, b] \to \mathbb{C}$. Then the following are equivalent:*

(a) F is absolutely continuous on $[a, b]$.

(b) There exists $f \in L^1([a,b],\lambda)$ such that for all $x \in [a,b]$, we have

$$F(x) - F(a) = \int_a^x f(t)\,dt.$$

(c) F is differentiable a.e. on $[a,b]$, $F' \in L^1([a,b],\lambda)$, and

$$F(x) - F(a) = \int_a^x F'(t)\,dt.$$

Exercises

3.10. Prove Proposition 3.18. (Hint: For Part ((c)), use the mean value theorem.)

3.11. Prove that if $F \in BV$ and $G(x) = F(x+) - F(-\infty)$, then $G \in NBV$ and $G - F$ is constant almost everywhere.

3.12. Construct an increasing function on $\mathbb{R}$ whose set of discontinuities is $\mathbb{Q}$.

3.13. Let $F(x) = x^2 \sin(x^{-1}) 1_{\{0\}^c}(x)$ and $G(x) = x^2 \sin(x^{-2}) 1_{\{0\}^c}(x)$. Prove that:

(a) $F'(x)$ and $G'(x)$ exist for all $x \in \mathbb{R}$.

(b) $F \in BV([-1,1])$, but $G \notin BV([-1,1])$.

3.14. Let $G : [a,b] \to \mathbb{R}$ be continuous and increasing, with $G(a) = c$ and $G(b) = d$. Prove that:

(a) If $E \subset [c,d]$ is a Borel set, then $\lambda(E) = \mu_G(G^{-1}(E))$.

(Hint: First consider the case where E is an interval.)

(b) If $f \in L^1([c,d],\lambda)$, then

$$\int_c^d f(y)\,dy = \int_a^b f(G(x))\,dG(x).$$

In particular, if G is absolutely continuous, then

$$\int_c^d f(y)\,dy = \int_a^b f(G(x))G'(x)\,dx.$$

(Remark: This result may fail if G is merely right-continuous.)

3.5 Functions with one-sided limits

3.5.1 Definition and basic properties

Let (X, d) be a metric space. A function $f : \mathbb{R} \to X$ is said to have **one-sided limits** if, for each $t \in \mathbb{R}$, the limits $f(t+) = \lim_{s \downarrow t} f(s)$ and $f(t-) = \lim_{s \uparrow t} f(s)$ both exist. These functions are more well-behaved than one might initially expect, as the following theorems demonstrate.

Theorem 3.24. *A function with one-sided limits is bounded on compact sets.*

Proof. Let f have one-sided limits and let $K \subset \mathbb{R}$ be compact. Fix any $p \in X$. We want to show that there exists $r > 0$ such that $f(K) \subset B_r(p)$.

Fix $t \in K$. Since $f(t+)$ exists, there exists $\delta_{t+} > 0$ such that $d(f(s), f(t+)) < 1$ for all $s \in (t, t + \delta_{t+})$. Thus, if $r_{t+} = 1 + d(f(t+), p)$, then

$$d(f(s), p) \leqslant d(f(s), f(t+)) + d(f(t+), p) < r_{t+}.$$

In other words, $f(s) \in B_{r_{t+}}(p)$, for all $s \in (t, t + \delta_{t+})$.

Similarly, since $f(t-)$ exists, there exists $\delta_{t-} > 0$ such that $f(s) \in B_{r_{t-}}(p)$ for all $s \in (t - \delta_{t-}, t)$, where $r_{t-} = 1 + d(f(t-), p)$. Thus, for all $s \in U_t = (t - \delta_{t-}, t + \delta_{t+})$, we have that $f(s) \in B_{r_t}(p)$, where $r_t = \max\{r_{t-}, r_{t+}, d(f(t), p) + 1\}$.

Since $\{U_t : t \in K\}$ is an open cover of K, there exists $\{t_1, \ldots, t_n\} \subset K$ such that $K \subset U_1 \cup \cdots \cup U_n$. It follows that, for all $s \in K$, we have $f(s) \in B_r(p)$, where $r = \max\{r_{t_1}, \ldots, r_{t_n}\}$. That is, $f(K) \subset B_r(p)$. □

This next theorem shows that a function with one-sided limits cannot have large discontinuities which accumulate.

Theorem 3.25. *Let f have one-sided limits. Then for all $t \in \mathbb{R}$ and $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$d(f(s+), f(s)) + d(f(s), f(s-)) < \varepsilon,$$

whenever $s \neq t$ and $|t - s| < \delta$.

Proof. Suppose not. Then there exists $t \in \mathbb{R}$, $\varepsilon > 0$, and a sequence $\{s_n\}$ of real numbers such that, for all n, we have $s_n \neq t$, $|t - s_n| < 1/n$, and

$$d(f(s_n+), f(s_n)) + d(f(s_n), f(s_n-)) \geqslant \varepsilon,$$

Consider the following four sets:

$$\begin{aligned}
S_1 &= \{n : s_n > t \text{ and } d(f(s_n+), f(s_n)) \geqslant \varepsilon/2\},\\
S_2 &= \{n : s_n > t \text{ and } d(f(s_n), f(s_n-)) \geqslant \varepsilon/2\},\\
S_3 &= \{n : s_n < t \text{ and } d(f(s_n+), f(s_n)) \geqslant \varepsilon/2\},\\
S_4 &= \{n : s_n < t \text{ and } d(f(s_n), f(s_n-)) \geqslant \varepsilon/2\}.
\end{aligned}$$

Since these sets cover $\mathbb{N}$, at least one of them is infinite. By passing to a subsequence, we may assume that the entire sequence $\{s_n\}$ is contained in one of these sets.

First assume that each $s_n \in S_1$. For each n, choose $u_n \in (s_n, s_n + 1/n)$ such that $d(f(u_n), f(s_n+)) < \varepsilon/4$. Then

$$\begin{aligned}\frac{\varepsilon}{2} \leqslant d(f(s_n+), f(s_n)) \leqslant d(f(s_n+), f(u_n)) + d(f(u_n), f(s_n))\\ < \frac{\varepsilon}{4} + d(f(u_n), f(s_n)).\end{aligned}$$

But $s_n \to t^+$ and $u_n \to t^+$, so $d(f(u_n), f(s_n)) \to d(f(t+), f(t+)) = 0$, a contradiction.

Next assume that each $s_n \in S_2$. For each n, choose $u_n \in (t, s_n)$ such that $d(f(u_n), f(s_n-)) < \varepsilon/4$. Then

$$\begin{aligned}\frac{\varepsilon}{2} \leqslant d(f(s_n), f(s_n-)) \leqslant d(f(s_n), f(u_n)) + d(f(u_n), f(s_n-))\\ < d(f(u_n), f(s_n)) + \frac{\varepsilon}{4}.\end{aligned}$$

But $s_n \to t^+$ and $u_n \to t^+$, so $d(f(u_n), f(s_n)) \to d(f(t+), f(t+)) = 0$, a contradiction.

Next assume that each $s_n \in S_3$. For each n, choose $u_n \in (s_n, t)$ such that $d(f(u_n), f(s_n+)) < \varepsilon/4$. Then

$$\begin{aligned}\frac{\varepsilon}{2} \leqslant d(f(s_n+), f(s_n)) \leqslant d(f(s_n+), f(u_n)) + d(f(u_n), f(s_n))\\ < \frac{\varepsilon}{4} + d(f(u_n), f(s_n)).\end{aligned}$$

But $s_n \to t^-$ and $u_n \to t^-$, so $d(f(u_n), f(s_n)) \to d(f(t-), f(t-)) = 0$, a contradiction.

Finally assume that each $s_n \in S_4$. For each n, choose $u_n \in (s_n - 1/n, s_n)$ such that $d(f(u_n), f(s_n-)) < \varepsilon/4$. Then

$$\begin{aligned}\frac{\varepsilon}{2} \leqslant d(f(s_n), f(s_n-)) \leqslant d(f(s_n), f(u_n)) + d(f(u_n), f(s_n-))\\ < d(f(u_n), f(s_n)) + \frac{\varepsilon}{4}.\end{aligned}$$

But $s_n \to t^-$ and $u_n \to t^-$, so $d(f(u_n), f(s_n)) \to d(f(t-), f(t-)) = 0$, a contradiction. □

Theorem 3.26. *A function with one-sided limits has at most countably many discontinuities.*

Proof. Let f have one-sided limit. Then f is continuous at t if and only if $f(t-) = f(t+) = f(t)$, which happens if and only if

$$d(f(t+), f(t)) + d(f(t), f(t-)) = 0.$$

Let

$$A_n = \{t \in \mathbb{R} : d(f(t+), f(t)) + d(f(t), f(t-)) \geqslant 1/n\}.$$

Then $A = \bigcup_{n=1}^{\infty} A_n$ is the set of discontinuities of f.

Fix $M, n \in \mathbb{N}$. Fix $t \in [-M, M]$. By Theorem 3.25 with $\varepsilon = 1/n$, there exists $\delta_t > 0$ such that $((t - \delta_t, t) \cup (t, t + \delta_t)) \cap A_n = \varnothing$. Thus, if $U_t = (t - \delta_t, t + \delta_t)$, then $U_t \cap A_n \subset \{t\}$. Since $[-M, M]$ is compact, and $\{U_t : t \in [-M, M]\}$ is an open cover of $[-M, M]$, it follows that there exists $\{t_1, \ldots, t_k\} \subset [-M, M]$ such that $[-M, M] \subset U_{t_1} \cup \cdots \cup U_{t_k}$. Hence, $[-M, M] \cap A_n \subset \{t_1, \ldots, t_k\}$. In particular, $[-M, M] \cap A_n$ is finite.

Therefore,

$$A = \bigcup_{n=1}^{\infty} \bigcup_{M=1}^{\infty} [-M, M] \cap A_n$$

is a countable set. □

3.5.2 Cadlag functions

If f has one-sided limits, we define $f_+ : \mathbb{R} \to \mathbb{R}$ and $f_- : \mathbb{R} \to \mathbb{R}$ by $f_+(t) = f(t+)$ and $f_-(t) = f(t-)$. Note that a function f with one-sided limits is right-continuous if and only if $f(t+) = f(t)$ for all $t \in \mathbb{R}$, which is equivalent to saying that $f_+ = f$. If f has one-sided limits and is right-continuous, then we say that f is **cadlag**. This is an acronym for the French phrase, "continu à droite, limite à gauche". If f has one-sided limits and is left-continuous, that is, if $f(t-) = f(t)$ for all $t \in \mathbb{R}$ (which is equivalent to $f_- = f$), then we say that f is **caglad**.

If f has one-sided limits, we also define the function $\Delta f : \mathbb{R} \to \mathbb{R}$ by $\Delta f = f_+ - f_-$. Note that, by Theorem 3.26, the set $\{t : \Delta f(t) \neq 0\}$ is countable.

Given any $f : \mathbb{R} \to \mathbb{R}$, let us define $Rf : \mathbb{R} \to \mathbb{R}$ by $Rf(t) = f(-t)$.

Lemma 3.27. *If f has one-sided limits, then so does Rf. Moreover, $(Rf)_+ = Rf_-$ and $(Rf)_- = Rf_+$.*

Proof. Let f have one-sided limits. Then

$$\lim_{s \downarrow t} Rf(s) = \lim_{s \downarrow t} f(-s) = \lim_{z \uparrow (-t)} f(z) = f_-(-t),$$

and

$$\lim_{s \uparrow t} Rf(s) = \lim_{s \uparrow t} f(-s) = \lim_{z \downarrow (-t)} f(z) = f_+(-t),$$

which shows that Rf has one-sided limits, and that $(Rf)_+ = Rf_-$ and $(Rf)_- = Rf_+$. □

Lemma 3.28. *If $f : \mathbb{R} \to \mathbb{R}$ is increasing, then f has one-sided limits, and f_+ and f_- are both increasing.*

Proof. Fix $t \in \mathbb{R}$ and fix some strictly increasing sequence $\{t_n\}$ with $t_n \to t$. Then $\{f(t_n)\}$ is an increasing sequence of real numbers, bounded above by $f(t)$. Hence, there exists $L \in \mathbb{R}$ such that $f(t_n) \to L$.

Now let $\{s_n\}$ be any other strictly increasing sequence with $s_n \to t$. As above, $f(s_n) \to L'$ for some $L' \in \mathbb{R}$. Now fix $m \in \mathbb{N}$. Since $s_m < t$ and $t_n \to t$, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, we have $s_m < t_n$. This implies $f(s_m) \leqslant f(t_n)$. Letting $n \to \infty$, we have $f(s_m) \leqslant L$. But this holds for all m, so letting $m \to \infty$, we have $L' \leqslant L$. A similar argument shows that $L \leqslant L'$. Thus, $L' = L$, so that $f(s_n) \to L$. Since this holds for any such sequence $\{s_n\}$, we have that $L = \lim_{s \to t^-} f(s)$, and so $f(t-)$ exists.

A similar argument shows that $f(t+)$ exists for all $t \in \mathbb{R}$.

Now let $s < t$. Choose a strictly decreasing sequence $\{s_n\} \subset (s, t)$ such that $s_n \to s$, and choose a strictly decreasing sequence $\{t_n\}$ such that $t_n \to t$. Then $s_n < t < t_n$ for all n. Hence, $f(s_n) \leqslant f(t_n)$ for all n. Letting $n \to \infty$ gives $f(s+) \leqslant f(t+)$, showing that f_+ is increasing. A similar argument shows that f_- is increasing. □

Theorem 3.29. *If f has one-sided limits, then*

$$\begin{aligned} f_+(t+) &= f_-(t+) = f(t+), \text{ and} \\ f_+(t-) &= f_-(t-) = f(t-), \end{aligned}$$

for all $t \in \mathbb{R}$. In other words, $(f_+)_+ = (f_-)_+ = f_+$ and $(f_+)_- = (f_-)_- = f_-$. In particular, f_+ is cadlag and f_- is caglad.

Proof. Fix $t \in \mathbb{R}$ and let $\{t_n\}$ be a strictly decreasing sequence of real numbers such that $t_n \to t$. Let $\varepsilon > 0$ be arbitrary. Using Theorem 3.25 and the fact that $f(t_n) \to f(t+)$, we may choose $N \in \mathbb{N}$ such that for all $n \geqslant N$, we have $d(f(t_n+), f(t_n)) < \varepsilon$, $d(f(t_n-), f(t_n)) < \varepsilon$, and $d(f(t_n), f(t+)) < \varepsilon$. By the triangle inequality, this implies that $d(f(t_n+), f(t+)) < 2\varepsilon$ and $d(f(t_n-), f(t+)) < 2\varepsilon$. Since ε was arbitrary, this shows that $f_+(t_n) = f(t_n+) \to f(t+)$ and $f_-(t_n) = f(t_n-) \to f(t+)$. Since the sequence $\{t_n\}$ was arbitrary, this shows that $f_+(t+) = f(t+)$ and $f_-(t+) = f(t+)$. Since this holds for all $t \in \mathbb{R}$, we have $(f_+)_+ = (f_-)_+ = f_+$.

Now let $g = Rf$. We have already shown that $(g_+)_+ = (g_-)_+ = g_+$. By Lemma 3.27, we have $g_+ = Rf_-$. Therefore, $(g_+)_+ = R(f_-)_-$ and similarly, $(g_-)_+ = R(f_+)_-$. Hence, $R(f_+)_- = R(f_-)_- = Rf_-$, which implies $(f_+)_- = (f_-)_- = f_-$.

Lastly, since $(f_+)_+ = f_+$, it follows that f_+ is cadlag, and since $(f_-)_- = f_-$, it follows that f_- is caglad. □

Remark 3.30. By Theorem 3.29, if f is cadlag (or any function with one-sided limits), then $g = f_-$ is caglad. Conversely, if g is any caglad function, then $g = g_- = (g_+)_-$. In other words, $g = f_-$, where $f = g_+$ is a cadlag function. What this shows is that a function $g : \mathbb{R} \to \mathbb{R}$ is caglad if and only if $g = f_-$ for some cadlag function f.

3.5.3 Relation to BV functions

By Theorem 3.19, a function $G : \mathbb{R} \to \mathbb{R}$ is in BV if and only if G can be written as $G = G_1 - G_2$, where each G_j is a bounded, increasing function. Since increasing functions have one-sided limits, every BV function has one-sided limits. Moreover, by Lemma 3.28, this shows that G_+ and G_- are both BV functions.

If $G \in BV$, then G_+ is right-continuous and BV, and $G_+ - G(-\infty) \in NBV$. As in Theorem 3.20, there exists a unique signed Borel measure μ_{G_+} on $\mathbb{R}$ such that $\mu_{G_+}((s,t]) = G_+(t) - G_+(s)$ for all $s < t$. Note that

$$\begin{aligned}
\mu_{G+}(\{t\}) &= \lim_{s\uparrow t} \mu_{G_+}((s,t]) \\
&= \lim_{s\uparrow t}(G_+(t) - G_+(s)) \\
&= G_+(t) - G_+(t-) \\
&= G_+(t) - G_-(t),
\end{aligned}$$

by Theorem 3.29. If we recall that $\Delta G = G_+ - G_-$, then $\mu_{G_+}(\{t\}) = \Delta G(t)$.

In (3.1), we defined the Lebesgue-Stieltjes integral for integrators which are NBV. We now extend this by defining the **Lebesgue-Stieltjes integral** of a Borel measurable function f with respect to a BV function G by

$$\int_A f\, dG = \int_A f\, d\mu_{G_+}.$$

The following theorem illustrates a relationship between the Lebesgue-Stieltjes integral and classical Riemann sums.

Theorem 3.31. *Let $G \in BV$ and let f be a function with one-sided limits. Fix $a < b$. For each $m \in \mathbb{N}$, let $\mathcal{P}_m = \{t_j^{(m)}\}_{j=0}^{n(m)}$ be a strictly increasing, finite sequence of real numbers with $a = t_0^{(m)} < \ldots < t_{n(m)}^{(m)} = b$. Assume that $\|\mathcal{P}_m\| = \max\{|t_j^{(m)} - t_{j-1}^{(m)}| : 1 \leqslant j \leqslant n(m)\} \to 0$ as $m \to \infty$. Let*

$$\begin{aligned}
I_-^{(m)} &= \sum_{j=1}^{n(m)} f(t_{j-1}^{(m)})(G(t_j^{(m)}) - G(t_{j-1}^{(m)})), \text{ and} \\
I_+^{(m)} &= \sum_{j=1}^{n(m)} f(t_j^{(m)})(G(t_j^{(m)}) - G(t_{j-1}^{(m)})).
\end{aligned}$$

Then:

(i) If G is cadlag, then $I_-^{(m)} \to \int_{(a,b]} f_-\, dG$ as $m \to \infty$.

(ii) If f and G are both cadlag, then $I_+^{(m)} \to \int_{(a,b]} f_+\, dG = \int_{(a,b]} f\, dG$ as $m \to \infty$.

(iii) If G is caglad, then $I_+^{(m)} \to \int_{[a,b)} f_+ \, dG$ as $m \to \infty$.

(iv) If f and G are both caglad, then $I_-^{(m)} \to \int_{[a,b)} f_- \, dG = \int_{[a,b)} f \, dG$ as $m \to \infty$.

Proof. In this proof, for notational simplicity, we will suppress the dependence of n, t_j, and $I_\pm$ on m.

Let us first assume that G is cadlag. Then $G = G_+$, and so

$$I_- = \sum_{j=1}^n f(t_{j-1}) \mu_{G+}((t_{j-1}, t_j]) = \int_{(a,b]} f_m^{(1)} \, d\mu_{G_+},$$

where

$$f_m^{(1)}(t) = \sum_{j=1}^n f(t_{j-1}) 1_{(t_{j-1}, t_j]}(t).$$

For each fixed t, we have $f_m^{(1)}(t) = f(t_{j-1})$, where $t_{j-1} < t$ and $|t - t_{j-1}| \leqslant \|\mathcal{P}_m\|$. Thus, $f_m^{(1)}(t) \to f_-(t)$. By Theorem 3.24, there exists $M < \infty$ such that $|f_m^{(1)}| \leqslant M$ for all m. Thus, by dominated convergence, $I_- \to \int_{(a,b]} f_- \, d\mu_{G_+} = \int_{(a,b]} f_- \, dG$, and this proves (i).

Similarly,

$$I_+ = \sum_{j=1}^n f(t_j) \mu_{G+}((t_{j-1}, t_j]) = \int_{(a,b]} f_m^{(2)} \, d\mu_{G_+},$$

where

$$f_m^{(2)}(t) = \sum_{j=1}^n f(t_j) 1_{(t_{j-1}, t_j]}(t).$$

For each fixed t, we have $f_m^{(2)}(t) = f(t_j)$, where $t \leqslant t_j$ and $|t - t_j| \leqslant \|\mathcal{P}_m\|$. Because of the possibility that $t = t_j$, we cannot conclude that $f_m^{(2)}(t) \to f_+(t)$ as $m \to \infty$. However, if we make the further assumption that f is cadlag, so that $f_+ = f$, then we do obtain $f_m^{(2)}(t) \to f(t)$ as $m \to \infty$, and again by dominated convergence, we have $I_+ \to \int_{(a,b]} f \, d\mu_{G_+} = \int_{(a,b]} f \, dG$, and this proves (ii).

Next assume that G is caglad. Then $G = G_- = G_+ - \Delta G$, and so for any $s < t$, we have

$$\begin{aligned} G(t) - G(s) &= G_+(t) - G_+(s) - \Delta G(t) + \Delta G(s) \\ &= \mu_{G_+}((s,t]) - \mu_{G_+}(\{t\}) + \mu_{G_+}(\{s\}) \\ &= \mu_{G_+}([s,t)). \end{aligned}$$

Hence,

$$I_+ = \sum_{j=1}^n f(t_j) \mu_{G+}([t_{j-1}, t_j)) = \int_{[a,b)} f_m^{(3)} \, d\mu_{G_+},$$

where

$$f_m^{(3)}(t) = \sum_{j=1}^n f(t_j) 1_{[t_{j-1},t_j)}(t).$$

For each fixed t, we have $f_m^{(3)}(t) = f(t_j)$, where $t < t_j$ and $|t - t_j| \leqslant \|\mathcal{P}_m\|$. Thus, $f_m^{(3)}(t) \to f_+(t)$. Again, by dominated convergence, $I_+ \to \int_{[a,b)} f_+ \, d\mu_{G_+} = \int_{[a,b)} f_+ \, dG$, and this proves (iii).

Similarly,

$$I_- = \sum_{j=1}^n f(t_{j-1}) \mu_{G+}([t_{j-1}, t_j)) = \int_{[a,b)} f_m^{(4)} \, d\mu_{G_+},$$

where

$$f_m^{(4)}(t) = \sum_{j=1}^n f(t_{j-1}) 1_{[t_{j-1},t_j)}(t).$$

For each fixed t, we have $f_m^{(4)}(t) = f(t_{j-1})$, where $t_{j-1} \leqslant t$ and $|t - t_j| \leqslant \|\mathcal{P}_m\|$. Because of the possibility that $t = t_{j-1}$, we cannot conclude that $f_m^{(4)}(t) \to f_-(t)$ as $m \to \infty$. However, if we make the further assumption that f is caglad, so that $f_- = f$, then we do obtain $f_m^{(4)}(t) \to f(t)$ as $m \to \infty$, and again by dominated convergence, we have $I_- \to \int_{[a,b)} f \, d\mu_{G_+} = \int_{[a,b)} f \, dG$, and this proves (iv). □

Remark 3.32. In (ii) and (iv) of Theorem 3.31, the assumptions on f cannot be omitted. For example, let $f = 1_{(0,\infty)}$, $G = 1_{[0,\infty)}$, $a = -1$, and $b = 1$. In this case, G is cadlag and f is caglad, but $I_+^{(m)}$ need not converge to anything.

To see this, let $\{\mathcal{P}_m\}$ be a sequence of partitions with $\|\mathcal{P}_m\| \to 0$, satisfying the following conditions:

(i) If m is even, then there exists $k = k(m)$ such that $t_k^{(m)} = 0$.

(ii) If m is odd, then then there exists $k = k(m)$ such that $t_{k-1}^{(m)} < 0 < t_k^{(m)}$.

In this case, $G(t_j^{(m)}) - G(t_{j-1}^{(m)}) = 1$ if $j = k(m)$, and 0 otherwise. Thus,

$$I_+^{(m)} = f(t_k^{(m)}) = \begin{cases} 0 & \text{if } m \text{ is even,} \\ 1 & \text{if } m \text{ is odd,} \end{cases}$$

and so $I_+^{(m)}$ does not converge. Similarly, if $f = 1_{[0,\infty)}$, $G = 1_{(0,\infty)}$, and $\{\mathcal{P}_m\}$ are the above partitions, then $I_-^{(m)}$ does not converge.

Remark 3.33. If f and G are both cadlag, then

$$\int_{(a,b]} f \, dG - \int_{(a,b]} f_- \, dG = \int_{(a,b]} (f - f_-) \, dG = \int_{(a,b]} \Delta f \, d\mu_{G_+}.$$

Since Δf vanishes outside a countable set, we have

$$\int_{(a,b]} \Delta f \, d\mu_{G_+} = \sum_{t \in (a,b]} \Delta f(t) \mu_{G_+}(\{t\}) = \sum_{t \in (a,b]} \Delta f(t) \Delta G(t),$$

where this sum is, in fact, a countable sum. In particular, this shows that $I_-^{(m)}$ and $I_+^{(m)}$ need not converge to the same limit. More specifically,

$$I_+^{(m)} - I_-^{(m)} = \sum_{j=1}^{n} (f(t_j) - f(t_{j-1}))(G(t_j) - G(t_{j-1})) \to \sum_{t \in (a,b]} \Delta f(t) \Delta G(t),$$

as $m \to \infty$. This quantity is called the **covariation** of f and G.

If f has one-sided limits, but is not of bounded variation, then the integral $\int_{(a,b]} G \, df$ is undefined. More specifically, the map $(s,t] \mapsto f_+(t) - f_+(s)$ cannot be extended to a signed measure. But, even though the integral is undefined, we can still obtain convergence of the Riemann sums in Theorem 3.31, provided that the integrand is of bounded variation.

Theorem 3.34. *Let $G \in BV$ and let f be a function with one-sided limits. Assume f and G are both cadlag. Fix $a < b$. For each $m \in \mathbb{N}$, let $\mathcal{P}_m = \{t_j^{(m)}\}_{j=0}^{n(m)}$ be a strictly increasing, finite sequence of real numbers with $a = t_0^{(m)} < \ldots < t_{n(m)}^{(m)} = b$. Assume that $\|\mathcal{P}_m\| = \max\{|t_j^{(m)} - t_{j-1}^{(m)}| : 1 \leqslant j \leqslant n(m)\} \to 0$ as $m \to \infty$. Let*

$$J_-^{(m)} = \sum_{j=1}^{n(m)} G(t_{j-1}^{(m)})(f(t_j^{(m)}) - f(t_{j-1}^{(m)})), \text{ and}$$

$$J_+^{(m)} = \sum_{j=1}^{n(m)} G(t_j^{(m)})(f(t_j^{(m)}) - f(t_{j-1}^{(m)})).$$

Then

$$J_-^{(m)} \to f(b)G(b) - f(a)G(b) - \int_{(a,b]} f_- \, dG - \sum_{t \in (a,b]} \Delta f(t) \Delta G(t), \text{ and} \tag{3.2}$$

$$J_+^{(m)} \to f(b)G(b) - f(a)G(b) - \int_{(a,b]} f \, dG + \sum_{t \in (a,b]} \Delta f(t) \Delta G(t), \tag{3.3}$$

as $m \to \infty$.

Proof. As before, for notational simplicity, we will suppress the dependence of n, t_j, and $J_\pm$ on m.

We begin by observing that

$$
\begin{aligned}
J_- &= \sum_{j=1}^{n} G(t_{j-1})f(t_j) - \sum_{j=0}^{n-1} G(t_j)f(t_j) \\
&= f(b)G(b) - f(a)G(a) - \sum_{j=1}^{n} f(t_j)(G(t_j) - G(t_{j-1})).
\end{aligned}
$$

By Theorem 3.31, we have $J_- \to f(b)G(b) - f(a)G(b) - \int_{(a,b]} f\,dG$. By Remark 3.33, this prove (3.2).

Next, we write

$$
\begin{aligned}
J_+ &= \sum_{j=2}^{n+1} G(t_{j-1})f(t_{j-1}) - \sum_{j=1}^{n} G(t_j)f(t_{j-1}) \\
&= f(b)G(b) - f(a)G(a) - \sum_{j=1}^{n} f(t_{j-1})(G(t_j) - G(t_{j-1})).
\end{aligned}
$$

By Theorem 3.31, we have $J_+ \to f(b)G(b) - f(a)G(b) - \int_{(a,b]} f_-\,dG$. By Remark 3.33, this prove (3.3). □

As a corollary, we obtain the following generalizations of Theorem 3.22.

Corollary 3.35. *If f and G are both cadlag functions of bounded variation, then*

$$
\int_{(a,b]} G_-\,df = f(b)G(b) - f(a)G(a) - \int_{(a,b]} f_-\,dG - \sum_{t\in(a,b]} \Delta f(t)\Delta G(t), \text{ and}
$$

$$
\int_{(a,b]} G\,df = f(b)G(b) - f(a)G(a) - \int_{(a,b]} f\,dG + \sum_{t\in(a,b]} \Delta f(t)\Delta G(t).
$$

Proof. Combine Theorem 3.34 with Theorem 3.31. □

3.5.4 The Stratonovich integral for cadlag functions

If g and h are cadlag, with $h \in BV$, then let us define the **Stratonovich integral** of g with respect to h as

$$
\int_0^t g(s) \circ dh(s) := \int_{(0,t]} \frac{g_- + g}{2}\,dh
$$

By Theorem 3.31, we have

$$
\sum_{j=1}^{n} \frac{g(t_{j-1}) + g(t_j)}{2}(h(t_j) - h(t_{j-1})) \to \int_0^t g(s) \circ dh(s),
$$

as the mesh of the partition tends to zero.

Theorem 3.36. *Let f, g, and h be cadlag functions, with $h \in BV$. Let*

$$k(t) = \int_0^t g(s) \circ dh(s).$$

Then k is cadlag, $k \in BV$, and

$$\int_0^t f(s) \circ dk(s) = \int_0^t f(s)g(s) \circ dh(s) - \frac{1}{4} \sum_{s \in (0,t]} \Delta f(s) \Delta g(s) \Delta h(s).$$

Proof. Since

$$k(t) = \int_{(0,t]} \frac{g_- + g}{2} \, dh,$$

we have

$$\begin{aligned}
\int_0^t f(s) \circ dk(s) &= \int_{(0,t]} \frac{f_- + f}{2} \, dk \\
&= \int_{(0,t]} \left(\frac{f_- + f}{2} \right) \left(\frac{g_- + g}{2} \right) dh \\
&= \int_{(0,t]} \left(\frac{f_- g_- + fg}{2} - \frac{(f - f_-)(g - g_-)}{4} \right) dh \\
&= \int_{(0,t]} \frac{f_- g_- + fg}{2} \, dh - \frac{1}{4} \int_{(0,t]} \Delta f \Delta g \, dh \\
&= \int_0^t f(s)g(s) \circ dh(s) - \frac{1}{4} \sum_{s \in (0,t]} \Delta f(s) \Delta g(s) \Delta h(s),
\end{aligned}$$

and we are done. □

Remark 3.37. This theorem shows that as long as f, g, and h have no simultaneous discontinuities, then the Stratonovich integral satisfies the usual transformation rule of calculus that if $dk = g \circ dh$, then $f \circ dk = fg \circ dh$. In general, however, the transformation rule involves a correction term which represents the **triple covariation** of the three functions, f, g, and h.

Part II

The Foundations of Probability Theory

Chapter 4

The Principal Definitions

Probability theory is based on two things.

The first is the notion of a probability space. Recall that a probability space is a measure space, $(\Omega, \mathcal{F}, P)$, such that $P(\Omega) = 1$. If $A \in \mathcal{F}$, then $P(A)$ is called the **probability of** A.

The second is the notion of conditional probability. If $A, B \in \mathcal{F}$ and $P(B) > 0$, then the **conditional probability of** A **given** B is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Since $P(A) = P(A \mid \Omega)$ for all $A \in \mathcal{F}$, all probabilities are in fact conditional probabilities.

Proposition 4.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $A \in \mathcal{F}$ with $P(A) > 0$. Define $P^A : \mathcal{F} \to [0, 1]$ by $P^A(B) = P(B \mid A)$. Then P^A is a probability measure on $(\Omega, \mathcal{F})$ with $P^A(A) = 1$. Moreover, if $B \in \mathcal{F}$ and $P(A \cap B) > 0$, then $P^A(B) > 0$ and $P^{A \cap B}(C) = P^A(C \mid B)$ for all $C \in \mathcal{F}$.*

Proof. The proof that P^A is a probability measure with $P^A(A) = 1$ is left to the reader. For the second part, suppose $B \in \mathcal{F}$ and $P(A \cap B) > 0$. Then

$$P^A(B) = P(B \mid A) = \frac{P(B \cap A)}{P(A)} > 0,$$

so $P^A(C \mid B)$ is well-defined. We now have

$$\begin{aligned} P^{A \cap B}(C) = P(C \mid A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)} &= \frac{P(A \cap B \cap C)}{P(A)} \cdot \frac{P(A)}{P(A \cap B)} \\ &= \frac{P^A(B \cap C)}{P^A(B)} = P^A(C \mid B), \end{aligned}$$

for all $C \in \mathcal{F}$. ∎

Corollary 4.2. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. Let* $A, B, C \in \mathcal{F}$ *with* $P(C) > 0$.

(a) If $P(A \cap B \mid C) = 0$, *then* $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C)$.

(b) If $P(A \cap C) > 0$, *then* $P(A \cap B \mid C) = P(A \mid C)P(B \mid A \cap C)$.

Proof. Since P^C is a finite measure, we have

$$P^C(A \cup B) = P^C(A) + P^C(B) - P^C(A \cap B),$$

and (a) follows immediately.

Suppose $P(A \cap C) > 0$. Then

$$P^{A \cap C}(B) = P^C(B \mid A) = \frac{P^C(A \cap B)}{P^C(A)}.$$

Multiplying by $P^C(A)$ gives (b). □

The statements expressed in (a) and (b) are sometimes called the addition rule and the multiplication rule, respectively, and are often regarded as the foundational principles of probability. The addition rule comes primarily from the finite additivity[1] of the measure P, whereas the multiplication rule comes primarily from the definition of conditional probability. As such, the probability space structure and the definition of conditional probability together form the basis for probability theory. In Chapter 5, we will focus on the probability space structure. In Chapter 6, we will turn our attention to conditional probability.

Exercises

4.1. Complete the proof of Proposition 4.1.

[1]The countable additivity of P is needed so that we may take limits, which is especially important when we study stochastic processes later in these notes.

Chapter 5

Probability Spaces and Random Variables

5.1 Probability spaces

This section corresponds to [2, Section 1.1].

Recall that a probability space is a measure space, $(\Omega, \mathcal{F}, P)$, such that $P(\Omega) = 1$. Traditionally, elements $\omega \in \Omega$ are referred to as "outcomes" and measurable sets $A \in \mathcal{F}$ are referred to as "events".

Example 5.1. Let Ω be a countable set and $\mathcal{F} = 2^\Omega$. Let $p : \Omega \to [0, 1]$ be such that $\sum_{\omega \in \Omega} p(\omega) = 1$. Define $P : \mathcal{F} \to [0, 1]$ by $P(A) = \sum_{\omega \in A} p(\omega)$. Then $(\Omega, \mathcal{F}, P)$ is a probability space.

A common special case is where Ω is a finite set and $p(\omega) = 1/|\Omega|$ for all $\omega \in \Omega$. The resulting measure P, in this case, is often called the uniform probability measure on Ω.

For instance, if $\Omega = \{1, \ldots, 6\}$ and P is the uniform measure, then $(\Omega, \mathcal{F}, P)$ is a common model for the roll of a fair 6-sided die. The subset $A = \{2, 3, 5\}$ corresponds to the event that the die lands on a prime number.

Or if $\Omega = \{00, 0, 1, \ldots, 36\}$ and P is the uniform measure, then $(\Omega, \mathcal{F}, P)$ is a common model for the spin of a balanced American roulette wheel. To model the spin of an imbalanced American roulette wheel, we could use the same $(\Omega, \mathcal{F})$, but generate our probability measure P using a nonconstant function $p : \Omega \to [0, 1]$.

Remark 5.2. The elements $\omega \in \Omega$ can be thought to correspond to possible "states of the world." In the fair die example, with $\Omega = \{1, \ldots, 6\}$, the element $\omega = j$ can be thought to correspond to the aggregate of physical conditions that results in the die landing on the number j. Suppose S is the sentence, "The die lands on a prime number." Then the set $A = \{2, 3, 5\}$ consists of all states of the world that result in S being true. In this way, we can think of probabilities applying to propositions rather than sets. When thought of in this way, the set operations of union, intersection, and complement correspond to the logical operations of disjunction, conjunction, and negation, respectively. This point of view provides an intuitive foundation that is not only useful in applications, but also in constructing proofs and solving purely mathematical problems in probability theory. For a somewhat mild example, see Example 5.10.

5.2 Real-valued random variables

This section corresponds to [2, Section 1.2].

A **random variable** on a probability space $(\Omega, \mathcal{F}, P)$ is an $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$. Recall that $\mathcal{R}$ is the Borel σ-algebra on $\mathbb{R}$. Since X is $(\mathcal{F}, \mathcal{R})$-measurable, it follows that $X^{-1} : \mathcal{R} \to \mathcal{F}$. (In fact, X^{-1} is a Boolean σ-algebra homomorphism from $\mathcal{R}$ to $\mathcal{F}$.)

If $B \in \mathcal{R}$, then

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

It is very common practice in probability to omit ω from our notation, so instead of the above, we usually write

$$X^{-1}(B) = \{X \in B\} \in \mathcal{F}.$$

The set $\{X \in B\}$ is the event that the random variable takes a value in B. Since $\{X \in B\} \in \mathcal{F}$, it has a probability, $P(\{X \in B\})$. It is also common practice in probability to omit the curly brackets inside of our probability measure, so we usually just write this as $P(X \in B)$, which is read as "the probability that X is in B."

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. Define the function $\mu : \mathcal{R} \to [0, 1]$ by $\mu(A) = P(X \in A)$. Then μ is a Borel probability measure on $\mathbb{R}$, and is called the **distribution of** X, or the **law of** X. Note that $\mu = P \circ X^{-1}$. We write $X \sim \mu$ to indicate that X has distribution μ.

The function $F : \mathbb{R} \to [0, 1]$ given by $F(x) = \mu((-\infty, x])$ is called the **distribution function of** X. Note that $F(x) = P(X \leqslant x)$.

By Theorem 1.22, F is increasing and right-continuous. It is also easy to check, as in Exercise 1.12, that $F(\infty) = 1$, $F(-\infty) = 0$, $F(x-) = P(X < x)$, and $P(X = x) = F(x) - F(x-)$.

Sometimes, μ and F are denoted by μ_X and F_X to indicate their relationship to X.

Example 5.3. Let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}_{(0,1)}$, and let P be Lebesgue measure. Define $X : \Omega \to \mathbb{R}$ by $X(\omega) = \omega$. Then X is a random variable with distribution function $F : \mathbb{R} \to \mathbb{R}$. If $x \in (0, 1)$, then

$$F(x) = P(X \leqslant x) = P((0, x]) = x.$$

If $x \leqslant 0$, then

$$F(x) = P(X \leqslant x) = P(\varnothing) = 0,$$

and if $x \geqslant 1$, then

$$F(x) = P(X \leqslant x) = P((0,1)) = 1.$$

Thus,

$$F(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ x & \text{if } 0 < x < 1, \\ 1 & \text{if } x \geqslant 1. \end{cases}$$

Or, more compactly, $F(x) = x 1_{(0,1)}(x) + 1_{[1,\infty)}$.

The distribution of X is the Borel measure μ on $\mathbb{R}$ associated with F according to Theorem 1.22. In this case, $\mu(B) = \lambda(B \cap (0,1))$, where λ is Lebesgue measure. The measure μ in this example is called the **uniform distribution on** $(0,1)$.

More generally, if $C \in \mathcal{R}$, $0 < \lambda(C) < \infty$, and $\mu : \mathcal{R} \to [0,1]$ is given by $\mu(B) = \lambda(B \cap C)/\lambda(C)$, then μ is a Borel probability measure on $\mathbb{R}$ and is called the **uniform distribution on** C. We sometimes write $U(C)$ for this distribution. If C is an interval with endpoints $a < b$, then we write $U(a,b)$ for this distribution.

Any random variable on any probability space whose distribution is the uniform distribution on C is said to be **uniformly distributed on** C.

Theorem 5.4. *Let $F : \mathbb{R} \to [0,1]$ be increasing and right-continuous, with $F(\infty) = 1$ and $F(-\infty) = 0$. Then there exists a random variable X on some probability space $(\Omega, \mathcal{F}, P)$ such that $F = F_X$.*

Proof. If F is strictly increasing and continuous, then F has an inverse function $F^{-1} : [0,1] \to \mathbb{R}$. More generally, if F is merely increasing and right-continuous, then let us define, for $x \in (0,1)$,

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

If F is strictly increasing and continuous, then this definition agrees with the usual definition. In the general case, this function F^{-1} is sometimes called a pseudo-inverse. In the proof of [2, Theorem 1.2.2], it is shown that for all $(x,y) \in (0,1) \times \mathbb{R}$, we have $F^{-1}(x) \leqslant y$ if and only if $x \leqslant F(y)$.

Let U be a random variable on some probability space such that $U \sim U(0,1)$. Let $X = F^{-1}(U)$. We will show that $F = F_X$. Let $x \in \mathbb{R}$. Then

$$\begin{aligned} F_X(x) = P(X \leqslant x) = P(F^{-1}(U) \leqslant x) &= P(\{\omega : F^{-1}(U(\omega)) \leqslant x\}) \\ &= P(\{\omega : U(\omega) \leqslant F(x)\}) = P(U \leqslant F(x)) = F(x), \end{aligned}$$

which shows that $F = F_X$. □

Remark 5.5. There is an easier way to construct X from F. Namely, take $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{R}$, $P = \mu_F$, and define $X : \Omega \to \mathbb{R}$ by $X(\omega) = \omega$. However, the above proof is very useful. It shows, in a constructive fashion, how any distribution can be created from a uniform distribution. This is helpful for both theoretical and applied purposes. It can aid in the construction of certain proofs, and it is helpful in designing simulations.

On a probability space, $(\Omega, \mathcal{F}, P)$, the phrase "almost everywhere" is typically replaced by the phrase "**almost surely**" with the abbreviation a.s. For example, if X and Y are random variables defined on $(\Omega, \mathcal{F}, P)$, then $X = Y$ a.s. means there exists $N \in \mathcal{F}$ such that $P(N) = 0$ and $X(\omega) = Y(\omega)$ for all $\omega \in N^c$.

Let X be a random variable defined on a probability space, $(\Omega, \mathcal{F}, P)$, and let Y be a random variable defined on a possibly different probability space, $(\Omega', \mathcal{F}', P')$. Then both μ_X and μ_Y are probability measure on $(\mathbb{R}, \mathcal{R})$. If $\mu_X = \mu_Y$, then we say X and Y are **equal in distribution**, or **in law**, and we write

$$X \overset{d}{=} Y,$$

or $X =_d Y$. Note that $X =_d Y$ if and only if $F_X = F_Y$, that is, $X =_d Y$ if and only if $P(X \leqslant x) = P'(Y \leqslant x)$ for all $x \in \mathbb{R}$.

Note that the statement, $X = Y$ a.s., implies that X and Y are defined on the same probability space, but the statement $X =_d Y$ does not.

Let X be a random variable with distribution μ and distribution function F. Suppose that $\mu \ll \lambda$, where λ is Lebesgue measure, and let f be a version of the Radon-Nikodym derivative, $d\mu/d\lambda$. Then $d\mu = f\,d\lambda$ and, in particular,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

In this case, we say that f is the **density function of** X, and we sometimes denote it by f_X. Note that f is nonnegative, integrable, and $\int f(x)\,dx = 1$.

Conversely, if f is nonnegative and $\int f(x)\,dx = 1$, then we may define a probability measure by $d\mu = f\,d\lambda$. By Theorem 5.4, we have that μ is the distribution of some random variable, X. It then follows that f is the density of X.

By Theorem 3.23, the fundamental theorem of calculus for Lebesgue integrals,

$$\begin{aligned}\lim_{h \to 0^+} h^{-1} P(X \in [x, x+h]) &= \lim_{h \to 0^+} \frac{F(x+h) - F(x)}{h} \\ &= F'(x) = f(x),\end{aligned}$$

for Lebesgue a.e. $x \in \mathbb{R}$. Thus, if Δx is small, then

$$P(X \in [x, x + \Delta x]) \approx f(x)\Delta x.$$

You will sometimes see authors write $P(X = x) = f(x)$. Of course, this is false. If X has a density function, then $P(X = x) = 0$ for all $x \in \mathbb{R}$. The most gracious, and perhaps the only sensible, interpretation in such a situation is to assume the authors meant to refer to the above approximation.

Example 5.6. Fix $r > 0$ and let $f(x) = re^{-rx} 1_{(0,\infty)}(x)$. Then $f \geqslant 0$ and $\int f(x)\,dx = 1$, so there exists a random variable X that has density f. The distribution function of X is then

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x^+} re^{-rt}\,dt = (1 - e^{-rx})1_{(0,\infty)}(x).$$

The distribution of X is called the **exponential distribution with rate** r, and is denoted by $\mathrm{Exp}(r)$.

Note that $X \sim \mathrm{Exp}(r)$ if and only if $P(X > t) = e^{-rt}$ for all $t > 0$. Thus, if $X \sim \mathrm{Exp}(r)$ and $s > 0$, then $sX \sim \mathrm{Exp}(s^{-1}r)$.

Example 5.7. Let

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Then f is nonnegative and $\int f(x)\,dx = 1$, so there exists a random variable X that has density f. The distribution function of X is then

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt,$$

which has no closed form expression. The distribution of X is called the **standard normal distribution**, or **standard Gaussian distribution**, and is denoted by $N(0,1)$. The distribution function of a standard normal is typically denoted by Φ.

In [2, Theorem 1.2.3], it is shown that

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-x^2/2} \leqslant \int_{x}^{\infty} e^{-t^2/2}\,dt \leqslant \frac{1}{x} e^{-x^2/2},$$

for all $x > 0$. Thus, if $X \sim N(0,1)$, then

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-x^2/2} \leqslant P(X > x) \leqslant \frac{1}{\sqrt{2\pi}\,x} e^{-x^2/2},$$

for all $x > 0$.

Let X be a random variable with $X \sim \mu$. Recall from Proposition 3.15 that μ has a unique decomposition $\mu = \mu_d + \mu_{ac} + \mu_{sc}$, where μ_d is discrete, $\mu_{ac} \ll \lambda$, and μ_{sc} is continuous with $\mu_{sc} \perp \lambda$. Note that X has a density function if and only if $\mu \ll \lambda$, which holds if and only if $\mu_d = \mu_{sc} = 0$. Also note that F_X is continuous if and only if μ is continuous, which holds if and only if $\mu_d = 0$.

Example 5.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Choose $c \in \mathbb{R}$ and define $X : \Omega \to \mathbb{R}$ by $X(\omega) = c$ for all $\omega \in \Omega$. Then $X \sim \delta_c$, where δ_c is the point mass measure at c. In this case, X does not have a density function. In order for X to have a density function, the distribution of X must be absolutely continuous with respect to Lebesgue measure. But in this case, $\delta_c \perp \lambda$, that is, the distribution of X is singular with respect to Lebesgue measure. Note that in this example, the distribution of X is discrete.

Example 5.9. Let F be the Cantor function and let X be a random variable with distribution function F. Then $X \sim \mu_F$. Since F is continuous, the measure μ_F is continuous. However, $\mu_F \perp \lambda$, so X does not have a density function.

Example 5.10. Suppose $\{X_n\}_{n=1}^\infty$ is a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Let

$$A = \left\{\omega : \lim_{n\to\infty} X_n(\omega) = \infty\right\},$$

or, as we would more commonly write it in probability,

$$A = \left\{\lim_{n\to\infty} X_n = \infty\right\}.$$

Note that $X_n \to \infty$ if and only if

$$\forall M \in \mathbb{N}, \exists N \in \mathbb{N}, \forall n \geqslant N, X_n > M.$$

Guided by the intuitive understanding described in Remark 5.2, we are led immediately to

$$A = \bigcap_{M\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{n=N}^{\infty} \{X_n > M\}.$$

Written in this way, we immediately see that $A \in \mathcal{F}$. Moreover, we may now be able to say something about $P(A)$ using what we may know about $P(X_n > M)$. We will have many opportunities to work with examples like this, especially when discussing limit theorems for discrete-time stochastic processes, such as the law of large numbers and the central limit theorem.

Exercises

5.1. [2, Exercise 1.2.2] Let $X \sim N(0,1)$. Use [2, Theorem 1.2.3] to find upper and lower bounds on $P(X \geqslant 4)$.

5.2. [2, Exercise 1.2.4] Let X be a random variable, and let $Y = F_X(X)$. Show that if F_X is continuous, then $Y \sim U(0,1)$.

5.3. [2, Exercise 1.2.6] Let $X \sim N(0,1)$. Find the density of $Y = e^X$.

5.4. [2, Exercise 1.2.7(i)] Let X be a random variable with density function f. Find the density function of X^2 in terms of f.

5.5. [2, Exercise 1.2.5] Let $-\infty \leqslant \alpha < \beta \leqslant \infty$ and let X be a random variable with $P(X \in (\alpha,\beta)) = 1$. Assume X has a continuous density function f. Let $g : (\alpha,\beta) \to \mathbb{R}$ be strictly increasing and differentiable, and define $Y = g(X)$.

(a) Prove that Y has a density function,

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} 1_{(g(\alpha),g(\beta))}(y).$$

(b) Show that in the case $g(x) = ax + b$, where $a > 0$, the density reduces to

$$h(y) = \frac{1}{a} f\left(\frac{y-b}{a}\right).$$

5.3 General random variables

This section corresponds to [2, Section 1.3].

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $(S, \mathcal{S})$ a measurable space. An S**-valued random variable** is an $(\mathcal{F}, \mathcal{S})$-measurable function $X : \Omega \to S$. Since X is $(\mathcal{F}, \mathcal{S})$-measurable, it follows that $X^{-1} : \mathcal{S} \to \mathcal{F}$. (In fact, X^{-1} is a Boolean σ-algebra homomorphism from $\mathcal{S}$ to $\mathcal{F}$.)

If $B \in \mathcal{S}$, then

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F},$$

which we usually write as

$$X^{-1}(B) = \{X \in B\} \in \mathcal{F}.$$

The set $\{X \in B\}$ is the event that the X takes a value in B, and $P(X \in B)$ is the probability of this event.

Let X be an S-valued random variable. Define the function $\mu : \mathcal{S} \to [0, 1]$ by $\mu(A) = P(X \in A)$. Then μ is a probability measure on $(S, \mathcal{S})$, and is called the **distribution of** X, or the **law of** X. Note that $\mu = P \circ X^{-1}$.

We write $X \sim \mu$ to indicate that X has distribution μ. More generally, if ν is a finite, nontrivial measure on $(S, \mathcal{S})$, then we write $X \sim \nu$ to indicate that X has distribution $\nu/\nu(S)$.

Sometimes μ is denoted by μ_X to indicate its relationship to X. If X and Y are S-valued random variables defined on possibly different probability spaces, then we write

$$X \overset{d}{=} Y,$$

or $X =_d Y$, to mean that $\mu_X = \mu_Y$, and we say X and Y are **equal in distribution**, or **in law**.

If μ is any probability measure on $(S, \mathcal{S})$, then we can create a random variable with distribution μ simply by taking $\Omega = S$, $\mathcal{F} = \mathcal{S}$, $P = \mu$, and defining $X : \Omega \to S$ by $X(\omega) = \omega$.

Typically, when we say X is a random variable, we will mean X is a real-valued random variable, unless otherwise specified, either explicitly or by context. An S-valued random variable may sometimes be called a random element of S. If elements of S have a particular names, we may use that instead. For example, an $\mathbb{R}^n$-valued random variable may sometimes be called a random vector.

Example 5.11. Let $S = C[0, 1]$, the set of all continuous functions from $[0, 1]$ to $\mathbb{C}$. The mapping $f \mapsto \|f\| = \sup_{x \in [0,1]} |f(x)|$ defines a norm on S, which yields a metric. We can therefore let $\mathcal{S} = \mathcal{B}_S$ be the Borel σ-algebra on S. An S-valued random variable is an $\mathcal{F}$-measurable function $X : \Omega \to S$.

For each $\omega \in \Omega$, we have that $X(\omega)$ is a function from $[0, 1]$ to $\mathbb{C}$. We typically write $X(t, \omega) = (X(\omega))(t)$. The object X is a random continuous function on $[0, 1]$. An example of such a $C[0, 1]$-valued random variable, which we will learn about much later on, is Brownian motion.

5.4 Expected value

This section corresponds to [2, Section 1.6].

If X is a real-valued random variable on $(\Omega, \mathcal{F}, P)$, then the **expected value of** X is defined as

$$E[X] = \int_\Omega X\,dP,$$

provided this integral exists. We frequently drop the brackets, writing EX when non confusion is likely to arise. The expected value of X is also called the **mean of** X**.**

On a probability space, constant random variables are integrable, and the expected value of a constant is that constant itself. So in addition to the usual properties of integration (linearity, monotonicity, etc.), we also have that $E[aX + b] = aE[X] + b$.

You will occasionally see the notation $E[X; A] = E[X1_A]$, where $A \in \mathcal{F}$. We will try to avoid this notation in these notes, writing simply $E[X1_A]$.

Since the expected value is an integral, we have at our disposal all the usual inequalities and limit theorems from measure and integration, such as Jensen's inequality, Hölder's inequality, Fatou's lemma, and so on. In addition to these, here are two theorems that will be especially useful to us.

The following theorem is [2, Theorem 1.6.4].

Theorem 5.12 (Chebyshev's inequality)**.** *Let X be a real-valued random variable and $\varphi : \mathbb{R} \to [0, \infty)$. For $A \in \mathcal{R}$, let $m_A = \inf\{\varphi(x) : x \in A\}$. Then*

$$P(X \in A) \leqslant \frac{E[\varphi(X)]}{m_A},$$

for all $A \in \mathcal{R}$ such that $m_A > 0$.

Proof. Since $\varphi \geqslant 0$, we have

$$E[\varphi(X)] \geqslant E[\varphi(X)1_{\{X\in A\}}] \geqslant E[m_A 1_{\{X\in A\}}] = m_A E[1_{\{X\in A\}}] = m_A P(X \in A).$$

Dividing by m_A finishes the proof. □

Remark 5.13. Many authors use the phrase "Chebyshev's inequality" to refer to the special case, $\varphi(x) = |x|^r$ and $A = (-a, a)^c$. In this case, the inequality reduces to $P(|X| \geqslant a) \leqslant E[|X|^r]/a^r$.

If $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = L$, then we say $f(x) \to L$ as $|x| \to \infty$. The following is [2, Theorem 1.6.8]. See the book for the proof.

Theorem 5.14. *Let X_n, X be real-valued random variables with $X_n \to X$ a.s. Let $g : \mathbb{R} \to [0, \infty)$ and $h : \mathbb{R} \to \mathbb{R}$ be continuous. Assume that*

(i) $g(x) \to \infty$ *as* $|x| \to \infty$*,*

(ii) $|h(x)|/g(x) \to 0$ *as* $|x| \to \infty$*, and*

(iii) $\sup_n Eg(X_n) < \infty$.

Then $Eh(X_n) \to Eh(X)$ *as* $n \to \infty$.

Corollary 5.15. *Let* X_n, X *be real-valued random variables with* $X_n \to X$ *a.s. Let* $p > 1$ *and suppose* $\sup_n E|X_n|^p < \infty$*. Then* $EX_n \to EX$ *as* $n \to \infty$.

Proof. Take $g(x) = |x|^p$ and $h(x) = x$. □

Let X be an S-valued random variable, where $(S, \mathcal{S})$ is a measurable space. If $g : S \to \mathbb{R}$ is $\mathcal{S}$-measurable, then $g(X)$ is a real-valued random variable. Assume either $g(X) \geqslant 0$ a.s. or $E|g(X)| < \infty$. Then by Theorem 2.32, we have

$$Eg(X) = \int_\Omega g \circ X \, dP = \int_S g \, d(P \circ X^{-1}) = \int_S g \, d\mu,$$

where μ is the distribution of X. For example, if $(S, \mathcal{S}) = (\mathbb{R}^n, \mathcal{R}^n)$, then this formula allows us to compute expected values by performing integrals on the more familiar space $\mathbb{R}^n$, rather than some abstract probability space $(\Omega, \mathcal{F}, P)$.

Example 5.16. Let X be a real-valued random variable with distribution μ and distribution function F. If $X \geqslant 0$ or $E|X| < \infty$, then taking $g(x) = x$ in the above, we have

$$EX = \int_{\mathbb{R}} x \, \mu(dx).$$

Since μ is the Lebesgue-Stieltjes measure associated with F, we can also write this as

$$EX = \int_{\mathbb{R}} x \, dF(x).$$

More generally,

$$Eg(X) = \int_{\mathbb{R}} g(x) \, dF(x),$$

whenever $g : \mathbb{R} \to \mathbb{R}$ is measurable and either $g(X) \geqslant 0$ a.s. or $E|g(X)| < \infty$.

Example 5.17. Let X be a real-valued random variable with density f. Then $X \sim \mu$, where $d\mu = f(x) \, dx$. As a special case of Example 5.16, if $X \geqslant 0$ or $E|X| < \infty$, then

$$EX = \int_{\mathbb{R}} x f(x) \, dx.$$

More generally,

$$Eg(X) = \int_{\mathbb{R}} g(x) f(x) \, dx,$$

whenever $g : \mathbb{R} \to \mathbb{R}$ is measurable and either $g(X) \geqslant 0$ a.s or $E|g(X)| < \infty$.

Example 5.18. Let X be a real-valued random variable with a discrete distribution. That is, $X \sim \mu$ and there exists a countable $S \subset \mathbb{R}$ such that $\mu(S) = 1$. As a special case of Example 5.16, if $X \geqslant 0$ or $E|X| < \infty$, then

$$EX = \sum_{x \in S} x\mu(\{x\}) = \sum_{x \in S} xP(X = x).$$

More generally,

$$Eg(X) = \sum_{x \in S} g(x)P(X = x),$$

whenever $g : \mathbb{R} \to \mathbb{R}$ is measurable and either $g(X) \geqslant 0$ a.s. or $E|g(X)| < \infty$.

If $k \in \mathbb{N}$, then $E[X^k]$ is called the **k-th moment** of X. Suppose X is a square-integrable random variable, that is, X has a finite second moment. By Exercise 2.16, this implies X is integrable. Let $\mu = EX \in \mathbb{R}$. Then the **variance** of X is defined to be $\operatorname{var}(X) = E[|X - \mu|^2]$. Note that

$$\operatorname{var}(X) = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2.$$

From this, we get $\operatorname{var}(X) \leqslant E[X^2]$. Also, suppose $Y = aX + b$, where $a, b \in \mathbb{R}$. Then $EY = a\mu + b$, so $Y - EY = a(X - \mu)$. This gives

$$\operatorname{var}(Y) = E[|Y - EY|^2] = a^2 E[|X - \mu|^2] = a^2 \operatorname{var}(X).$$

The **standard deviation** of X is defined as $\sqrt{\operatorname{var}(X)}$.

Example 5.19. Recall Example 5.6 and let $X \sim \operatorname{Exp}(1)$. Then $X \geqslant 0$ a.s. and

$$E[X^k] = \int_0^\infty x^k e^{-x}\,dx,$$

for all $k \in \mathbb{N}$. Using integration by parts and induction, one finds that $E[X^k] = k!$. Thus, $EX = 1$ and $EX^2 = 2$, which gives $\operatorname{var}(X) = 2 - 1^2 = 1$.

Let $r > 0$ and $Y = X/r$. Then $Y \sim \operatorname{Exp}(r)$ and $E[Y^k] = E[X^k]/r^k = k!/r^k$ for all $k \in \mathbb{N}$. In particular, $EY = 1/r$ and $EY^2 = 2/r^2$, which gives $\operatorname{var}(Y) = 1/r^2$. In other words, if Y is exponentially distributed with parameter r, then the mean and standard deviation of Y are both $1/r$.

Example 5.20. Let us define $0!! = 1!! = 1$ and, for integers $n > 1$, let $n!! = n((n-2)!!)$. The number $n!!$ is called the **double factorial** of n.

Recall Example 5.7 and let $X \sim N(0, 1)$. Then

$$EX^k = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^k e^{-x^2/2}\,dx,$$

for all $k \in \mathbb{N}$. Using integration by parts and symmetry, one can show that

$$EX^k = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ (k-1)!! & \text{if } k \text{ is even.} \end{cases}$$

In particular, $EX = 0$ and $EX^2 = 1$, which gives $\operatorname{var}(X) = 1$.

Let $\sigma, \mu \in \mathbb{R}$ with $\sigma \neq 0$, and let $Y = \sigma X + \mu$. Then $EY = \mu$ and $\operatorname{var}(Y) = \sigma^2$. By Exercise 5.5, the random variable Y has density,

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}.$$

The distribution of Y is called the **normal** (or **Gaussian**) **distribution** with mean μ and variance σ^2, and is denoted by $N(\mu, \sigma^2)$. Note that

$$e^{-(y-\mu)^2/2\sigma^2} = \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

Although this is a violation of the order of operations, it is a standard abuse of notation when writing the density of the normal distribution.

Example 5.21. Let $p \in [0,1]$ and let X be a random variable with $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The distribution of X is $p\delta_1 + (1-p)\delta_0$, which is a discrete distribution. This distribution is called the **Bernoulli distribution** with parameter p, and is denoted by $\operatorname{Bernoulli}(p)$. Note that

$$Eg(X) = \sum_{x \in \{0,1\}} g(x)P(X = x) = g(0)(1-p) + g(1)p.$$

In particular, we have $EX^k = p$ for all $k \in \mathbb{N}$. Thus, $\operatorname{var}(X) = p - p^2 = p(1-p)$. A frequently useful observation is that $X^k = X$ a.s., for any $k \in \mathbb{N}$.

Example 5.22. Let $r > 0$ and let X be a random variable such that

$$P(X = k) = e^{-r}\frac{r^k}{k!},$$

for all nonnegative integers k. The distribution of X is called the **Poisson distribution** with parameter r, and is denoted by $\operatorname{Poisson}(r)$.

Let $n \in \mathbb{N}$ and note that

$$\begin{aligned} E\left[\prod_{j=0}^{n-1}(X-j)\right] &= \sum_{k \in \mathbb{N}\cup\{0\}} \left(\prod_{j=0}^{n-1}(k-j)\right) P(X = k) = \sum_{k=0}^{\infty}\left(\prod_{j=0}^{n-1}(k-j)\right) e^{-r}\frac{r^k}{k!} \\ &= e^{-r}\sum_{k=n}^{\infty}\left(\frac{k!}{(k-n)!}\right)\frac{r^k}{k!} = e^{-r}r^n\sum_{k=0}^{\infty}\frac{r^k}{k!} = r^n. \end{aligned}$$

Taking $n = 1$ gives $EX = r$, and taking $n = 2$ gives $E[X(X-1)] = r^2$. Thus,

$$EX^2 = E[X(X-1)] + EX = r^2 + r,$$

and, therefore, $\operatorname{var}(X) = r^2 + r - r^2 = r$.

Example 5.23. Let $p \in (0,1)$ and let X be a real-valued random variable with

$$P(X = k) = p(1-p)^{k-1},$$

for each $k \in \mathbb{N}$. The distribution of X is called the **geometric distribution** with parameter p, and is denoted by $\mathrm{Geom}(p)$.

By Theorem 2.31,

$$\begin{aligned} EX = \sum_{k \in \mathbb{N}} kP(X = k) &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} \\ &= -p\frac{d}{dp}\sum_{k=0}^{\infty}(1-p)^k = -p\frac{d}{dp}\left(\frac{1}{p}\right) = \frac{1}{p}, \end{aligned}$$

and, similarly,

$$E[X(X-1)] = \sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1} = p(1-p)\frac{d^2}{dp^2}\left(\frac{1}{p}\right) = \frac{2(1-p)}{p^2}.$$

Thus,

$$\begin{aligned} \mathrm{var}(X) = E[X^2] - (EX)^2 &= E[X(X-1)] + EX - (EX)^2 \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}. \end{aligned}$$

A similar technique could be used to derive higher moments of X.

As a closing remark, let us mention that expected values can be used to derive the inclusion-exclusion formula:

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) &+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \\ &- \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right). \end{aligned}$$

See [2, Exercise 1.6.9].

Exercises

5.6. [2, Exercise 1.6.7] Let $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}_{(0,1)}$, and let P be Lebesgue measure. Fix $\alpha \in (1,2)$ and define $X_n = n^{\alpha} 1_{(1/(n+1),1/n)}$.

(a) Show that there does not exist an integrable random variable Y such that $|X_n| \leqslant Y$ a.s. for each $n \in \mathbb{N}$. In other words, the dominated convergence theorem does not apply to this sequence of random variables.

(b) Use Theorem 5.14 to show that $EX_n \to 0$ as $n \to \infty$.

Hint: Use $g(x) = |x|^{2/\alpha}$.

5.7. [2, Exercise 1.6.6] Let X be a nonnegative random variable with $EX^2 < \infty$. Prove that

$$P(X > 0) \geqslant \frac{(EX)^2}{EX^2}.$$

Hint: Apply Cauchy-Schwarz to $X1_{\{X>0\}}$.

5.8. [2, Exercise 1.6.14] Let X be a nonnegative random variable. Prove that

$$\lim_{x\to\infty} xE[X^{-1}1_{\{X>x\}}] = \lim_{x\downarrow 0} xE[X^{-1}1_{\{X>x\}}] = 0.$$

Warning: Be careful not to assume that $E[X^{-1}] < \infty$.

Chapter 6

Independence and Conditional Expectation

6.1 Conditional probability and independence

This section corresponds to [2, Section 2.1].

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose $B \in \mathcal{F}$ and $P(B) > 0$. Recall that the conditional probability of A given B is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Clearly, if $B \subset A$, then $P(A \mid B) = 1$. The following lemma provides a converse result.

Lemma 6.1. *Let $A, B \in \mathcal{F}$ with $P(B) > 0$. Then the following are equivalent.*

(a) $P(A \mid B) = 1$.

(b) $P(B \backslash A) = 0$.

(c) $P(B^c \cup A) = 1$.

(d) There exists $N \in \mathcal{F}$ such that $P(N) = 0$ and $B \subset A \cup N$.

Remark 6.2. Recall Remark 5.2, in which we discussed the heuristic identification of measurable sets with propositions. We noted that the set operations of union, intersection, and complement correspond to the logical operations of disjunction, conjunction, and negation, respectively. We now note that the subset relation corresponds to the relation of logical implication.

When we think of A and B as representing propositions, we usually interpret $P(A \mid B) = 1$ as meaning that B logically implies A. While it is not quite true that $P(A \mid B) = 1$ if and only if $B \subset A$, it is true up to null sets, as expressed by the equivalence between (a) and (d).

It is also interesting to point out that in propositional logic, there is a difference between logical implication and material implication (sometimes just called implication). Material implication is an operation, denoted by $\to$. If S and T are propositions, then $T \to S$ is the statement, "T implies S", which is false when T is true and S is false, and is true otherwise. The statement $T \to S$ is logically equivalent to $(\sim T) \vee S$.

To say that T logically implies S is to say that $T \to S$ is a tautology, or equivalently, that $(\sim T) \vee S$ is a tautology. The analogue for events in a probability space is the equivalence between (a) and (c).

As Example 5.10 demonstrates, the events we work with in probability are frequently cumbersome to write down, making it sometimes difficult to tell at a glance when one event is a subset of another. The heuristic identification between the subset relation and logical implication is frequently useful in making this determination. We often think in terms of logical relations and operations, while working with set relations and operations.

Proof of Lemma 6.1. Let $A, B \in \mathcal{F}$ with $P(B) > 0$. Since

$$P(B) = P(A \cap B) + P(A^c \cap B),$$

it follows that $P(A \cap B) = P(B)$ if and only if $P(A^c \cap B) = 0$. Thus, (a) and (b) are equivalent. Also, since $(B \backslash A)^c = B^c \cup A$, it follows that (b) and (c) are equivalent.

Now assume (b). Let $N = B \backslash A$, so that $P(N) = 0$. Since $A \cup N = A \cup B$, this gives (d).

Finally, assume (d). Then

$$B \cap A^c \subset (A \cup N) \cap A^c = N \cap A^c \subset N.$$

Thus, $P(B \cap A^c) \leqslant P(N) = 0$, which gives (b). □

If $A, B \in \mathcal{F}$, and $P(A \cap B) = P(A)P(B)$, then we say A and B are **independent**.

Lemma 6.3. *Let $A, B \in \mathcal{F}$ with $P(A) > 0$ and $P(B) > 0$. Then the following are equivalent.*

(a) $P(A \mid B) = P(A)$.

(b) $P(B \mid A) = P(B)$.

(c) A and B are independent.

Proof. Let $A, B \in \mathcal{F}$ with $P(A) > 0$ and $P(B) > 0$.

Suppose $P(A \mid B) = P(A)$. Then

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B)}{P(B)} \frac{P(B)}{P(A)} = P(A \mid B) \frac{P(B)}{P(A)} = P(B).$$

By reversing the roles of A and B, we have that (a) and (b) are equivalent. Since $P(A \cap B) = P(A)P(B \mid A)$ and $P(A) > 0$, we have that (b) and (c) are equivalent. □

In words, $P(A \mid B) = P(A)$ means that the probability of A remains unaffected, whether we are given B or not. In other words, A is independent of B. The lemma shows that this relationship between A and B is symmetric, so we need only say that A and B are independent.

The lemma only applies when $P(A)$ and $P(B)$ are both positive, whereas the definition of independence makes sense even when one of these probabilities is zero.

Lemma 6.4. *Let $A, B \in \mathcal{F}$. If $P(A) \in \{0, 1\}$, then A and B are independent.*

Proof. If $P(A) = 0$, then $P(A \cap B) \leqslant P(A) = 0$, so $P(A \cap B) = P(A)P(B)$, and A and B are independent.

Suppose $P(A) = 1$. Then $P(A^c \cap B) \leqslant P(A^c) = 0$, so

$$P(A)P(B) = P(B) = P(A \cap B) + P(A^c \cap B) = P(A \cap B),$$

and A and B are independent. □

Example 6.5. Let $\Omega = \{1, \dots, 6\}$, $\mathcal{F} = 2^\Omega$, and let P be given by $P(A) = |A|/6$. Let $A = \{2, 3, 5\}$ and $B = \{1, 2, 3, 4\}$. If we interpret this probability space as modeling a single roll of a fair 6-sided die, then A is the event that we roll a prime number, and B is the event that we roll a number less than 5.

Note that $P(A \cap B) = P(\{2, 3\}) = 1/3$, whereas $P(A)P(B) = (1/2)(2/3) = 1/3$. Thus, $P(A \cap B) = P(A)P(B)$, and so A and B are independent.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be measurable spaces. Let X_j be an S_j-valued random variable. We say that X_1 and X_2 are **independent** if

$$P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2),$$

for all $B_j \in \mathcal{S}_j$. Note that $P(X_1 \in B_1, X_2 \in B_2)$ is shorthand for $P(\{X_1 \in B_1\} \cap \{X_2 \in B_2\})$. The use of the comma to mean intersection is common practice in probability.

Proposition 6.6. *Let $A, B \in \mathcal{F}$. The events A and B are independent if and only if the random variables 1_A and 1_B are independent.*

Proof. Exercise 6.2. □

Let $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{H} \subset \mathcal{F}$ be σ-algebras. We say that $\mathcal{G}$ and $\mathcal{H}$ are **independent** if, for all $A \in \mathcal{G}$ and $B \in \mathcal{H}$, the events A and B are independent.

Proposition 6.7. *Let X and Y be random variables, and $\mathcal{G}$ and $\mathcal{H}$ σ-algebras. If X and Y are independent, then so are $\sigma(X)$ and $\sigma(Y)$. Conversely, if $\mathcal{G}$ and $\mathcal{H}$ are independent, and $X \in \mathcal{G}$ and $Y \in \mathcal{H}$, then X and Y are independent.*

Proof. Exercise 6.1. □

6.1.1 Independence of a set

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A_1, \ldots, A_n \in \mathcal{F}$. Then $A_1, \ldots, A_n$ are **independent** if, for all $I \subset \{1, \ldots, n\}$,

$$P\left(\bigcap_{j \in I} A_j\right) = \prod_{j \in I} P(A_j). \tag{6.1}$$

We say that $A_1, \ldots, A_n$ are **pairwise independent** if A_i and A_j are independent whenever $i \neq j$. By taking $I = \{i, j\}$, we see that independence implies pairwise independence, but the converse is not true. See [2, Example 2.1.1] for an example of three events that are pairwise independent, but not independent.

For each $j \in \{1, \ldots, n\}$, let $(S_j, \mathcal{S}_j)$ be a measurable space, and let X_j be an S_j-valued random variable. Then $X_1, \ldots, X_n$ are **independent** if

$$P\left(\bigcap_{j=1}^{n} \{X_j \in B_j\}\right) = \prod_{j=1}^{n} P(X_j \in B_j), \tag{6.2}$$

for all $B_j \in \mathcal{S}_j$.

On the surface, it looks like the structure of (6.2) is less general than (6.1), because it does not explicitly mention subsets of $\{1, \ldots, n\}$. However, with a small trick, we can see that it does, in fact, cover such subsets. Suppose $I \subset \{1, \ldots, n\}$. For $j \in I$, let $B_j \in \mathcal{S}_j$ be arbitrary. For $j \notin I$, let $B_j = S_j$. Since $\{X_j \in S_j\} = \Omega$ and $P(X_j \in S) = 1$, the above equality becomes

$$P\left(\bigcap_{j \in I} \{X_j \in B_j\}\right) = \prod_{j \in I} P(X_j \in B_j),$$

for all $B_j \in \mathcal{S}_j$.

Proposition 6.8. *Let $A_1, \ldots, A_n \in \mathcal{F}$. The events $A_1, \ldots, A_n$ are independent if and only if the random variables $1_{A_1}, \ldots, 1_{A_n}$ are independent.*

Proof. Exercise 6.3. □

For $j \in \{1, \ldots, n\}$, let $\mathcal{G}_j \subset \mathcal{F}$ be a σ-algebra. Then $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are **independent** if

$$P\left(\bigcap_{j=1}^{n} A_j\right) = \prod_{j=1}^{n} P(A_j),$$

whenever $A_j \in \mathcal{G}_j$. As above, suppose $I \subset \{1, \ldots, n\}$. For $j \in I$, let $A_j \in \mathcal{G}_j$ be arbitrary, and for $j \notin I$, let $A_j = \Omega \in \mathcal{G}_j$. Then the above equality becomes $P(\bigcap_{j \in I} A_j) = \prod_{j \in I} P(A_j)$.

Proposition 6.9. *Let $X_1, \ldots, X_n$ be random variables, and let $\mathcal{G}_1, \ldots, \mathcal{G}_n$ be σ-algebras. If $X_1, \ldots, X_n$ are independent, then so are $\sigma(X_1), \ldots, \sigma(X_n)$. Conversely, if $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent, and $X_j \in \mathcal{G}_j$, then $X_1, \ldots, X_n$ are independent.*

Proof. Exercise 6.4. □

More generally, if $\{A_\alpha\}_{\alpha\in\Lambda} \subset \mathcal{F}$ is any collection of events, then $\{A_\alpha\}_{\alpha\in\Lambda}$ is independent if, for all $F \subset \Lambda$ such that F is finite, $\{A_\alpha\}_{\alpha\in F}$ is independent. The analogous definitions for infinite collections of random variables or σ-algebras also holds.

6.1.2 Sufficient conditions for independence

Our first order of business is to show that when checking independence of σ-algebras, it is sufficient to check the product formula for collections of events that generate the σ-algebras, provided those collections are closed under intersections. (See Theorem 6.11 below.) The proof of this theorem is an excellent example of the use of the π-λ theorem (Theorem 2.3.) To state and prove the theorem, however, we first need a definition and a lemma.

For each $j \in \{1, \ldots, n\}$, suppose $\mathcal{E}_j \subset \mathcal{F}$. Note that we are not assuming that $\mathcal{E}_j$ is a σ-algebra. Then $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are **independent** if, for all $I \subset \{1, \ldots, n\}$,

$$P\bigg(\bigcap_{j\in I} A_j\bigg) = \prod_{j\in I} P(A_j),$$

whenever $A_j \in \mathcal{E}_j$.

Lemma 6.10. *For each $j \in \{1, \ldots, n\}$, let $\mathcal{E}_j \subset \mathcal{F}$, and let $\overline{\mathcal{E}}_j = \mathcal{E}_j \cup \{\Omega\}$. Then $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent if and only if $\overline{\mathcal{E}}_1, \ldots, \overline{\mathcal{E}}_n$ are independent, which holds if and only if*

$$P\bigg(\bigcap_{j=1}^{n} A_j\bigg) = \prod_{j=1}^{n} P(A_j),$$

whenever $A_j \in \overline{\mathcal{E}}_j$.

Proof. The second equivalence holds by taking $A_j = \Omega$ whenever $j \notin I$. In the first equivalence, the "if" part is trivial. For the "only if" part, suppose $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent. For each $j \in \{1, \ldots, n\}$, let $A_j \in \overline{\mathcal{E}}_j$. Let $I = \{j : A_j \neq \Omega\}$. Then $A_j \in \mathcal{E}_j$ whenever $j \in I$, and so we have

$$P\bigg(\bigcap_{j=1}^{n} A_j\bigg) = P\bigg(\bigcap_{j\in I} A_j\bigg) = \prod_{j\in I} P(A_j) = \prod_{j=1}^{n} P(A_j),$$

which shows that $\overline{\mathcal{E}}_1, \ldots, \overline{\mathcal{E}}_n$ are independent. □

Recall the π-λ theorem (Theorem 2.3) from Section 2.1.

Theorem 6.11. *Suppose $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent and each $\mathcal{E}_j$ is a π-system. Then $\sigma(\mathcal{E}_1), \ldots, \sigma(\mathcal{E}_n)$ are independent.*

Proof. Let $\mathcal{E}_j \subset \mathcal{F}$ and suppose each $\mathcal{E}_j$ is a π-system. It suffices to prove that if $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent, then $\sigma(\mathcal{E}_1), \mathcal{E}_2, \ldots, \mathcal{E}_n$ are independent. Indeed, suppose for the moment that this implication holds. Then applying it to $\mathcal{E}_2, \ldots, \mathcal{E}_n, \sigma(\mathcal{E}_1)$ shows that $\sigma(\mathcal{E}_1), \sigma(\mathcal{E}_2), \mathcal{E}_3, \ldots, \mathcal{E}_n$ are independent. Iterating this argument yields the result of the theorem.

So assume $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent. By Lemma 6.10, $\overline{\mathcal{E}}_1, \ldots, \overline{\mathcal{E}}_n$ are independent. We want to prove $\sigma(\mathcal{E}_1), \mathcal{E}_2, \ldots, \mathcal{E}_n$ are independent. Since $\overline{\sigma(\mathcal{E}_1)} = \sigma(\mathcal{E}_1) = \sigma(\overline{\mathcal{E}}_1)$, it suffices by Lemma 6.10 to prove that $\sigma(\overline{\mathcal{E}}_1), \overline{\mathcal{E}}_2, \ldots, \overline{\mathcal{E}}_n$ are independent.

For $j \in \{2, \ldots, n\}$, let $A_j \in \overline{\mathcal{E}}_j$, and let $F = \bigcap_{j=2}^n A_j$. Let

$$\mathcal{L} = \{A \in \sigma(\overline{\mathcal{E}}_1) : P(A \cap F) = P(A)P(F)\}.$$

We will now prove, using the π-λ theorem (Theorem 2.3), that $\mathcal{L} = \sigma(\overline{\mathcal{E}}_1)$.

Since $P(\Omega \cap F) = P(F) = P(\Omega)P(F)$, we have $\Omega \in \mathcal{L}$. Suppose $A, B \in \mathcal{L}$ with $A \subset B$. Then

$$\begin{aligned} P((B \backslash A) \cap F) &= P((B \cap F) \backslash (A \cap F)) \\ &= P(B \cap F) - P(A \cap F) \\ &= P(B)P(F) - P(A)P(F) \\ &= (P(B) - P(A))P(F) = P(B \backslash A)P(F). \end{aligned}$$

Thus, $B \backslash A \in \mathcal{L}$. Lastly, suppose $\{B_n\}_{n \in \mathbb{N}} \subset \mathcal{L}$ with $B_n \subset B_{n+1}$, and let $B = \bigcup_{n \in \mathbb{N}} B_n$. Then

$$\begin{aligned} P(B \cap F) &= P\left(\bigcup_{n=1}^{\infty} (B_n \cap F) \right) \\ &= \lim_{n \to \infty} P(B_n \cap F) \\ &= \lim_{n \to \infty} P(B_n)P(F) \\ &= P(F) \lim_{n \to \infty} P(B_n) = P(B)P(F). \end{aligned}$$

Thus, $B \in \mathcal{L}$, and this shows that $\mathcal{L}$ is a λ-system.

Since $\overline{\mathcal{E}}_1, \ldots, \overline{\mathcal{E}}_n$ are independent, it follows that $\overline{\mathcal{E}}_1 \subset \mathcal{L}$. Therefore, by the π-λ theorem (Theorem 2.3), we have $\sigma(\overline{\mathcal{E}}_1) \subset \mathcal{L}$. Since $\mathcal{L} \subset \sigma(\overline{\mathcal{E}}_1)$ by the definition of $\mathcal{L}$, we have $\mathcal{L} = \sigma(\overline{\mathcal{E}}_1)$.

We have thus proven that for all $A \in \sigma(\overline{\mathcal{E}}_1)$,

$$P(A \cap A_2 \cap \cdots \cap A_n) = P(A)P(A_2 \cap \cdots \cap A_n) = P(A)P(A_2) \cdots P(A_n).$$

Since $A_2, \ldots, A_n$, were arbitrary, this shows that $\sigma(\overline{\mathcal{E}}_1), \overline{\mathcal{E}}_2, \ldots, \overline{\mathcal{E}}_n$ are independent. ∎

Theorem 6.12. *Let $X_1, \ldots, X_n$ be real-valued random variables. Suppose that for any $x_1, \ldots, x_n \in (-\infty, \infty]$, we have*

$$P(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{j=1}^n P(X_j \leqslant x_j). \tag{6.3}$$

Then $X_1, \ldots, X_n$ are independent.

Proof. Let $\mathcal{E}_j = \{\{X_j \leqslant x\} : x \in (-\infty, \infty]\}$. Then $\Omega \in \mathcal{E}_j$ for each j. Thus, (6.3) implies $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent. Since each $\mathcal{E}_j$ is a π-system, we have that $\sigma(\mathcal{E}_1), \ldots, \sigma(\mathcal{E}_n)$. It follows from Exercise 2.5 and Proposition 1.3(e) that $\sigma(\mathcal{E}_j) = \sigma(X_j)$, so that $X_1, \ldots, X_n$ are independent. □

Proposition 6.13. *Let $X = (X_1, \ldots, X_n)$ be an $\mathbb{R}^n$-valued random variable with a density function $f : \mathbb{R}^n \to [0, \infty)$. Suppose there exist nonnegative, measurable functions $g_j : \mathbb{R} \to \mathbb{R}$ such that $f(x) = \prod_{j=1}^n g_j(x_j)$ for all $x \in \mathbb{R}^n$. Then $X_1, \ldots, X_n$ are independent, and, for each j, the function*

$$x \mapsto \frac{g_j(x)}{\int_{\mathbb{R}} g_j(y)\, dy}$$

is a density for X_j.

Proof. Exercise 6.5. □

Proposition 6.14. *Let $X_1, \ldots, X_n$ be random variables and assume that, for each j, there exists a countable set S_j such that $P(X_j \in S_j) = 1$. If*

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{j=1}^n P(X_j = x_j),$$

whenever $x_j \in S_j$, then $X_1, \ldots, X_n$ are independent.

Proof. Exercise 6.6. □

Theorem 6.15. *Suppose $\mathcal{F}_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m(i)$, are independent σ-algebras. Let $\mathcal{G}_i = \sigma(\bigcup_j \mathcal{F}_{ij})$. Then $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent.*

Proof. For each i, let $\mathcal{E}_i = \{\bigcap_j A_{ij} : A_{ij} \in \mathcal{F}_{ij}\}$. Then $\Omega \in \mathcal{E}_i$ and $\mathcal{E}_i$ is a π-system. Also, $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent. Thus, $\sigma(\mathcal{E}_1), \ldots, \sigma(\mathcal{E}_n)$ are independent. Now, $\mathcal{E}_i \subset \mathcal{G}_i$ implies $\sigma(\mathcal{E}_i) \subset \mathcal{G}_i$. Conversely, $\bigcup_j \mathcal{F}_{ij} \subset \mathcal{E}_i$ implies $\mathcal{G}_i \subset \sigma(\mathcal{E}_i)$. Hence, $\sigma(\mathcal{E}_i) = \mathcal{G}_i$, and so $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent. □

Corollary 6.16. *Suppose $X_{i,j}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m(i)$, are independent random variables, with $X_{i,j}$ taking values in a measurable space $(S_{ij}, \mathcal{S}_{ij})$. Let $f_i : \prod_{j=1}^{m(i)} S_{ij} \to \mathbb{R}$ be $\bigotimes_{j=1}^{m(i)} \mathcal{S}_{ij}$-measurable, and let $Y_i = f_i(X_{i,1}, \ldots, X_{i,m(i)})$. Then $Y_1, \ldots, Y_n$ are independent.*

Proof. Let $\mathcal{F}_{ij} = \sigma(X_{i,j})$ and $\mathcal{G}_i = \sigma(\bigcup_j \mathcal{F}_{ij})$, so that $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent. Since $Y_i \in \mathcal{G}_i$, the result follows from Proposition 6.9. □

Remark 6.17. This corollary is very fundamental and will frequently be used, typically without citing the corollary.

6.1.3 Independence, distribution, and expectation

Theorem 6.18. *Let $X_1, \ldots, X_n$ be independent random variables with $X_i \sim \mu_i$, and let $X = (X_1, \ldots, X_n)$. Then $X \sim \mu_1 \times \cdots \times \mu_n$.*

Proof. Let $\mathcal{E} = \{A_1 \times \cdots \times A_n : A_j \in \mathcal{R}\}$, so that $\mathcal{R}^n = \sigma(\mathcal{E})$. Let $X \sim \nu$, and define $\mathcal{L} = \{A \in \mathcal{R}^n : \nu(A) = (\mu_1 \times \cdots \times \mu_n)(A)\}$. We wish to show that $\mathcal{R}^n \subset \mathcal{L}$. It is left to the reader to verify that $\mathcal{L}$ is a λ-system. Since $\mathcal{E}$ is clearly a π-system, it remains only to show that $\mathcal{E} \subset \mathcal{L}$.

Let $A = A_1 \times \cdots \times A_n \in \mathcal{E}$. Then

$$\begin{aligned}\nu(A) = P(X \in A) &= P(X_1 \in A_1, \ldots, X_n \in A_n) \\ &= \prod_{j=1}^n P(X_j \in A_j) = \prod_{j=1}^n \mu_j(A_j) = (\mu_1 \times \cdots \times \mu_n)(A),\end{aligned}$$

and so $A \in \mathcal{L}$. □

Theorem 6.19. *Let $X_1, \ldots, X_n$ be independent random variables. Suppose that either (a) each $X_j \geqslant 0$ a.s., or (b) $E|X_j| < \infty$ for each j. Then*

$$E\left[\prod_{j=1}^n X_j\right] = \prod_{j=1}^n EX_j.$$

That is, the expected value on the left exists, and has the value given on the right.

Proof. First assume $n = 2$. Let $|X_1| \sim \mu$ and $|X_2| \sim \nu$. Since $|X_1|$ and $|X_2|$ are independent, Tonelli's theorem gives

$$\begin{aligned}E|X_1X_2| &= \int |xy|\,(\mu \times \nu)(dx\,dy) = \int\int |xy|\,\mu(dx)\,\nu(dy) \\ &= \left(\int |x|\,\mu(dx)\right)\left(\int |y|\,\nu(dy)\right) = E|X_1| \cdot E|X_2|.\end{aligned}$$

If each $X_j \geqslant 0$ a.s., then we are done. Otherwise, the above expression is finite, so using Fubini's theorem as above gives $E[X_1X_2] = EX_1 \cdot EX_2$.

Now assume the theorem is true for some n. To prove the theorem for $n+1$, we apply the above to X_1 and $\prod_{j=2}^{n+1} X_j$. □

Let $\{X_\alpha\}_{\alpha \in A} \subset L^2(\Omega)$. We say that $\{X_\alpha\}$ are **uncorrelated** if $E[X_\alpha X_\beta] = (EX_\alpha)(EX_\beta)$ whenever $\alpha \neq \beta$. As shown above, independent random variables are uncorrelated. The converse, however, is not true. See [2, Example 2.1.2] for an elementary counterexample.

Theorem 6.20. *If $\{X_j\}_{j=1}^n$ are uncorrelated, then*

$$\operatorname{var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \operatorname{var}(X_j).$$

Proof. Let $\mu_j = EX_j$. Then

$$\begin{aligned}\operatorname{var}\left(\sum_{j=1}^n X_j\right) &= E\left|\sum_{j=1}^n X_j - \sum_{j=1}^n \mu_j\right|^2 \\ &= E\left|\sum_{j=1}^n (X_j - \mu_j)\right|^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n E[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \sum_{j=1}^n E|X_j - \mu_j|^2 + 2\sum_{i=1}^n \sum_{j=1}^{i-1} E[(X_i - \mu_i)(X_j - \mu_j)].\end{aligned}$$

Since $E|X_j - \mu_j|^2 = \operatorname{var}(X_j)$, it suffices to show that $E[(X_i - \mu_i)(X_j - \mu_j)] = 0$ whenever $i \neq j$. For this, we calculate

$$\begin{aligned}E[(X_i - \mu_i)(X_j - \mu_j)] &= E[X_i X_j] - \mu_j EX_i - \mu_i EX_j + \mu_i \mu_j \\ &= E[X_i X_j] - \mu_i \mu_j.\end{aligned}$$

But X_i and X_j are uncorrelated, so $E[X_i X_j] = (EX_i)(EX_j) = \mu_i \mu_j$. □

6.1.4 Sums of independent random variables

Theorem 6.21. *Let X and Y be independent random variables with distribution functions F and G, respectively. Then the distribution function of $X + Y$ is*

$$H(z) = \int_{\mathbb{R}} F(z - y)\, dG(y).$$

Proof. Fix $z \in \mathbb{R}$. Let $A = \{(x, y) \in \mathbb{R}^2 : x + y \leqslant z\}$. Then

$$\begin{aligned}H(z) = P(X + Y \leqslant z) &= E[1_{\{X+Y \leqslant z\}}] \\ &= E[1_A(X, Y)] = \int\int 1_A(x, y)\, \mu_F(dx)\, \mu_G(dy).\end{aligned}$$

Note that $1_A(x, y) = 1_{(-\infty, z-y]}(x)$. Thus,

$$H(z) = \int \mu_F((-\infty, z - y])\, \mu_G(dy) = \int F(z - y)\, \mu_G(dy),$$

which is what we wanted to prove. □

For a proof of the following special cases, see [2, Theorem 2.1.11].

Theorem 6.22. *Let X and Y be independent random variables with distribution functions F and G, respectively. If X has density f, then $X + Y$ has density*

$$h(z) = \int f(z - y)\, dG(y).$$

If, in addition, Y has density g, then this can be rewritten as

$$h(z) = \int f(z-y)g(y)\,dy.$$

The **gamma function** is the function $\Gamma : (0,\infty) \to \mathbb{R}$ defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$. It can be shown that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for all $\alpha > 0$, and that $\Gamma(n) = (n-1)!$ for all $n \in \mathbb{N}$.

We say that a real-valued random variable has the **gamma distribution** with parameters α and λ, written $X \sim \text{Gamma}(\alpha,\lambda)$, if X has density

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}e^{-\lambda x}\,1_{(0,\infty)}(x).$$

Note that $\text{Gamma}(1,\lambda) = \text{Exp}(\lambda)$.

The proofs of the following results use the above theorems, as well as a lot of tedious calculus. For details, see [2, Theorems 2.1.12 and 2.1.13].

Theorem 6.23. *If $X_1,\ldots,X_n$ are independent with $X_j \sim Gamma(\alpha_j,\lambda)$, then $\sum_{j=1}^n X_j \sim Gamma(\sum_{j=1}^n \alpha_j,\lambda)$.*

Theorem 6.24. *If $X \sim N(\mu,\sigma^2)$ and $Y \sim N(\nu,\tau^2)$ are independent, then $X+Y \sim N(\mu+\nu,\sigma^2+\tau^2)$.*

6.1.5 Constructing independent random variables

A measurable space $(S,\mathcal{S})$ is a **standard Borel space** if there exists a bijection $\varphi : S \to \mathbb{R}$ such that φ is $(\mathcal{S},\mathcal{R})$-measurable and φ^{-1} is $(\mathcal{R},\mathcal{S})$-measurable. For a proof of the following theorem, see [2, Theorem 2.1.15].

Theorem 6.25. *Let (M,d) be a complete, separable metric space. Let $S \in \mathcal{B}_M$ and $\mathcal{S} = \{A \cap S : A \in \mathcal{B}_M\}$. Then $(S,\mathcal{S})$ is a standard Borel space.*

The main result of this subsection is following theorem.

Theorem 6.26. *For each $j \in \mathbb{N}$, let $(S_j,\mathcal{S}_j)$ be a standard Borel space, and ν_j a probability measure on $(S_j,\mathcal{S}_j)$. Then there exists a probability space $(\Omega,\mathcal{F},P)$ and a sequence of independent random variables $\{X_j\}_{j=1}^\infty$ defined on $(\Omega,\mathcal{F},P)$ such that X_j takes values in S_j, and $X_j \sim \nu_j$.*

Proof. For each j, choose $\varphi_j : S_j \to \mathbb{R}$ such that φ_j and φ_j^{-1} are both measurable. Let ν_j' be the probability measure on $(\mathbb{R},\mathcal{R})$ defined by $\nu_j' = \nu_j \circ \varphi_j^{-1}$, and let $\mu_n = \prod_{j=1}^n \nu_j'$. Then μ_n is a probability measure on $(\mathbb{R}^n,\mathcal{R}^n)$ and the measures $\{\mu_n\}_{n=1}^\infty$ are consistent. (See Theorem 2.52.) Let $\Omega = \mathbb{R}^\infty$ and $\mathcal{F} = \mathcal{R}^\infty$. Let P be the probability measure on $(\Omega,\mathcal{F})$ described in Theorem 2.52. Let $X_j(\omega) = \varphi_j^{-1}(\omega_j)$.

Since $X_j = \varphi_j^{-1} \circ \pi_j$, where $\pi_j : \mathbb{R}^\infty \to \mathbb{R}$ is the projection map, it follows that X_j is measurable. Let $A_j \in \mathcal{S}_j$. Note that

$$\{X_j \in A_j\} = X_j^{-1}(A_j) = \pi_j^{-1}(\varphi_j(A_j)) = \{\pi_j \in \varphi_j(A_j)\}.$$

Thus, By Theorem 2.52,

$$\begin{aligned}P(X_1 \in A_1, \ldots, X_n \in A_n) &= P(\pi_1 \in \varphi_1(A_1), \ldots, \pi_n \in \varphi_n(A_n)) \\ &= P(\{\omega : \omega_j \in \varphi_j(A_j) \text{ for } 1 \leqslant j \leqslant n\}) \\ &= \mu_n(\varphi_1(A_1) \times \cdots \times \varphi_n(A_n)) \\ &= \prod_{j=1}^{n} \nu_j'(\varphi_j(A_j)).\end{aligned}$$

Taking $A_j = S_j$ for $j < n$ shows that $X_n \sim \nu_n' \circ \varphi_n = \nu_n$, and this is valid for every $n \in \mathbb{N}$. Therefore, the above shows that $X_1, \ldots, X_n$ are independent. Since n was arbitrary, this implies $X_1, X_2, \ldots$ are independent. □

Exercises

6.1. [2, Exercise 2.1.1] Prove Proposition 6.7.

6.2. [2, Exercise 2.1.2(i)] Prove Proposition 6.6. (Hint: First show that if A and B are independent, then so are A^c and B, A and B^c, and A^c and B^c.)

6.3. [2, Exercise 2.1.3(ii)+] Prove Proposition 6.8.

6.4. Prove Proposition 6.9.

6.5. [2, Exercise 2.1.4+] Prove Proposition 6.13.

6.6. [2, Exercise 2.1.5] Prove Proposition 6.14.

6.7. [2, Exercise 2.1.8(i)] Let X and Y be real-valued random variables on a probability space, $(\Omega, \mathcal{F}, P)$, with $X \sim \mu$ and $Y \sim \nu$. Prove that if X and Y are independent, then

$$P(X + Y = 0) = \sum_{y \in \mathbb{R}} \mu(\{-y\})\nu(\{y\}).$$

[Recall: See Remark 2.36 for the definition of infinite sums of this type.]

6.8. [2, Exercise 2.1.13] Let X and Y be integer-valued random variables on a probability space, $(\Omega, \mathcal{F}, P)$. Prove that if X and Y are independent, then

$$P(X + Y = n) = \sum_{m \in \mathbb{Z}} P(X = m)P(Y = n - m),$$

for all $n \in \mathbb{Z}$.

6.9. [2, Exercise 2.1.14] Let X and Y be real-valued random variables on a probability space, $(\Omega, \mathcal{F}, P)$, with $X \sim \text{Poisson}(r)$ and $Y \sim \text{Poisson}(s)$, where $r, s > 0$. Use Exercise 6.8 to prove that if X and Y are independent, then $X + Y \sim \text{Poisson}(r + s)$.

6.2 Conditional expectation

This section corresponds to [2, Subsections 5.1.1 and 5.1.2].

6.2.1 The general definition

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$ with $P(A) > 0$. Recall that $P^A := P(\cdot \mid A)$ is a probability measure on $(\Omega, \mathcal{F})$. If X is a random variable, we define the **conditional expectation of X given A** as

$$E[X \mid A] = \int_\Omega X \, dP^A, \tag{6.4}$$

whenever this integral is well-defined. Note that $E[1_B \mid A] = P(B \mid A)$.

Theorem 6.27. *We have that X is P^A-integrable if and only if $E[|X|1_A] < \infty$. If $X \geqslant 0$ or $E[|X|1_A] < \infty$, then*

$$E[X \mid A] = \frac{E[X1_A]}{P(A)}. \tag{6.5}$$

Remark 6.28. Note that (6.5) may be written as

$$E[X \mid A] = \frac{\alpha(A)}{P(A)}, \tag{6.6}$$

where $d\alpha = X \, dP$. Also note that (6.5) gives us the formula $E[X1_A] = P(A)E[X \mid A]$. If $X = 1_B$, then this reduces to the familiar multiplication rule, $P(A \cap B) = P(A)P(B \mid A)$.

Proof of Theorem 6.27. Note that if $P(B) = 0$, then $P^A(B) = 0$. Hence $P^A \ll P$. Also note that

$$P^A(B) = \int_B \frac{1_A}{P(A)} \, dP, \qquad \text{for all } B \in \mathcal{F}.$$

Thus, $dP^A/dP = 1_A/P(A)$. It follows that if $X \geqslant 0$, then

$$E[X \mid A] = \int X \, dP^A = \int X \frac{dP^A}{dP} \, dP = E\left[X \frac{1_A}{P(A)}\right] = \frac{E[X1_A]}{P(A)}.$$

Therefore, X is P^A-integrable if and only if $E[|X|1_A] < \infty$, and in this case, the same formula holds. ∎

Lemma 6.29. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Suppose $A \in \mathcal{G}$ and X is a $\mathcal{G}$-measurable random variable. If $X \geqslant 0$ a.s., then*

$$E[X1_A] = \int_A X \, dP = \int_A X \, d(P|_\mathcal{G}).$$

Also, if $E|X| < \infty$, then $X \in L^1(\Omega, \mathcal{G}, P|_\mathcal{G})$ and the above equality holds.

Proof. Exercise 6.10. □

Theorem 6.30. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, X an integrable random variable, and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Then there exists a random variable Z such that*

(i) $Z \in \mathcal{G}$, and

(ii) $E[X 1_A] = E[Z 1_A]$, for all $A \in \mathcal{G}$.

If Z' is any other such random variable, then $Z = Z'$ a.s. Moreover, Z is integrable with $E|Z| \leqslant E|X|$.

Remark 6.31. The random variable Z in the above theorem is called the **conditional expectation of X given $\mathcal{G}$**, written $E[X \mid \mathcal{G}]$. Since it is only unique up to a P-null set, we may sometimes refer to Z as a version of $E[X \mid \mathcal{G}]$.

The **conditional probability of A given $\mathcal{G}$** is defined by $P(A \mid \mathcal{G}) = E[1_A \mid \mathcal{G}]$.

Subsection 6.2.2 may provide help in developing an intuitive understanding of this definition.

Proof of Theorem 6.30. Let α be the complex measure on $(\Omega, \mathcal{F})$ given by $d\alpha = X\,dP$. Note that $\alpha|_{\mathcal{G}}$ is a complex measure on $(\Omega, \mathcal{G})$ and $P|_{\mathcal{G}}$ is a probability measure on $(\Omega, \mathcal{G})$. Suppose $A \in \mathcal{G}$ and $(P|_{\mathcal{G}})(A) = 0$. Then $P(A) = 0$ and

$$(\alpha|_{\mathcal{G}})(A) = \alpha(A) = \int_A X\,dP = 0.$$

Thus, $\alpha|_{\mathcal{G}} \ll P|_{\mathcal{G}}$, so we may define $Z = d(\alpha|_{\mathcal{G}})/d(P|_{\mathcal{G}})$, the Radon-Nikodym derivative of $\alpha|_{\mathcal{G}}$ with repsect to $P|_{\mathcal{G}}$.

By definition, $Z \in \mathcal{G}$. Let $A \in \mathcal{G}$. Then

$$E[X 1_A] = \int_A X\,dP = \alpha(A) = (\alpha|_{\mathcal{G}})(A).$$

By Lemma 6.29,

$$E[Z 1_A] = \int_A Z\,dP = \int_A Z\,d(P|_{\mathcal{G}}) = \int_A \frac{d(\alpha|_{\mathcal{G}})}{d(P|_{\mathcal{G}})}\,d(P|_{\mathcal{G}}) = (\alpha|_{\mathcal{G}})(A).$$

Thus, $E[X 1_A] = E[Z 1_A]$ for all $A \in \mathcal{G}$.

Suppose Z' is another random variable satisfying (i) and (ii). Then, as above,

$$\int_A Z'\,d(P|_{\mathcal{G}}) = E[Z' 1_A] = E[X 1_A] = (\alpha|_{\mathcal{G}})(A),$$

for all $A \in \mathcal{G}$. By the uniqueness of the Radon-Nikodym derivative, this implies $Z' = Z$ a.s.

Finally, since $Z^+ = Z 1_A$, where $A = \{Z > 0\} \in \mathcal{G}$, we have

$$\int Z^+\,dP = \int Z 1_A\,dP = E[Z 1_A] = E[X 1_A] \leqslant E[|X| 1_A] < \infty.$$

Similarly, $\int Z^- \, dP \leqslant E[|X| 1_{A^c}] < \infty$. Thus,

$$E|Z| = \int Z^+ \, dP + \int Z^- \, dP \leqslant E[|X| 1_A] + E[|X| 1_{A^c}] = E|X| < \infty,$$

and so Z is integrable with $E|Z| \leqslant E|X|$. □

Remark 6.32. From the proof, we see that

$$E[X \mid \mathcal{G}] = \frac{d(\alpha|_{\mathcal{G}})}{d(P|_{\mathcal{G}})},$$

where $d\alpha = X \, dP$. Note the similarity between this and (6.6).

Lemma 6.33. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(S, \mathcal{S})$ a measurable space, and Y an S-valued random variable. Let $(S', \mathcal{S}')$ be another measurable space and suppose X is an S'-valued random variable that is $\sigma(Y)$-measurable. Then there exists an $(\mathcal{S}, \mathcal{S}')$-measurable function $h : S \to S'$ such that $X = h(Y)$ a.s. If h' is another such function, then $h = h'$ μ_Y-a.e.*

Proof. Exercise 6.11. □

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X an integrable random variable. Let Y be an arbitrary random variable. The **conditional expectation of X given Y** is defined by $E[X \mid Y] = E[X \mid \sigma(Y)]$. We also define the **conditional probability of A given Y** as $P(A \mid Y) = E[1_A \mid Y]$. Note that $E[X \mid Y]$ is $\sigma(Y)$-measurable. By Lemma 6.33, there exists a measurable function h (which depends on X) such that $E[X \mid Y] = h(Y)$, and this function is unique μ_Y-a.e.

6.2.2 Elementary special cases

Lemma 6.34. *Let $(\Omega, \mathcal{G})$ be a measurable space and assume that $\mathcal{G}$ is a finite set. Then there exists a unique partition $\mathcal{E} = \{A_j\}_{j=1}^n$ of Ω such that $\mathcal{G} = \sigma(\{A_j\}_{j=1}^n)$.*

Proof. For each $\omega \in \Omega$, let A_ω be the smallest measurable set containing ω. That is, $A_\omega = \bigcap \mathcal{G}_\omega$, where $\mathcal{G}_\omega = \{A \in \mathcal{G} : \omega \in A\}$. Since this is a finite intersection, $A_\omega \in \mathcal{G}$. In particular, $\mathcal{E} = \{A_\omega : \omega \in \Omega\}$ is a finite set. We claim that $\mathcal{E}$ is a partition of Ω and that $\mathcal{G} = \sigma(\mathcal{E})$.

Clearly, $\Omega = \bigcup_{\omega \in \Omega} A_\omega$, so to show that $\mathcal{E}$ is a partition, it suffices to show that this is a disjoint union. More specifically, we wish to show that if $\omega, \omega' \in \Omega$, then either $A_\omega = A_{\omega'}$ or $A_\omega \cap A_{\omega'} = \varnothing$. Let $\omega, \omega' \in \Omega$. Note that for any $A \in \mathcal{G}$, if $\omega \in A$, then $A \in \mathcal{G}_\omega$, which implies $A_\omega \subset A$. Hence, if $\omega \in A_{\omega'}$, then $A_\omega \subset A_{\omega'}$; and if $\omega \in A_{\omega'}^c$, then $A_\omega \subset A_{\omega'}^c$. That is, either $A_\omega \subset A_{\omega'}$ or $A_\omega \subset A_{\omega'}^c$. By symmetry, either $A_{\omega'} \subset A_\omega$ or $A_{\omega'} \subset A_\omega^c$. Taken together, this shows that either $A_\omega = A_{\omega'}$ or $A_\omega \cap A_{\omega'} = \varnothing$.

To see that $\mathcal{G} = \sigma(\mathcal{E})$, simply note that any $A \in \mathcal{G}$ can be written as $A = \bigcup_{\omega \in A} A_\omega$, and that this is a finite union.

For uniqueness, suppose that $\mathcal{G} = \sigma(\{B_j\}_{j=1}^n)$, where $\Omega = \biguplus_{j=1}^n B_j$. If $\omega \in B_j$, then $A_\omega = B_j$. Therefore, $\mathcal{E} = \{B_j\}_{j=1}^n$. □

Remark 6.35. The technique in this proof can be used in Exercise 1.6.

Proposition 6.36. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X an integrable random variable. Let $\mathcal{G} \subset \mathcal{F}$ be a finite σ-algebra. Write $\mathcal{G} = \sigma(\{A_j\}_{j=1}^n)$, where $\{A_j\}_{j=1}^n$ is a partition of Ω. Then*

$$E[X \mid \mathcal{G}](\omega) = \begin{cases} E[X \mid A_1] & \text{if } \omega \in A_1, \\ E[X \mid A_2] & \text{if } \omega \in A_2, \\ \quad\vdots & \\ E[X \mid A_n] & \text{if } \omega \in A_n. \end{cases}$$

Proof. Exercise 6.12. □

Remark 6.37. As a special case of the above, if $\mathcal{G} = \{\varnothing, \Omega\}$ is the trivial σ-algebra, then $E[X \mid \mathcal{G}] = EX$.

Remark 6.38. To illustrate the idea of the preceding proposition, consider the following heuristic example. Imagine my friend is at the local bar, and is about to throw a dart at a dartboard. If I model the dart board by a unit circle, which I call Ω, then his dart will land at some point $\omega \in \Omega$.

Unfortunately, I am not there with him and will not be able to observe the exact location of ω. But after he throws the dart, he is going to call me on the phone and tell me what his score for that throw was. This information will not be enough for me to determine ω. It will, however, narrow it down. Before I receive his call, I can partition the dartboard Ω into several pieces, $A_1, \ldots, A_n$, with each piece corresponding to a unique score. Once he calls me, I will know which piece contains his dart.

Let X be the distance from his dart to the bullseye. Suppose he calls me and I determine that his dart is somewhere inside A_j. I can then compute $E[X \mid A_j]$. However, before he calls, I can get prepared by computing $E[X \mid A_j]$ for all j, and then encoding all this information into the single random variable $E[X \mid \mathcal{G}]$.

In probability theory, we model information by σ-algebras. In this example, the σ-algebra $\mathcal{G}$ generated by the partition $\{A_j\}$ models the information I will receive from my friend's phone call. Imagine that while I am waiting for my friend's phone call, an interviewer starts asking me questions. For various events A, the interviewer asks me, "After your friend calls, will you know with certainty whether or not A has occurred?" Depending on the event A, my answer will be "yes", "no", or "it depends on what he says". The events $A \in \mathcal{G}$ are precisely those events for which my answer is yes.

Proposition 6.39. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X an integrable random variable. Let Y be a random variable, S a countable set, and assume $P(Y \in S) = 1$ and $P(Y = k) > 0$ for all $k \in S$. For $k \in S$, define*

$$h(k) = E[X \mid Y = k] = \frac{E[X 1_{\{Y=k\}}]}{P(Y = k)}.$$

Then $E[X \mid Y] = h(Y)$.

Proof. Exercise 6.13. □

Example 6.40. Suppose X and Y are random variables with a joint density function $f(x,y)$. That is, for any $A \in \mathcal{R}^2$, we have $P((X,Y) \in A) = \int_A f\,d\lambda$. Note that $P(X \in A \mid Y = k)$ is undefined, since $P(Y = k) = 0$. Nonetheless, it may be intuitively clear to some that the following equation ought to hold:

$$P(X \in A \mid Y = k) = \frac{\displaystyle\int_A f(x,k)\,dx}{\displaystyle\int_{\mathbb{R}} f(x,k)\,dx}. \tag{6.7}$$

The integral in the denominator is necessary in order to make the function $x \mapsto f(x,k)$ a probability density function. In this example, we will explore the sense in which this formula is rigorously valid. (In an undergraduate class, it may be rigorously valid *by definition.* But for us, as usual, it is a special case of something more general.)

Proposition 6.41. *Let X and Y have joint density $f(x,y)$. Let g be a measurable function such that $E|g(X)| < \infty$. Define*

$$h(k) = \frac{\displaystyle\int_{\mathbb{R}} g(x)f(x,k)\,dx}{\displaystyle\int_{\mathbb{R}} f(x,k)\,dx},$$

whenever $\int_{\mathbb{R}} f(x,k)\,dx > 0$, and $h(k) = 0$ otherwise. Then $E[g(X) \mid Y] = h(Y)$.

Remark 6.42. If $g(x) = 1_A(x)$, then $h(k)$ agrees with the right-hand side of (6.7). Also, as can be seen from the proof below, we could have defined $h(k)$ arbitrarily when $\int_{\mathbb{R}} f(x,k)\,dx = 0$.

Proof of Theorem 6.41. Since $h(Y)$ is $\sigma(Y)$-measurable, it will suffice for us to show that $E[h(Y)1_A] = E[g(X)1_A]$ for all $A \in \sigma(Y)$. Let $A \in \sigma(Y)$. Then $A = \{Y \in B\}$ for some $B \in \mathcal{R}$. We now have

$$\begin{aligned} E[h(Y)1_A] = E[h(Y)1_B(Y)] &= \int_B \int_{\mathbb{R}} h(y)f(x,y)\,dx\,dy \\ &= \int_B \left(h(y) \int_{\mathbb{R}} f(x,y)\,dx \right) dy = \int_{B\cap C} \left(h(y) \int_{\mathbb{R}} f(x,y)\,dx \right) dy, \end{aligned}$$

where $C = \{y : \int_{\mathbb{R}} f(x,y)\,dx > 0\}$. Note that for all $y \in C$, we have

$$h(y) \int_{\mathbb{R}} f(x,y)\,dx = \int_{\mathbb{R}} g(x)f(x,y)\,dx.$$

Also, for all $y \in C^c$, we have $f(x,y) = 0$ for Lebesgue almost every x. Thus, $y \in C^c$ implies $\int_{\mathbb{R}} g(x)f(x,y)\,dx = 0$. It therefore follows that

$$\begin{aligned} E[h(Y)1_A] = \int_{B\cap C} \int_{\mathbb{R}} g(x)f(x,y)\,dx\,dy &= \int_B \int_{\mathbb{R}} g(x)f(x,y)\,dx\,dy \\ &= E[g(X)1_B(Y)] = E[g(X)1_A], \end{aligned}$$

which was what we needed to prove. $\square$

In general, we interpret $E[X \mid Y = y]$ to mean $g(y)$, where g is a measurable function such that $E[X \mid Y] = g(Y)$. Some caution is needed in these cases, though, since such a function g is only defined μ_Y-a.e.

6.2.3 Basic properties

For the remainder of this section, unless otherwise noted, $(\Omega, \mathcal{F}, P)$ is a probability space, X is an integrable random variable, and $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra.

Proposition 6.43. *$E[E[X \mid \mathcal{G}]] = EX$.*

Proof. By the definition of conditional expectation, we have $E[E[X \mid \mathcal{G}]1_A] = E[X1_A]$ for all $A \in \mathcal{G}$. Take $A = \Omega$. $\square$

Proposition 6.44. *If X is $\mathcal{G}$-measurable, then $E[X \mid \mathcal{G}] = X$.*

Proof. It follows trivially, since X is $\mathcal{G}$-measurable and $E[X1_A] = E[X1_A]$ for all $A \in \mathcal{G}$. $\square$

If X is a random variable and $\mathcal{G}$ is a σ-algebra, then we say that X and $\mathcal{G}$ are independent if $\sigma(X)$ and $\mathcal{G}$ are independent, which in turn means that $P(A \cap B) = P(A)P(B)$ whenever $A \in \sigma(X)$ and $B \in \mathcal{G}$. Hence, X and $\mathcal{G}$ are independent if and only if $P(\{X \in C\} \cap B) = P(X \in C)P(B)$ for all $C \in \mathcal{R}$ and $B \in \mathcal{G}$.

Proposition 6.45. *If X and $\mathcal{G}$ are independent, then $E[X \mid \mathcal{G}] = E[X]$. In particular, $E[X \mid \{\emptyset, \Omega\}] = E[X]$.*

Proof. A constant random variable is measurable with respect to every σ-algebra, so $E[X]$ is trivially $\mathcal{G}$-measurable. Also, for all $A \in \mathcal{G}$, we have $E[X1_A] = E[X]E[1_A] = E[E[X]1_A]$. The final claim holds since every random variable is independent of the trivial σ-algebra. $\square$

Theorem 6.46. *If $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$, then $E[E[X \mid \mathcal{G}_1] \mid \mathcal{G}_2] = E[E[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = E[X \mid \mathcal{G}_1]$.*

Remark 6.47. In words, this says that in a battle between nested σ-algebras, the smallest σ-algebra always wins.

Proof of Theorem 6.46. Since $\mathcal{G}_1 \subset \mathcal{G}_2$ and $E[X \mid \mathcal{G}_1]$ is $\mathcal{G}_1$-measurable, it is also $\mathcal{G}_2$-measurable. Hence, by Proposition 6.44, $E[E[X \mid \mathcal{G}_1] \mid \mathcal{G}_2] = E[X \mid \mathcal{G}_1]$. The other equality holds since $E[X \mid \mathcal{G}_1]$ is $\mathcal{G}_1$-measurable and, for all $A \in \mathcal{G}_1 \subset \mathcal{G}_2$, we have $E[E[X \mid \mathcal{G}_1]1_A] = E[X1_A] = E[E[X \mid \mathcal{G}_2]1_A]$. $\square$

Theorem 6.48. *If Y and XY are integrable, and X is $\mathcal{G}$-measurable, then*

$$E[XY \mid \mathcal{G}] = XE[Y \mid \mathcal{G}] \text{ a.s.}$$

Proof. Let $d\alpha = Y\,dP$, so that $E[Y \mid \mathcal{G}] = d(\alpha|_{\mathcal{G}})/d(P|_{\mathcal{G}})$. Let $d\beta_1 = X\,d\alpha^+$, $d\beta_2 = X\,d\alpha^-$, and $\beta = \beta_1 - \beta_2$, so that $d\beta = XY\,dP$ and $E[XY \mid \mathcal{G}] = d(\beta|_{\mathcal{G}})/d(P|_{\mathcal{G}})$. Since X is $\mathcal{G}$-measurable, we have

$$d(\beta_1|_{\mathcal{G}})/d(\alpha^+|_{\mathcal{G}}) = d(\beta_2|_{\mathcal{G}})/d(\alpha^-|_{\mathcal{G}}) = X.$$

Hence,

$$\begin{aligned} E[XY \mid \mathcal{G}] = \frac{d(\beta|_{\mathcal{G}})}{d(P|_{\mathcal{G}})} &= \frac{d(\beta_1|_{\mathcal{G}})}{d(P|_{\mathcal{G}})} - \frac{d(\beta_2|_{\mathcal{G}})}{d(P|_{\mathcal{G}})} \\ &= \frac{d(\beta_1|_{\mathcal{G}})}{d(\alpha^+|_{\mathcal{G}})} \cdot \frac{d(\alpha^+|_{\mathcal{G}})}{d(P|_{\mathcal{G}})} - \frac{d(\beta_2|_{\mathcal{G}})}{d(\alpha^-|_{\mathcal{G}})} \cdot \frac{d(\alpha^-|_{\mathcal{G}})}{d(P|_{\mathcal{G}})} \\ &= X\left(\frac{d(\alpha^+|_{\mathcal{G}})}{d(P|_{\mathcal{G}})} - \frac{d(\alpha^-|_{\mathcal{G}})}{d(P|_{\mathcal{G}})}\right) = X\frac{d(\alpha|_{\mathcal{G}})}{d(P|_{\mathcal{G}})} = XE[Y \mid \mathcal{G}], \ P\text{-a.s.,} \end{aligned}$$

and we are done. □

Theorem 6.49 (linearity)**.** $E[aX + Y \mid \mathcal{G}] = aE[X \mid \mathcal{G}] + E[Y \mid \mathcal{G}]$.

Proof. The right-hand side is clearly $\mathcal{G}$-measurable. Let $A \in \mathcal{G}$. Then

$$\begin{aligned} E[(aE[X \mid \mathcal{G}] + E[Y \mid \mathcal{G}])1_A] &= aE[E[X \mid \mathcal{G}]1_A] + E[E[Y \mid \mathcal{G}]1_A] \\ &= aE[X1_A] + E[Y1_A] = E[(aX + Y)1_A], \end{aligned}$$

and we are done. □

Lemma 6.50. *Let U and V be $\mathcal{G}$-measurable random variables. If $E[U1_A] \leqslant E[V1_A]$ for all $A \in \mathcal{G}$, then $U \leqslant V$ a.s. If $E[U1_A] = E[V1_A]$ for all $A \in \mathcal{G}$, then $U = V$ a.s.*

Proof. By reversing the roles of U and V, the second claim follows from the first. To prove the first, suppose $E[U1_A] \leqslant E[V1_A]$ for all $A \in \mathcal{G}$. Let $A = \{U > V\} \in \mathcal{G}$ and define $Z = (U - V)1_A$, so that $Z \geqslant 0$. Note that $EZ = E[U1_A] - E[V1_A] \leqslant 0$. Hence, $EZ = 0$, so $Z = 0$ a.s., which implies $P(A) = 0$. □

Theorem 6.51 (monotonicity)**.** *If $X \leqslant Y$ a.s., then $E[X \mid \mathcal{G}] \leqslant E[Y \mid \mathcal{G}]$ a.s.*

Proof. For all $A \in \mathcal{G}$, we have $E[E[X \mid \mathcal{G}]1_A] = E[X1_A] \leqslant E[Y1_A] = E[E[Y \mid \mathcal{G}]1_A]$. Hence, by Lemma 6.50, $E[X \mid \mathcal{G}] \leqslant E[Y \mid \mathcal{G}]$ a.s. □

Theorem 6.52. *Suppose X and Y are independent and φ is a measurable function such that $E|\varphi(X,Y)| < \infty$, then $E[\varphi(X,Y) \mid X] = g(X)$, where $g(x) = E[\varphi(x,Y)]$.*

Remark 6.53. It is important here that X and Y are independent. This result is not true when X and Y are dependent.

Proof of Theorem 6.52. Clearly, $g(X)$ is $\sigma(X)$-measurable. Let $A \in \mathcal{R}$. Then

$$\begin{aligned} E[\varphi(X,Y)1_{\{X\in A\}}] &= \int\int \varphi(x,y)1_A(x)\mu_Y(dy)\mu_X(dx) \\ &= \int 1_A(x)\left(\int \varphi(x,y)\mu_Y(dy)\right)\mu_X(dx) \\ &= \int 1_A(x)g(x)\mu_X(dx) = E[g(X)1_{\{X\in A\}}], \end{aligned}$$

and we are done. □

Example 6.54. Let X, Y, Z be i.i.d. (independent and identically distributed), uniformly distributed on $(0,1)$. We shall compute the distribution of $(XY)^Z$. We begin by computing the distribution of $W = XY$. Let $w \in (0,1)$. Then

$$P(W \leqslant w) = P(XY \leqslant w) = E[1_{\{XY\leqslant w\}}] = E[E[1_{\{XY\leqslant w\}} \mid X]].$$

By Theorem 6.52, $E[1_{\{XY\leqslant w\}} \mid X] = f(X)$, where

$$f(x) = E[1_{\{xY\leqslant w\}}] = P(xY \leqslant w) = P\left(Y \leqslant \frac{w}{x}\right) = 1_{\{x<w\}} + \frac{w}{x}1_{\{x\geqslant w\}}.$$

Thus,

$$P(W \leqslant w) = E[f(X)] = E\left[1_{\{X<w\}} + \frac{w}{X}1_{\{X\geqslant w\}}\right] = w + \int_w^1 \frac{w}{x}\,dx = w - w\log w.$$

Differentiating, we find that W has density $f_W(w) = (-\log w)1_{(0,1)}(w)$.

Similarly, for $x \in (0,1)$, we now compute

$$P((XY)^Z \leqslant x) = E[P(W^Z \leqslant x \mid W)] = E[g(W)],$$

where

$$g(w) = P(w^Z \leqslant x) = P\left(Z \geqslant \frac{\log x}{\log w}\right) = \left(1 - \frac{\log x}{\log w}\right)1_{\{w\leqslant x\}}.$$

Thus,

$$\begin{aligned} P((XY)^Z \leqslant x) = E\left[\left(1 - \frac{\log x}{\log W}\right)1_{\{W\leqslant x\}}\right] &= \int_0^x \left(1 - \frac{\log x}{\log w}\right)(-\log w)\,dw \\ &= -\int_0^x \log w\,dw + x\log x = x. \end{aligned}$$

In other words, $(XY)^Z$ is uniformly distributed on $(0,1)$.

6.2.4 Limit theorems and inequalities

Theorem 6.55 (monotone convergence)**.** *If $0 \leqslant X_n \uparrow X$ a.s. and X is integrable, then $E[X_n \mid \mathcal{G}] \uparrow E[X \mid \mathcal{G}]$ a.s.*

Proof. By monotonicity, there exists a $\mathcal{G}$-measurable random variable Z such that $E[X_n \mid \mathcal{G}] \uparrow Z$ a.s. Let $A \in \mathcal{G}$. Using monotone convergence,

$$E[Z1_A] = \lim_{n\to\infty} E[E[X_n \mid \mathcal{G}]1_A] = \lim_{n\to\infty} E[X_n 1_A] = E[X1_A],$$

which shows $Z = E[X \mid \mathcal{G}]$. □

Theorem 6.56. (Fatou's lemma) *If $X_n \geqslant 0$ a.s., each X_n is integrable, and $\liminf_{n\to\infty} X_n$ is integrable, then*

$$E[\liminf_{n\to\infty} X_n \mid \mathcal{G}] \leqslant \liminf_{n\to\infty} E[X_n \mid \mathcal{G}] \ a.s.$$

Proof. Let $\overline{X}_n = \inf_{j\geqslant n} X_j$ and $X = \liminf_{n\to\infty} X_n$. Note that $0 \leqslant \overline{X}_n \uparrow X$. In particular, $\overline{X}_n$ is integrable. For each $j \geqslant n$, we have $\overline{X}_n \leqslant X_j$ a.s. Hence, by monotonicity, $E[\overline{X}_n \mid \mathcal{G}] \leqslant E[X_j \mid \mathcal{G}]$ a.s. It follows that

$$E[\overline{X}_n \mid \mathcal{G}] \leqslant \inf_{j\geqslant n} E[X_j \mid \mathcal{G}] \text{ a.s.}$$

Monotone convergence implies

$$E[X \mid \mathcal{G}] = \lim_{n\to\infty} E[\overline{X}_n \mid \mathcal{G}] \leqslant \liminf_{n\to\infty} E[X_n \mid \mathcal{G}] \text{ a.s.,}$$

and we are done. □

Theorem 6.57 (dominated convergence)**.** *Let X_n be random variables with $X_n \to X$ a.s. Suppose there exists an integrable random variable Y such that $|X_n| \leqslant Y$ a.s. for all n. Then*

$$\lim_{n\to\infty} E[X_n \mid \mathcal{G}] = E[X \mid \mathcal{G}] \ a.s.$$

Proof. Exercise 6.14. □

Lemma 6.58. *Show that if $\varphi : \mathbb{R} \to \mathbb{R}$ is convex, then the left-hand derivative,*

$$\varphi'_-(c) = \lim_{h\downarrow 0} \frac{\varphi(c) - \varphi(c-h)}{h}$$

exists for all c. Moreover,

$$\varphi(x) - \varphi(c) - (x-c)\varphi'_-(c) \geqslant 0, \tag{6.8}$$

for all x and c.

Proof. Exercise 6.15. □

Theorem 6.59. (Jensen's inequality) *If φ is convex and X and $\varphi(X)$ are integrable, then $\varphi(E[X \mid \mathcal{G}]) \leqslant E[\varphi(X) \mid \mathcal{G}]$.*

Proof. Let $Z = (X - E[X \mid \mathcal{G}])\varphi'_-(E[X \mid \mathcal{G}])$, so that by (6.8), $\varphi(X) - \varphi(E[X \mid \mathcal{G}]) - Z \geqslant 0$, which implies

$$0 \leqslant E[\varphi(X) - \varphi(E[X \mid \mathcal{G}]) - Z \mid \mathcal{G}] = E[\varphi(X) \mid \mathcal{G}] - \varphi(E[X \mid \mathcal{G}]) - E[Z \mid \mathcal{G}].$$

It therefore suffices to show that $E[Z \mid \mathcal{G}] = 0$. To see this, we calculate

$$\begin{aligned} E[Z \mid \mathcal{G}] &= E[(X - E[X \mid \mathcal{G}])\varphi'_-(E[X \mid \mathcal{G}]) \mid \mathcal{G}] \\ &= \varphi'_-(E[X \mid \mathcal{G}])E[X - E[X \mid \mathcal{G}] \mid \mathcal{G}] \\ &= \varphi'_-(E[X \mid \mathcal{G}])(E[X \mid \mathcal{G}] - E[E[X \mid \mathcal{G}] \mid \mathcal{G}]) \\ &= \varphi'_-(E[X \mid \mathcal{G}])(E[X \mid \mathcal{G}] - E[X \mid \mathcal{G}]) = 0, \end{aligned}$$

and we are done. □

Theorem 6.60 (Hölder's inequality)**.** *Let $p, q \in (1, \infty)$ be conjugate exponents, so that $1/p + 1/q = 1$. Suppose that $|X|^p$ and $|Y|^q$ are integrable. Then*

$$E[|XY| \mid \mathcal{G}] \leqslant (E[|X|^p \mid \mathcal{G}])^{1/p}(E[|Y|^q \mid \mathcal{G}])^{1/q} \text{ a.s.}$$

Proof. Note that by the ordinary Hölder's inequality, XY is integrable, so that $E[|XY| \mid \mathcal{G}]$ is well-defined. Let $U = (E[|X|^p \mid \mathcal{G}])^{1/p}$ and $V = (E[|Y|^q \mid \mathcal{G}])^{1/q}$. Note that both U and V are $\mathcal{G}$-measurable. Observe that

$$\begin{aligned} E[|X|^p 1_{\{U=0\}}] &= E[E[|X|^p 1_{\{U=0\}} \mid \mathcal{G}]] \\ &= E[1_{\{U=0\}} E[|X|^p \mid \mathcal{G}]] = E[1_{\{U=0\}} U^p] = 0. \end{aligned}$$

Hence, $|X| 1_{\{U=0\}} = 0$ a.s., which implies

$$E[|XY| \mid \mathcal{G}] 1_{\{U=0\}} = E[|XY| 1_{\{U=0\}} \mid \mathcal{G}] = 0.$$

Similarly, $E[|XY| \mid \mathcal{G}] 1_{\{V=0\}} = 0$. It therefore suffices to show that $E[|XY| \mid \mathcal{G}] 1_H \leqslant UV$, where $H = \{U > 0, V > 0\}$. For this, we will use Lemma 6.50 to prove that

$$\frac{E[|XY| \mid \mathcal{G}]}{UV} 1_H \leqslant 1 \text{ a.s.}.$$

Note that the left-hand side is defined to be zero on H^c.

Let $A \in \mathcal{G}$ be arbitrary and define $G = H \cap A$. Then

$$\begin{aligned} E\left[\frac{E[|XY| \mid \mathcal{G}]}{UV} 1_H 1_A\right] &= E\left[E\left[\frac{|XY|}{UV} 1_G \,\middle|\, \mathcal{G}\right]\right] \\ &= E\left[\frac{|X|}{U} 1_G \cdot \frac{|Y|}{V} 1_G\right] \\ &\leqslant \left(E\left[\frac{|X|^p}{U^p} 1_G\right]\right)^{1/p} \left(E\left[\frac{|Y|^q}{V^q} 1_G\right]\right)^{1/q} \\ &= \left(E\left[\frac{E[|X|^p \mid \mathcal{G}]}{U^p} 1_G\right]\right)^{1/p} \left(E\left[\frac{E[|Y|^q \mid \mathcal{G}]}{V^q} 1_G\right]\right)^{1/q} \\ &= (E[1_G])^{1/p}(E[1_G])^{1/q} = E[1_G] \leqslant E[1_A]. \end{aligned}$$

Applying Lemma 6.50 finishes the proof. □

6.2.5 Minimizing the mean square error

We say that a random variable X is **square integrable** if $E|X|^2 < \infty$. Let X be square integrable and consider the function $f(a) = E|X-a|^2 = a^2 - 2(EX)a + E|X|^2$. This function has a minimum at $a = EX$. In other words, if we wish to approximate X by a constant, then the constant EX is the one which minimizes our mean square error.

The conditional expectation has a similar property. If we wish to approximate X by a square integrable, $\mathcal{G}$-measurable random variable, then $E[X \mid \mathcal{G}]$ is the random variable which minimizes our mean square error. This is made precise in the following theorem.

Theorem 6.61. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X be square integrable. Let $\mathcal{G} \subset \mathcal{F}$ and define $Z = E[X \mid \mathcal{G}]$. If Y is any square integrable, $\mathcal{G}$-measurable random variable, then $E|X-Z|^2 \leqslant E|X-Y|^2$.*

Proof. First note that by Jensen's inequality,

$$|Z|^2 = |E[X \mid \mathcal{G}]|^2 \leqslant E[|X|^2 \mid \mathcal{G}] \text{ a.s.}$$

Hence, $E|Z|^2 \leqslant E[E[|X|^2 \mid \mathcal{G}]] = E|X|^2 < \infty$ and Z is square integrable. Let $W = Z - Y$. Since W is $\mathcal{G}$-measurable,

$$E[WZ] = E[WE[X \mid \mathcal{G}]] = E[E[WX \mid \mathcal{G}]] = E[WX].$$

Hence, $E[W(X-Z)] = 0$, which implies

$$\begin{aligned} E|X-Y|^2 = E|X-Z+W|^2 &= E|X-Z|^2 + 2E[W(X-Z)] + E|W|^2 \\ &= E|X-Z|^2 + E|W|^2 \geqslant E|X-Z|^2, \end{aligned}$$

and we are done. □

Remark 6.62. In the language of Hilbert spaces and L^p spaces, this theorem says the following: X is an element of the Hilbert space $L^2(\Omega, \mathcal{F}, P)$, and $E[X \mid \mathcal{G}]$ is the orthogonal projection of X onto the subspace $L^2(\Omega, \mathcal{G}, P)$.

Exercises

6.10. Prove Lemma 6.29.

6.11. Prove Lemma 6.33. [Hint: First prove it when X is an indicator function, then a simple function, then a nonnegative function, then a general random variable.]

6.12. Prove Proposition 6.36. [Hint: It may be notationally convenient to write $E[X \mid \mathcal{G}] = \sum_{j=1}^n E[X \mid A_j] 1_{A_j}$.]

6.13. Prove Proposition 6.39.

6.14. Prove Theorem 6.57.

6.15. Prove Lemma 6.58.

6.16. [2, Exercise 5.1.1] Let X, Y be integrable random variables on a probability space, $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra, and let $A \in \mathcal{G}$. Assume that $X1_A = Y1_A$ a.s.

(a) Fix $\varepsilon > 0$ and let $B = \{E[X \mid \mathcal{G}] - E[Y \mid \mathcal{G}] \geqslant \varepsilon\}$. Use the definition of conditional expectation to prove that

$$E[E[X \mid \mathcal{G}]1_{A \cap B} - E[Y \mid \mathcal{G}]1_{A \cap B}] = 0.$$

(b) Use the definition of B to prove that

$$E[E[X \mid \mathcal{G}]1_{A \cap B} - E[Y \mid \mathcal{G}]1_{A \cap B}] \geqslant \varepsilon P(A \cap B),$$

and conclude that $P(A \cap B) = 0$.

(c) Use Parts (a) and (b) to prove that $E[X \mid \mathcal{G}]1_A \leqslant E[Y \mid \mathcal{G}]1_A$ a.s.

(d) Prove that $E[X \mid \mathcal{G}]1_A = E[Y \mid \mathcal{G}]1_A$ a.s.

6.17. [2, Exercise 5.1.2] Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Let $A \in \mathcal{G}$ and $B \in \mathcal{F}$.

(a) Prove that

$$P(A \mid B) = \frac{E[P(B \mid \mathcal{G})1_A]}{E[P(B \mid \mathcal{G})]}.$$

(b) (Bayes' theorem) Suppose $\mathcal{G}$ is generated by a partition, $\{A_j\}_{j=1}^n$. Use Part (a) to show that

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_{j=1}^n P(A_j)P(B \mid A_j)},$$

for all $i \in \{1, \ldots, n\}$.

6.18. [2, Exercise 5.1.3] Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Let $X \in L^2(\Omega)$ and $a > 0$. Prove that

$$P(|X| > a \mid \mathcal{G}) \leqslant \frac{E[X^2 \mid \mathcal{G}]}{a^2} \text{ a.s.}$$

6.19. [2, Exercise 5.1.6] Let $\Omega = \{a, b, c\}$ and $\mathcal{F} = 2^\Omega$. Show by example that there exists a probability measure P on $(\Omega, \mathcal{F})$, a random variable X on Ω, and σ-algebras $\mathcal{F}_j \subset \mathcal{F}$ for which it is not the case that

$$E[E[X \mid \mathcal{F}_1] \mid \mathcal{F}_2] = E[E[X \mid \mathcal{F}_2] \mid \mathcal{F}_1] \text{ a.s.}$$

6.20. [2, Exercise 5.1.7] Let X and Y be random variables such that $E|X|$, $E|Y|$, and $E|XY|$ are finite. Consider the following statements: (i) X and Y are independent. (ii) $E[Y \mid X] = EY$ a.s. (iii) $E[XY] = (EX)(EY)$.

(a) Prove that (i) implies (ii), and (ii) implies (iii).

(b) Find $\{-1, 0, 1\}$-valued random variables X and Y such that (ii) holds and (i) fails.

(c) Find $\{-1, 0, 1\}$-valued random variables X and Y such that (iii) holds and (ii) fails.

6.21. [2, Exercise 5.1.8] Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ σ-algebras. Let X be a square-integrable random variable. Prove that if $\mathcal{G} \subset \mathcal{H}$, then

$$E|X - E[X \mid \mathcal{H}]|^2 + E|E[X \mid \mathcal{H}] - E[X \mid \mathcal{G}]|^2 = E|X - E[X \mid \mathcal{G}]|^2 \text{ a.s.}$$

[Remark: This implies $E|X - E[X \mid \mathcal{H}]|^2 \leqslant E|X - E[X \mid \mathcal{G}]|^2$. In other words, $\mathcal{G} \subset \mathcal{H}$ implies $E[X \mid \mathcal{H}]$ is closer to X in L^2 than $E[X \mid \mathcal{G}]$. If $\mathcal{G} = \{\varnothing, \Omega\}$, then we obtain $E|X - E[X \mid \mathcal{H}]|^2 \leqslant \operatorname{var}(X)$.]

6.22. [2, Exercise 5.1.9] Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Let X be a square-integrable random variable. Define

$$\operatorname{var}(X \mid \mathcal{G}) = E[|X - E[X \mid \mathcal{G}]|^2 \mid \mathcal{G}].$$

(a) Prove that $\operatorname{var}(X \mid \mathcal{G}) = E[X^2 \mid \mathcal{G}] - (E[X \mid \mathcal{G}])^2$ a.s.

(b) Prove that $\operatorname{var}(X) = E[\operatorname{var}(X \mid \mathcal{G})] + \operatorname{var}(E[X \mid \mathcal{G}])$.

6.23. [2, Exercise 5.1.11] Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Let X be a square-integrable random variable, and let $Z = E[X \mid \mathcal{G}]$. Suppose that $EX^2 = EZ^2$. Prove that $X = Z$ a.s.

6.3 Regular conditional distributions

This section corresponds to [2, Subsection 5.1.3].

6.3.1 Introduction

If X is a real-valued random variable, then $\mu_X(A) = P(X \in A)$ defines a measure μ_X on the real line which we call the distribution (or law) of X. One feature of the distribution is that it provides us with a way to calculate expectations:

$$E[f(X)] = \int_{\mathbb{R}} f(x)\, \mu_X(dx).$$

Likewise, if B is an event, then $\mu_{X,B}(A) = P(X \in A \mid B)$ defines a measure $\mu_{X,B}$ on $\mathbb{R}$ which is the conditional distribution of X given B, and we have

$$E[f(X) \mid B] = \int_{\mathbb{R}} f(x)\, \mu_{X,B}(dx).$$

If $\mathcal{G}$ is a finite σ-algebra, so that $\mathcal{G} = \sigma(\{A_j\}_{j=1}^n)$, where $\{A_j\}_{j=1}^n$ is a partition of Ω, then

$$P(X \in A \mid \mathcal{G})(\omega) = \begin{cases} P(X \in A \mid A_1) & \text{if } \omega \in A_1, \\ P(X \in A \mid A_2) & \text{if } \omega \in A_2, \\ \quad \vdots & \\ P(X \in A \mid A_n) & \text{if } \omega \in A_n. \end{cases}$$

In other words, $P(X \in \cdot \mid \mathcal{G})$ is just a conditional probability distribution that happens to depend on ω. Another way of saying it is that $P(X \in \cdot \mid \mathcal{G})$ is a random probability measure on the real line.

Conditional expectations can be computed by integrating against this random measure. That is, if we define $\mu_{X,\mathcal{G}}(\omega, A) = \mu_{X,A_j}(A)$ for $\omega \in A_j$, then

$$E[f(X) \mid \mathcal{G}](\omega) = \int_{\mathbb{R}} f(x)\, \mu_{X,\mathcal{G}}(\omega, dx).$$

With a structure such as this, expectations conditioned on σ-algebras behave very much like ordinary expectations. When this happens, we are able to make valuable intuitive connections to mathematical ideas that we are already familiar with. It would be nice if $P(X \in \cdot \mid \mathcal{G})$ was always a random measure, even when $\mathcal{G}$ is infinite. The following theorem is a step in this direction.

Theorem 6.63. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ a measurable space. Let X be an S-valued random variable and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Then*

(i) $P(X \in A \mid \mathcal{G}) \in [0,1]$ *a.s., for all* $A \in \mathcal{S}$.

(ii) $P(X \in \varnothing \mid \mathcal{G}) = 0$ *a.s.*

(iii) $P(X \in S \mid \mathcal{G}) = 1$ *a.s.*

(iv) $P(X \in \biguplus_{n=1}^\infty A_n \mid \mathcal{G}) = \sum_{n=1}^\infty P(X \in A_n \mid \mathcal{G})$ *a.s., for all disjoint collections* $\{A_n\} \subset \mathcal{S}$.

Proof. Exercise 6.24. □

Unfortunately, Theorem 6.63 does not show that $A \mapsto P(X \in A \mid \mathcal{G})(\omega)$ is a measure for P-a.e. $\omega \in \Omega$. This is because the null set in (iv) can depend on the collection $\{A_n\}$. So there may not exist a single event of probability one on which (iv) holds simultaneously for all disjoint collections.

However, when X takes values in a standard Borel space, such as the real line, it is possible to express $P(X \in \cdot \mid \mathcal{G})$ as a genuine random measure. The remainder of this section elaborates on this topic.

6.3.2 Random measures

Let $(S, \mathcal{S})$ be a measurable space and let $M(S)$ be the set of all σ-finite measures on $(S, \mathcal{S})$. Let $\mathcal{M}(S)$ be the σ-algebra on $M(S)$ generated by sets of the form $\{\nu : \nu(A) \in B\}$, where $A \in \mathcal{S}$ and $B \in \mathcal{R}$. Note that $\mathcal{M}(S)$ is the smallest σ-algebra such that the projection functions $\pi_A : M(S) \to \mathbb{R}$, defined by $\pi_A(\nu) = \nu(A)$, are $\mathcal{M}(S)$-measurable for all $A \in \mathcal{S}$. Taking $A = S$ and $B = \{1\}$ shows that $M_1(S)$, the set of all probability measures on $(S, \mathcal{S})$ is measurable. Let $\mathcal{M}_1(S)$ denote $\mathcal{M}(S)$ restricted to $M_1(S)$.

Let $(T, \mathcal{T})$ be another measurable space. If $\mu : T \to M(S)$, we will write $\mu(t, A) = (\mu(t))(A)$. Note that μ is $(\mathcal{T}, \mathcal{M}(S))$-measurable if and only if $\pi_A \circ \mu = \mu(\cdot, A)$ is $(\mathcal{T}, \mathcal{R})$-measurable for all $A \in \mathcal{S}$. Any such measurable function is called a **kernel** from T to S. If μ takes values in $M_1(S)$, then μ is a **probability kernel**.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A **random measure** on S is an $M(S)$-valued random variable. In other words, it is a kernel from Ω to S. If a random measure takes values in $M_1(S)$, then it is a **random probability measure** on S. Note that if μ is a kernel from T to S and Y is a T-valued random variables, then $\mu(Y)$ is a random measure.

6.3.3 Regular conditional distributions

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ and $(T, \mathcal{T})$ measurable spaces. Let X and Y be S- and T-valued random variables, respectively. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. If there exists a random measure $\mu = \mu_{X,\mathcal{G}}$ on S such that $P(X \in A \mid \mathcal{G}) = \mu(\cdot, A)$ a.s. for every $A \in \mathcal{S}$, then μ is a **regular conditional distribution** for X given $\mathcal{G}$, and we write $X \mid \mathcal{G} \sim \mu$. Similarly, if there exists a probability kernel $\mu = \mu_{X,Y}$ from T to S such that $P(X \in A \mid Y) = \mu(Y, A)$ a.s. for every $A \in \mathcal{S}$, then $\mu(Y)$ is a regular conditional distribution for X given $\sigma(Y)$, and we write $X \mid Y \sim \mu(Y)$.

The following theorem is an expanded version of [2, Theorem 5.1.9]. The version that appears here, along with its proof, can be found in [7, Theorem 5.3].

Theorem 6.64. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ and $(T, \mathcal{T})$ measurable spaces. Let X and Y be S- and T-valued random variables, respectively. If S is a standard Borel space, then there exists a probability kernel $\mu = \mu_{X,Y}$ from T to S such that $X \mid Y \sim \mu(Y)$. If $\widetilde{\mu}$ is another such probability kernel, then $\mu = \widetilde{\mu}$, μ_Y-a.e.*

Corollary 6.65. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ a measurable space. Let X be an S-valued random variable and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. If S is a standard Borel space, then there exists a $\mathcal{G}$-measurable random probability measure $\mu = \mu_{X,\mathcal{G}}$ such that $X \mid \mathcal{G} \sim \mu$. If $\widetilde{\mu}$ is another such random probability measure, then $\mu = \widetilde{\mu}$ a.s.*

Proof. Apply Theorem 6.64 with $(T, \mathcal{T}) = (\Omega, \mathcal{G})$ and Y the identity function. □

The first example of what we can do with regular conditional distributions is the following theorem, which can be regarded as a generalized version of Theorem 6.52.

Theorem 6.66. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ a measurable space. Let X be an S-valued random variable, $\mathcal{G} \subset \mathcal{F}$ a σ-algebra, and suppose $X \mid \mathcal{G} \sim \mu$. Let $(T, \mathcal{T})$ be a measurable space and Y a T-valued random variable. Let $f : S \times T \to \mathbb{R}$ be $(\mathcal{S} \times \mathcal{T}, \mathcal{R})$-measurable with $E|f(X,Y)| < \infty$. If $Y \in \mathcal{G}$, then*

$$E[f(X,Y) \mid \mathcal{G}] = \int_S f(x,Y)\, \mu(\cdot, dx) \quad a.s.$$

Proof. If $f = 1_{A\times B}$, where $A \in \mathcal{S}$ and $B \in \mathcal{T}$, then

$$\begin{aligned} E[1_{A\times B}(X,Y) \mid \mathcal{G}] &= 1_B(Y) P(X \in A \mid \mathcal{G}) \\ &= 1_B(Y)\mu(\cdot, A) = \int_S 1_{A\times B}(x,Y)\, \mu(\cdot, dx) \quad \text{a.s.} \end{aligned}$$

By the π-λ theorem, this proves the result for $f = 1_C$, where $C \in \mathcal{S} \times \mathcal{T}$. By linearity (Theorem 6.49), the result holds for all simple functions f. By monotone convergence (Theorem 6.55), the result holds for all nonnegative functions f satisfying $E|f(X,Y)| < \infty$. And finally, by considering the positive and negative parts, the result holds for all measurable functions f satisfying $E|f(X,Y)| < \infty$. □

Corollary 6.67. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ a measurable space. Let X be an S-valued random variable, $\mathcal{G} \subset \mathcal{F}$ a σ-algebra, and suppose $X \mid \mathcal{G} \sim \mu$. If $f : S \to \mathbb{R}$ is $(\mathcal{S}, \mathcal{R})$-measurable with $E|f(X)| < \infty$, then*

$$E[f(X) \mid \mathcal{G}] = \int_S f(x)\, \mu(\cdot, dx) \quad a.s.$$

Proof. Apply Theorem 6.66 with Y a constant random variable. □

For our second example, we give a simple proof of Theorem 6.60 (Hölder's inequality).

Proof of Theorem 6.60. Since $(\mathbb{R}^2, \mathcal{R}^2)$ is a standard Borel space, there exists a random measure μ on $\mathbb{R}^2$ such that $(X,Y) \mid \mathcal{G} \sim \mu$. Since $|X|^p$ and $|Y|^q$ are integrable, the ordinary Hölder's inequality implies $|XY|$ is integrable. Thus, by Corollary 6.67,

$$E[|XY| \mid \mathcal{G}] = \int_{\mathbb{R}^2} |xy|\, \mu(\cdot, dx\, dy) \quad \text{a.s.}$$

For P-a.e. $\omega \in \Omega$, we may apply the ordinary Hölder's inequality to the measure $\mu(\omega, \cdot)$, yielding

$$E[|XY| \mid \mathcal{G}] \leqslant \left(\int_{\mathbb{R}^2} |x|^p\, \mu(\cdot, dx\, dy) \right)^{1/p} \left(\int_{\mathbb{R}^2} |y|^q\, \mu(\cdot, dx\, dy) \right)^{1/q} \quad \text{a.s.}$$

Applying Theorem 6.66 once again finishes the proof. □

For our final example, let us first consider a property of unconditioned expectations. If X is a real-valued random variable, and $h : \mathbb{R} \to [0, \infty)$ is absolutely continuous with $h' \geqslant 0$ a.e. and $h(x) \downarrow 0$ as $x \to -\infty$, then

$$E[h(X)] = \int_{\mathbb{R}} h'(t) P(X > t)\, dt. \tag{6.9}$$

This is the content of [2, Exercise 2.2.7], and one way to see it is to use Fubini's theorem:

$$\begin{aligned} E[h(X)] = \int_{\Omega} h(X)\, dP = \int_{\Omega} \int_{-\infty}^{X} h'(t)\, dt\, dP = \int_{\Omega} \int_{\mathbb{R}} 1_{\{X>t\}} h'(t)\, dt\, dP \\ = \int_{\mathbb{R}} h'(t) \int_{\Omega} 1_{\{X>t\}}\, dP\, dt = \int_{\mathbb{R}} h'(t) P(X > t)\, dt. \end{aligned}$$

It is then natural to ask whether a similar thing is true for conditional expectations, and whether a similar proof can demonstrate it. We will answer both questions in the affirmative by using regular conditional probabilities.

Theorem 6.68. *Let X be a real-valued random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Let $h : \mathbb{R} \to [0, \infty)$ be absolutely continuous with $h' \geqslant 0$ a.e. and $h(x) \downarrow 0$ as $x \to -\infty$. Suppose that $E|h(X)| < \infty$. Then it is possible to choose, for each $t \in \mathbb{R}$, a version of $P(X > t \mid \mathcal{G})$ so that the function $t \mapsto h'(t)P(X > t \mid \mathcal{G})$ is almost surely Lebesgue integrable on $\mathbb{R}$, and satisfies*

$$E[h(X) \mid \mathcal{G}] = \int_{\mathbb{R}} h'(t) P(X > t \mid \mathcal{G})\, dt \quad a.s.$$

Proof. Since $(\mathbb{R}, \mathcal{R})$ is a standard Borel space, there exists a random measure μ such that $X \mid \mathcal{G} \sim \mu$. For each $t \in \mathbb{R}$, let us choose the version of $P(X > t \mid \mathcal{G})$ determined by μ, that is, $P(X > t \mid \mathcal{G})(\omega) = \mu(\omega, (t, \infty))$. Then for P-a.e. $\omega \in \Omega$, we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}} 1_{\{x>t\}} h'(t)\, dt\, \mu(\omega, dx) = \int_{\mathbb{R}} h(x)\, \mu(\omega, dx) = E[h(X) \mid \mathcal{G}](\omega) < \infty.$$

By Fubini's theorem, the function

$$t \mapsto \int_{\mathbb{R}} 1_{\{x>t\}} h'(t)\, \mu(\omega, dx) = h'(t)\mu(\omega, (t, \infty)) = h'(t) P(X > t \mid \mathcal{G})(\omega)$$

is Lebesgue integrable on $\mathbb{R}$ and

$$\begin{aligned} \int_{\mathbb{R}} h'(t) P(X > t \mid \mathcal{G})(\omega)\, dt &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{\{x>t\}} h'(t)\, \mu(\omega, dx)\, dt \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{\{x>t\}} h'(t)\, dt\, \mu(\omega, dx) = E[h(X) \mid \mathcal{G}](\omega), \end{aligned}$$

which proves the theorem. ∎

Exercises

6.24. Prove Theorem 6.63.

6.25. [2, Exercise 5.1.13] Let X and Y have joint density function $f(x, y)$. For $y \in \mathbb{R}$ and $A \in \mathcal{R}$, define

$$\mu(y, A) = \frac{\int_A f(x, y)\, dx}{\int_{\mathbb{R}} f(x, y)\, dx},$$

if $\int_{\mathbb{R}} f(x, y)\, dx \in (0, \infty)$, and $\mu(y, A) = 1_A(0)$ otherwise. Prove that μ is a probability kernel from $\mathbb{R}$ to $\mathbb{R}$ and that $X \mid Y \sim \mu(Y)$.

6.4 A preview of stochastic processes

A **stochastic process** is a collection of random variable $\{X(t) : t \in T\}$ indexed by some set T, defined on a common probability space, $(\Omega, \mathcal{F}, P)$, and taking values in a common measurable space, $(S, \mathcal{S})$. We usually think of T as time. A **discrete time stochastic process** is where $T = \mathbb{N}$, in which case the process is just a sequence of random variables.

Let $\{X_n : n \in \mathbb{N}\}$ be a discrete time stochastic process. Define $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. The σ-algebra $\mathcal{F}_n$ represents all the information at time n that we would have from observing the values $X_1, \ldots, X_n$. Note that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$.

More generally, a **filtration** is a sequence of σ-algebras $\{\mathcal{F}_n\}_{n=1}^\infty$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$. A stochastic process $\{X_n : n \in \mathbb{N}\}$ is said to be **adapted** to the filtration $\{\mathcal{F}_n\}_{n=1}^\infty$ if X_n is $\mathcal{F}_n$-measurable for all n. The special case $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is called the **filtration generated by** X, and is denoted by $\{\mathcal{F}_n^X\}_{n=1}^\infty$.

An important class of discrete time stochastic processes is the martingales. A real-valued stochastic process $\{X_n : n \in \mathbb{N}\}$ is a **martingale** with respect to the filtration $\{\mathcal{F}_n\}_{n=1}^\infty$ if

(i) X_n is integrable for all n,

(ii) $\{X_n : n \in \mathbb{N}\}$ is adapted to $\{\mathcal{F}_n\}_{n=1}^\infty$, and

(iii) $E[X_{n+1} \mid \mathcal{F}_n] = X_n$ for all n.

The critical item is (iii). Imagine that X_n models our cumulative wealth as we play a sequence of gambling games. Condition (iii) says that, given all the information up to time n, our expected wealth at time $n + 1$ is the same as our wealth at time n. In other words, a martingale models a "fair" game.

Another important class of discrete time stochastic processes is the Markov chains. A stochastic process $\{X_n : n \in \mathbb{N}\}$ is a **Markov chain** with respect to the filtration $\{\mathcal{F}_n\}_{n=1}^\infty$ if

(i) $\{X_n : n \in \mathbb{N}\}$ is adapted to $\{\mathcal{F}_n\}_{n=1}^\infty$, and

(ii) $P(X_{n+1} \in B \mid \mathcal{F}_n) = P(X_{n+1} \in B \mid X_n)$ for all $B \in \mathcal{S}$.

Here, the critical item is (ii). It is called the **Markov property**. In words, it says that the conditional distribution of X_{n+1} given all the information up to time n is the same as if we were only given X_n. In other words, the future behavior of a Markov chain depends only on the present location of the chain, and not on how it got there.

The canonical example of a Markov chain is a random walk. If $\{\xi_j\}_{j=1}^{\infty}$ are i.i.d., $\mathbb{R}^d$-valued random variables, and $X_n = \xi_1 + \cdots + \xi_n$, then $\{X_n : n \in \mathbb{N}\}$ is a **random walk**. The random walk is a Markov chain with respect to the filtration generated by X. Moreover, if each ξ_j is real-valued and integrable with mean zero, then the random walk is also a martingale.

A continuous time stochastic process has the form $\{X(t) : t \in [0, \infty)\}$. Examples include the Poisson process and Brownian motion. Concepts such as filtrations, adaptedness, martingales, and the Markov property can all be extended to continuous time. Care is needed however, because (for one thing) the time domain is uncountable. Brownian motion is the continuous time analog of a random walk. It is the canonical example in continuous time of both a martingale and a Markov process. It can be realized as the limit of a sequence of random walks, where the step sizes are becoming smaller and the steps are occurring more frequently.

More specifically, let $\{X_n : n \in \mathbb{N}\}$ be a mean zero random walk. Let $X(t) = X_{\lfloor t \rfloor}$, where $\lfloor \cdot \rfloor$ denotes the greatest integer function. Then the sequence of processes

$$\left\{ \frac{X(nt)}{\sqrt{n}} : t \in [0, \infty) \right\}$$

converges (in a certain sense) as $n \to \infty$ to a continuous time stochastic process called Brownian motion. This is the conclusion of Donsker's theorem, which is a kind of central limit theorem for stochastic processes.

Differential equations involving Brownian motion are referred to as stochastic differential equations (SDEs). SDEs are used to model dynamical systems that involve randomness, and are very common in scientific applications. In order to understand SDEs, one must first understand the stochastic integral (with respect to Brownian motion), which behaves quite differently from the ordinary Lebesgue-Stieltjes integral. In particular, the classical fundamental theorem of calculus no longer applies when one is working with stochastic integrals. It must be replaced by a new rule called Itô's rule. Itô's rule gives rise to a whole new calculus called stochastic calculus.

Chapter 7

Modes of Convergence

7.1 Convergence in probability

This section corresponds to [2, pp. 53–54, 65–66].

Let $\{X_n\}_{n=1}^{\infty}$ and X be random variables on a probability space $(\Omega, \mathcal{F}, P)$. Then $X_n \to X$ **in probability** if $X_n \to X$ in measure. In other words, $X_n \to X$ in probability if and only if, for all $\varepsilon > 0$, we have

$$P(|X_n - X| \geqslant \varepsilon) \to 0,$$

as $n \to \infty$.

Lemma 7.1. *If $p > 0$ and $E|X_n - X|^p \to 0$ as $n \to \infty$, then $X_n \to X$ in probability.*

Proof. By Chebyshev's inequality,

$$P(|X_n - X| \geqslant \varepsilon) \leqslant \frac{E|X_n - X|^p}{\varepsilon^p},$$

which tends to 0. □

Also recall the following from Section 2.4. If $X_n \to X$ a.s., then $X_n \to X$ in probability. Conversely, if $X_n \to X$ in probability, then there exists a subsequence such that $X_{n_k} \to X$ a.s.

Also recall Exercise 2.13, which shows that convergence in probability is metrizable. That is, there exists a metric ρ on $L^0(\Omega, \mathcal{F}, P)$ such that $X_n \to X$ in probability if and only if $\rho(X_n, X) \to 0$. In fact, in can be shown that this metric space is complete. (See [2, Exercise 2.3.9].)

Recall the following fact about metric spaces ([2, Theorem 2.3.3]).

Theorem 7.2. *Let $\{x_n\}$ be a sequence in a metric space (M, ρ) and let $x \in M$. Then $x_n \to x$ as $n \to \infty$ if and only if every subsequence $\{x_{n(m)}\}$ has a further subsequence $\{x_{n(m_k)}\}$ such that $x_{n(m_k)} \to x$ as $k \to \infty$.*

We can use this to prove the following result about convergence in probability.

Theorem 7.3. *Let $\{X_n\}_{n=1}^\infty$ and X be random variables on a probability space, $(\Omega, \mathcal{F}, P)$. Then $X_n \to X$ in probability as $n \to \infty$ if and only if every subsequence $\{X_{n(m)}\}$ has a further subsequence $\{X_{n(m_k)}\}$ such that $X_{n(m_k)} \to X$ a.s. as $k \to \infty$.*

Proof. The "only if" part follows from the results in Section 2.4. For the "if" part, fix $\varepsilon > 0$ and define $x_n := P(|X_n - X| \geqslant \varepsilon)$. Then $\{x_n\}$ is a sequence in the metric space $\mathbb{R}$ with the Euclidean metric. Let $\{x_{n(m)}\}$ be a subsequence. By hypothesis, the subsequence $\{X_{n(m)}\}$ has a further subsequence $\{X_{n(m_k)}\}$ such that $X_{n(m_k)} \to X$ a.s. This implies that $X_{n(m_k)} \to X$ in probability. By the definition of convergence in probability, this gives $x_{n(m_k)} \to 0$. By Theorem 7.2, therefore, we have $x_n \to 0$, and so $X_n \to X$ in probability. □

Remark 7.4. Since convergence in probability does not imply almost sure convergence, this theorem shows that almost sure convergence is not metrizable.

Theorem 7.5. *Let $X_n \to X$ in probability and let $f : \mathbb{R} \to \mathbb{R}$. If f is continuous, then $f(X_n) \to f(X)$ in probability. If f is continuous and bounded, then $E[f(X_n)] \to E[f(X)]$.*

Proof. First suppose f is continuous. Let $\{n(m)\}$ be a strictly increasing sequence of natural numbers. Choose a subsequence $\{X_{n(m_k)}\}$ such that $X_{n(m_k)} \to X$ a.s. Since f is continuous, we have $f(X_{n(m_k)}) \to f(X)$ a.s. By Theorem 7.3, this implies $f(X_n) \to f(X)$ in probability.

Now assume f is continuous and bounded. Define $x_n = E[f(X_n)]$. Let $\{n(m)\}$ be a strictly increasing sequence of natural numbers. Choose a subsequence $\{X_{n(m_k)}\}$ such that $X_{n(m_k)} \to X$ a.s. Since f is continuous, we have $f(X_{n(m_k)}) \to f(X)$ a.s. Since f is bounded, by dominated convergence, we have $x_{n(m_k)} \to E[f(X)]$. Therefore, by Theorem 7.2, we have $x_n \to E[f(X)]$. □

Theorem 7.6. *Let X_n, X be nonnegative random variables with $X_n \to X$ in probability. Then $\liminf_{n\to\infty} EX_n \geqslant EX$.*

Proof. Exercise 7.1. □

Theorem 7.7. *Let X_n, X be real-valued random variables with $X_n \to X$ in probability. If there exists an integrable random variable Y such that $|X_n| \leqslant Y$ a.s. for all n, then $EX_n \to EX$.*

Proof. Exercise 7.2. □

Theorem 7.8. *Let X_n, X be real-valued random variables with $X_n \to X$ in probability. Let $g : \mathbb{R} \to [0, \infty)$ and $h : \mathbb{R} \to \mathbb{R}$ be continuous. Assume that*

(i) $g(x) \to \infty$ *as* $|x| \to \infty$,

(ii) $|h(x)|/g(x) \to 0$ *as* $|x| \to \infty$, *and*

(iii) $\sup_n Eg(X_n) < \infty$.

Then $Eh(X_n) \to Eh(X)$ *as* $n \to \infty$.

Proof. Exercise 7.3. □

Exercises

7.1. [2, Exercise 2.3.6] Prove Theorem 7.6.

7.2. [2, Exercise 2.3.7(a)] Prove Theorem 7.7.

7.3. [2, Exercise 2.3.7(b)] Prove Theorem 7.8.

7.2 The Borel-Cantelli lemmas

This section corresponds to [2, Section 2.3].

Let Ω be a set and for each $n \in \mathbb{N}$, let $A_n \subset \Omega$. Define $B_n = \bigcup_{k=n}^\infty A_k$. Note that $B_n \supset B_{n+1}$ for all n. We then define

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^\infty B_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k.$$

We may similarly define

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k.$$

It can be shown that if $A = \limsup_{n\to\infty} A_n$, then

$$\limsup_{n\to\infty} 1_{A_n}(\omega) = 1_A(\omega),$$

for all $\omega \in \Omega$, and similarly for $\liminf$.

Also note that

$$\begin{aligned}\limsup_{n\to\infty} A_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k &= \{\omega \in \Omega : \forall n \in \mathbb{N}, \exists k \geqslant n, \omega \in A_k\}\\ &= \{\omega \in \Omega : \omega \in A_k \text{ for infinitely many } k \in \mathbb{N}\}.\end{aligned}$$

We therefore adopt the notation $\{A_n \text{ i.o.}\} := \limsup_{n\to\infty} A_n$, where i.o. stands for "infinitely often".

Theorem 7.9. *Let* $\{X_n\}$ *be a sequence of random variables on a probability space* $(\Omega, \mathcal{F}, P)$. *Then* $X_n \to 0$ *a.s. if and only if, for every* $\varepsilon > 0$ *we have*

$$P(|X_n| \geqslant \varepsilon \textit{ i.o.}) = 0.$$

Proof. First assume $X_n \to 0$ a.s. and fix $\varepsilon > 0$. Let $A_n = \{|X_n| \geqslant \varepsilon\}$. Then

$$\{A_n \text{ i.o.}\}^c = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} A_j^c = \{\omega \in \Omega : \exists n \in \mathbb{N}, \forall j \geqslant n, |X_j(\omega)| < \varepsilon\}$$

It follows that $\{X_n \to 0\} \subset \{A_n \text{ i.o.}\}^c$, and so $P(A_n \text{ i.o.}) = 0$.

For the converse, let

$$\Omega^* = \bigcap_{\varepsilon \in \{1/k : k \in \mathbb{N}\}} \{|X_n| \geqslant \varepsilon \text{ i.o.}\}^c.$$

By hypothesis, we have $P(\Omega^*) = 1$. As above, $X_n(\omega) \to 0$ for each $\omega \in \Omega^*$. Thus, $X_n \to 0$ a.s. □

Theorem 7.10 (Borel-Cantelli lemma). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n$ i.o.$) = 0$.*

Proof. Let $N = \sum_{n=1}^{\infty} 1_{A_n}$. Then $E[N] = \sum_{n=1}^{\infty} P(A_n) < \infty$. Thus, $N < \infty$ a.s. But $\{N = \infty\} = \{A_n \text{ i.o.}\}$. Thus, $P(A_n \text{ i.o.}) = 0$. □

Theorem 7.11 (the second Borel-Cantelli lemma). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$. Suppose $\{A_n\}$ are independent. If $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(A_n$ i.o.$) = 1$.*

Proof. Let $B_n = \bigcup_{k=n}^{\infty} A_k$ so that $B_n \downarrow \{A_n \text{ i.o.}\}$. Thus, $P(B_n) \to P(A_n \text{ i.o.})$. Fix $n \in \mathbb{N}$. Fix $\varepsilon > 0$. Since $\sum_{k=n}^{\infty} P(A_n) = \infty$, we may choose $N \geqslant n$ such that $\sum_{k=n}^{N} P(A_n) > \log(1/\varepsilon)$. Using independence and the inequality, $1 - x \leqslant e^{-x}$, we have

$$\begin{aligned} P(B_n) &\geqslant P\left(\bigcup_{k=n}^{N} A_k\right) = 1 - P\left(\bigcap_{k=n}^{N} A_k^c\right) = 1 - \prod_{k=n}^{N}(1 - P(A_k)) \\ &\geqslant 1 - \prod_{k=n}^{N} e^{-P(A_k)} = 1 - e^{-\sum_{k=n}^{N} P(A_n)} > 1 - e^{-\log(1/\varepsilon)} = 1 - \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this shows $P(B_n) = 1$. Thus, $P(A_n \text{ i.o.}) = 1$. □

Exercises

7.4. [2, Exercise 2.3.2 (modified)] Recall that $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$. Let X_n be random variables and assume $EX_n \sim an^{\alpha}$, where $a > 0$ and $\alpha > 0$. Also assume $\text{var}(X_n) \leqslant Bn^{\beta}$ for some $\beta < 2\alpha - 1$. Prove that $n^{-\alpha} X_n \to a$ a.s.

7.5. [2, Exercise 2.3.10] Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real-valued random variables on a probability space, $(\Omega, \mathcal{F}, P)$. Prove that there exist real constants c_n such that $X_n/c_n \to 0$ a.s.

7.6. [2, Exercise 2.3.12] Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$ a sequence of independent events. Assume that $P(A_n) < 1$ for all n, and that $P(\bigcup_n A_n) = 1$. Prove that $P(A_n \text{ i.o.}) = 1$.

7.3 Weak convergence

This section corresponds to [2, Section 3.2].

Let (M, ρ) be a metric space. For each $n \in \mathbb{N} \cup \{\infty\}$, let μ_n be a probability measure on $(M, \mathcal{B}_M)$, where $\mathcal{B}_M$ is the Borel σ-algebra. We say that μ_n **converges weakly** to μ_∞, written $\mu_n \Rightarrow \mu_\infty$, if

$$\int_M f\,d\mu_n \to \int_M f\,d\mu_\infty,$$

as $n \to \infty$, for every bounded, continuous $f : M \to \mathbb{R}$.

For each $n \in \mathbb{N} \cup \{\infty\}$, let $(\Omega_n, \mathcal{F}_n, P_n)$ be a probability space and let $X_n : \Omega_n \to M$ be an M-valued random variable. We say that X_n **converges in distribution** (or **converges in law**) to X_∞, written $X_n \Rightarrow X_\infty$, if $\mu_n \Rightarrow \mu_\infty$, where $X_n \sim \mu_n$. In other words, $X_n \Rightarrow X_\infty$ if and only if

$$E_n[f(X_n)] \to E_\infty[f(X_\infty)],$$

as $n \to \infty$, for every bounded, continuous $f : M \to \mathbb{R}$.

Remark 7.12. Let $\overline{C}(M)$ denote the set of all bounded, continuous $f : M \to \mathbb{R}$ and let μ be a probability measure on $(M, \mathcal{B}_M)$. Then $\overline{C}(M)$ is a vector space over the reals and the map $f \mapsto \int_M f\,d\mu$ is a linear functional on $\overline{C}(M)$. For this reason, one often sees alternative notation such as $\mu(f)$ or $\langle \mu, f \rangle$ for the integral $\int_M f\,d\mu$. Thought of in this way, weak convergence of probability measures is just pointwise convergence as functions on $\overline{C}(M)$.

Remark 7.13. For the remainder of this section, unless otherwise indicated, we will focus on the case $M = \mathbb{R}$ with the Euclidean metric.

Lemma 7.14. *For each $n \in \mathbb{N} \cup \{\infty\}$, let X_n be a real-valued random variable on some probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Let F_n be the distribution function of X_n. If $X_n \Rightarrow X_\infty$, then $F_n(x) \to F_\infty(x)$ for all x such that F_∞ is continuous at x.*

Proof. Suppose $X_n \Rightarrow X_\infty$ and fix $x \in \mathbb{R}$ such that F_∞ is continuous at x. Fix $\varepsilon > 0$. Define $f_{x,\varepsilon} : \mathbb{R} \to \mathbb{R}$ by

$$f_{x,\varepsilon}(t) = 1_{(-\infty,x]}(t) - \left(\frac{t-x}{\varepsilon}\right) 1_{(x,x+\varepsilon]}(t).$$

Then $f_{x,\varepsilon}$ is bounded and continuous, so $E_n[f_{x,\varepsilon}(X_n)] \to E_\infty[f_{x,\varepsilon}(X_\infty)]$. Note that $1_{(-\infty,x]} \leqslant f_{x,\varepsilon} \leqslant 1_{(-\infty,x+\varepsilon]}$. Thus,

$$\begin{aligned}\limsup_{n\to\infty} F_n(x) &= \limsup_{n\to\infty} E_n[1_{(-\infty,x]}(X_n)] \leqslant \limsup_{n\to\infty} E_n[f_{x,\varepsilon}(X_n)] \\ &= E_\infty[f_{x,\varepsilon}(X_\infty)] \leqslant E_\infty[1_{(-\infty,x+\varepsilon]}(X_\infty)] = F_\infty(x+\varepsilon).\end{aligned}$$

Letting $\varepsilon \to 0$ gives $\limsup_{n\to\infty} F_n(x) \leqslant F_\infty(x)$.

Similarly, $1_{(-\infty,x-\varepsilon]} \leqslant f_{x-\varepsilon,\varepsilon} \leqslant 1_{(-\infty,x]}$. Thus,

$$\begin{aligned}\liminf_{n\to\infty} F_n(x) = \liminf_{n\to\infty} E_n[1_{(\infty,x]}(X_n)] &\geqslant \liminf_{n\to\infty} E_n[f_{x-\varepsilon,\varepsilon}(X_n)] \\ &= E_\infty[f_{x-\varepsilon,\varepsilon}(X_\infty)] \geqslant E_\infty[1_{(-\infty,x-\varepsilon]}(X_\infty)] = F_\infty(x-\varepsilon).\end{aligned}$$

Since F_∞ is continuous at x, letting $\varepsilon \to 0$ gives $\liminf_{n\to\infty} F_n(x) \geqslant F_\infty(x)$. □

Theorem 7.15 (Skorohod representation theorem)**.** *For each $n \in \mathbb{N} \cup \{\infty\}$, let X_n be a real-valued random variable on some probability space $(\Omega_n, \mathcal{F}_n, P_n)$. If $X_n \Rightarrow X_\infty$, then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence $\{Y_n\}$ of random variables on $(\Omega, \mathcal{F}, P)$ such that $X_n =_d Y_n$ for all n, and $Y_n \to Y_\infty$ a.s.*

Proof sketch. Suppose $X_n \Rightarrow X_\infty$. Let F_n be the distribution function of X_n. By Lemma 7.14, we have $F_n(x) \to F_\infty(x)$ for all x such that F_∞ is continuous at x. For a proof that this implies the conclusion of the theorem, see the proof of [2, Theorem 3.2.2]. □

Remark 7.16. The Skorohod representation theorem does not require the X_n's to be real-valued. In fact, it is still true when all we assume is that the X_n's take values in a separable metric space (see [3, Theorem 3.1.8]). Moreover, by Exercise 7.9, the converse of the Skorohod representation theorem is also true.

Theorem 7.17. *For each $n \in \mathbb{N} \cup \{\infty\}$, let X_n be a real-valued random variable on some probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Let F_n be the distribution function of X_n. Then $X_n \Rightarrow X_\infty$ if and only if $F_n(x) \to F_\infty(x)$ for all x such that F_∞ is continuous at x.*

Proof. By Lemma 7.14, we need only prove the "if" part. Assume $F_n(x) \to F_\infty(x)$ for all x such that F_∞ is continuous at x. As mentioned in the proof sketch for Theorem 7.15, this is sufficient for us to infer the conclusion of the Skorohod representation theorem. Thus, there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence $\{Y_n\}$ of random variables on $(\Omega, \mathcal{F}, P)$ such that $X_n =_d Y_n$ for all n, and $Y_n \to Y_\infty$ a.s.

Let $f : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Then $f(Y_n) \to f(Y_\infty)$ a.s. and, by dominated convergence, $E[f(Y_n)] \to E[f(Y_\infty)]$. But $X_n =_d Y_n$ for all n. Thus, $E_n[f(X_n)] = E[f(Y_n)]$ for all n. Therefore, $E_n[f(X_n)] \to E_\infty[f(X_\infty)]$ and $X_n \Rightarrow X_\infty$. □

It is an exercise to prove the following version of Fatou's lemma. This exercise provides practice using the technique in the previous proof.

Theorem 7.18. *Let X_n, X be real-valued random variables with $X_n \Rightarrow X$, and $g : \mathbb{R} \to [0, \infty)$ continuous. Then $\liminf_{n\to\infty} Eg(X_n) \geqslant Eg(X)$.*

Proof. Exercise 7.7. □

Theorem 7.19 (Continuous mapping theorem)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be measurable and $D_g \subset \mathbb{R}$ the set of its discontinuities. Suppose $X_n \Rightarrow X_\infty$ and $X_\infty \in D_g^c$ a.s. Then $g(X_n) \Rightarrow g(X_\infty)$. Moreover, if g is bounded, then $Eg(X_n) \to E[g(X_\infty)]$.*

Proof. By the Skorohod representation theorem, choose Y_n such that $X_n =_d Y_n$ for all n and $Y_n \to Y$ a.s. Let $f : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Since $Y_\infty \in D_g^c$ a.s. and f is continuous, it follows that $f(g(Y_n)) \to f(g(Y_\infty))$ a.s. Since f is bounded, dominated convergence implies

$$E[f(g(X_n))] = E[f(g(Y_n))] \to E[f(g(Y_\infty))] = E[f(g(X_\infty))].$$

Since f was arbitrary, this implies $g(X_n) \Rightarrow g(X_\infty)$.

Now suppose g is bounded. Then, as above, $g(Y_n) \to g(Y_\infty)$ a.s. and dominated convergence give $E[f(g(X_n))] \to E[f(g(X_\infty))]$. □

Remark 7.20. By Remark 7.16, we see that the proof of the continuous mapping theorem is still valid when we only assume the X_n's take values in a separable metric space.

For a general metric space, (M, ρ), a function $f : M \to \mathbb{R}$ is **Lipschitz continuous** if there exists $C > 0$ such that $|f(x) - f(y)| \leqslant C\rho(x, y)$ for all $x, y \in M$. The following theorem is valid in a general metric space.

Theorem 7.21 (Portmanteau theorem)**.** *The following are equivalent:*

(i) $X_n \Rightarrow X_\infty$,

(ii) $E[f(X_n)] \to E[f(X_\infty)]$ *for all bounded, Lipschitz continuous* $f : M \to \mathbb{R}$.

(iii) $\liminf_{n\to\infty} P(X_n \in G) \geqslant P(X_\infty \in G)$ *for all open* G,

(iv) $\limsup_{n\to\infty} P(X_n \in K) \leqslant P(X_\infty \in K)$ *for all closed* K, *and*

(v) $P(X_n \in A) \to P(X_\infty \in A)$ *whenever* $P(X_\infty \in \partial A) = 0$.

Proof. Uses Skorohod representation theorem. See [2, Theorems 3.2.5 and 3.9.1] for the full proof. □

Remark 7.22. To remember the order of the inequalities, keep in mind the following example. Let $X_n = 1/n$ and $X_\infty = 0$. With $G = (0, \infty)$, we have $P(X_n \in G) = 1$ for all $n < \infty$ and $P(X_\infty \in G) = 0$.

The following lemma is sometimes useful.

Lemma 7.23. *Let $X_n \Rightarrow X$ and $x_n \to x$. Let F_n and F be the distribution functions of X_n and X, respectively. If F is continuous at x, then $F_n(x_n) \to F(x)$ as $n \to \infty$.*

Proof. Fix $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ for all $n \geqslant N$. Thus, for any $n \geqslant N$, we have

$$F_n(x_n) = P(X_n \leqslant x_n) \leqslant P(X_n \leqslant x + \varepsilon).$$

By Theorem 7.21(iii),

$$\limsup_{n\to\infty} F_n(x_n) \leqslant P(X \leqslant x + \varepsilon) = F(x + \varepsilon).$$

Similarly, for any such n, we have

$$F_n(x_n) \geqslant P(X_n < x - \varepsilon).$$

By Theorem 7.21(ii),

$$\liminf_{n\to\infty} F_n(x_n) \geqslant P(X < x - \varepsilon) \geqslant F(x - 2\varepsilon).$$

Since F is continuous at x, letting $\varepsilon \to 0$ finishes the proof. □

Theorem 7.24. *Let $M_1(\mathbb{R})$ be the set of probability measures on $(\mathbb{R}, \mathcal{R})$. For $\mu, \nu \in M_1(\mathbb{R})$, let*

$$\rho(\mu, \nu) = \inf\{\varepsilon : F_\mu(x - \varepsilon) - \varepsilon \leqslant F_\nu(x) \leqslant F_\mu(x + \varepsilon) + \varepsilon \textit{ for all } x\},$$

where $F_\mu(x) = \mu((-\infty, x])$. Then ρ is a metric on $M_1(\mathbb{R})$ and $\mu_n \Rightarrow \mu_\infty$ if and only if $\rho(\mu_n, \mu_\infty) \to 0$.

Proof. Exercise 7.8. □

Remark 7.25. The metric ρ is called the **Lévy metric**.

Remark 7.26. Weak convergence of probability measures on $(M, \mathcal{B}_M)$, where M is an arbitrary metric space, is also metrizable. See [3, Section 3.1] for details.

Let $\{\mu_n\}_{n=1}^\infty$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{R})$. We say that $\{\mu_n\}_{n=1}^\infty$ is **tight** if, for all $\varepsilon > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} \mu_n((-M, M]^c) \leqslant \varepsilon.$$

A sequence of random variables, $\{X_n\}_{n=1}^\infty$, is **tight** if $\{\mu_n\}_{n=1}^\infty$ is tight, where $X_n \sim \mu_n$. That is, if, for all $\varepsilon > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} P(|X_n| > M) \leqslant \varepsilon.$$

In a tight sequence, mass cannot "escape" to $\pm\infty$.

The following theorem is a combination of [2, Theorems 3.2.6 and 3.2.7]. See the book for the proofs.

Theorem 7.27. *A sequence of random variables, $\{X_n\}_{n=1}^\infty$, is tight if and only if it is relatively compact, that is, every subsequence has a further subsequence that converges in distribution.*

Remark 7.28. Suppose we wish to prove that $X_n \Rightarrow X$. Since convergence in distribution is metrizable, we could take an arbitrary subsequence, $\{X_{n(m)}\}$, and try to prove that there exists a further subsequence $\{X_{n(m_k)}\}$ such that $X_{n(m_k)} \Rightarrow X$ as $k \to \infty$. Typically, one first proves that $\{X_n\}$ is tight. Then, when trying to prove that $X_{n(m_k)} \Rightarrow X$, we may assume that $X_{n(m_k)} \Rightarrow Y$ for some Y, and reduce our task to showing that $Y = X$.

The proof method described above is a two-step procedure: first prove tightness, and then prove that every subsequential limit has the same distribution. Soon, we will learn to use characteristic functions to prove convergence in distribution. This amounts to a kind of "shortcut" that subsumes both steps of this procedure. As such, you will not have much opportunity to use it. But later, when studying stochastic processes, this two-step proof method will be very important.

Theorem 7.29. *Suppose* $\sup_n E[\varphi(X_n)] < \infty$*, where* $\varphi \geqslant 0$ *and* $\varphi \to \infty$ *as* $|x| \to \infty$*. Then* $\{X_n\}$ *is tight.*

Proof. Let $\varepsilon > 0$. Since $\varphi \to \infty$ as $|x| \to \infty$, we may choose $M > 0$ such that

$$\inf_{|x|\geqslant M} \varphi(x) \geqslant \frac{1}{\varepsilon} \sup_n E[\varphi(X_n)].$$

Thus,

$$P(|X_n| > M) = E[1_{\{|X_n|>M\}}] \leqslant \frac{E[\varphi(X_n)1_{\{|X_n|>M\}}]}{\inf_{|x|\geqslant M} \varphi(x)} \leqslant \frac{E[\varphi(X_n)]}{\inf_{|x|\geqslant M} \varphi(x)} \leqslant \varepsilon,$$

which shows $\{X_n\}$ is tight. □

Exercises

7.7. [2, Exercise 3.2.4] Prove Theorem 7.18.

7.8. [2, Exercise 3.2.6] Prove Theorem 7.24.

7.9. [2, Exercise 3.2.12] Let X_n, X be real-valued random variables defined on a common probability space, $(\Omega, \mathcal{F}, P)$.

(a) Prove that if $X_n \to X$ in probability, then $X_n \Rightarrow X$.

(b) Prove that if $X_n \Rightarrow c$, where c is a constant, then $X_n \to c$ in probability.

7.10. [2, Exercise 3.2.13] Let X_n, Y_n, X be real-valued random variables defined on a common probability space, $(\Omega, \mathcal{F}, P)$. Prove that if $X_n \Rightarrow X$ and $Y_n \Rightarrow c$, where c is a constant, then $X_n + Y_n \Rightarrow X + c$.

7.4 Characteristic functions

This section corresponds to [2, Sections 3.3.1-3.3.3].

The **characteristic function (ch.f.)** of a random variable X is the function $\varphi_X : \mathbb{R} \to \mathbb{C}$ given by

$$\varphi_X(t) = E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)].$$

(Note that this is well-defined for all $t \in \mathbb{R}$.)

Theorem 7.30. *If φ is the ch.f. of X, then*

(a) $\varphi(0) = 1$,

(b) $\varphi(-t) = \overline{\varphi(t)}$,

(c) $|\varphi(t)| = |E[e^{itX}]| \leqslant E|e^{itX}| = 1$,

(d) $|\varphi(t+h) - \varphi(t)| \leqslant E|e^{ihX} - 1|$, *so that φ is uniformly continuous on $\mathbb{R}$, and*

(e) $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.

Proof. Straightforward. See book for details. □

Theorem 7.31. *If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.*

Proof. For all $t \in \mathbb{R}$, we have

$$E[e^{it(X+Y)}] = E[e^{itX}e^{itY}] = E[e^{itX}]E[e^{itY}],$$

since e^{itX} and e^{itY} are independent. □

Example 7.32. If $P(X = 1) = P(X = -1) = 1/2$, then

$$\varphi_X(t) = E[e^{itX}] = \frac{e^{it} + e^{-it}}{2} = \cos t.$$

Example 7.33. If $X \sim \text{Poisson}(\lambda)$, then

$$\begin{aligned}\varphi_X(t) = E[e^{itX}] &= \sum_{k=0}^{\infty} e^{itk} P(X = k) \\ &= \sum_{k=0}^{\infty} e^{itk} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!} = \exp(\lambda(e^{it} - 1)).\end{aligned}$$

Theorem 7.34. *If $X \sim N(\mu, \sigma^2)$, then $\varphi_X(t) = \exp(i\mu t - \sigma^2 t^2/2)$. In particular, the ch.f. of a standard normal is $e^{-t^2/2}$.*

Proof. First assume X is a standard normal and let $\varphi = \varphi_X$. Since $x \mapsto \sin(x)$ is an odd function,

$$\int \sin(tx)e^{-x^2/2}\,dx = 0.$$

Thus,

$$\varphi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos(tx)e^{-x^2/2}\,dx.$$

By Theorem 2.31,

$$\begin{aligned}\varphi'(t) &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x\sin(tx)e^{-x^2/2}\,dx\\ &= -\frac{1}{\sqrt{2\pi}}\left(-\sin(tx)e^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} t\cos(tx)e^{-x^2/2}\,dx\right)\\ &= -t\varphi(t).\end{aligned}$$

Thus,

$$\frac{d}{dt}(\varphi(t)e^{t^2/2}) = (\varphi'(t) + t\varphi(t))e^{t^2/2} = 0,$$

which implies $\varphi(t)e^{t^2/2} = \varphi(0) = 1$.

Now assume $X \sim N(\mu, \sigma^2)$. Then $X = \mu + \sigma Z$, where $Z \sim N(0,1)$, so $\varphi_X(t) = e^{i\mu t}\varphi_Z(\sigma t) = \exp(i\mu t - \sigma^2 t^2/2)$. □

Theorem 7.35. *If $X \sim \mu$, then*

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

for all $a < b$.

Proof. See [2, Theorem 3.3.4]. □

Proposition 7.36. *If $X \sim \mu$, then*

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-ita}\varphi_X(t)\,dt = \mu(\{a\}),$$

for all $a \in \mathbb{R}$.

Proof. See [2, Exercise 3.3.2]. □

Corollary 7.37. *If $\varphi_X = \varphi_Y$, then $X \overset{d}{=} Y$.*

Proof. Let $X \sim \mu$ and $Y \sim \nu$. By the two preceding results, $\mu((a,b]) = \nu((a,b])$ for all $a < b$, which implies $\mu = \nu$. □

Theorem 7.38. *If $\varphi_X \in L^1(\mathbb{R})$, then X has a density function f which is bounded and continuous, and satisfies*

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_X(t)\,dt.$$

Proof. See [2, Theorem 3.3.5]. □

Example 7.39. Let $X = (X_1, \dots, X_n)$ be an $\mathbb{R}^n$-valued random variable and N a $\{1, \dots, n\}$ valued random variable. Assume X and N are independent. Let $X_j \sim \mu_j$ and $p_j = P(N = j)$. We claim that

$$X_N \sim p_1\mu_1 + \cdots + p_n\mu_n.$$

To see this, we calculate

$$P(X_N \in A) = E[P(X_N \in A \mid N)] = E[h(N)],$$

where $h(j) = P(X_j \in A) = \mu_j(A)$. Thus,

$$E[h(N)] = p_1 h(1) + \cdots + p_n h(n) = p_1\mu_1(A) + \cdots + p_n\mu_n(A),$$

which proves the claim.

Similarly,

$$\varphi_{X_N}(t) = E[e^{itX_N}] = E[E[e^{itX_N} \mid N]] = \sum_{j=1}^{n} p_j E[e^{itX_j}] = \sum_{j=1}^{n} p_j \varphi_{X_j}(t).$$

Example 7.40. Let $X \sim \mathrm{Exp}(1)$. Then

$$\varphi_X(t) = E[e^{itX}] = \int_0^\infty e^{itx}e^{-x}\,dx = \left.\frac{e^{itx}e^{-x}}{it-1}\right|_0^\infty = \frac{1}{1-it}.$$

Example 7.41. Let X_1, X_2, N be independent with $X_1 \sim \mathrm{Exp}(1)$, $-X_2 \sim \mathrm{Exp}(1)$, and $P(N = 1) = P(N = 2) = 1/2$. Let $X = X_N$. Then X has density $\frac{1}{2}e^{-|x|}$ (check). By Example 7.39, we have

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{2}\varphi_{X_1}(t) + \frac{1}{2}\varphi_{X_2}(t)\\
&= \frac{1}{2}\varphi_{X_1}(t) + \frac{1}{2}\varphi_{X_1}(-t)\\
&= \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{1+t^2}.
\end{aligned}$$

Example 7.42. Let X have the **Cauchy distribution**, that is, suppose X has density

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

This density is "bell"-shaped and symmetric about the origin, but is not integrable, so X does not have a mean. The Cauchy distribution is a source of many counterexamples to things that might otherwise seem intuitively true. Here, we present it simply to illustrate the use of the inversion theorem.

Note that

$$\varphi_X(t) = E[e^{itX}] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}\,dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{-ity}}{1+y^2}\,dy.$$

Using the previous example, we then have

$$\varphi_X(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-ity}\varphi_Y(y)\,dy,$$

where Y has density $g(y) = \frac{1}{2}e^{-|y|}$. But by Theorem 7.38,

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\varphi_Y(y)\,dy,$$

from which we obtain $\varphi_X(t) = e^{-|t|}$.

Lemma 7.43. *Let X be a random variable. Then*

$$P\left(|X| > \frac{2}{u}\right) \leqslant \frac{1}{u}\int_{-u}^{u}(1-\varphi_X(t))\,dt,$$

for all $u > 0$.

Proof. Fix $u > 0$. Note that

$$E\left[\int_{\mathbb{R}} |1_{(-u,u)}(t)(1-e^{itX})|\,dt\right] \leqslant E\left[\int_{-u}^{u} 2\,dt\right] = 4u < \infty.$$

Hence, by Fubini's theorem,

$$\begin{aligned} E\left[\int_{\mathbb{R}} 1_{(-u,u)}(t)(1-e^{itX})\,dt\right] &= \int_{\mathbb{R}} E[1_{(-u,u)}(t)(1-e^{itX})]\,dt \\ &= \int_{-u}^{u}(1-E[e^{itX}])\,dt = \int_{-u}^{u}(1-\varphi_X(t))\,dt. \end{aligned}$$

On the other hand,

$$\begin{aligned} E\left[\int_{\mathbb{R}} 1_{(-u,u)}(t)(1-e^{itX})\,dt\right] &= E\left[\int_{-u}^{u}(1-e^{itX})\,dt\right] \\ &= E\left[2u - \frac{e^{iuX}-e^{-iuX}}{iX}\right] = E\left[2u - \frac{2\sin(uX)}{X}\right]. \end{aligned}$$

Thus,

$$\frac{1}{u}\int_{-u}^{u}(1-\varphi_X(t))\,dt = 2E\left[1 - \frac{\sin(uX)}{uX}\right].$$

Since $(\sin y)/y \leqslant 1$ for all y, we have

$$\begin{aligned} \frac{1}{u}\int_{-u}^{u}(1-\varphi_X(t))\,dt &\geqslant 2E\left[\left(1-\frac{\sin(uX)}{uX}\right)1_{\{|uX|\geqslant 2\}}\right] \\ &\geqslant 2E\left[\left(1-\frac{1}{|uX|}\right)1_{\{|uX|\geqslant 2\}}\right] \geqslant 2E\left[\frac{1}{2}1_{\{|uX|\geqslant 2\}}\right] = P(|uX| \geqslant 2), \end{aligned}$$

which is equivalent to what we wanted to prove. □

Theorem 7.44 (Continuity theorem)**.** *Let $\{X_n\}$ be a sequence of random variables. Let φ_n be the ch.f. of X_n.*

(i) If $X_n \Rightarrow X_\infty$, then $\varphi_n \to \varphi_\infty$ pointwise.

(ii) Suppose there exists $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi_n \to \varphi$ pointwise. If φ is continuous at 0, then there exists a random variable X such that $X_n \Rightarrow X$ and $\varphi_X = \varphi$.

Remark 7.45. The condition of continuity at 0 cannot be omitted. If $X_n \sim N(0, n)$, then $\varphi_n \to 1_{\{0\}}$ pointwise, but $\{X_n\}$ is not even tight (check).

Proof of Theorem 7.44. Suppose $X_n \Rightarrow X_\infty$. Fix $t \in \mathbb{R}$. Let $f(x) = e^{itx}$. Then f is bounded and continuous, so

$$\varphi_n(t) = E[f(X_n)] \to E[f(X_\infty)] = \varphi_\infty(t),$$

and this proves (i).

Now suppose $\varphi_n \to \varphi$ pointwise and φ is continuous at 0. Since $\varphi_n(0) = 1$ for all n, we have $\varphi(0) = 1$. Since φ is continuous at 0, it follows that

$$\frac{1}{u} \int_{-u}^{u} (1 - \varphi(t))\, dt \to 0$$

as $u \to 0$. Let $\varepsilon > 0$. Choose u so that $u^{-1} \int_{-u}^{u} (1 - \varphi(t))\, dt < \varepsilon$, and let $M = 2/u$. By Lemma 7.43,

$$\limsup_{n\to\infty} P(|X_n| > M) \leqslant \limsup_{n\to\infty} \frac{1}{u} \int_{-u}^{u} (1 - \varphi_n(t))\, dt.$$

Since $1-\varphi_n \to 1-\varphi$ pointwise and $|1-\varphi_n| \leqslant 2$ for all n, it follows by dominated convergence that

$$\limsup_{n\to\infty} P(|X_n| > M) \leqslant \frac{1}{u} \int_{-u}^{u} (1 - \varphi(t))\, dt < \varepsilon.$$

Since ε was arbitrary, $\{X_n\}$ is tight.

Since $\{X_n\}$ is tight, there exists a subsequence, $\{X_{\tilde{n}(m)}\}$, and a random variable X such that $X_{\tilde{n}(m)} \Rightarrow X$ as $m \to \infty$. By (i), this implies $\varphi_{\tilde{n}(m)} \to \varphi_X$ pointwise. But by hypothesis, $\varphi_{\tilde{n}(m)} \to \varphi$ pointwise. Thus, $\varphi_X = \varphi$. It remains only to show that $X_n \Rightarrow X$ as $n \to \infty$.

Let $\{X_{n(m)}\}$ be an arbitrary subsequence. Since $\{X_n\}$ is tight, there exists a further subsequence, $\{X_{n(m_k)}\}$, and a random variable Y such that $X_{n(m_k)} \Rightarrow Y$ as $k \to \infty$. By (i), this implies $\varphi_{n(m_k)} \to \varphi_Y$ pointwise. But by hypothesis, $\varphi_{n(m_k)} \to \varphi$ pointwise. Thus, $\varphi_Y = \varphi$. Combined with the above, this gives $\varphi_Y = \varphi_X$, which implies $Y =_d X$. Hence, $X_{n(m_k)} \Rightarrow X$. Since the subsequence $\{X_{n(m)}\}$ was arbitrary, this shows $X_n \Rightarrow X$. □

Remark 7.46. When dealing with real-valued random variables, this theorem will be our main tool for proving convergence in distribution. As you can see from the proof, when we use this theorem, we are implicitly using the proof method described in the comments following Theorem 7.27. We will not explicitly use that proof method again until we must deal with the convergence of function-valued random variables (that is, stochastic processes).

Theorem 7.47. *If $E|X|^n < \infty$, then φ_X has a continuous derivative of order n given by $\varphi_X^{(n)}(t) = E[(iX)^n e^{itX}]$.*

Proof. Exercise 7.16. □

Lemma 7.48. *Let n be a nonnegative integer. Then*

$$\left| e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} \right| \leqslant \frac{|x|^{n+1}}{(n+1)!} \wedge \frac{2|x|^n}{n!}$$

for all $x \in \mathbb{R}$.

Proof. See [2, Lemma 3.3.7]. □

Theorem 7.49. *If $E|X|^2 < \infty$, then*

$$\varphi_X(t) = 1 + itE[X] - \frac{t^2}{2} E[X^2] + o(t^2).$$

Proof. Let

$$r(t) = \varphi_X(t) - 1 - itE[X] + \frac{t^2}{2} E[X^2] = E\left[e^{itX} - \sum_{m=0}^{2} \frac{(itX)^m}{m!} \right].$$

By the lemma,

$$t^{-2}|r(t)| \leqslant t^{-2} E\left[\frac{|tX|^3}{6} \wedge |tX|^2 \right] \leqslant E[|tX| \wedge |X|^2].$$

We have $|tX| \wedge |X|^2 \to 0$ a.s. as $t \to 0$. Also, $|tX| \wedge |X|^2 \leqslant |X|^2$ for all t, and $|X|^2$ is integrable. Thus, by dominated convergence, $t^{-2}|r(t)| \to 0$ as $t \to 0$. □

Exercises

7.11. [2, Exercise 3.3.1] Let X be a real-valued random variable. Prove that there exist real-valued random variables, Y and Z, such that $\varphi_Y = \mathrm{Re}\varphi_X$ and $\varphi_Z = |\varphi_X|^2$.

7.12. [2, Exercise 3.3.3] Let X be a random variable such that φ_X is real-valued. Prove that $-X =_d X$.

7.13. [2, Exercise 3.3.8] Let $X_1, \ldots, X_n$ be independent real-valued random variables. Assume that each X_j has a Cauchy distribution with density

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Prove that

$$\frac{X_1 + \cdots + X_n}{n} \stackrel{d}{=} X_1.$$

7.14. [2, Exercise 3.3.9] Let X_n, X be real-valued random variables. Assume $X_n \sim N(0, \sigma_n^2)$ and $X_n \Rightarrow X$. Prove that there exists $\sigma \in [0, \infty)$ such that $\sigma_n^2 \to \sigma$.

7.15. [2, Exercise 3.3.10] For each $n \in \mathbb{N} \cup \{\infty\}$, let (X_n, Y_n) be an $\mathbb{R}^2$-valued random variable defined on a probability space, $(\Omega_n, \mathcal{F}_n, P_n)$. Suppose that X_n and Y_n are independent, for each n. Prove that if $X_n \Rightarrow X_\infty$ and $Y_n \Rightarrow Y_\infty$, then $X_n + Y_n \Rightarrow X_\infty + Y_\infty$.

7.16. [2, Exercise 3.3.14] Prove Theorem 7.47.

Part III

Discrete-time Stochastic Processes I: Random Walks

Chapter 8

Introduction

As discussed in Section 6.4, a discrete-time stochastic process is a sequence of random variables, $\{X_n : n \in \mathbb{N}\}$, all defined on the same common probability space, $(\Omega, \mathcal{F}, P)$, and all taking values in the same common measurable space, $(S, \mathcal{S})$.

Recall that a random walk is an $\mathbb{R}^d$-valued, discrete-time stochastic process such that $X_n = \xi_1 + \cdots + \xi_n$, where $\{\xi_j\}$ is an i.i.d. sequence. In Chapters 9 and 10, we will only be concerned with the asymptotic behavior of the random walk as $n \to \infty$, primarily when $d = 1$.

We will first show that, under suitable conditions, $X_n \approx n\mu$, where μ is the common mean of the ξ_j's. This result is known as the law of large numbers (LLN), and we will look at multiple incarnations of this theorem. The LLN gives us a crude, first-order approximation of X_n when n is large, and it so happens that this approximation is deterministic. (There is nothing random about the quantity $n\mu$.)

For a finer approximation, we turn to the central limit theorem (CLT), which shows that $X_n \approx n\mu + n^{1/2}\sigma Z$, where σ^2 is the common variance of the ξ_j's and Z is a standard normal. Here, the approximation is not deterministic, so we must be careful about the meaning of "$\approx$". In fact, it is only the distributions that are close when n is large. As with the LLN, we will look at multiple incarnations of the CLT. Some of the versions of the LLN and CLT that we consider will be quite general and even apply to certain discrete-time stochastic processes that are not random walks.

In the final chapter of this part of the notes, we will address the issue of recurrence. That is, does the random walk return infinitely often to (or close to) the origin?

Chapter 9

Laws of Large Numbers

9.1 Weak laws of large numbers

This section corresponds to [2, Section 2.2].

Theorem 9.1 (L^2 weak law)**.** *Let $\{\xi_j\}_{j=1}^\infty$ be uncorrelated, with $\mu = E\xi_j$ and $\sup_j \operatorname{var}(\xi_j) < \infty$. Let $X_n = \xi_1 + \cdots + \xi_n$. Then $X_n/n \to \mu$ in probability and in L^2.*

Proof. Since $EX_n = n\mu$, we have

$$E\left|\frac{X_n}{n} - \mu\right|^2 = \operatorname{var}\left(\frac{X_n}{n}\right) = \frac{1}{n^2}\operatorname{var}(X_n) = \frac{1}{n^2}\operatorname{var}\left(\sum_{j=1}^n \xi_j\right) = \frac{1}{n^2}\sum_{j=1}^n \operatorname{var}(\xi_j),$$

where the last equality follows from the fact that $\{\xi_j\}$ are uncorrelated. Let $C = \sup_j \operatorname{var}(\xi_j) < \infty$. Then $E|X_n/n - \mu|^2 \leqslant C/n \to 0$ as $n \to \infty$. Hence $X_n/n \to \mu$ in L^2. Finally, L^2 convergence implies convergence in probability. □

Theorem 9.2. *Let $\{X_n\}_{n=1}^\infty \subset L^2(\Omega)$. Let $\mu_n = EX_n$, $\sigma_n^2 = \operatorname{var}(X_n)$, and let $\{b_n\}_{n=1}^\infty \subset \mathbb{R}$. If $\sigma_n^2/b_n^2 \to 0$ as $n \to \infty$, then*

$$\frac{X_n - \mu_n}{b_n} \to 0$$

in probability and in L^2 as $n \to \infty$.

Proof. The result follows immediately since $E|b_n^{-1}(X_n - \mu_n)|^2 = \sigma_n^2/b_n^2$. □

Example 9.3 (Coupon collector's problem)**.** Suppose we have a sequence of independent trials, and on the m-th trial we collect a random object from among n different possible objects. Let $\{\xi_m : m \in \mathbb{N}\}$ be i.i.d. and uniform on $\{1, \ldots, n\}$, so that ξ_m represents our collected object on the m-th trial. Let

$$\tau_k = \tau_k^n = \inf\{m : |\{\xi_1, \ldots, \xi_m\}| = k\}.$$

Then τ_k represents the first time at which we have collected k distinct objects. Note that $\tau_1 = 1$ and let us adopt the convention that $\tau_0 = 0$. Let $X_n = \tau_n^n$, which represents the time it takes us to obtain a complete collection of all n objects. We wish to understand the asymptotic behavior of X_n as $n \to \infty$. In fact, we will show that

$$\frac{X_n}{n \log n} \to 1 \tag{9.1}$$

in probability as $n \to \infty$, so that $X_n \approx n \log n$ for large n.

For $1 \leqslant k \leqslant n$, let $\xi_{n,k} = \tau_k^n - \tau_{k-1}^n$. In words, after we have collected $k - 1$ objects, $\xi_{n,k}$ represents the number of additional trials we need in order to collect a new object, distinct from the ones we already have. We should therefore have $\xi_{n,k} \sim \text{Geom}(p_{n,k})$, where

$$p_{n,k} = 1 - \frac{k-1}{n},$$

and $\xi_{n,1}, \dots, \xi_{n,n}$ are independent. Note that $X_n = \xi_{n,1} + \cdots + \xi_{n,n}$.

Let $\mu_n = EX_n$, $\sigma_n^2 = \text{var}(X_n)$, and let $b_n = n \log n$. Then

$$\frac{\sigma_n^2}{b_n^2} = \frac{1}{(n \log n)^2} \sum_{k=1}^n \text{var}(\xi_{n,k}) = \frac{1}{(n \log n)^2} \sum_{k=1}^n \frac{1 - p_{n,k}}{p_{n,k}^2} \leqslant \frac{1}{(n \log n)^2} \sum_{k=1}^n \frac{1}{p_{n,k}^2}.$$

Thus,

$$\begin{aligned} \frac{\sigma_n^2}{b_n^2} \leqslant \frac{1}{(n \log n)^2} \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right)^{-2} &= \frac{1}{(n \log n)^2} \sum_{k=1}^n \left(\frac{n}{n-k+1}\right)^2 \\ &= \frac{1}{(\log n)^2} \sum_{m=1}^n \frac{1}{m^2} \leqslant \frac{1}{(\log n)^2} \sum_{m=1}^\infty \frac{1}{m^2} \to 0, \end{aligned}$$

as $n \to \infty$. By Theorem 9.2,

$$\frac{X_n - \mu_n}{n \log n} \to 0$$

in probability as $n \to \infty$.

Note that, as above,

$$\mu_n = \sum_{k=1}^n E\xi_{n,k} = \sum_{k=1}^n \frac{1}{p_{n,k}} = n \sum_{m=1}^n \frac{1}{m}.$$

Since

$$\log(n+1) = \int_1^{n+1} \frac{1}{x}\,dx \leqslant \sum_{m=1}^n \frac{1}{m} \leqslant 1 + \int_1^n \frac{1}{x}\,dx = 1 + \log n,$$

it follows that $\mu_n/(n \log n) \to 1$. Thus, using Exercise 2.14, we have

$$\frac{X_n}{n \log n} = \frac{X_n - \mu_n}{n \log n} + \frac{\mu_n}{n \log n} \to 1$$

in probability as $n \to \infty$, proving (9.1).

Theorem 9.4 (Weak law for triangular arrays)**.** *For each $n \in \mathbb{N}$, let the random variables $\xi_{n,1}, \dots, \xi_{n,n}$ be independent. Let $\{b_n\}_{n=1}^\infty \subset (0,\infty)$ with $b_n \to \infty$, and let*

$$a_n = \sum_{k=1}^n E[\xi_{n,k} 1_{\{|\xi_{n,k}| \leqslant b_n\}}].$$

Let $X_n = \xi_{n,1} + \cdots + \xi_{n,n}$. Suppose that

(i) $\sum_{k=1}^n P(|\xi_{n,k}| > b_n) \to 0$ as $n \to \infty$, and

(ii) $b_n^{-2} \sum_{k=1}^n E[\xi_{n,k}^2 1_{\{|\xi_{n,k}| \leqslant b_n\}}] \to 0$ as $n \to \infty$.

Then $(X_n - a_n)/b_n \to 0$ in probability as $n \to \infty$.

Proof. Let $\overline{\xi}_{n,k} = \xi_{n,k} 1_{\{|\xi_{n,k}| \leqslant b_n\}}$ and $\overline{X}_n = \overline{\xi}_{n,1} + \cdots + \overline{\xi}_{n,n}$, so that $a_n = E\overline{X}_n$. Fix $\varepsilon > 0$. We begin by observing that

$$\begin{aligned} P\left(\left|\frac{X_n - a_n}{b_n}\right| > \varepsilon\right) &= P\left(\left\{\left|\frac{X_n - a_n}{b_n}\right| > \varepsilon\right\} \cap \{X_n \neq \overline{X}_n\}\right) \\ &\quad + P\left(\left\{\left|\frac{X_n - a_n}{b_n}\right| > \varepsilon\right\} \cap \{X_n = \overline{X}_n\}\right) \\ &\leqslant P(X_n \neq \overline{X}_n) + P\left(\left\{\left|\frac{\overline{X}_n - a_n}{b_n}\right| > \varepsilon\right\} \cap \{X_n = \overline{X}_n\}\right) \\ &\leqslant P(X_n \neq \overline{X}_n) + P\left(\left|\frac{\overline{X}_n - a_n}{b_n}\right| > \varepsilon\right). \end{aligned}$$

We then have

$$P(X_n \neq \overline{X}_n) \leqslant P\bigg(\bigcup_{k=1}^n \{\overline{\xi}_{n,k} \neq \xi_{n,k}\}\bigg) \leqslant \sum_{k=1}^n P(|\xi_{n,k}| > b_n) \to 0$$

by (i). For the second term, Chebyshev gives

$$\begin{aligned} P\left(\left|\frac{\overline{X}_n - a_n}{b_n}\right| > \varepsilon\right) &\leqslant \varepsilon^{-2} E\left|\frac{\overline{X}_n - a_n}{b_n}\right|^2 = \varepsilon^{-2} b_n^{-2} \operatorname{var}(\overline{X}_n) \\ &= \varepsilon^{-2} b_n^{-2} \sum_{k=1}^n \operatorname{var}(\overline{\xi}_{n,k}) \leqslant \varepsilon^{-2} b_n^{-2} \sum_{k=1}^n E|\overline{\xi}_{n,k}|^2 \to 0 \end{aligned}$$

by (ii). □

Theorem 9.5. *Let $\{\xi_k\}_{k=1}^\infty$ be i.i.d. and let $X_n = \xi_1 + \cdots + \xi_n$. There exist constant μ_n such that $X_n/n - \mu_n \to 0$ in probability if and only if*

$$xP(|\xi_1| > x) \to 0 \tag{9.2}$$

as $x \to \infty$. Moreover, we can take these constants to be $\mu_n = E[\xi_1 1_{\{|\xi_1| \leqslant n\}}]$.

Proof. For the "only if" part see [4, pp. 234-236]. Assume (9.2). We wish to apply Theorem 9.4. Let $\xi_{n,k} = \xi_k$ for all $n \geqslant k$, and let $b_n = n$. Since $\{\xi_k\}$ are i.i.d., we have

$$\sum_{k=1}^{n} P(|\xi_{n,k}| > b_n) = \sum_{k=1}^{n} P(|\xi_k| > n) = nP(|\xi_1| > n) \to 0$$

by (9.2), and so Theorem 9.4(i) holds.

For (ii), we consider

$$b_n^{-2} \sum_{k=1}^{n} E[\xi_{n,k}^2 1_{\{|\xi_{n,k}| \leqslant b_n\}}] = n^{-2} \sum_{k=1}^{n} E[\xi_k^2 1_{\{|\xi_k| \leqslant n\}}] = n^{-1} E[\xi_1^2 1_{\{|\xi_1| \leqslant n\}}].$$

By (6.9),

$$E[\xi_1^2 1_{\{|\xi_1| \leqslant n\}}] = \int_0^\infty 2tP(|\xi_1| 1_{\{|\xi_1| \leqslant n\}} > t)\, dt \leqslant \int_0^n 2tP(|\xi_1| > t)\, dt$$

Let $\varepsilon > 0$. By (9.2), we may choose $t_0 > 0$ such that $2tP(|\xi_1| > t) < \varepsilon$ for all $t > t_0$. Thus,

$$\begin{aligned}
&\limsup_{n\to\infty} n^{-1} E[\xi_1^2 1_{\{|\xi_1| \leqslant n\}}] \\
&\qquad \leqslant \limsup_{n\to\infty} \frac{1}{n} \left(\int_0^{t_0} 2tP(|\xi_1| > t)\, dt + \int_{t_0}^n 2tP(|\xi_1| > t)\, dt \right) \\
&\qquad\qquad \leqslant \limsup_{n\to\infty} \frac{1}{n} \left(\int_0^{t_0} 2tP(|\xi_1| > t)\, dt + n\varepsilon \right) = \varepsilon.
\end{aligned}$$

Letting $\varepsilon \to 0$ verifies (ii).

Thus, by Theorem 9.4, we have $(X_n - a_n)/n \to 0$ in probability, where

$$a_n = \sum_{k=1}^{n} E[\xi_{n,k} 1_{\{|\xi_{n,k}| \leqslant b_n\}}] = \sum_{k=1}^{n} E[\xi_k 1_{\{|\xi_k| \leqslant n\}}] = n\mu_n,$$

and it follows that $X_n/n - \mu_n \to 0$ in probability. □

Suppose (9.2) holds. Let $0 < \varepsilon < 1$. By (6.9), we have

$$E|\xi_1|^{1-\varepsilon} = \int_0^\infty (1-\varepsilon) t^{-\varepsilon} P(|\xi_1| > t)\, dt \leqslant \int_0^1 t^{-\varepsilon}\, dt + C \int_1^\infty t^{-(1+\varepsilon)}\, dt < \infty.$$

The following theorem is the familiar form of the weak LLN. It has a somewhat stronger hypothesis than (9.2), in that it assumes $E|\xi_1| < \infty$.

Theorem 9.6. *Let $\{\xi_k\}_{k=1}^\infty$ be i.i.d. with $E|\xi_1| < \infty$. Let $X_n = \xi_1 + \cdots + \xi_n$ and $\mu = E\xi_1$. Then $X_n/n \to \mu$ in probability.*

Proof. We have

$$xP(|\xi_1| > x) \leqslant E[|\xi_1| 1_{\{|\xi_1|>x\}}],$$

which goes to 0 by dominated convergence. Thus, by Theorem 9.5, we have $X_n/n - \mu_n \to 0$ in probability, where $\mu_n = E[|\xi_1| 1_{\{|\xi_n|\leqslant n\}}]$.

But $\mu_n \to \mu$, also by dominated convergence. Hence, $X_n/n = (X_n/n - \mu_n) + \mu_n \to \mu$ in probability. □

Example 9.7. Let $\{\xi_j\}_{j=1}^\infty$ be i.i.d. Cauchy distributed random variables. That is,

$$P(\xi_1 \in dx) = \frac{1}{\pi}\frac{1}{1+x^2}\,dx.$$

Then

$$xP(|\xi_1| > x) = \frac{2}{\pi} x \int_x^\infty \frac{1}{1+t^2}\,dt = \frac{2}{\pi}\frac{\int_x^\infty \frac{1}{1+t^2}\,dt}{x^{-1}},$$

so by L'Hôpital,

$$\lim_{x\to\infty} xP(|\xi_1| > x) = \lim_{x\to\infty} \frac{2}{\pi}\frac{x^2}{1+x^2} = \frac{2}{\pi}.$$

By Theorem 9.5, if $X_n = \xi_1 + \cdots + \xi_n$, then there are no constants μ_n such that $X_n/n - \mu_n \to 0$. In fact, by Exercise 7.13, we have $X_n/n =_d \xi_1$ for all n.

Example 9.8 (St. Petersburg paradox)**.** Consider a game in which you flip a fair coin until the first time you get a head. If it takes you j flips, then you will receive a prize of 2^j dollars. You may play this game repeatedly, as many times as you like. How much would you be willing to pay per trial to play this game?

Let $\{\xi_k\}_{k=1}^\infty$ be i.i.d. with $P(\xi_1 = 2^j) = 2^{-j}$ for $j \in \mathbb{N}$, so that ξ_j represents the prize you receive on the j-th time you play the game. Note that

$$E\xi_1 = \sum_{j=1}^\infty 2^j P(\xi_1 = 2^j) = \infty.$$

There is a general principle among some gamblers that optimizing your expected value is always what you want to do. If you have a positive expectation, then the game is profitable for you, and you should make the most of it. According to this principle, we should be willing to pay *any* price to play this game.

But the expected value is just one of many parameters associated with a random variable. It has no intrinsic "meaning" beyond its mathematical definition. Any meaning it has for gamblers is meaning that it inherits from the familiar form of the LLN. Hence, if the LLN doesn't apply, then the expected value is just another mathematical parameter.

In the example we are considering, the traditional LLN, Theorem 9.6, does not apply. So we must use something else to understand the asymptotic behavior of $X_n = \xi_1 + \cdots + \xi_n$. As we did with the coupon collector's problem, we will use the weak law for triangular arrays, Theorem 9.4, to prove that

$$\frac{X_n}{n\log_2 n} \to 1$$

in probability as $n \to \infty$.

Let $\xi_{n,k} = \xi_k$. Let

$$K(n) = \lfloor \log_2 n + \log_2 \log_2 n \rfloor - \log_2 n.$$

Note that

$$\log_2 \log_2 n - 1 \leqslant K(n) \leqslant \log_2 \log_2 n, \tag{9.3}$$

and, if we define $m(n) = \log_2 n + K(n)$, then $m(n) \in \mathbb{N}$. Define $b_n = 2^{m(n)}$.

We first check that

$$\sum_{k=1}^{n} P(|\xi_{n,k}| > b_n) = \sum_{k=1}^{n} P(|\xi_k| > 2^{m(n)}) = n2^{-m(n)} = 2^{-K(n)} \to 0,$$

which verifies Theorem 9.4(i).

We next check that

$$E[\xi_{n,k}^2 1_{\{|\xi_{n,k}| \leqslant b_n\}}] = E[\xi_k^2 1_{\{|\xi_k| \leqslant 2^{m(n)}\}}] = \sum_{j=1}^{m(n)} 2^{2j} 2^{-j} = 2^{m(n)+1} - 2 \leqslant 2b_n.$$

Thus,

$$b_n^{-2} \sum_{k=1}^{n} E[\xi_{n,k}^2 1_{\{|\xi_{n,k}| \leqslant b_n\}}] \leqslant \frac{2n}{b_n} = \frac{2}{2^{K(n)}} \to 0,$$

and this verifies Theorem 9.4(ii).

It therefore follows that $2^{-m(n)}(X_n - a_n) \to 0$ in probability, where

$$\begin{aligned} a_n = \sum_{k=1}^{n} E[\xi_{n,k} 1_{\{|\xi_{n,k}| \leqslant b_n\}}] &= \sum_{k=1}^{n} E[\xi_k 1_{\{|\xi_k| \leqslant 2^{m(n)}\}}] \\ &= nE[\xi_1 1_{\{|\xi_1| \leqslant 2^{m(n)}\}}] = n \sum_{j=1}^{m(n)} 2^j 2^{-j} = nm(n) = n \log_2 n + nK(n). \end{aligned}$$

Note that

$$\frac{a_n}{n \log_2 n} = 1 + \frac{K(n)}{\log_2 n} \to 1,$$

by (9.3). Also note that

$$\frac{2^{m(n)}}{n \log_2 n} = 2^{K(n) - \log_2 \log_2 n} \in [1/2, 1],$$

for all n. Hence, using Exercises 9.3 and 2.14, we have

$$\frac{X_n}{n \log_2 n} = \frac{2^{m(n)}}{n \log_2 n} \frac{X_n - a_n}{2^{m(n)}} + \frac{a_n}{n \log_2 n} \to 1$$

in probability as $n \to \infty$.

So for large n, we have $X_n \approx n \log_2 n$, meaning that after playing this game n times, our average winnings per game will be about $\log_2 n$. For example, if we plan to play the game 1024 times, our average winnings per game will be about \$10. In this case, we should pay less than \$10 per game if it is to be profitable for us. If the price of each game were \$40, we would need to play the game 2^{40}, or about one trillion, times before we start to break even.

Exercises

9.1. [2, Exercise 2.2.4] Let $\xi_1, \xi_2, \dots$ be i.i.d. random variables with

$$P(\xi_j = (-1)^k k) = \frac{C}{k^2 \log k},$$

for $k \geqslant 2$, where $C = \left(\sum_{k=2}^{\infty} 1/(k^2 \log k)\right)^{-1}$. Let $X_n = \xi_1 + \cdots + \xi_n$.

(a) Prove that $E|\xi_j| = \infty$.

(b) Prove that there exists $\mu \in \mathbb{R}$ such that $X_n/n \to \mu$ in probability.

9.2. [2, Exercise 2.2.8] For $j \in \mathbb{N}$, let

$$p_j = \frac{1}{2^j j(j+1)}.$$

Note that $\sum_{j \in \mathbb{N}} p_j \in (0,1)$. Define $p_0 = 1 - \sum_{j \in \mathbb{N}} p_j$. Let $\xi_1, \xi_2, \dots$ be i.i.d. with $P(\xi_1 = -1) = p_0$ and

$$P(\xi_1 = 2^j - 1) = p_j,$$

for $j \geqslant 1$, and let $X_n = \xi_1 + \cdots + \xi_n$.

(a) Prove that ξ_1 is integrable and that $E\xi_1 = 0$. [In other words, by some people's standards, this is a model of a fair game.]

(b) For $n \in \mathbb{N}$, let $m(n) = \min\{m \in \mathbb{N} : n \leqslant 2^m m^{3/2}\}$. Prove that $m(n) \sim \log_2 n$, i.e. prove that $m(n)/\log_2 n \to 1$ as $n \to \infty$.

(c) Let $b_n = 2^{m(n)}$. Let $a_n = nE[\xi_1 1_{\{|\xi_1| \leqslant b_n\}}]$. Prove that $a_n \sim -n/(\log_2 n)$.

(d) Prove that $b_n/(n/(\log_2 n)) \to 0$ as $n \to \infty$.

(e) Use Theorem 9.4 to prove that

$$\frac{X_n}{n/(\log_2 n)} \to -1$$

in probability as $n \to \infty$. [In other words, if n is large, then after n plays, with high probability, you will have lost $n/(\log_2 n)$ units.]

9.3. Suppose $X_n \to 0$ in probability and $\{c_n\}$ is a bounded sequence of real numbers. Prove that $c_n X_n \to 0$ in probability. [Hint: Use Theorem 7.3.]

9.2 Strong law of large numbers

This section corresponds to [2, Section 2.4].

Lemma 9.9. *If $X_n \to X_\infty$ a.s., and $N(n) \to \infty$ a.s., then $X_{N(n)} \to X_\infty$ a.s.*

Proof. Choose $\Omega_1 \in \mathcal{F}$ such that $P(\Omega_1) = 1$ and $X_n(\omega) \to X_\infty(\omega)$ for all $\omega \in \Omega_1$. Choose $\Omega_2 \in \mathcal{F}$ such that $P(\Omega_2) = 1$ and $N(n, \omega) \to \infty$ for all $\omega \in \Omega_2$. Let $\Omega^* = \Omega_1 \cap \Omega_2$. Then $P(\Omega^*) = 1$. Fix $\omega \in \Omega^*$. Then $X_{N(n,\omega)}(\omega) \to X_\infty(\omega)$, and so it follows that $X_{N(n)} \to X_\infty$ a.s. □

Remark 9.10. If we only assume that $X_n \to X_\infty$ in probability, then the result is false. (See Exercise 9.4.) This demonstrates one (mathematical) reason why convergence in probability may not be sufficient. Whereas the weak LLNs provide convergence in probability, the strong LLNs provide convergence a.s.

Lemma 9.11. *If $y \geqslant 0$, then*

$$2y \sum_{\{k:k>y\}} \frac{1}{k^2} \leqslant 4.$$

Proof. See [2, Lemma 2.4.4]. □

Theorem 9.12 (Strong law of large numbers)**.** *Let $\{\xi_k\}_{k=1}^\infty$ be pairwise independent and identically distributed, with $E|\xi_1| < \infty$. Let $X_n = \xi_1 + \cdots + \xi_n$ and $\mu = E\xi_1$. Then $X_n/n \to \mu$ a.s.*

Proof. First assume $\xi_1 \geqslant 0$ a.s. Let $\zeta_k = \xi_k 1_{\{|\xi_k| \leqslant k\}}$ and $Y_n = \zeta_1 + \cdots + \zeta_n$. Fix $\alpha > 1$ and let $k(n) = \lfloor \alpha^n \rfloor$. We begin by showing that

$$\frac{Y_{k(n)} - EY_{k(n)}}{k(n)} \to 0 \quad \text{a.s.} \tag{9.4}$$

as $n \to \infty$. We will use Theorem 7.9.

Fix $\varepsilon > 0$. Then, by Chebyshev,

$$\begin{aligned} \sum_{n=1}^\infty P\left(\left| \frac{Y_{k(n)} - EY_{k(n)}}{k(n)} \right| > \varepsilon \right) &\leqslant \frac{1}{\varepsilon^2} \sum_{n=1}^\infty \frac{\operatorname{var}(Y_{k(n)})}{k(n)^2} \\ &= \frac{1}{\varepsilon^2} \sum_{n=1}^\infty \sum_{m=1}^{k(n)} \frac{\operatorname{var}(\zeta_m)}{k(n)^2} = \frac{1}{\varepsilon^2} \sum_{m=1}^\infty \sum_{\{n:k(n)\geqslant m\}} \frac{\operatorname{var}(\zeta_m)}{k(n)^2}. \end{aligned} \tag{9.5}$$

Since $\lfloor \alpha^n \rfloor \geqslant \alpha^n/2$, we have

$$\sum_{\{n:k(n)\geqslant m\}} \frac{1}{k(n)^2} = \sum_{\{n:k(n)\geqslant m\}} \lfloor \alpha^n \rfloor^{-2} \leqslant 4 \sum_{\{n:k(n)\geqslant m\}} \alpha^{-2n}.$$

Note that if $k(n) \geqslant m$, then $m \leqslant \lfloor \alpha^n \rfloor \leqslant \alpha^n$, which implies $\alpha^{-2n} \leqslant m^{-2}$. Thus, the above is a geometric series whose first term is bounded by above by m^{-2}, so that

$$\sum_{\{n:k(n)\geqslant m\}} \frac{1}{k(n)^2} \leqslant \frac{4m^{-2}}{1-\alpha^{-2}}$$

Substituting this into (9.5), we have

$$\sum_{n=1}^{\infty} P\left(\left|\frac{Y_{k(n)} - EY_{k(n)}}{k(n)}\right| > \varepsilon\right) \leqslant \frac{4}{(1-\alpha^{-2})\varepsilon^2} \sum_{m=1}^{\infty} \frac{\operatorname{var}(\zeta_m)}{m^2}. \tag{9.6}$$

We now turn our attention to the sum on the right.

By (6.9), we have

$$\begin{aligned}
\sum_{m=1}^{\infty} \frac{\operatorname{var}(\zeta_m)}{m^2} \leqslant \sum_{m=1}^{\infty} \frac{E\zeta_m^2}{m^2} &= \sum_{m=1}^{\infty} \frac{1}{m^2} \int_0^{\infty} 2tP(|\zeta_m| > t)\,dt \\
&\leqslant \sum_{m=1}^{\infty} \frac{1}{m^2} \int_0^{\infty} 2tP(|\xi_1| > t)1_{\{t<m\}}\,dt \\
&= \int_0^{\infty} \left(2t \sum_{m=1}^{\infty} \frac{1}{m^2} 1_{\{t<m\}}\right) P(|\xi_1| > t)\,dt.
\end{aligned}$$

By [2, Lemma 2.4.4], we have $2t\sum_{m=1}^{\infty} \frac{1}{m^2} 1_{\{t<m\}} \leqslant 4$ for all $t \geqslant 0$. Thus,

$$\sum_{m=1}^{\infty} \frac{\operatorname{var}(\zeta_m)}{m^2} \leqslant 4 \int_0^{\infty} P(|\xi_1| > t)\,dt = 4E|\xi_1|.$$

Substituting this into (9.6), we have

$$\sum_{n=1}^{\infty} P\left(\left|\frac{Y_{k(n)} - EY_{k(n)}}{k(n)}\right| > \varepsilon\right) \leqslant \frac{16E|\xi_1|}{(1-\alpha^{-2})\varepsilon^2} < \infty.$$

Thus, by Borel-Cantelli,

$$P\left(\left|\frac{Y_{k(n)} - EY_{k(n)}}{k(n)}\right| > \varepsilon \text{ i.o.}\right) = 0.$$

Applying Theorem 7.9 proves (9.4).

Next, note that $E\zeta_m \to \mu$ as $m \to \infty$, by dominated convergence. Therefore, using the fact that $a_m \to L$ implies $\frac{1}{n}\sum_{m=1}^{n} a_m \to L$, we have

$$\frac{EY_{k(n)}}{k(n)} = \frac{1}{k(n)} \sum_{m=1}^{k(n)} E\zeta_m \to \mu.$$

Combining this with (9.4), we now have

$$\frac{Y_{k(n)}}{k(n)} \to \mu \quad \text{a.s.}$$

as $n \to \infty$.

Since we assumed $\xi_1 \geqslant 0$ a.s., we can now use monotonicity to extend this convergence from the subsequence $\{Y_{k(n)}\}$ to the entire sequence. Note that $Y_m \leqslant Y_{m+1}$ a.s., for all m. Also note that, for each m, there exists n_m such that $k(n_m) \leqslant m < k(n_m+1)$. Hence,

$$\frac{Y_{k(n_m)}}{k(n_m+1)} \leqslant \frac{Y_m}{m} \leqslant \frac{Y_{k(n_m+1)}}{k(n_m)},$$

which implies

$$\frac{k(n_m)}{k(n_m+1)} \frac{Y_{k(n_m)}}{k(n_m)} \leqslant \frac{Y_m}{m} \leqslant \frac{Y_{k(n_m+1)}}{k(n_m+1)} \frac{k(n_m+1)}{k(n_m)}.$$

Since $k(n+1)/k(n) \to \alpha > 1$, we can let $m \to \infty$ and obtain

$$\frac{1}{\alpha} E\xi_1 \leqslant \liminf_{m\to\infty} \frac{Y_m}{m} \leqslant \limsup_{m\to\infty} \frac{Y_m}{m} \leqslant \alpha E\xi_1,$$

for P-a.e. ω. Since $\alpha > 1$ was arbitrary, this shows that $Y_m/m \to \mu$ a.s.

Finally, we need to remove the truncation and show that $X_n/n \to \mu$ a.s. Note that

$$\begin{aligned}\sum_{k=1}^{\infty} P(\xi_k \neq \zeta_k) \leqslant \sum_{k=1}^{\infty} P(|\xi_k| > k) \leqslant \sum_{k=1}^{\infty} \int_{k-1}^{k} P(|\xi_1| > t)\, dt \\ = \int_0^{\infty} P(|\xi_1| > t)\, dt = E|\xi_1| < \infty.\end{aligned}$$

By Borel-Cantelli, $P(\xi_k \neq \zeta_k \text{ i.o.}) = 0$. Thus, there exists $\Omega^* \in \mathcal{F}$ with $P(\Omega^*) = 1$ such that, for all $\omega \in \Omega^*$, there exists $N(\omega)$ such that $\xi_k(\omega) = \zeta_k(\omega)$ whenever $k > N(\omega)$. Since $Y_n/n \to \mu$ a.s., we may also assume that $Y_n(\omega)/n \to \mu$ for all $\omega \in \Omega^*$.

Let $\omega \in \Omega^*$ be arbitrary. Then

$$\begin{aligned}\frac{X_n(\omega)}{n} &= \frac{1}{n}\sum_{k=1}^{n} \xi_k(\omega) = \frac{1}{n}\left(\sum_{k=1}^{N(\omega)} \xi_k(\omega) + \sum_{k=N(\omega)+1}^{n} \zeta_k(\omega)\right) \\ &= \frac{1}{n}\left(\sum_{k=1}^{N(\omega)} \xi_k(\omega) + Y_n(\omega) - Y_{N(\omega)}(\omega)\right) \to \mu\end{aligned}$$

as $n \to \infty$. Hence, $X_n/n \to \mu$ a.s. and we have proved the theorem under the assumption that $\xi_1 \geqslant 0$ a.s.

We now drop the assumption of nonnegativity on ξ_1. Since $\{\xi_k^+\}_{k=1}^{\infty}$ satisfy the assumptions of the theorem and are nonnegative, we have

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k^+ \to E\xi_1^+ \quad \text{a.s.}$$

as $n \to \infty$. The analogous statement for ξ_k^- also holds. Thus,

$$\frac{X_n}{n} = \frac{1}{n}\sum_{k=1}^n \xi_k^+ - \frac{1}{n}\sum_{k=1}^n \xi_k^- \to E\xi_1^+ - E\xi_1^- = E\xi_1 \quad \text{a.s.}$$

as $n \to \infty$. □

Theorem 9.13. *Let $\{\xi_k\}_{k=1}^\infty$ be i.i.d. and $X_n = \xi_1 + \cdots + \xi_n$. Suppose $E\xi_1$ exists. Then $X_n/n \to E\xi_1$ a.s.*

Proof. Theorem 9.12 covers the case where ξ_1 is integrable. We may therefore assume $E\xi_1 \in \{-\infty, \infty\}$. First suppose $E\xi_1 = \infty$, so that $E\xi_1^+ = \infty$ and $E\xi_1^- < \infty$. Let $M \in \mathbb{N}$ and define $\xi_k^M = \xi_k \wedge M$. Let $X_n^M = \xi_1^M + \cdots + \xi_n^M$. By Theorem 9.12, we have $X_n^M/n \to E\xi_1^M$ a.s., as $n \to \infty$. Thus, there exists $\Omega^* \in \mathcal{F}$ such that $P(\Omega^*) = 1$ and, for all $\omega \in \Omega^*$, we have $\lim_{n\to\infty} X_n^M(\omega)/n = E\xi_1^M$ for all $M \in \mathbb{N}$. Note that $X_n \geqslant X_n^M$. Thus, for all $\omega \in \Omega^*$ and all $M \in \mathbb{N}$, we have

$$\liminf_{n\to\infty} \frac{X_n(\omega)}{n} \geqslant \lim_{n\to\infty} \frac{X_n^M(\omega)}{n} = E\xi_1^M.$$

Let $M \to \infty$. The monotone convergence theorem implies that $E[(\xi_1^M)^+] \to E[\xi_1^+] = \infty$ and $E[(\xi_1^M)^-] \to E[\xi_1^-] < \infty$, as $M \to \infty$. Thus, $E\xi_1^M \to \infty$ as $M \to \infty$, and it follows that $X_n(\omega)/n \to \infty$ for all $\omega \in \Omega^*$. In other words, $X_n/n \to E\xi_1$ a.s.

For the case $E\xi_1 = -\infty$, simply apply the previous result to $\{-\xi_k\}_{k=1}^\infty$. □

Example 9.14. Imagine a janitor whose only job is to watch a single light bulb. The moment it burns out, he replaces it. Let $\{\xi_k\}_{k=1}^\infty$ be i.i.d., $(0, \infty)$-valued random variables, so that ξ_k represents the lifetime of the k-th light bulb. Let $X_n = \xi_1 + \cdots + \xi_n$, so that X_n represents the time at which the janitor replaces the n-th light bulb. Let

$$N(t) = \sup\{n : X_n \leqslant t\},$$

so that $N(t)$ represents the number of light bulbs that have burned out by time t. Note that $N(t) < \infty$ a.s. This is because $\{N(t) = \infty\} \subset \{\xi_n \to 0\}$ and $P(\xi_n \to 0) = 0$ (check; use Theorem 7.11).

We will show that

$$\frac{N(t)}{t} \to \frac{1}{E\xi_1} \quad \textit{a.s.},$$

as $t \to \infty$, where we interpret $1/\infty$ as 0.

To see this, first observe that from the definition of $N(t)$, we have

$$X_{N(t)} \leqslant t < X_{N(t)+1},$$

which gives

$$\frac{X_{N(t)}}{N(t)} \leqslant \frac{t}{N(t)} < \frac{X_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}.$$

Next observe that $X_n(\omega) \leqslant t$ implies $n \leqslant N(t,\omega)$. Thus, $n > N(t,\omega)$ implies $X_n(\omega) > t$. Hence, if there exists $\omega \in \Omega$ and $n \in \mathbb{N}$ such that $N(t,\omega) < n$ for all $t > 0$, then it would follow that $X_n(\omega) = \infty$. But this is impossible since each ξ_k is a $(0,\infty)$-valued random variable. Therefore, $N(t) \to \infty$ a.s., as $t \to \infty$. The result now follows from Theorem 9.13 and Lemma 9.9.

Exercises

9.4. [2, Exercise 2.4.1] Give an example of a sequence, $\{X_n\}_{n=1}^{\infty}$, of $\{0,1\}$-valued random variables, and a sequence $\{N(n)\}_{n=1}^{\infty}$ of $\mathbb{N}$-valued random variables, such that $X_n \to 0$ in probability, $N(n) \uparrow \infty$ a.s., and $X_{N(n)} \to 1$ a.s.

9.5. [2, Exercise 2.4.3] Let $U_1, U_2, \ldots$ be i.i.d., $\mathbb{R}^2$-valued random variables such that U_1 is uniformly distributed on the unit disk. Let $X_0 = (1,0) \in \mathbb{R}^2$ a.s., and for $n \in \mathbb{N}$, let $X_n = |X_{n-1}|U_n$. Find an explicit constant c such that $n^{-1} \log |X_n| \to c$ a.s.

Chapter 10

Central Limit Theorems

10.1 Limit theorems in $\mathbb{R}$

This section corresponds to [2, Sections 3.4.1–3.4.2].

Lemma 10.1. *If* $c_n \to c \in \mathbb{C}$, *then* $(1 + c_n/n)^n \to e^c$.

Proof. Basic analysis; see book. □

Theorem 10.2 (classical central limit theorem)**.** *Let* $\xi_1, \xi_2, \dots$ *be i.i.d. with* $E\xi_1 = \mu$ *and* $\operatorname{var}(\xi_1) = \sigma^2 \in (0, \infty)$. *Let* $X_n = \xi_1 + \cdots + \xi_n$. *Then*

$$\frac{X_n - n\mu}{\sigma n^{1/2}} \Rightarrow Z,$$

where $Z \sim N(0, 1)$.

Proof. Let $\xi_j' = \xi_j - \mu$ and $X_n' = \xi_1' + \cdots + \xi_n'$. Let

$$Y_n = \frac{X_n - n\mu}{\sigma n^{1/2}} = \frac{X_n'}{\sigma n^{1/2}}.$$

Let $\varphi_n = \varphi_{Y_n}$ and $\varphi = \varphi_{\xi_1'}$. By Theorem 7.49,

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + r(t),$$

where $t^{-2} r(t) \to 0$ as $t \to 0$. Thus,

$$\varphi_n(t) = \varphi\left(\frac{t}{\sigma n^{1/2}}\right)^n = \left(1 - \frac{t^2}{2n} + r\left(\frac{t}{\sigma n^{1/2}}\right)\right)^n.$$

Fix $t \in \mathbb{R}$ and let

$$c_n = -\frac{t^2}{2} + nr\left(\frac{t}{\sigma n^{1/2}}\right),$$

so that $\varphi_n(t) = (1 + c_n/n)^n$. Note that

$$nr\left(\frac{t}{\sigma n^{1/2}}\right) = \frac{\sigma^2}{t^2}\left(\frac{t}{\sigma n^{1/2}}\right)^{-2} r\left(\frac{t}{\sigma n^{1/2}}\right) \to 0,$$

as $n \to \infty$. Thus, by Lemma 10.1, $\varphi_n(t) \to e^{-t^2/2}$, which implies $Y_n \Rightarrow Z$. $\square$

Example 10.3. Let $\xi_1, \xi_2, \dots$ be i.i.d. with $P(\xi_1 = 1) = 18/28$ and $P(\xi_1 = -1) = 20/28$, and $X_n = \xi_1 + \cdots + \xi_n$. Then X_n represents your total winnings after playing n games of roulette, if in each game you bet \$1 on black. We are interested in the probability that your winnings are positive after $n = 361 = 19^2$ plays.

First note that $\mu = E\xi_1 = -1/19$ and

$$\sigma^2 = \mathrm{var}(\xi_1) = E\xi_1^2 - \mu^2 = 1 - (1/19)^2 = \frac{360}{361}.$$

By the central limit theorem,

$$\frac{X_n - n\mu}{\sigma n^{1/2}} \Rightarrow Z,$$

where $Z \sim N(0,1)$. Thus, for any fixed $x \in \mathbb{R}$,

$$P\left(\frac{X_n - n\mu}{\sigma n^{1/2}} > x\right) \to 1 - \Phi(x) \tag{10.1}$$

as $n \to \infty$. Let us take $x = 19/(6\sqrt{10})$. Then, for large n,

$$P\left(\frac{X_n - n\mu}{\sigma n^{1/2}} > \frac{19}{6\sqrt{10}}\right) \approx 1 - \Phi\left(\frac{19}{6\sqrt{10}}\right).$$

Taking $n = 361$ and noting that $x = \sigma^{-1}$, this gives

$$P\left(\frac{X_{361} + 19}{19} > 1\right) = P(X_{361} > 0) \approx 1 - \Phi\left(\frac{19}{6\sqrt{10}}\right) \approx 0.1583196.$$

Thus, the probability that your winnings are positive after 361 plays is about 16%.

One thing to be careful of is the following. In the book, they begin their analysis with

$$P(X_n > 0) = P\left(\frac{X_n - n\mu}{\sigma n^{1/2}} > -\frac{n\mu}{\sigma n^{1/2}}\right). \tag{10.2}$$

However, the central limit theorem only tells us (10.1) for a fixed x. It does not say anything about the asymptotic relationship between (10.2) and

$$1 - \Phi\left(-\frac{n\mu}{\sigma n^{1/2}}\right).$$

On the other hand, if you are in the mood to not be careful, there is another approach. Heuristically, the central limit theorem tells us that

$$X_n \overset{d}{\approx} n\mu + n^{1/2}\sigma Z.$$

Thus,

$$P(X_n > 0) \approx P(n\mu + n^{1/2}\sigma Z > 0) = P(Z > -n^{1/2}\mu/\sigma),$$

and the computations proceed as before.

Example 10.4. Let $\xi_1, \xi_2, \dots$ be i.i.d. with $P(\xi_1 = 1) = P(\xi_1 = 0) = 1/2$, and $X_n = \xi_1 + \cdots + \xi_n$. Then X_n represents the number of heads in n tosses of a fair coin.

We have $\mu = E\xi_1 = 1/2$ and $\sigma^2 = \mathrm{var}(\xi_1) = E\xi^2 - 1/4 = 1/2 - 1/4 = 1/4$. Thus, the central limit theorem gives

$$X_n \overset{d}{\approx} \frac{n}{2} + \frac{\sqrt{n}}{2} Z.$$

Taking $n = 10000$, we have

$$X := X_{10000} \approx 5000 + 5\sqrt{50}\, Z.$$

Let $a \in (0, 5000)$. Then

$$\begin{aligned} P(X \in [5000 - a, 5000 + a]) &\approx P(5000 + 5\sqrt{50}\, Z \in [5000 - a, 5000 + a]) \\ &= P\left(|Z| \leqslant \frac{a}{5\sqrt{50}}\right) \\ &= 2\Phi\left(\frac{a}{5\sqrt{50}}\right) - 1 \end{aligned}$$

Suppose we want this probability to be 95%. Then we need

$$\Phi\left(\frac{a}{5\sqrt{50}}\right) = \frac{1.95}{2} = 0.975,$$

which gives

$$\frac{a}{5\sqrt{50}} = 1.959964,$$

or $a \approx 69.29519$. In other words, in 10000 flips of a fair coin, the chance of getting between 4931 and 5069 heads is about 95%.

Example 10.5. Let $\{X_n\}$ be as in the previous example, and let us try to approximate $P(X_{16} = 8)$. If we proceed as before, we have $X_{16} \overset{d}{\approx} 8 + 2Z$, so that

$$P(X_{16} = 8) \approx P(8 + 2Z = 8) = P(Z = 0) = 0.$$

Of course, this is a terrible approximation, and it results from using a continuous random variable to approximate a discrete random variable. To deal with this situation, we use something called the *histogram correction*. Note that since $X_{16} \in \mathbb{Z}$, we have

$$P(X_{16} = 8) = P(X_{16} \in [7.5, 8.5]).$$

Thus, we have

$$\begin{aligned} P(X_{16} = 8) &\approx P(8 + 2Z \in [7.5, 8.5]) \\ &= P(|Z| \leqslant 1/4) = 2\Phi(1/4) - 1 \approx 0.1974127. \end{aligned}$$

The exact probability is

$$\binom{16}{8} 2^{-16} \approx 0.1963806.$$

More generally, if $A \subset \mathbb{Z}$, then the histogram correction asks us to use the identity

$$P(X_n \in A) = P\bigg(X_n \in \bigcup_{k \in A} [k - 1/2, k + 1/2]\bigg).$$

For example, $P(X_n \leqslant 11) = P(X_n \leqslant 11.5)$ and $P(X_n < 11) = P(X_n \leqslant 10.5)$.

In the preceding examples we approximated the exact probabilities with probabilities from a normal distribution, and used the central limit theorem to justify this approximation. If we need to quantify the error we make when using such an approximation, the following theorem is helpful.

Theorem 10.6 (Berry-Esseen theorem)**.** *Let $\xi_1, \xi_2, \ldots$ be i.i.d. with $E\xi_1 = 0$, $E\xi_1^2 = \sigma^2$, and $E|\xi_1|^3 = \rho < \infty$. Let $X_n = \xi_1 + \cdots + \xi_n$. Then*

$$\left| P\left(\frac{X_n}{\sigma n^{1/2}} \leqslant x\right) - \Phi(x)\right| \leqslant \frac{3\rho}{\sigma^3 n^{1/2}},$$

for all $x \in \mathbb{R}$.

Proof. See [2, Theorem 3.4.9]. □

The following theorem is our final example of using the classical central limit theorem.

Theorem 10.7. *For each $\lambda > 0$, let $N_\lambda \sim$ Poisson(λ). Then*

$$\frac{N_\lambda - \lambda}{\lambda^{1/2}} \Rightarrow Z,$$

as $\lambda \to \infty$, where $Z \sim N(0, 1)$.

Proof. Let $\xi_1, \xi_2, \ldots$ be i.i.d. with $\xi_1 \sim$ Poisson(1). Let $X_n = \xi_1 + \cdots + \xi_n$. Then $X_n \sim$ Poisson(n). Since $X_n \overset{d}{=} N_n$, it follows from the central limit theorem that

$$\frac{N_n - n}{n^{1/2}} \Rightarrow Z. \tag{10.3}$$

Now let $\{\lambda_n\}_{n=1}^\infty \subset (0, \infty)$ satisfy $\lambda_n \to \infty$ as $n \to \infty$. Let $\{U_n\}$ and $\{V_n\}$ be i.i.d. sequences such that

(i) $\{X_n\}$, $\{U_n\}$, and $\{V_n\}$ are independent,

(ii) $U_n \sim \text{Poisson}(\lambda_n - \lfloor \lambda_n \rfloor)$, and

(iii) $V_n \sim \text{Poisson}(\lfloor \lambda_n \rfloor + 1 - \lambda_n)$.

Then

$$X_{\lfloor \lambda_n \rfloor} \leqslant X_{\lfloor \lambda_n \rfloor} + U_n \leqslant X_{\lfloor \lambda_n \rfloor} + U_n + V_n.$$

Also,

$$X_{\lfloor \lambda_n \rfloor} + U_n \overset{d}{=} N_{\lambda_n},$$

and

$$X_{\lfloor \lambda_n \rfloor} + U_n + V_n \overset{d}{=} X_{\lfloor \lambda_n \rfloor + 1}.$$

Fix $x \in \mathbb{R}$. Then

$$\begin{aligned} P\left(\frac{N_{\lambda_n} - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) &= P\left(\frac{X_{\lfloor \lambda_n \rfloor} + U_n - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) \\ &\leqslant P\left(\frac{X_{\lfloor \lambda_n \rfloor} - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) \\ &= P\left(\frac{\lfloor \lambda_n \rfloor^{1/2}}{\lambda_n^{1/2}}\left(\frac{X_{\lfloor \lambda_n \rfloor} - \lfloor \lambda_n \rfloor}{\lfloor \lambda_n \rfloor^{1/2}} + \frac{\lfloor \lambda_n \rfloor - \lambda_n}{\lfloor \lambda_n \rfloor^{1/2}}\right) \leqslant x\right) \\ &= P\left(\frac{X_{\lfloor \lambda_n \rfloor} - \lfloor \lambda_n \rfloor}{\lfloor \lambda_n \rfloor^{1/2}} \leqslant \frac{\lambda_n^{1/2} x}{\lfloor \lambda_n \rfloor^{1/2}} - \frac{\lfloor \lambda_n \rfloor - \lambda_n}{\lfloor \lambda_n \rfloor^{1/2}}\right). \end{aligned}$$

By (10.3) and Lemma 7.23, we have

$$\limsup_{n \to \infty} P\left(\frac{N_{\lambda_n} - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) \leqslant P(Z \leqslant x).$$

Conversely,

$$\begin{aligned} P\left(\frac{N_{\lambda_n} - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) &= P\left(\frac{X_{\lfloor \lambda_n \rfloor} + U_n - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) \\ &\geqslant P\left(\frac{X_{\lfloor \lambda_n \rfloor} + U_n + V_n - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) \\ &= P\left(\frac{X_{\lfloor \lambda_n \rfloor + 1} - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right), \end{aligned}$$

and a similar argument gives

$$\liminf_{n \to \infty} P\left(\frac{N_{\lambda_n} - \lambda_n}{\lambda_n^{1/2}} \leqslant x\right) \geqslant P(Z \leqslant x),$$

which finishes the proof. □

The following generalization of the classical central limit theorem is much more versatile and allows us to analyze sequences whose rate of convergence differs from $\sqrt{n}$.

Theorem 10.8 (Lindeberg-Feller theorem)**.** *For $n \in \mathbb{N}$, let $\xi_{n,1}, \xi_{n,2} \ldots, \xi_{n,n}$ be independent with $E\xi_{n,m} = 0$, and let $\overline{X}_n = \xi_{n,1} + \cdots + \xi_{n,n}$. Suppose*

(i) $\sum_{m=1}^{n} E\xi_{n,m}^2 \to \sigma^2 > 0$, *and*

(ii) for all $\varepsilon > 0$, *we have* $\sum_{m=1}^{n} E[\xi_{n,m}^2 1_{\{|\xi_{n,m}|>\varepsilon\}}] \to 0$.

Then $\overline{X}_n \Rightarrow \sigma Z$.

Proof. Uses characteristic functions. See the book for details. □

Example 10.9. Let $\xi_1, \xi_2, \ldots$ be i.i.d. with $\xi_1 \stackrel{d}{=} -\xi_1$ and $P(|\xi_1| > x) = x^{-2}$ for all $x \geqslant 1$. Let $X_n = \xi_1 + \cdots + \xi_n$. Notice that

$$E|\xi_1| = \int_0^\infty P(|\xi_1| > x)\,dx = 1 + \int_1^\infty \frac{1}{x^2}\,dx < \infty,$$

and $E\xi_1 = 0$. However,

$$E|\xi_1|^2 = \int_0^\infty 2xP(|\xi_1| > x)\,dx = \int_0^1 2x\,dx + \int_1^\infty \frac{2}{x}\,dx = \infty.$$

Thus, the classical central limit theorem does not apply. We will use the Lindeberg-Feller theorem to show that

$$\frac{X_n}{\sqrt{n \log n}} \Rightarrow Z, \tag{10.4}$$

as $n \to \infty$, where $Z \sim N(0,1)$.

Let $\zeta_{n,m} = \xi_m 1_{\{|\xi_m| \leqslant c_n\}}$, where $c_n = n^{1/2} \log\log n$, and $Y_n = \zeta_{n,1} + \cdots + \zeta_{n,n}$. Define

$$\xi_{n,m} = \frac{\zeta_{n,m}}{\sqrt{n \log n}}.$$

To verify Theorem 10.8(i), observe that

$$\begin{aligned}
E\zeta_{n,m}^2 &= \int_0^\infty 2xP(|\zeta_{n,m}| > x)\,dx \\
&= \int_0^{c_n} 2xP(|\zeta_{n,m}| > x)\,dx \\
&\leqslant \int_0^{c_n} 2xP(|\xi_m| > x)\,dx \\
&= \int_0^1 2x\,dx + \int_1^{c_n} \frac{2}{x}\,dx \\
&= 1 + 2\log c_n \\
&= 1 + \log n + 2\log\log\log n.
\end{aligned}$$

On the other hand,

$$\begin{aligned} E\zeta_{n,m}^2 &\geqslant \int_1^{\sqrt{n}} 2xP(|\zeta_{n,m}| > x)\,dx \\ &= \int_1^{\sqrt{n}} 2x(P(|\xi_m| > x) - P(|\xi_m| > c_n))\,dx. \end{aligned}$$

For $x \leqslant \sqrt{n}$, we have

$$P(|\xi_m| > c_n) = \frac{x^2}{c_n^2} P(|\xi_m| > x) \leqslant \frac{1}{(\log\log n)^2} P(|\xi_m| > x).$$

Thus,

$$E\zeta_{n,m}^2 \geqslant \left(1 - \frac{1}{(\log\log n)^2}\right) \int_1^{\sqrt{n}} 2xP(|\xi_m| > x)\,dx = \left(1 - \frac{1}{(\log\log n)^2}\right) \log n.$$

It follows that

$$1 - \frac{1}{(\log\log n)^2} \leqslant \sum_{m=1}^n E\xi_{n,m}^2 \leqslant \frac{1}{\log n} + 1 + \frac{2\log\log\log n}{\log n},$$

and Theorem 10.8(i) follows immediately with $\sigma = 1$.

Now fix $\varepsilon > 0$. Since $|\zeta_{n,m}| \leqslant c_n$ a.s. and $c_n/\sqrt{n\log n} \to 0$, we may choose $N \in \mathbb{N}$ such that for all $n \geqslant N$, we have $|\xi_{n,m}| \leqslant \varepsilon$ a.s. For any such n, we have $\xi_{n,m}^2 1_{\{|\xi_{n,m}| > \varepsilon\}} = 0$ a.s., which verifies Theorem 10.8(ii).

By Theorem 10.8,

$$\overline{Y}_n := \frac{Y_n}{\sqrt{n\log n}} = \xi_{n,1} + \cdots + \xi_{n,n} \Rightarrow Z,$$

where $Z \sim N(0,1)$. Fix $x \in \mathbb{R}$. Then

$$P\left(\frac{X_n}{\sqrt{n\log n}} \leqslant x\right) \leqslant P(X_n \neq Y_n) + P\left(\frac{Y_n}{\sqrt{n\log n}} \leqslant x.\right)$$

Since

$$P(X_n \neq Y_n) \leqslant \sum_{m=1}^n P(|\xi_m| > c_n) = \frac{2n}{c_n^2} = \frac{2}{(\log\log n)^2} \to 0,$$

it follows that

$$\limsup_{n\to\infty} P\left(\frac{X_n}{\sqrt{n\log n}} \leqslant x\right) \leqslant \Phi(x).$$

Similarly,

$$P\left(\frac{X_n}{\sqrt{n\log n}} \leqslant x\right) \geqslant P\left(\frac{Y_n}{\sqrt{n\log n}} \leqslant x\right) - P(X_n \neq Y_n),$$

and so

$$\liminf_{n\to\infty} P\left(\frac{X_n}{\sqrt{n\log n}} \leqslant x\right) \geqslant \Phi(x).$$

Putting them together proves (10.4).

Exercises

10.1. [2, Exercise 3.4.4] Let $\xi_1, \xi_2, \ldots$ be i.i.d., $[0, \infty)$-valued random variables with $E\xi_1 = 1$ and $\mathrm{var}(\xi_1) = \sigma^2 \in (0, \infty)$. Let $X_n = \xi_1 + \cdots + \xi_n$. Prove that $2(\sqrt{X_n} - \sqrt{n}) \Rightarrow \sigma Z$, where $Z \sim N(0,1)$.

10.2. [2, Exercise 3.4.13(ii)] Let $\beta \in (0,1)$. Let $\xi_1, \xi_2, \ldots$ be independent with

$$\begin{aligned} P(\xi_k = -k) &= \frac{1}{2} k^{-\beta}, \\ P(\xi_k = 0) &= 1 - k^{-\beta}, \\ P(\xi_k = k) &= \frac{1}{2} k^{-\beta}. \end{aligned}$$

Let $X_n = \xi_1 + \cdots + \xi_n$. Find an explicit constant c (in terms of β) such that

$$\frac{X_n}{n^{(3-\beta)/2}} \Rightarrow cZ,$$

where $Z \sim N(0,1)$.

10.2 Poisson convergence

This section corresponds to [2, Section 3.6].

10.2.1 Convergence to a Poisson distribution

The following is the basic Poisson limit theorem, which includes a bound on the rate of convergence.

Theorem 10.10. *For each $n \in \mathbb{N}$, let $\xi_{n,1} \ldots, \xi_{n,n}$ be independent $\{0,1\}$-valued random variables with $P(\xi_{n,m} = 1) = p_{n,m}$. Let $\lambda_n = \sum_{m=1}^n p_{n,m}$. Define $X_n = \xi_{n,1} + \cdots + \xi_{n,n}$ and let $N_n \sim \mathrm{Poisson}(\lambda_n)$. Then*

$$\sup_{A \subset \mathbb{N} \cup \{0\}} |P(X_n \in A) - P(N_n \in A)| \leqslant \sum_{m=1}^n p_{n,m}^2.$$

Suppose that

(i) $\lambda_n \to \lambda \in (0, \infty)$ *and*

(ii) $\max_{1 \leqslant m \leqslant n} p_{n,m} \to 0$.

Then $X_n \Rightarrow N$, where $N \sim \mathrm{Poisson}(\lambda)$.

Proof. In the textbook, see the second proof of [2, Theorem 3.6.1], which begins on p. 150. The proof uses [2, Lemmas 3.6.2-3.6.4] and involves only basic calculations with discrete measures. □

The following is a slight generalization, but without the rate of convergence.

Theorem 10.11. *For each $n \in \mathbb{N}$, let $\xi_{n,1} \dots, \xi_{n,n}$ be independent, nonnegative integer valued random variables with $P(\xi_{n,m} = 1) = p_{n,m}$ and $P(\xi_{n,m} \geqslant 2) = \varepsilon_{n,m}$. Suppose that*

(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$,

(ii) $\max_{1 \leqslant m \leqslant n} p_{n,m} \to 0$, *and*

(iii) $\sum_{m=1}^{n} \varepsilon_{n,m} \to 0$.

Let $X_n = \xi_{n,1} + \cdots + \xi_{n,n}$. Then $X_n \Rightarrow N$, where $N \sim \text{Poisson}(\lambda)$.

Proof. Let $\xi'_{n,m} = 1_{\{\xi_{n,m}=1\}}$ and $X'_n = \xi'_{n,1} + \cdots + \xi'_{n,n}$. By Theorem 10.10, we have $X'_n \Rightarrow N$. Fix $\varepsilon > 0$. Then

$$P(|X_n - X'_n| > \varepsilon) \leqslant P(X_n \neq X'_n) \leqslant \sum_{m=1}^{n} P(|\xi_{n,m}| \geqslant 2) \to 0,$$

by (iii). Thus, $X_n - X'_n \to 0$ in probability, and it follows from Exercise 7.10 that $X_n = X'_n + (X_n - X'_n) \Rightarrow N$. □

Example 10.12. Let $n = 400$. For $1 \leqslant m \leqslant 400$, let $\xi_{n,1}, \dots, \xi_{n,n}$ be independent with $P(\xi_{n,m} = 1) = 1/365$ and $P(\xi_{n,m} = 0) = 364/365$. Then $X_n = \xi_{n,1} + \cdots + \xi_{n,n}$ represents the number of students in a class of 400 that have their birthday on the day of the final. By Theorem 10.10, we have $X_n \approx_d N$, where $N \sim \text{Poisson}(400/365)$. Thus, the probability that no one has their birthday on the day of the final is

$$P(X_n = 0) \approx P(N = 0) = e^{-400/365} \approx 0.3342419.$$

In general, when there are many chances for something rare to happen, the number of occurrences is approximately Poisson distributed.

Theorem 10.10 also gives us a bound on the error:

$$|P(X_n = 0) - P(N = 0)| \leqslant \sum_{m=1}^{400} \frac{1}{365^2} = \frac{400}{365^2} \approx 0.003002439.$$

Example 10.13. Let $\zeta_{n,1}, \dots, \zeta_{n,n}$ be independent $U(-n/2, n/2)$. That is, we place n points randomly on the interval of length n centered at the origin. Fix $a < b$ and let $\xi_{n,m} = 1_{\{\zeta_{n,m} \in (a,b)\}}$. Then $X_n = \xi_{n,1} + \cdots + \xi_{n,n}$ represents the number of points which land in (a, b).

Then, in Theorem 10.10, we have $p_{n,m} = (b - a)/n$ and $\lambda_n = b - a$. Thus, $X_n \approx_d N$, where $N \sim \text{Poisson}(b - a)$.

Example 10.14 (occupancy problem)**.** Let $Y_1, \dots, Y_r$ be independent and uniform on $\{1, \dots, n\}$. We are modeling here the random placement of r balls into n boxes. For $1 \leqslant i \leqslant n$, let $\zeta_i = \sum_{\ell=1}^{r} 1_{\{Y_\ell = i\}}$, which represents the number of balls in the ith box.

Let $\xi_{n,r,i} = 1_{\{\zeta_i=0\}}$, so that $X_{n,r} = \xi_{n,r,1} + \cdots + \xi_{n,r,n}$ represents the number of empty boxes. We will show in Proposition 10.20 that if $ne^{-r(n)/n} \to \lambda \in (0,\infty)$, then $X_{n,r(n)} \Rightarrow N$, where $N \sim \text{Poisson}(\lambda)$.

Lemma 10.15. *Let $p_m(r,n) = P(X_{n,r} = m)$. Then*

$$p_0(r,n) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(1 - \frac{k}{n}\right)^r.$$

Proof. Note that

$$P(\zeta_{i_1} = 0, \zeta_{i_2} = 0, \ldots, \zeta_{i_k} = 0) = \left(1 - \frac{k}{n}\right)^r.$$

Then, using inclusion-exclusion,

$$\begin{aligned}
p_0(r,n) &= 1 - P(X_{n,r} > 0) \\
&= 1 - P\left(\bigcup_{k=1}^{n} \{\zeta_k = 0\}\right) \\
&= 1 - n\left(1 - \frac{1}{n}\right)^r + \binom{n}{2}\left(1 - \frac{2}{n}\right)^r - \cdots + (-1)^n \binom{n}{n}\left(1 - \frac{n}{n}\right)^r \\
&= \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(1 - \frac{k}{n}\right)^r,
\end{aligned}$$

and we are done. □

Lemma 10.16. *We also have*

$$p_m(r,n) = \binom{n}{m}\left(1 - \frac{m}{n}\right)^r p_0(r, n-m).$$

Proof. Let

$$\begin{aligned}
A &= \{\zeta_1 > 0, \zeta_2 > 0, \ldots, \zeta_{n-m} > 0\}, \\
B &= \{\zeta_{n-m+1} = 0, \zeta_{n-m+2} = 0, \ldots, \zeta_n = 0\}.
\end{aligned}$$

Conditioned on B, the random variables $Y_1, \ldots, Y_r$ are independent and uniform on $\{1, \ldots, n-m\}$. Thus, $P(A \mid B) = p_0(r, n-m)$ and it follows that

$$P(A \cap B) = P(B)P(A \mid B) = \left(1 - \frac{m}{n}\right)^r p_0(r, n-m).$$

Therefore,

$$p_m(r,n) = \binom{n}{m}\left(1 - \frac{m}{n}\right)^r p_0(r, n-m),$$

and we are done. □

Now suppose $\{r(n)\}_{n=1}^{\infty}$ is a sequence satisfying $ne^{-r(n)/n} \to \lambda \in (0,\infty)$. For notational simplicity, we will suppress the dependence of $r(n)$ on n and simply write r.

Lemma 10.17. *There exists $C > 0$ such that*

$$\binom{n}{m}\left(1-\frac{m}{n}\right)^r \leqslant \frac{C^m}{m!}$$

for all m and n. Moreover,

$$\limsup_{n\to\infty}\binom{n}{m}\left(1-\frac{m}{n}\right)^r \leqslant \frac{\lambda^m}{m!}.$$

Proof. Choose $C > 0$ such that $ne^{-r/n} \leqslant C$ for all n. Using $1 - x \leqslant e^{-x}$ and $n!/(n-m)! \leqslant n^m$, we have

$$\binom{n}{m}\left(1-\frac{m}{n}\right)^r \leqslant \frac{n^m}{m!}e^{-mr/n} \leqslant \frac{C^m}{m!}.$$

Moreover,

$$\limsup_{n\to\infty}\binom{n}{m}\left(1-\frac{m}{n}\right)^r \leqslant \lim_{n\to\infty}\frac{n^m}{m!}e^{-mr/n} = \frac{\lambda^m}{m!},$$

and we are done. □

Lemma 10.18. *For each fixed m,*

$$\lim_{n\to\infty}\binom{n}{m}\left(1-\frac{m}{n}\right)^r = \frac{\lambda^m}{m!}.$$

Proof. First note that $n!/(n-m)! \geqslant (n-m)^m$. Also, for $t \in [0,1/2]$,

$$\begin{aligned}\log(1-t) = -\sum_{j=1}^{\infty}\frac{t^j}{j} \geqslant -t - \frac{1}{2}\sum_{j=2}^{\infty}t^j = -t - \frac{t^2}{2}\sum_{j=0}^{\infty}t^j \\ \geqslant -t - \frac{t^2}{2}\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j = -t - t^2.\end{aligned}$$

It follows that

$$\begin{aligned}\binom{n}{m}\left(1-\frac{m}{n}\right)^r &\geqslant \frac{(n-m)^m}{m!}\left(1-\frac{m}{n}\right)^r \\ &= \frac{1}{m!}\left(1-\frac{m}{n}\right)^m n^m\left(1-\frac{m}{n}\right)^r, \qquad (10.5)\end{aligned}$$

and

$$\begin{aligned}\log\left(n^m\left(1-\frac{m}{n}\right)^r\right) &= m\log n + r\log\left(1-\frac{m}{n}\right) \\ &\geqslant m\log n - \frac{rm}{n} - \frac{rm^2}{n^2} \\ &= m\left(\log n - \frac{r}{n}\right) + \frac{m^2}{n}\left(\log n - \frac{r}{n}\right) - \frac{m^2\log n}{n}.\end{aligned}$$

Note that $\log n - r/n = \log(ne^{-r/n}) \to \log \lambda$. Thus,

$$\liminf_{n\to\infty} \log\left(n^m\left(1-\frac{m}{n}\right)^r\right) \geqslant m\log\lambda,$$

which implies

$$\liminf_{n\to\infty}\left(n^m\left(1-\frac{m}{n}\right)^r\right) \geqslant \lambda^m.$$

Applying this to (10.5) gives

$$\liminf_{n\to\infty}\binom{n}{m}\left(1-\frac{m}{n}\right)^r \geqslant \frac{\lambda^m}{m!}.$$

Combining this with Lemma 10.17 finishes the proof. □

Lemma 10.19. *For each fixed m, we have $p_0(r, n-m) \to e^{-\lambda}$.*

Proof. First assume $m = 0$. By Lemma 10.15, we can write

$$p_0(r,n) = \sum_{k=0}^{\infty} f_n(k),$$

where

$$f_n(k) = (-1)^k\binom{n}{k}\left(1-\frac{k}{n}\right)^r 1_{[0,n]}(k).$$

By Lemma 10.18, for each fixed k, we have $f_n(k) \to (-1)^k\lambda^k/k!$ as $n \to \infty$. By Lemma 10.17 and the fact that $\sum_{k=0}^{\infty} C^k/k! < \infty$, we may apply dominated convergence to conclude that

$$\lim_{n\to\infty} p_0(r,n) = \sum_{k=0}^{\infty}(-1)^k\frac{\lambda^k}{k!} = e^{-\lambda},$$

proving the case $m = 0$.

Now suppose m is arbitrary. In the previous case, we proved that $ne^{-r/n} \to \lambda$ implies $p_0(r,n) \to e^{-\lambda}$. Thus, it will suffice to prove that $(n-m)e^{-r/(n-m)} \to \lambda$. First, note that

$$-\frac{r}{n^2} = \frac{1}{n}\left(\log n - \frac{r}{n}\right) - \frac{\log n}{n} \to 0,$$

since $\log n - r/n \to \log\lambda$. Therefore,

$$(n-m)e^{-r/(n-m)} = \left(1-\frac{m}{n}\right)(ne^{-r/n})\exp\left(-\frac{rm}{n(n-m)}\right) \to \lambda,$$

and we are done. □

Proposition 10.20. *For each fixed m, we have $P(X_{n,r(n)} = m) \to e^{-\lambda}\lambda^m/m!$ as $n \to \infty$. That is, $X_{n,r(n)} \Rightarrow N$, where $N \sim \text{Poisson}(\lambda)$.*

Proof. By Lemmas 10.16, 10.18, and 10.19,

$$P(X_{n,r} = m) = p_m(r, n) = \binom{n}{m} \left(1 - \frac{m}{n}\right)^r p_0(r, n - m) \to \frac{\lambda^m}{m!} e^{-\lambda}.$$

Thus, for any $x > 0$,

$$P(X_{n,r} \leqslant x) = \sum_{m=0}^{\lfloor x \rfloor} P(X_{n,r} = m) \to \sum_{m=0}^{\lfloor x \rfloor} \frac{\lambda^m}{m!} e^{-\lambda} = P(N \leqslant x),$$

and we are done. □

Example 10.21 (coupon collector's problem)**.** Let $\{Y_m : m \in \mathbb{N}\}$ be i.i.d. and uniform on $\{1, \ldots, n\}$ and $T_n = \inf\{m : |\{Y_1, \ldots, Y_m\}| = n\}$. Recall from Example 9.3 that $T_n/(n \log n) \to 1$ in probability. In this example, we will show that

$$\frac{T_n - n \log n}{n} \Rightarrow X,$$

where $P(X \leqslant x) = \exp(-e^{-x})$ for all $x \in \mathbb{R}$.

With the notation of Example 10.14, we have $\{T_n \leqslant r\} = \{X_{n,r} = 0\}$. Fix $x \in \mathbb{R}$. Then

$$P\left(\frac{T_n - n \log n}{n} \leqslant x\right) = P(T_n \leqslant r(n)) = P(X_{n,r(n)} = 0),$$

where $r(n) = \lfloor n \log n + nx \rfloor$. Note that

$$\begin{aligned} \log(ne^{-r(n)/n}) &= \log n - \frac{\lfloor n \log n + nx \rfloor}{n} \\ &= -x + \frac{n \log n + nx - \lfloor n \log n + nx \rfloor}{n} \to -x, \end{aligned}$$

as $n \to \infty$. Thus, $ne^{-r(n)/n} \to \lambda := e^{-x}$, so by Proposition 10.20, we have

$$P\left(\frac{T_n - n \log n}{n} \leqslant x\right) \to e^{-\lambda} = \exp(-e^{-x}),$$

which is what we wanted to show.

10.2.2 The Poisson process

Let $N = \{N(t) : t \geqslant 0\}$ be a nonnegative integer valued stochastic process with $N(0) = 0$ a.s. Suppose that N is increasing, that is, for P-a.e. $\omega \in \Omega$, the function $t \mapsto N(t, \omega)$ is increasing. As an application, $N(t)$ could represent the number of occurrences of a certain type of event during the time interval $(0, t]$. In this case, $N(t) - N(s)$ is the number occurrences in $(s, t]$.

Theorem 10.22. *Let $\lambda > 0$. Suppose that* *(i) for all* $0 = t_0 < t_1 < \cdots < t_n$*, the random variables* $\{N(t_k) - N(t_{k-1})\}_{k=1}^n$ *are independent,*

(ii) $N(t_1) - N(s_1) =_d N(t_2) - N(s_2)$ *whenever* $t_1 - s_1 = t_2 - s_2$*,*

(iii) $P(N(h) = 1) = \lambda h + o(h)$*, and*

(iv) $P(N(h) \geqslant 2) = o(h)$*.*

Then $N(t) \sim \text{Poisson}(\lambda t)$ *for all* $t > 0$*.*

Proof. Fix $t > 0$. For $n \in \mathbb{N}$ and $m \in \{1, \ldots, n\}$, define

$$\xi_{m,n} = N\left(\frac{(m-1)t}{n}\right) - N\left(\frac{mt}{n}\right).$$

Then $X_n := \xi_{n,1} + \cdots + \xi_{n,n} = N(t)$ for all n.

Note that

$$p_{n,m} := P(\xi_{n,m} = 1) = P(N(t/n) = 1) = \lambda t/n + r_1(t/n),$$

where $r_1(h)/h \to 0$ as $h \to 0$. Thus,

$$\sum_{m=1}^n p_{n,m} = \lambda t + n r_1(t/n) \to \lambda t$$

as $n \to \infty$, and

$$\max_{1 \leqslant m \leqslant n} p_{n,m} = p_{n,1} \to 0$$

as $n \to \infty$. Similarly,

$$\varepsilon_{n,m} = P(\xi_{n,m} \geqslant 2) = P(N(t/n) \geqslant 2) = r_2(t/n),$$

where $r_2(h)/h \to 0$. Thus,

$$\sum_{m=1}^n \varepsilon_{n,m} = n r_2(t/n) \to 0$$

as $n \to \infty$. Therefore, by Theorem 10.11, we have $X_n \Rightarrow N$, where $N \sim \text{Poisson}(\lambda t)$. But $X_n = N(t)$ for all n. So $N(t) \sim \text{Poisson}(\lambda t)$. □

If N satisfies the hypotheses of Theorem 10.22 and $t \mapsto N(t, \omega)$ is right-continuous for P-a.e. $\omega \in \Omega$, then N is a **Poisson process with rate** λ.

A constructive characterization of the Poisson process is given by the following theorem.

Theorem 10.23. *Let* $\lambda > 0$*. Let* $\xi_1, \xi_2, \ldots$ *be i.i.d. with* $\xi_1 \sim \text{Exp}(\lambda)$*. Let* $T_0 = 0$ *and, for* $n \in \mathbb{N}$*, let* $T_n = \xi_1 + \cdots + \xi_n$*. For* $t \geqslant 0$*, define*

$$N(t) = \sup\{n : T_n \leqslant t\}.$$

Then $N = \{N(t) : t \geqslant 0\}$ *is a Poisson process with rate* λ*.*

Proof. See [2, Section 3.6.3]. □

In the preceding theorem, the ξ_j's are often called the interarrival times, T_n is the time of the nth arrival, and $N(t)$ is the number of arrivals by time t.

Lemma 10.24. *Let $\lambda > 0$ and $N \sim \text{Poisson}(\lambda)$. Let $\zeta_1, \zeta_2, \dots$ be i.i.d., $\{0, 1, \dots, k\}$-valued random variables, with $N, \zeta_1, \zeta_2, \dots$ independent. For $0 \leqslant j \leqslant k$, let*

$$N_j = |\{m \leqslant N : \zeta_m = j\}| = \sum_{m=1}^{N} 1_{\{\zeta_m = j\}},$$

so that $N_0 + N_1 + \cdots + N_k = N$. Then $N_0, N_1, \dots, N_k$ are independent and $N_j \sim \text{Poisson}(\lambda p_j)$, where $p_j = P(\zeta_1 = j)$.

Proof. Exercise 10.3. □

Example 10.25 (compound Poisson process)**.** Let $\lambda > 0$ and let N be a Poisson process with rate λ. Let $\zeta = \{\zeta_j\}_{j=1}^{\infty}$ be an i.i.d. sequence of random variables. Assume N and ζ are independent. Let

$$X(t) = \sum_{m=1}^{N(t)} \zeta_m.$$

Then $X = \{X(t) : t \geqslant 0\}$ is called a **compound Poisson process**.

Suppose $\zeta_1 \in \{0, 1, \dots, k\}$ a.s. and $p_j = P(\zeta_1 = j)$. For $0 \leqslant j \leqslant k$, define

$$N^j(t) = \sum_{m=1}^{N(t)} 1_{\{\zeta_m = j\}}.$$

Then $N^0, N^1, \dots, N^k$ are independent Poisson processes and

$$X(t) = \sum_{j=1}^{k} j N^j(t),$$

for all $t \geqslant 0$ (check, use Lemma 10.24).

Example 10.26 (a Poisson process on a measure space)**.** Let $(S, \mathcal{S})$ be a measurable space. Recall that $M(S)$ is the set of all σ-finite measures on $(S, \mathcal{S})$, and that $\mathcal{M}(S)$ is the smallest σ-algebra on $M(S)$ such that $\nu \mapsto \nu(A)$ is measurable for all $A \in \mathcal{S}$. Also recall that a random measure on S is an $M(S)$-valued random variable.

Let $\alpha \in M(S)$. A **Poisson process on S with mean measure α** is a random measure μ such that, for all disjoint $A_1, \dots, A_n \in \mathcal{S}$ with $\alpha(A_j) < \infty$, we have that

$$\mu(A_1), \dots, \mu(A_n)$$

are independent with $\mu(A_j) \sim \text{Poisson}(\alpha(A_j))$.

Suppose $\alpha(S) < \infty$. Let $N \sim \text{Poisson}(\alpha(S))$. Let $\zeta_1, \zeta_2, \dots$ be i.i.d. S-valued random variables with $\zeta_1 \sim \alpha/\alpha(S)$. Suppose $N, \zeta_1, \zeta_2, \dots$ are independent. Define

$$\mu = \sum_{m=1}^{N} \delta_{\zeta_m}.$$

Then μ is a Poisson process on S with mean measure α (check, use Lemma 10.24).

Suppose $\alpha(S) = \infty$. Let $\{S_i\}_{i=1}^\infty$ be a disjoint sequence in $\mathcal{S}$ with $\bigcup_{i=1}^\infty S_i = S$ and $\alpha(S_i) < \infty$. Let $\mathcal{S}_i = \{A \in \mathcal{S} : A \subset S_i\}$ and $\alpha_i = \alpha|_{\mathcal{S}_i}$. Let $\mu_1, \mu_2, \dots$ be independent, where μ_i is a Poisson process on S_i with mean measure α_i. For $A \in \mathcal{S}$ and $\omega \in \Omega$, define

$$\mu(A, \omega) = \sum_{i=1}^{\infty} \mu_i(A \cap S_i, \omega).$$

Then μ is a Poisson process on S with mean measure α (check).

Finally, let $\lambda > 0$ and let N^+ and N^- be independent Poisson processes with rate λ. Let $\widetilde{N}^-(t) = N^-(t-)$. For $t \in \mathbb{R}$, define

$$N(t) = N^+(t)1_{(0,\infty)}(t) - \widetilde{N}^-(-t)1_{(-\infty,0)}(t).$$

The process $N = \{N(t) : t \in \mathbb{R}\}$ is sometimes called a two-sided Poisson process. Fix $\omega \in \Omega$. Note that $t \mapsto N(t, \omega)$ is an increasing, right-continuous function from $\mathbb{R}$ to $\mathbb{R}$. Let $\mu(\omega)$ be the Lebesgue-Stieltjes measure on $\mathbb{R}$ associated with $N(\cdot, \omega)$. That is, $\mu((a, b]) = N(b) - N(a)$. Then μ is a Poisson process on $\mathbb{R}$ with mean measure λm, where m is Lebesgue measure (check).

Exercises

10.3. [2, Exercise 3.6.12] Prove Lemma 10.24.

10.4. [2, Exercise 3.6.3] Let $\{Y_m : m \in \mathbb{N}\}$ be i.i.d. and uniform on $\{1, \dots, n\}$. For $k \leqslant n$, let

$$\tau_k^n = \inf\{m : |\{Y_1, \dots, Y_m\}| = k\}.$$

Recall that $\tau_1^n = 1$ and, for $2 \leqslant k \leqslant n$, the random variables

$$\tau_k^n - \tau_{k-1}^n \sim \text{Geom}(p_k)$$

are independent. Here, $p_k = 1 - (k-1)/n$. Suppose $k(n)/n^{1/2} \to \lambda \in (0, \infty)$. Prove that $\tau_{k(n)}^n - k(n) \Rightarrow N$, where $N \sim \text{Poisson}(\lambda^2/2)$.

10.5. [2, Exercise 3.6.5] Let T be a $(0, \infty)$-valued random variable. Suppose that

$$P(T > t + s \mid T > t) = P(T > s),$$

for all $s, t > 0$. Prove that there exists $\lambda > 0$ such that $T \sim \text{Exp}(\lambda)$.

10.3 Limit theorems in $\mathbb{R}^d$

This section corresponds to [2, Section 3.9].

Let $X = (X_1, \ldots, X_d)^T$ be a random vector. In this section, we shall be careful about indicating whether or vectors are row vectors or column vectors, since we will be making use of matrix multiplication. In general, all vectors will be column vectors unless stated otherwise. The **distribution function** of X is the function $F_X : \mathbb{R}^d \to \mathbb{R}$ given by

$$F_X(x) = P(X \leqslant x) = P(X_1 \leqslant x_1, \ldots, X_d \leqslant x_d).$$

It is a consequence of [2, Theorem 3.9.1] that $X_n \Rightarrow X$ if and only if $F_{X_n}(x) \to F_X(x)$ whenever F_X is continuous at x.

A sequence, $\{\mu_n\}$, of Borel probability measures on $\mathbb{R}^d$ is **tight** if, for any $\varepsilon > 0$, there exists an $M > 0$ such that

$$\liminf_{n\to\infty} \mu_n([-M, M]^d) \geqslant 1 - \varepsilon.$$

A sequence, $\{X_n\}$, of random vectors is **tight** if their corresponding distributions are tight. Theorem 7.27 is still valid in $\mathbb{R}^d$, so that $\{X_n\}$ is tight if an only if every subsequence has a further subsequence that converges in distribution. (See [2, Theorem 3.9.2] for the proof.)

The **characteristic function (ch.f.)** of a random (row) vector X in $\mathbb{R}^d$ is the function $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ given by $\varphi_X(t) = E[e^{i\langle t, X\rangle}]$, where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product on $\mathbb{R}^d$, or dot product. Theorem 7.44 is still valid in $\mathbb{R}^d$, so that convergence in distribution is equivalent to pointwise convergence of characteristic functions. (See [2, Theorem 3.9.4] for the proof.)

Theorem 10.27 (Cramér-Wold device). *If $\langle \theta, X_n \rangle \Rightarrow \langle \theta, X_\infty \rangle$ for all fixed $\theta \in \mathbb{R}^d$, then $X_n \Rightarrow X_\infty$.*

Proof. Let $\varphi_n = \varphi_{X_n}$. Fix $t \in \mathbb{R}^d$. Since $\langle t, X_n \rangle \Rightarrow \langle t, X_\infty \rangle$ and $x \mapsto e^{ix}$ is bounded and continuous, it follows that $E[e^{i\langle t, X_n\rangle}] \to E[e^{i\langle t, X_\infty\rangle}]$. That is, $\varphi_n(t) \to \varphi_\infty(t)$. □

Given (row) vectors $x, y \in \mathbb{R}^d$, note that $x^T y = \langle x, y \rangle$ and xy^T is the $d \times d$ matrix whose ijth entry is $x_i y_j$. The **mean** of a random (row) vector X in $\mathbb{R}^d$ is $E[X] = \mu$, where $\mu_j = EX_j$. The **covariance** of X is

$$\Gamma = \Gamma^X = E[(X - \mu)(X - \mu)^T] = (\Gamma_{ij}),$$

where $\Gamma_{ij} = \operatorname{cov}(X_i, X_j)$. The matrix Γ is symmetric. It is also nonnegative definite, that is, $\langle x, \Gamma x \rangle \geqslant 0$ for all $x \in \mathbb{R}^d$. To see this, note that

$$\begin{aligned}\langle x, \Gamma x \rangle = x^T \Gamma x = x^T E[(X-\mu)(X-\mu)^T]x &= E[x^T (X-\mu)(X-\mu)^T x] \\ &= E[\langle (X-\mu)^T x, (X-\mu)^T x \rangle] = E[\|(X-\mu)^T x\|^2] \geqslant 0,\end{aligned}$$

and this shows that Γ is nonnegative definite.

A random vector X is said to have a **multivariate (or joint) normal (or Gaussian) distribution** if there exists $\mu \in \mathbb{R}^d$ and $\Gamma \in \mathbb{R}^{d\times d}$ such that

$$\varphi_X(t) = \exp\left(i\langle t,\mu\rangle - \frac{1}{2}\langle t,\Gamma t\rangle.\right)$$

In this case, it necessarily follows that μ and Γ are the mean and covariance of X, respectively, and we write $X \sim N(\mu,\Gamma)$. In other words, the distribution of a jointly normal random variable is determined by its mean and covariance.

Let us adopt the convention that the real-valued random variable which is identically zero has a normal distribution, so that in $d = 1$, the multivariate normal distribution and the (previously defined) normal distribution are the same. In general, if $X \sim N(\mu,\Gamma)$ and Γ is invertible, then we say the distribution of X is **nondegenerate**, and otherwise it is **degenerate**.

Theorem 10.28. *The random vector X is jointly Gaussian if and only if every linear combination of $X_1,\ldots,X_d$ is Gaussian.*

Proof. Exercise 10.6. □

Let $Z_1,\ldots,Z_d$ be i.i.d. standard normals, and let $Z = (Z_1,\ldots,Z_d)$. By Theorems 6.24 and 10.28, it follows that Z is jointly normal. Since $E = 0$ and $\Gamma^Z = I$, we have $Z \sim N(0,I)$. This distribution is called the **standard normal distribution in $\mathbb{R}^d$**.

Let $\mu \in \mathbb{R}^d$ and let $\Gamma \in \mathbb{R}^{d\times d}$ be symmetric and nonnegative definite. Then there exists an orthogonal matrix U (that is, $U^TU = I$) and a diagonal matrix V with $V_{ij} \geqslant 0$ for all i,j such that $\Gamma = U^TVU$. Choose diagonal W such that $W^2 = V$ and let $A = WU$. Then $A^TA = \Gamma$.

Suppose $Z \sim N(0,I)$, and let $X = \mu + A^TZ$. For any $\theta \in \mathbb{R}^d$, the random variable $\theta^TX = \theta^T\mu + (A\theta)^TZ$ is normal, since $(A\theta)^TZ$ is a linear combination of $Z_1,\ldots,Z_d$. Thus, X is jointly normal. Moreover, $EX = \mu$ and the covariance of X is

$$\Gamma^X = E[(X-\mu)(X-\mu)^T] = E[A^TZZ^TA] = A^TA = \Gamma.$$

Thus, $X \sim N(\mu,\Gamma)$.

The preceding idea is used to prove the following.

Theorem 10.29. *Let $X \sim N(\mu,\Gamma)$. Then X has a density if and only if its distribution is nondegenerate, and in this case, its density is*

$$f_X(x) = \frac{1}{(2\pi)^{d/2}(\det\Gamma)^{1/2}}\exp\left(-\frac{1}{2}\left\langle x-\mu,\Gamma^{-1}(x-\mu)\right\rangle.\right)$$

Proof. See [6, Corollary 16.2]. □

Theorem 10.30 (central limit theorem in $\mathbb{R}^d$)**.** *Let $\xi_1,\xi_2,\ldots$ be i.i.d. random vectors in $\mathbb{R}^d$ with mean μ and covariance Γ. Let $X_n = \xi_1+\cdots+\xi_n$. Then*

$$n^{-1/2}(X_n - n\mu) \Rightarrow Z_\Gamma,$$

where $Z \sim N(0,\Gamma)$.

Proof. Let $Y_n = n^{-1/2}(X_n - n\mu)$ and $\theta \in \mathbb{R}^d$. By the central limit theorem, $\theta^T Y_n \Rightarrow Z_\theta \sim N(0, \sigma_\theta^2)$, where

$$\sigma_\theta^2 = \mathrm{var}(\theta^T \xi_1) = E[\theta^T(\xi_1 - \mu)(\xi_1 - \mu)^T \theta] = \theta^T \Gamma \theta.$$

Note that $\theta^T Z_\Gamma$ is normal with mean zero and variance

$$E[\theta^T Z_\Gamma Z_\Gamma^T \theta] = \theta^T \Gamma \theta = \sigma_\theta^2.$$

Thus, $\theta^T Y_n \Rightarrow \theta^T Z_\Gamma$. It follows that $Y_n \Rightarrow Z_\Gamma$. □

Exercises

10.6. [2, Exercise 3.9.8] Prove Theorem 10.28.

10.7. [2, Exercise 3.9.4] For $n \in \mathbb{N} \cup \{\infty\}$, let $X_n = (X_n(1), \ldots, X_n(d))$ be a random vector in $\mathbb{R}^d$. Assume that $X_n \Rightarrow X_\infty$. Prove that $X_n(j) \Rightarrow X_\infty(j)$ as $n \to \infty$ for all j.

10.8. [2, Exercise 3.9.7] Let $(X_1, \ldots, X_d)$ have a multivariate Gaussian distribution. Prove that $X_1, \ldots, X_d$ are independent if and only if they are uncorrelated.

Chapter 11

Further Properties of Random Walks

11.1 Stopping times

This section corresponds to [2, Section 4.1].

Let us recall some of the definitions from 6.4. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(S, \mathcal{S})$ a measurable space, and for each $n \in \mathbb{N}$, let $X_n : \Omega \to S$ be $(\mathcal{F}, \mathcal{S})$-measurable. In other words, $X = \{X_n\}$ is an indexed collection of S-valued random variables defined on a common probability space. That is, X is a stochastic process. Since X is indexed by $\mathbb{N}$, it is a discrete-time stochastic process.

A filtration is a sequence $\{\mathcal{F}_n\}$ of σ-algebras such that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all n. We say that X is adapted to the filtration if $X_n \in \mathcal{F}_n$ for all n. The filtration generated by X is $\{\mathcal{F}_n^X\}$, where $\mathcal{F}_n^X = \sigma(X_1, \ldots, X_n)$. The process X is adapted to $\{\mathcal{F}_n^X\}$. Moreover, if X is adapted to $\{\mathcal{F}_n\}$, then $\mathcal{F}_n^X \subset \mathcal{F}_n$ for all n.

A **random time**, N, is an $\mathbb{N} \cup \{\infty\}$-valued random variable. If $X = \{X_n : n \in \mathbb{N} \cup \{\infty\}\}$ is a stochastic process indexed by $\mathbb{N} \cup \{\infty\}$ and N is a random time, then X_N is a well-defined random variable given by $(X_N)(\omega) = X_{N(\omega)}(\omega)$. On the other hand, if $X = \{X_n : n \geqslant 1\}$ is indexed by $\mathbb{N}$, then the most we can say is that

$$X_N : \{N < \infty\} \to S$$

is $(\mathcal{F}|_{\{N<\infty\}}, \mathcal{S})$-measurable. If $N < \infty$ a.s., then after modification on a null set, X_N is again a well-defined random variable.

In the case that X is indexed by $\mathbb{N}$ and $P(N = \infty) > 0$, the notation $\sigma(X_N)$ technically refers to a σ-algebra on $\{N < \infty\}$. But we will abuse notation and, in this case, define

$$\sigma(X_N) = \{\{X_N \in B\} : B \in \mathcal{S}\} \cup \{\{X_N \in B\} \cup \{N = \infty\} : B \in \mathcal{S}\}.$$

The reader should verify that, in this case, $\sigma(X_N)$ is a sub-σ-algebra of $\mathcal{F}$.

If X is indexed by $\mathbb{N}$ and N represents the time at which something happens with the process X, then the event $\{N = \infty\}$ typically indicates that the event never occurs.

A **stopping time with respect to the filtration** $\{\mathcal{F}_n\}$ is a random time such that $\{N = n\} \in \mathcal{F}_n$ for all n. Note that constant random variables are stopping time. A **stopping time for a stochastic process** X is a stopping time with respect to $\mathcal{F}_n^X$. If N is a stopping time for X and X is adapted to $\{\mathcal{F}_n\}$, then N is a stopping time with respect to $\{\mathcal{F}_n\}$.

The **hitting time of** $A \in \mathcal{S}$ is $N = \inf\{n \in \mathbb{N} : X_n \in A\}$. Since

$$\{N = n\} = \left(\bigcap_{j=1}^{n=1} \{X_j \notin A\}\right) \cap \{X_n \in A\} \in \mathcal{F}_n^X,$$

it follows that N is a stopping time for X.

Intuitively, the defining condition of a stopping time, $\{N = n\} \in \mathcal{F}_n$, means that the occurrence or nonoccurrence of the event $\{N = n\}$ can be determined from the information at time n. That is, you can always tell whether or not the random time N has occurred (giving you the *option* to act in that moment or to *stop* some ongoing procedure).

Proposition 11.1. *Let S and T be stopping times with respect to a filtration $\{\mathcal{F}_n\}$, and let $n \in \mathbb{N}$. Then $S \wedge T$, $S \vee T$, $S \wedge n$, and $S \vee n$ are stopping times.*

Proof. Exercise 11.4. □

Given a stopping time N with respect to a filtration $\{\mathcal{F}_n\}$, we define

$$\mathcal{F}_N = \{A \in \mathcal{F} : A \cap \{N = n\} \in \mathcal{F}_n \text{ for all } n\}.$$

The set $\mathcal{F}_N$ is a σ-algebra (check). The σ-algebra $\mathcal{F}_N$ is interpreted as the information known at time N.

Proposition 11.2. *If M and N are stopping times with $M(\omega) \leqslant N(\omega)$ for all $\omega \in \Omega$, then $\mathcal{F}_M \subset \mathcal{F}_N$.*

Proof. Exercise 11.7. □

Proposition 11.3. *Let M and N be stopping times with $M(\omega) \leqslant N(\omega)$ for all $\omega \in \Omega$. If $A \in \mathcal{F}_M$ and $T = M1_A + N1_{A^c}$, then T is a stopping time.*

Proof. Exercise 11.8. □

In the remainder of this section, we assume $(S, \mathcal{S})$ is a standard Borel space, let $\xi_1, \xi_2, \ldots$ be i.i.d. S-valued random variables, and consider the process $\xi = \{\xi_j\}$.

We will frequently consider the special case of a random walk X on $\mathbb{R}^d$, which is the case where $S = \mathbb{R}^d$, $\mathcal{S} = \mathcal{R}^d$, $X_0 = 0$, $X_n = \xi_1 + \ldots + \xi_n$, and $X = \{X_n\}$.

If $S = \mathbb{R}^d$ and $X_n = \xi_1 + \cdots + \xi_n$, then $\{X_n\}$ is a random walk. Note that in this case $\mathcal{F}_n^X = \mathcal{F}_n^\xi$.

We also assume that the sequence $\{\xi_j\}$ has been constructed in the canonical way, using Kolmogorov's extension theorem. That is, if $\xi_1 \sim \mu$, then

$$\Omega = S^{\mathbb{N}} = \prod_{j=1}^{\infty} S, \qquad \mathcal{F} = \bigotimes_{j=1}^{\infty} \mathcal{S}, \qquad P = \prod_{j=1}^{\infty} \mu,$$

and $\xi_j(\omega) = \omega_j$ is the projection onto the jth component.

A **finite permutation** of $\mathbb{N}$ is a surjection $\pi : \mathbb{N} \to \mathbb{N}$ such that $\pi(j) \neq j$ for only finitely many j. A finite permutation is necessarily bijective. A finite permutation π induces a map $\pi_\Omega : \Omega \to \Omega$, where $(\pi_\Omega \omega)_j = \omega_{\pi(j)}$. Note that $\xi_j \circ \pi_\Omega = \xi_{\pi(j)}$.

An event $A \in \mathcal{F}$ is **permutable** if $\pi_\Omega^{-1} A = A$ for all finite permutations π. Let

$$\mathcal{E} = \{A \in \mathcal{F} : A \text{ is permutable}\}.$$

Then $\mathcal{E}$ is a σ-algebra (check), which is called the **exchangeable** σ-algebra.

Lemma 11.4. *Let X be a random walk on $\mathbb{R}$. Then*

(i) $\{X_n \in B \text{ i.o.}\} \in \mathcal{E}$,

(ii) $\left\{ \limsup_{n\to\infty} \frac{X_n}{c_n} \geqslant 1 \right\} \in \mathcal{E}$, *and*

(iii) $\mathcal{T} \subset \mathcal{E}$, *where* $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\xi_n, \xi_{n+1}, \ldots)$.

Proof. Exercise 11.1. □

Theorem 11.5 (Hewitt-Savage 0-1 law)**.** *If $A \in \mathcal{E}$, then $P(A) \in \{0,1\}$.*

Proof. See [2, Theorem 4.1.1]. The idea of the proof is to show that A is independent of itself, so that $P(A) = P(A \cap A) = P(A)^2$. □

Theorem 11.6. *Let X be a random walk on $\mathbb{R}$. Then exactly one of the following is true.*

(i) $X_n = 0$ *a.s. for all* n.

(ii) $X_n \to \infty$ *a.s.*

(iii) $X_n \to -\infty$ *a.s.*

(iv) $\liminf_{n\to\infty} X_n = -\infty$ *a.s. and* $\limsup_{n\to\infty} X_n = \infty$ *a.s.*

Proof. Let $X = \limsup_{n\to\infty} X_n$. Then $\{X \leqslant x\} \in \mathcal{E}$ for all $x \in \mathbb{R}$. By the 0-1 law, $P(X \leqslant x) \in \{0,1\}$ for all $x \in \mathbb{R}$. Let $c = \inf\{x \in \mathbb{R} : P(X \leqslant x) = 1\}$. Then $X = c$ a.s. (check).

Let $X_n' = X_{n+1} - \xi_1$. Since $\{X_n\}$ and $\{X_n'\}$ have the same distribution, it follows that

$$c = \limsup_{n\to\infty} X_n \overset{d}{=} \limsup_{n\to\infty} X_n' = c - \xi_1,$$

from which it follows that $c = c - \xi_1$ a.s.

Suppose (i) is false. Then it is not the case that $\xi_1 = 0$ a.s. Thus, $c = c - \xi_1$ a.s. implies that $c \in \{-\infty, \infty\}$. In other words, one of the following is true:

(a) $\limsup_{n\to\infty} X_n = -\infty$ a.s.

(b) $\limsup_{n\to\infty} X_n = \infty$ a.s.

A similar argument shows that one of the following is true:

(a′) $\liminf_{n\to\infty} X_n = -\infty$ a.s.

(b′) $\liminf_{n\to\infty} X_n = \infty$ a.s.

Since (a) and (b′) is impossible, there are three possibilities: (a) and (a′), which is (iii); (b) and (a′), which is (iv); and (b) and (b′), which is (ii). □

A **symmetric random walk on** $\mathbb{R}$ is a random walk on $\mathbb{R}$ in which $\xi_1 =_d -\xi_1$ and $P(\xi_1 = 0) < 1$. A **simple random walk on** $\mathbb{R}$ is a symmetric random walk on $\mathbb{R}$ with $\xi_1 \in \{-1, 1\}$ a.s. An **asymmetric simple random walk** is a random walk with $P(\xi_1 = 1) = p$ and $P(\xi_1 = -1) = 1 - p$, where $p \neq 1/2$. In this context, q is typically defined as $q := 1 - p$, so that $p \neq 1/2$ is equivalent to $p \neq q$.

The following proposition implies that a simple random walk visits every integer infinitely many times.

Proposition 11.7. *If X is a symmetric random walk on $\mathbb{R}$, then Theorem 11.6(iv) holds.*

Proof. Exercise 11.2. □

Theorem 11.8. *Let N be a stopping time for the stochastic process $\{\xi_j : j \in \mathbb{N}\}$ with $P(N < \infty) > 0$. Conditional on $\{N < \infty\}$, the process $\{\xi_{N+j} : j \in \mathbb{N}\}$ is independent of $\mathcal{F}_N$ and has the same distribution as $\{\xi_j : j \in \mathbb{N}\}$.*

More specifically, if $n \in \mathbb{N}$, $A_j \in \mathcal{S}$, and $A \in \mathcal{F}_N$, then

$$\begin{aligned} &P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \cap A \mid N < \infty) \\ &\qquad = P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \mid N < \infty) P(A \mid N < \infty), \end{aligned}$$

and

$$\begin{aligned} &P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \mid N < \infty) \\ &\qquad\qquad = P(\{\xi_1 \in A_1\} \cap \ldots \cap \{\xi_n \in A_n\}). \end{aligned}$$

Proof. Since $\Omega \in \mathcal{F}_N$ and $\xi_1, \xi_2, \ldots$ are i.i.d. with distribution μ, it suffices to prove

$$P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \cap A \mid N < \infty) = P(A \mid N < \infty) \prod_{j=1}^{n} \mu(A_j).$$

Multiplying both sides by $P(N < \infty)$, this is equivalent to proving

$$P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \cap A \cap \{N < \infty\}) = P(A \cap \{N < \infty\}) \prod_{j=1}^{n} \mu(A_j).$$

Fix $k \in \mathbb{N}$. Then

$$P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \cap A \cap \{N = k\}) = P(\{\xi_{k+1} \in A_1\} \cap \ldots \cap \{\xi_{k+n} \in A_n\} \cap A \cap \{N = k\}).$$

Since $A \in \mathcal{F}_N$, it follows that

$$A \cap \{N = k\} \in \mathcal{F}_k^\xi = \sigma(\xi_1, \ldots, \xi_k).$$

Thus, $\xi_{k+1}, \ldots, \xi_{k+n}$ and $A \cap \{N = k\}$ are independent, and we have

$$\begin{aligned} P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \cap A \cap \{N = k\}) \\ = P(\{\xi_{k+1} \in A_1\} \cap \ldots \cap \{\xi_{k+n} \in A_n\}) P(A \cap \{N = k\}) \\ = P(A \cap \{N = k\}) \prod_{j=1}^{n} \mu(A_j). \end{aligned}$$

Finally, summing over k, we obtain

$$\begin{aligned} P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \cap A \cap \{N < \infty\}) \\ = \sum_{k=1}^{\infty} P(\{\xi_{N+1} \in A_1\} \cap \ldots \cap \{\xi_{N+n} \in A_n\} \cap A \cap \{N = k\}) \\ = \sum_{k=1}^{\infty} P(A \cap \{N = k\}) \prod_{j=1}^{n} \mu(A_j) \\ = P(A \cap \{N < \infty\}) \prod_{j=1}^{n} \mu(A_j), \end{aligned}$$

and we are done. □

Recall that $\Omega = S^{\mathbb{N}}$. We define the **shift operator** $\theta : \Omega \to \Omega$ by $(\theta\omega)_n = \omega_{n+1}$. Let θ^0 be the identity, $\theta^1 = \theta$, and $\theta^{k+1} = \theta \circ \theta^k$ for $k \in \mathbb{N}$. That is, $(\theta^k \omega)_n = \omega_{n+k}$

Let Δ be any element such that $\Delta \notin \Omega$. (Such a Δ is sometimes referred to as a "cemetery point".) Given a stopping time N, we define $\theta^N : \Omega \to \Omega \cup \{\Delta\}$ by

$$\theta^N \omega = \left(\sum_{n=1}^{\infty} (\theta^n \omega) 1_{\{N=n\}}(\omega) \right) + \Delta 1_{\{N=\infty\}}(\omega).$$

(Although addition and multiplication are not necessarily defined for the objects $x \in \Omega \cup \{\Delta\}$, we shall assume that $x1 = x$ and $x + 0 = 0 + x = x$, so that the above definition makes sense.)

Example 11.9. Let $S = \mathbb{R}^d$ and $X_n = \xi_1 + \cdots + \xi_n$, so that $X = \{X_n : n \in \mathbb{N}\}$ is a random walk. Let $\tau = \inf\{n \in \mathbb{N} : X_n = 0\}$ be the hitting time of $\{0\}$. If we let $X_0 = 0$, then τ is the first time that X returns to the origin. As noted earlier, τ is a stopping time for the process X.

Note that $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$. We will extend the domain of τ to $\Omega \cup \{\Delta\}$ by setting $\tau(\Delta) = \infty$.

Let $\tau_2 = \tau + \tau \circ \theta^\tau$. Suppose $\tau(\omega) = \infty$. Then

$$\tau_2(\omega) = \tau(\omega) + \tau(\theta^\tau \omega) = \tau(\omega) + \tau(\Delta) = \infty.$$

Suppose $\tau(\omega) = m \in \mathbb{N}$. Then

$$\begin{aligned} \tau_2(\omega) &= m + \tau(\theta^m \omega) \\ &= m + \inf\{n \in \mathbb{N} : X_n(\theta^m \omega) = 0\} \\ &= m + \inf\{n \in \mathbb{N} : (\theta^m \omega)_1 + \cdots + (\theta^m \omega)_n = 0\} \\ &= m + \inf\{n \in \mathbb{N} : \omega_{m+1} + \cdots + \omega_{m+n} = 0\} \\ &= m + \inf\{n \in \mathbb{N} : X_{m+n} - X_m = 0\} \\ &= m + \inf\{n \in \mathbb{N} : X_{m+n} = 0\} \\ &= \inf\{k > m : X_k = 0\}. \end{aligned}$$

In other words, τ_2 is the time of the second return to the origin. In general, if $\tau_{n+1} = \tau_n + \tau \circ \theta^{\tau_n}$, then τ_n is the time of the nth return to the origin.

Proposition 11.10. *Let T be a stopping time for the process $\{\xi_j\}$. Let $T_0 = 0$ and $T_n = T_{n-1} + T \circ \theta^{T_{n-1}}$ for $n \in \mathbb{N}$. Then $P(T_n < \infty) = P(T < \infty)^n$.*

Proof. Since $T_0 = 0$ and θ^0 is the identity, the result follows for $n = 1$. Suppose it is true for some n. By Theorem 11.8, conditional on $\{T_n < \infty\}$, the process $\{\xi_{T_n+1}, \xi_{T_n+2}, \ldots\}$ is independent of $\mathcal{F}_{T_n}$ and has the same distribution as $\{\xi_1, \xi_2, \ldots\}$. Note that on $\{T_n < \infty\}$, we have

$$\begin{aligned} T \circ \theta^{T_n}(\omega) &= T(\omega_{T_n(\omega)+1}, \omega_{T_n(\omega)+2}, \ldots) \\ &= T(\xi_{T_n(\omega)+1}(\omega), \xi_{T_n(\omega)+2}(\omega), \ldots). \end{aligned}$$

Suppressing the ω's, we have

$$T \circ \theta^{T_n} = T(\xi_{T_n+1}, \xi_{T_n+2}, \ldots).$$

Thus,

$$\begin{aligned}
P(T_{n+1} < \infty) &= P(T_n < \infty, T \circ \theta^{T_n} < \infty) \\
&= P(T_n < \infty) P(T \circ \theta^{T_n} < \infty \mid T_n < \infty) \\
&= P(T_n < \infty) P(T(\xi_{T_n+1}, \xi_{T_n+2}, \ldots) < \infty \mid T_n < \infty) \\
&= P(T_n < \infty) P(T(\xi_1, \xi_2, \ldots) < \infty) \\
&= P(T_n < \infty) P(T < \infty) \\
&= P(T < \infty)^{n+1},
\end{aligned}$$

where the last equality comes from the inductive hypothesis. □

Theorem 11.11 (Wald's equation)**.** *Let $X = \{X_n\}$ be a random walk on $\mathbb{R}$ with $E|\xi_1| < \infty$. Let N be a stopping time for X with $EN < \infty$. Then $EX_N = (EN)(E\xi_1)$.*

Proof. First note that $EN < \infty$ implies $N < \infty$ a.s. In particular, this implies that X_N is well-defined.

Next, we observe that

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} E[|\xi_m| 1_{\{N=n\}}] 1_{\{n \geqslant m\}} = \sum_{m=1}^{\infty} E[|\xi_m| 1_{\{N \geqslant m\}}].$$

Since N is a stopping time, we have $\{N \geqslant m\} = \{N \leqslant m-1\}^c \in \mathcal{F}_{m-1}$. Also, ξ_m is independent of $\mathcal{F}_{m-1}$. Thus,

$$\begin{aligned}
\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} E[|\xi_m| 1_{\{N=n\}}] 1_{\{n \geqslant m\}} &= \sum_{m=1}^{\infty} E|\xi_m| P(N \geqslant m) \\
&= \sum_{m=1}^{\infty} E|\xi_1| P(N \geqslant m) \\
&= (E|\xi_1|)(EN) < \infty.
\end{aligned}$$

By Tonelli's theorem, $\xi_m 1_{\{N=n\}} 1_{\{n \geqslant m\}}$ is integrable on $\mathbb{N}^2 \times \Omega$ with respect to $\mu \times P$, where μ is counting measure on $\mathbb{N}^2$. Therefore, by Fubini's theorem,

$$\begin{aligned}
EX_N = E \sum_{n=1}^{\infty} X_n 1_{\{N=n\}} &= E \sum_{n=1}^{\infty} \sum_{m=1}^{n} \xi_m 1_{\{N=n\}} \\
= E \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \xi_m 1_{\{N=n\}} 1_{\{n \geqslant m\}} &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} E[\xi_m 1_{\{N=n\}}] 1_{\{n \geqslant m\}} \\
&= (E\xi_1)(EN),
\end{aligned}$$

where the last equality is calculated as above. □

Example 11.12. Let $X = \{X_n\}$ be a simple random walk on $\mathbb{R}$. Let $a, b \in \mathbb{Z}$ with $a < 0 < b$. Let $N = \inf\{n : X_n \notin (a, b)\}$. Since N is the hitting time of $\{a, b\}$, it follows that N is a stopping time. We will first show that $EN < \infty$.

First note that for any $x \in (a,b)$, we have

$$P(x + X_{b-a} \notin (a,b)) \geqslant P(\xi_1 = 1, \xi_2 = 1, \ldots, \xi_{b-a} = 1) = 2^{-(b-a)}. \qquad (11.1)$$

We will prove by induction that, for any $n \in \mathbb{N}$, we have

$$P(X_{b-a} \in (a,b), X_{2(b-a)} \in (a,b), \ldots, X_{n(b-a)} \in (a,b)) \leqslant (1 - 2^{-(b-a)})^n. \quad (11.2)$$

For $n = 1$, taking $x = 0$ in (11.1) gives (11.2). Now assume (11.2) is true for some n. Then

$$\begin{aligned} &P(X_{b-a} \in (a,b), \ldots, X_{(n+1)(b-a)} \in (a,b)) \\ &\quad = E[P(X_{b-a} \in (a,b), \ldots, X_{(n+1)(b-a)} \in (a,b) \mid \mathcal{F}_{n(b-a)})] \\ &\quad = E[1_{\{X_{b-a} \in (a,b)\}} \cdots 1_{\{X_{n(b-a)} \in (a,b)\}} P(X_{(n+1)(b-a)} \in (a,b) \mid \mathcal{F}_{n(b-a)})]. \end{aligned}$$

In general, if $m > n$, then using Theorem 6.52 (or more precisely, its more general form, Theorem 6.66), we have

$$\begin{aligned} E[f(X_m) \mid \mathcal{F}_n] &= E[f(X_n + \xi_{n+1} + \cdots + \xi_m) \mid \mathcal{F}_n] \\ &= E[f(x + \xi_{n+1} + \cdots + \xi_m)]|_{x = X_n} = E[f(x + X_{m-n})]|_{x = X_n} \end{aligned}$$

Thus, by (11.1) and the inductive hypothesis,

$$\begin{aligned} &P(X_{b-a} \in (a,b), \ldots, X_{(n+1)(b-a)} \in (a,b)) \\ &\quad = E[1_{\{X_{b-a} \in (a,b)\}} \cdots 1_{\{X_{n(b-a)} \in (a,b)\}} P(x + X_{b-a} \in (a,b))|_{x = X_{n(b-a)}}] \\ &\quad \leqslant P(X_{b-a} \in (a,b), \ldots, X_{n(b-a)} \in (a,b))(1 - 2^{-(b-a)}) \\ &\quad \leqslant (1 - 2^{-(b-a)})^{n+1}, \end{aligned}$$

and this proves (11.2).

It now follows that

$$P(N > n(b-a)) \leqslant P(X_{b-a} \in (a,b), \ldots, X_{n(b-a)} \in (a,b)) \leqslant (1 - 2^{-(b-a)})^n.$$

From this we have

$$P(N = \infty) = \lim_{n \to \infty} P(N > n(b-a)) = 0.$$

Thus, since $N < \infty$ a.s., we have

$$\begin{aligned} EN &= \sum_{m=0}^{\infty} P(N > m) = \sum_{n=0}^{\infty} \sum_{m=n(b-a)}^{(n+1)(b-a)-1} P(N > m) \\ &\leqslant (b-a) \sum_{n=0}^{\infty} P(N > n(b-a)) \leqslant (b-a)\left(1 + \sum_{n=1}^{\infty} (1 - 2^{-(b-a)})^n\right) < \infty, \end{aligned}$$

and we can therefore apply Wald's equation.

By Wald's equation, we have $EX_N = (EN)(E\xi_1) = 0$. On the other hand, $X_N \in \{a, b\}$, so

$$\begin{aligned} 0 = EX_N &= aP(X_N = a) + bP(X_N = b) \\ &= a(1 - P(X_N = b)) + bP(X_N = b) \\ &= a + (b - a)P(X_N = b), \end{aligned}$$

from which we obtain

$$P(X_N = b) = -\frac{a}{b-a} = \frac{|a|}{b-a}$$

and

$$P(X_N = a) = 1 + \frac{a}{b-a} = \frac{b}{b-a}.$$

For $x \in \mathbb{Z}$, let $T_x = \inf\{n \in \mathbb{N} : X_n = x\}$ be the hitting time of $\{x\}$, so that $N = T_a \wedge T_b$. Then $\{X_N = a\} = \{T_a < T_b\}$. Also, $\{T_a < T_b\} \uparrow \{T_a < \infty\}$ as $b \uparrow \infty$. Thus,

$$P(T_a < \infty) = \lim_{b \to \infty} P(T_a < T_b) = \lim_{b \to \infty} \frac{b}{b-a} = 1,$$

so that $T_a < \infty$ a.s. In particular, this implies that X_{T_a} is well-defined, and of course $X_{T_a} = a$ a.s.

Finally, we claim that $ET_a = \infty$. To see this, suppose that $ET_a < \infty$. Then by Wald's equation,

$$a = EX_{T_a} = (ET_a)(E\xi_1) = 0,$$

a contradiction.

Theorem 11.13 (Wald's second equation)**.** *Let $X = \{X_n\}$ be a random walk on $\mathbb{R}$ with $E\xi_1 = 0$ and $E\xi_1^2 = \sigma^2 < \infty$. Let N be a stopping time for X with $EN < \infty$. Then $EX_N^2 = (EN)\sigma^2$.*

Proof. Fix $m \in \mathbb{N} \cup \{0\}$. Define $Y = \{Y_n\}_{n=m}^\infty$ by $Y_n = X_{N \wedge n} - X_{N \wedge m}$. Let $n \geqslant m + 1$. If $N < n$, then $N \wedge n = N \wedge (n - 1) = N$, so that $Y_n = Y_{n-1}$. On the other hand, if $N \geqslant n > m$, then $Y_n = X_n - X_m$ and $Y_{n-1} = X_{n-1} - X_m$. Thus,

$$\begin{aligned} Y_n^2 &= Y_{n-1}^2 + ((X_n - X_m)^2 - (X_{n-1} - X_m)^2)1_{\{N \geqslant n\}} \\ &= Y_{n-1}^2 + ((X_{n-1} - X_m + \xi_n)^2 - (X_{n-1} - X_m)^2)1_{\{N \geqslant n\}} \\ &= Y_{n-1}^2 + (2\xi_n(X_{n-1} - X_m) + \xi_n^2)1_{\{N \geqslant n\}}. \end{aligned}$$

Since ξ_n and $X_{n-1} - X_m$ are both square-integrable, it follows that $(2\xi_n(X_{n-1} - X_m) + \xi_n^2)1_{\{N \geqslant n\}} \in L^1(\Omega)$, and so we may take its conditional expectation. Since $\{N \geqslant n\} = \{N \leqslant n - 1\}^c \in \mathcal{F}_{n-1}$ and $X_{n-1} - X_m \in \mathcal{F}_{n-1}$ and ξ_n is independent of $\mathcal{F}_{n-1}$, we have

$$\begin{aligned} &E[(2\xi_n(X_{n-1} - X_m) + \xi_n^2)1_{\{N \geqslant n\}} \mid \mathcal{F}_{n-1}] \\ &\qquad = 1_{\{N \geqslant n\}}(2(X_{n-1} - X_m)E[\xi_n] + E[\xi_n^2]) = \sigma^2 1_{\{N \geqslant n\}}, \end{aligned}$$

which gives

$$EY_n^2 = EY_{n-1}^2 + \sigma^2 P(N \geqslant n).$$

Since $Y_m = 0$, it follows by induction on n that for every $n \geqslant m+1$, we have

$$EY_n^2 = \sigma^2 \sum_{k=m+1}^{n} P(N \geqslant k).$$

In other words,

$$\|X_{N \wedge n} - X_{N \wedge m}\|_{L^2(\Omega)}^2 = \sigma^2 \sum_{k=m+1}^{n} P(N \geqslant k) \leqslant \sigma^2 P(N > m) \to 0,$$

as $n, m \to \infty$. Hence, $\{X_{N \wedge n}\}_{n=0}^{\infty}$ is a Cauchy sequence in $L^2(\Omega)$, and so there exists $Z \in L^2(\Omega)$ such that $X_{N \wedge n} \to Z$ in L^2 as $n \to \infty$. It follows that there is a subsequence with $X_{N \wedge n(j)} \to Z$ a.s. as $j \to \infty$. But $X_{N \wedge n} \to X_N$ a.s. Therefore, $Z = X_N$, and so $X_{N \wedge n} \to X_N$ in L^2. Consequently,

$$EX_N^2 = \lim_{n \to \infty} EX_{N \wedge n}^2 = \lim_{n \to \infty} \sigma^2 \sum_{k=1}^{n} P(N \geqslant k) = \sigma^2 \sum_{k=1}^{\infty} P(N \geqslant k) = \sigma^2 EN.$$

□

Example 11.14. Let $X = \{X_n\}$ be a simple random walk on $\mathbb{R}$, let $a, b \in \mathbb{Z}$ with $a < 0 < b$, and let $N = \inf\{n : X_n \notin (a, b)\}$. In Example 11.12, we showed that $EN < \infty$. Thus, by Theorem 11.13, we have $EX_N^2 = \sigma^2 EN$, where $\sigma^2 = E\xi_1^2 = 1$. Thus,

$$\begin{aligned} EN = EX_N^2 &= a^2 P(X_N = a) + b^2 P(X_N = b) \\ &= a^2 \frac{b}{b-a} + b^2 \frac{|a|}{b-a} = |a|b. \end{aligned}$$

In particular, if $a = -L$ and $b = L$, then $EN = L^2$.

Exercises

11.1. Prove Lemma 11.4.

11.2. [2, Exercise 4.1.1] Prove Proposition 11.7.

11.3. [2, Exercise 4.1.2] Let $\{X_n\}$ be a random walk on $\mathbb{R}$ with $E\xi_1 = 0$ and $\operatorname{var}(\xi_1) \in (0, \infty)$. Prove that Theorem 11.6(iv) holds.

11.4. [2, Exercise 4.1.3] Prove Proposition 11.1.

11.5. [2, Exercise 4.1.4] Let S and T be stopping times with respect to a filtration $\{\mathcal{F}_n\}$. Prove or disprove: $S + T$ is a stopping time.

11.6. [2, Exercise 4.1.5] Suppose $Y = \{Y_n\}$ is adapted to a filtration $\{\mathcal{F}_n\}$ and N is an $\{\mathcal{F}_n\}$ stopping time. Prove that $Y_N 1_{\{N<\infty\}} \in \mathcal{F}_N$.

11.7. [2, Exercise 4.1.6] Prove Proposition 11.2.

11.8. [2, Exercise 4.1.7] Prove Proposition 11.3.

11.2 Recurrence

This section corresponds to [2, Section 4.2].

Throughout this section, $X = \{X_n\}$ is a random walk in $\mathbb{R}^d$. The point $x \in \mathbb{R}^d$ is a **recurrent value for** X if, for all $\varepsilon > 0$, we have

$$P(\|X_n - x\| < \varepsilon \text{ i.o.}) = 1,$$

where $\|x\| = \max\{|x_1|, \ldots, |x_d|\}$. Note that by the Hewitt-Savage 0-1 law, for any $x \in \mathbb{R}^d$, we have $P(\|X_n - x\| < \varepsilon \text{ i.o.}) \in \{0, 1\}$.

A point $x \in \mathbb{R}^d$ is a **possible value of** X if, for all $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $P(\|X_n - x\| < \varepsilon) > 0$.

Let $\mathcal{V}$ be the set of recurrent values and $\mathcal{U}$ the set of possible values.

Theorem 11.15. *The set $\mathcal{V}$ is either empty or a closed subgroup of $(\mathbb{R}^d, +)$ (that is, closed under addition and additive inverses). In the latter case, $\mathcal{V} = \mathcal{U}$.*

Proof. See [2, Theorem 4.2.1]. □

If $\mathcal{V} = \varnothing$, then X is **transient**. Otherwise, X is **recurrent**. As in Example 11.9, let τ_n be the nth return to 0.

Theorem 11.16. *The following are equivalent:*

(i) $\tau_1 < \infty$ *a.s.,*

(ii) $P(X_n = 0 \text{ i.o.}) = 1$*, and*

(iii) $\sum_{m=0}^{\infty} P(X_m = 0) = \infty$.

Proof. By Proposition 11.10, we have that (i) implies $\tau_n < \infty$ a.s. for all n, which is equivalent to (ii). Let

$$N = \sum_{m=0}^{\infty} 1_{\{X_m = 0\}} = \sum_{n=0}^{\infty} 1_{\{\tau_n < \infty\}},$$

so that

$$EN = \sum_{m=0}^{\infty} P(X_m = 0) = \sum_{n=0}^{\infty} P(\tau_n < \infty),$$

which shows that (ii) implies (iii). Moreover, if $P(\tau_1 < \infty) < 1$, then

$$\sum_{m=0}^{\infty} P(X_m = 0) = \sum_{n=0}^{\infty} P(\tau_1 < \infty)^n < \infty,$$

so by contraposition, (iii) implies (i). □

Lemma 11.17. *For all $n \in \mathbb{N}$, we have*

$$\sum_{m=0}^{n} \binom{n}{m} \binom{n}{n-m} = \binom{2n}{n}.$$

Proof. Heuristically, consider an urn containing n black marbles and n white marbles. There are $\binom{2n}{n}$ ways to select n marbles from the urn. One way to make this selection is to first select an integer $0 \leqslant m \leqslant n$, and then select m black balls and $n-m$ white balls.

For an analytic proof, consider that $(x+1)^{2n} = (x+1)^n(x+1)^n$ implies

$$\sum_{j=0}^{2n} \binom{2n}{j} x^j = \left(\sum_{j=0}^{n} \binom{n}{j} x^j \right)^2 = \sum_{i=0}^{n} \sum_{j=0}^{n} \binom{n}{i} \binom{n}{j} x^{i+j}.$$

Equating the coefficient of x^n, we have

$$\binom{2n}{n} = \sum_{\substack{0 \leqslant i,j \leqslant n \\ i+j=n}} \binom{n}{i} \binom{n}{j} = \sum_{m=0}^{n} \binom{n}{m} \binom{n}{n-m},$$

which proves the identity. □

Theorem 11.18. *A simple random walk in $\mathbb{R}^d$ is recurrent if $d \leqslant 2$ and transient if $d \geqslant 3$.*

Proof. In this proof, let $\{e_1, \ldots, e_d\}$ be the standard basis in $\mathbb{R}^d$.

First assume $d = 1$. For $n \in \mathbb{N}$, we have $P(X_{2n-1} = 0) = 0$ and

$$P(X_{2n} = 0) = \binom{2n}{n} 2^{-2n} = \frac{(2n)! 2^{-2n}}{(n!)^2}.$$

By Stirling's formula,

$$n! \sim n^n e^{-n} \sqrt{2\pi n},$$

as $n \to \infty$. Thus,

$$P(X_{2n} = 0) \sim \frac{(2n)^{2n} e^{-2n} \sqrt{4\pi n}\, 2^{-2n}}{n^{2n} e^{-2n} (2\pi n)} = \frac{1}{\sqrt{\pi n}},$$

giving

$$\sum_{m=0}^{\infty} P(X_m = 0) = 1 + \sum_{n=1}^{\infty} P(X_{2n} = 0) \geqslant 1 + C \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty.$$

By Theorem 11.16, we have that 0 is a recurrent value for X, and so X is recurrent.

Now suppose $d = 2$. Again, for all $n \in \mathbb{N}$, we have $P(X_{2n-1} = 0) = 0$. Fix $n \in \mathbb{N}$. Let

$$M = |\{1 \leqslant j \leqslant 2n : \xi_j = e_1\}|,$$

and note that

$$\{X_{2n} = 0\} = \biguplus_{m=0}^{n} \{X_{2n} = 0\} \cap \{M = m\}.$$

Thus,

$$\begin{aligned}P(X_{2n}=0) &= \sum_{m=0}^{n} P(X_{2n}=0, M=m)\\ &= \sum_{m=0}^{n} \frac{(2n)!}{m!m!(n-m)!(n-m)!}4^{-2n}\\ &= \sum_{m=0}^{n} \frac{(2n)!}{(n!)^2}\left(\frac{n!}{m!(n-m)!}\right)^2 4^{-2n}\\ &= 4^{-2n}\binom{2n}{n}\sum_{m=0}^{n}\binom{n}{m}\binom{n}{n-m}.\end{aligned}$$

By Lemma 11.17,

$$P(X_{2n}=0) = 4^{-2n}\binom{2n}{n}^2.$$

Again, by Stirling's formula, this gives

$$P(X_{2n}=0) \sim \frac{1}{\pi n},$$

which implies

$$\sum_{m=0}^{\infty} P(X_m=0) = 1 + \sum_{n=1}^{\infty} P(X_{2n}=0) \geqslant 1 + C\sum_{n=1}^{\infty}\frac{1}{\pi n} = \infty.$$

And again, by Theorem 11.16, we have that 0 is a recurrent value for X, and so X is recurrent.

Next, suppose $d = 3$. Again, for all $n \in \mathbb{N}$, we have $P(X_{2n-1} = 0) = 0$. As in the case $d = 2$, if

$$\begin{aligned}J &= |\{1 \leqslant j \leqslant 2n : \xi_j = e_1\}|,\\ K &= |\{1 \leqslant k \leqslant 2n : \xi_k = e_2\}|,\end{aligned}$$

then

$$\begin{aligned}P(X_{2n}=0) &= \sum_{\substack{0\leqslant j,k\leqslant n\\ j+k\leqslant n}} P(X_{2n}=0, J=j, K=k)\\ &= \sum_{\substack{0\leqslant j,k\leqslant n\\ j+k\leqslant n}} \frac{(2n)!}{j!j!k!k!(n-j-k)!(n-j-k)!}6^{-2n}\\ &= 2^{-2n}\binom{2n}{n}\sum_{\substack{0\leqslant j,k\leqslant n\\ j+k\leqslant n}}\left(\frac{n!}{j!k!(n-j-k)!}3^{-n}\right)^2\\ &\leqslant C_n 2^{-2n}\binom{2n}{n}\sum_{\substack{0\leqslant j,k\leqslant n\\ j+k\leqslant n}}\frac{n!}{j!k!(n-j-k)!}3^{-n},\end{aligned}$$

where

$$C_n = \max_{\substack{0 \leqslant j,k \leqslant n \\ j+k \leqslant n}} \frac{n!}{j!k!(n-j-k)!} 3^{-n}.$$

Let

$$\widetilde{J} = |\{0 \leqslant j \leqslant n : \xi_j = \pm e_1\}|,$$
$$\widetilde{K} = |\{0 \leqslant k \leqslant n : \xi_k = \pm e_2\}|,$$

and note that

$$\sum_{\substack{0 \leqslant j,k \leqslant n \\ j+k \leqslant n}} \frac{n!}{j!k!(n-j-k)!} 3^{-n} = \sum_{\substack{0 \leqslant j,k \leqslant n \\ j+k \leqslant n}} P(\widetilde{J} = j, \widetilde{K} = k) = 1,$$

so that

$$P(X_{2n} = 0) \leqslant C_n 2^{-2n} \binom{2n}{n}.$$

Lastly, it can be shown that $C_n = O(n^{-1})$ (see the proof of [2, Theorem 4.2.3] for details), so that $P(X_{2n} = 0) = O(n^{-3/2})$, and this implies $\sum_{m=0}^{\infty} P(X_m = 0) < \infty$. By Theorem 11.16, we have $P(X_m = 0 \text{ i.o.}) < 1$, so that 0 is not a recurrent value for X. By Theorem 11.15, we have $\mathcal{V} = \varnothing$, and so X is transient.

Finally, assume $d > 3$. As described in the proof of [2, Theorem 4.2.3], a 3-dimensional simple random walk can be created from the first 3 coordinates of X, and the transience of this embedded random walk implies the transience of X. □

The remainder of [2, Section 4.2] is devoted to showing that Theorem 11.18 is still true, in some sense, for random walks that are not necessarily simple. We state the relevant theorems here and refer the reader to the text for more details.

Theorem 11.19 (Chung-Fuchs theorem)**.** *Let X be a random walk on $\mathbb{R}$. If $n^{-1}X_n \to 0$ in probability as $n \to \infty$, then X is recurrent.*

Theorem 11.20. *Let X be a random walk in $\mathbb{R}^2$. If $n^{-1/2}X_n$ converges in distribution to a nonconstant, normally distributed $\mathbb{R}^2$-valued random variable as $n \to \infty$, then X is recurrent.*

Theorem 11.21. *Let X be a random walk in $\mathbb{R}^3$. Suppose that for all $\theta \in \mathbb{R}^3$, we have $P(\langle \xi_1, \theta \rangle \neq 0) > 0$. Then X is transient.*

Part IV

Discrete-time Stochastic Processes II: Martingales and Markov Chains

Chapter 12

Martingales

12.1 Definitions and basic properties

This section corresponds to [2, Section 5.2].

Recall from Section 6.4 that a real-valued stochastic process $X = \{X_n\}$ is a martingale with respect a filtration $\{\mathcal{F}_n\}$ if X is adapted to $\{\mathcal{F}_n\}$, each X_n is integrable, and X satisfies the martingale property:

$$E[X_{n+1} \mid \mathcal{F}_n] = X_n,$$

for all n. More generally, X is a **supermartingale** if X is adapted, integrable, and $E[X_{n+1} \mid \mathcal{F}_n] \leqslant X_n$. And X is a **submartingale** if X is adapted, integrable, and $E[X_{n+1} \mid \mathcal{F}_n] \geqslant X_n$.

If X represents your changing wealth as you play a game, then X is a martingale if the game is fair, X is a supermartingale if the game is weighted against you, and X is a submartingale if the game is weighted in your favor.

If we say that X is a martingale, supermartingale, or submartingale without reference to a filtration, then the implied filtration is $\{\mathcal{F}_n^X\}$, the filtration generated by X.

Suppose X is a martingale with respect to $\{\mathcal{F}_n\}$ and $\mathcal{F}_n^X \subset \mathcal{G}_n \subset \mathcal{F}_n$. Then

$$E[X_{n+1} \mid \mathcal{G}_n] = E[E[X_{n+1} \mid \mathcal{F}_n] \mid \mathcal{G}_n] = E[X_n \mid \mathcal{G}_n] = X_n,$$

and it follows that X is a martingale with respect to $\mathcal{G}_n$. The same is true for supermartingales and submartingales.

Example 12.1. Let X be a simple random walk on $\mathbb{R}$. Since X_n is integrable and

$$E[X_{n+1} \mid \mathcal{F}_n] = E[X_n + \xi_{n+1} \mid \mathcal{F}_n] = X_n + E[\xi_{n+1}] = X_n,$$

it follows that X is martingale.

Theorem 12.2. *Let X be adapted to $\{\mathcal{F}_n\}$. Then* *(i) If X is a supermartingale with respect to $\{\mathcal{F}_n\}$, then for all $n > m$, we have $E[X_n \mid \mathcal{F}_m] \leqslant X_m$.*

(ii) If X is a submartingale with respect to $\{\mathcal{F}_n\}$, then for all $n > m$, we have $E[X_n \mid \mathcal{F}_m] \geqslant X_m$.

(iii) If X is a martingale with respect to $\{\mathcal{F}_n\}$, then for all $n > m$, we have $E[X_n \mid \mathcal{F}_m] = X_m$.

Proof. To prove (i), we will show that $E[X_{m+k} \mid \mathcal{F}_m] \leqslant X_m$ for all $k \in \mathbb{N}$. By definition, it is true for $k = 1$. Suppose it is true for some k. Then

$$E[X_{m+k+1} \mid \mathcal{F}_m] = E[E[X_{m+k+1} \mid \mathcal{F}_{m+k}] \mid \mathcal{F}_m] \leqslant E[X_{m+k} \mid \mathcal{F}_m] \leqslant X_m,$$

and induction completes the proof.

Note that if $\{X_n\}$ is a submartingale, then $\{-X_n\}$ is a supermartingale. Also note that a martingale is both a supermartingale and a submartingale. Thus, applying (i) to $\{-X_n\}$ proves (ii), and (i) and (ii) together imply (iii). □

Remark 12.3. Many of the upcoming results will be stated only for supermartingales, leaving it to the reader to prove them (when applicable) for submartingales and martingales.

Proposition 12.4. *If X is a supermartingale and $n > m$, then $EX_n \leqslant EX_m$.*

Proof. This follows since $EX_n = E[E[X_n \mid \mathcal{F}_m]] \leqslant E[X_m]$. □

Theorem 12.5. *Let X be a martingale with respect to $\{\mathcal{F}_n\}$ and φ a convex function. Suppose that $E|\varphi(X_n)| < \infty$ for all n. Then $\{\varphi(X_n)\}$ is a submartingale with respect to $\{\mathcal{F}_n\}$.*

Proof. Adaptedness is immediate and integrability is by hypothesis. By Jensen's inequality and the martingale property for X, we have

$$E[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geqslant \varphi(E[X_{n+1} \mid \mathcal{F}_n]) = \varphi(X_n),$$

and $\{\varphi(X_n)\}$ is a submartingale. □

Corollary 12.6. *If X is a martingale, $p \geqslant 1$, and $E|X_n|^p < \infty$ for all n, then $\{|X_n|^p\}$ is a submartingale.*

Proof. This follows since $x \mapsto |x|^p$ is a convex function. □

Theorem 12.7. *Let X be a submartingale with respect to $\{\mathcal{F}_n\}$ and φ an increasing convex function. Suppose that $E|\varphi(X_n)| < \infty$ for all n. Then $\{\varphi(X_n)\}$ is a submartingale with respect to $\{\mathcal{F}_n\}$.*

Proof. Adaptedness is immediate and integrability is by hypothesis. By Jensen's inequality, the submartingale property for X, and the fact that φ is increasing, we have

$$E[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geqslant \varphi(E[X_{n+1} \mid \mathcal{F}_n]) \geqslant \varphi(X_n),$$

and $\{\varphi(X_n)\}$ is a submartingale. □

Corollary 12.8. *Let X be adapted to $\{\mathcal{F}_n\}$ and $a \in \mathbb{R}$. Then*

(i) *If X is a submartingale with respect to $\{\mathcal{F}_n\}$, then $\{(X_n - a)^+\}$ is a submartingale with respect to $\{\mathcal{F}_n\}$.*

(ii) *If X is a supermartingale with respect to $\{\mathcal{F}_n\}$, then $\{X_n \wedge a\}$ is a supermartingale with respect to $\{\mathcal{F}_n\}$.*

Proof. In both cases, measurability is immediate. Integrability follows, since both $|(X_n - a)^+|$ and $|X_n \wedge a|$ are bounded above by $|X_n| + |a|$.

The submartingale property in (i) follows from the fact that $x \mapsto (x - a)^+$ is an increasing, convex function. For (ii), since $\{-X_n\}$ is a submartingale, and $x \mapsto -((-x) \wedge a) = x \vee (-a)$ is an increasing, convex function, it follows that $\{-(X_n \wedge a)\}$ is a submartingale, and so $\{X_n \wedge a\}$ is a supermartingale. □

Let $H = \{H_n : n \in \mathbb{N}\}$ be a stochastic process and $\{\mathcal{F}_n\}_{n=0}^{\infty}$ a filtration. Then H is **predictable with respect to** $\{\mathcal{F}_n\}$ if $H_n \in \mathcal{F}_{n-1}$ for all n. Let $X = \{X_n : n \geqslant 0\}$ be adapted to $\{\mathcal{F}_n\}$ and define the process $H \cdot X = \{(H \cdot X)_n\}$ by

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1}).$$

Note that $H \cdot X$ is adapted to $\{\mathcal{F}_n\}$. Also note that $(H \cdot X)_n$ is a discrete analogue of $\int_0^t H(s)\, dX(s)$.

Suppose we are playing a sequence of gambling games. Let $\xi_n = 1$ if we win the nth game and $\xi_n = -1$ if we lose. Let $X_0 = 0$ and $X_n = \xi_1 + \cdots + \xi_n$. If H_n denotes the amount we plan to wager on the nth game, then $(H \cdot X)_n$ denotes our wealth after the nth game. Note that since H is predictable, the amount we plan to wager on the nth game depends only on the information we have after the $(n-1)$th game, as it should.

For example, suppose $H_1 = 1$ and, for $n \in \mathbb{N}$, we have

$$H_{n+1} = 2H_n 1_{\{\xi_n = -1\}} + 1_{\{\xi_n = 1\}}.$$

Then H is predictable and represents the strategy wherein we double our wager every time we lose. This strategy is a famous gambling system called the "martingale".

Theorem 12.9. *Let $X = \{X_n : n \geqslant 0\}$ be a supermartingale with respect to a filtration $\{\mathcal{F}_n\}$. Let H be predictable with respect to $\{\mathcal{F}_n\}$. Suppose each H_n is nonnegative and bounded, that is, for all $n \in \mathbb{N}$, there exists $C_n > 0$ such that $0 \leqslant H_n \leqslant C_n$ a.s. Then $H \cdot X$ is a supermartingale with respect to $\{\mathcal{F}_n\}$.*

Proof. As noted earlier, $H \cdot X$ is adapted to $\{\mathcal{F}_n\}$. Since each H_n is bounded and each X_n is integrable, it follows that each $(H \cdot X)_n$ is integrable. Finally,

$$\begin{aligned} E[(H \cdot X)_{n+1} \mid \mathcal{F}_n] &= E[(H \cdot X)_n + H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n] \\ &= (H \cdot X)_n + H_{n+1} E[X_{n+1} - X_n \mid \mathcal{F}_n]. \end{aligned}$$

Since $H_{n+1} \geqslant 0$ and

$$E[X_{n+1} - X_n \mid \mathcal{F}_n] = E[X_{n+1} \mid \mathcal{F}_n] - X_n \leqslant 0,$$

it follows that $E[(H \cdot X)_{n+1} \mid \mathcal{F}_n] \leqslant (H \cdot X)_n$, and $H \cdot X$ is a supermartingale. □

Remark 12.10. Theorem 12.9 is also true for submartingales. For martingales, Theorem 12.9 is true without the restriction that each H is nonnegative. Moreover, the assumption that H is bounded is needed only to ensure the integrability of $(H \cdot X)_n$. Assuming the boundedness of X would also suffice.

Theorem 12.11. *Let X be a supermartingale and N a stopping time, both with respect to $\{\mathcal{F}_n\}$. Then $\{X_{N \wedge n}\}$ is a supermartingale with respect to $\{\mathcal{F}_n\}$.*

Proof. Let $H_n = 1_{\{N \geqslant n\}}$. Since $\{N \geqslant n\} = \{N \leqslant n-1\}^c \in \mathcal{F}_{n-1}$, it follows that $H = \{H_n\}$ is nonnegative and predictable, and so $H \cdot X$ is a supermartingale. Note that

$$\begin{aligned}
(H \cdot X)_n &= \sum_{m=1}^{n} H_m (X_m - X_{m-1}) \\
&= \sum_{m=1}^{n} (X_m - X_{m-1}) 1_{\{N \geqslant m\}} \\
&= \sum_{m=1}^{N \wedge n} (X_m - X_{m-1}) \\
&= X_{N \wedge n} - X_0.
\end{aligned}$$

Thus, $X_{N \wedge n} = (H \cdot X)_n + X_0$. Since the random variable X_0 is integrable and $\mathcal{F}_0$-measurable, it follows that the constant process $Y = \{Y_n\}$, where $Y_n = X_0$ for all n, is a martingale (and therefore a supermartingale). Therefore, since the sum of supermartingales is a supermartingale, it follows that $\{X_{N \wedge n}\}$ is a supermartingale. □

Theorem 12.12 (martingale convergence theorem)**.** *Let X be a submartingale with $\sup_n E[X_n^+] < \infty$. Then there exists an integrable random variable X_∞ such that $X_n \to X_\infty$ a.s.*

Proof. Uses “upcrossings”. See [2, Theorem 5.2.8]. □

Remark 12.13. The above martingale convergence theorem is analogous to the law of large numbers, at least insofar as the mode of convergence is almost sure. There is also a martingale central limit. See, for example, [3, Section 7.1].

Theorem 12.14. *Let X be a nonnegative supermartingale (that is, $X_n \geqslant 0$ a.s. for all n). Then there exists an integrable, nonnegative random variable X_∞ such that $X_n \to X_\infty$ a.s. Moreover, $EX_\infty \leqslant EX_0$.*

Proof. Applying the martingale convergence theorem to $\{-X_n\}$ implies there exists an integrable random variable X_∞ such that $X_n \to X_\infty$ a.s. Since each X_n is nonnegative, it follows that X_∞ is nonnegative. Moreover, $EX_n \leqslant EX_0$ for all n, so

$$EX_\infty \leqslant \liminf_{n\to\infty} EX_n \leqslant EX_0,$$

by Fatou's lemma. □

Example 12.15. Let X be a simple random walk on $\mathbb{R}$. Let $Y_n = 1 + X_n$. Let $N = \inf\{n \geqslant 0 : Y_n = 0\}$, so that $\{Y_{N\wedge n}\}$ is a nonnegative martingale (and hence, a nonnegative supermartingale). Therefore, there exists an integrable, nonnegative random variable Y such that $Y_{N\wedge n} \to Y$ a.s. But, as shown previously, $N < \infty$ a.s., which implies $Y_{N\wedge n} \to Y_N = 0$ a.s. Therefore, $Y = 0$.

Since $\{Y_{N\wedge n}\}$ is a martingale, we have $EY_{N\wedge n} = EY_{N\wedge 0} = EY_0 = 1$ for all n, whereas $EY = 0$. It follows that $Y_{N\wedge n}$ does not converge to Y in $L^1(\Omega)$.

Theorem 12.16 (Doob's decomposition)**.** *Let $X = \{X_n : n \geqslant 0\}$ be a submartingale with respect to a filtration $\{\mathcal{F}_n\}$. Then there exist processes $M = \{M_n\}$ and $A = \{A_n\}$ such that*

(i) $X_n = M_n + A_n$ *for all* n,

(ii) $M = \{M_n\}$ *is a martingale with respect to* $\{\mathcal{F}_n\}$,

(iii) $A = \{A_n\}$ *is predictable with respect to* $\{\mathcal{F}_n\}$,

(iv) A *is increasing, that is,* $A_n \leqslant A_{n+1}$ *a.s. for all* n, *and*

(v) $A_0 = 0$ *a.s.*

Moreover, if M' and A' are another pair of processes satisfying (i)-(v), then $M' = M$ and $A' = A$.

Proof. Define

$$A_n = \sum_{m=1}^{n} (E[X_m \mid \mathcal{F}_{m-1}] - X_{m-1})$$

and $M_n = X_n - A_n$. Then (i), (iii), and (v) are immediate. By the submartingale property,

$$A_{n+1} - A_n = E[X_{n+1} \mid \mathcal{F}_n] - X_n \geqslant 0,$$

and this gives (iv). Lastly,

$$\begin{aligned} E[M_{n+1} \mid \mathcal{F}_n] &= E[X_{n+1} \mid \mathcal{F}_n] - A_{n+1} \\ &= E[X_{n+1} \mid \mathcal{F}_n] - (A_n + E[X_{n+1} \mid \mathcal{F}_n] - X_n) \\ &= X_n - A_n \\ &= M_n, \end{aligned}$$

which gives (ii).

Now suppose M' and A' are another pair of processes satisfying (i)-(v). Then

$$\begin{aligned} E[X_n \mid \mathcal{F}_{n-1}] &= E[M'_n \mid \mathcal{F}_{n-1}] + E[A'_n \mid \mathcal{F}_{n-1}] \\ &= M'_{n-1} + A'_n \\ &= X_{n-1} - A'_{n-1} + A'_n, \end{aligned}$$

which implies $A'_n - A'_{n-1} = E[X_n \mid \mathcal{F}_{n-1}] - X_{n-1}$. Since $A'_0 = 0$, it follows by induction that $A' = A$, and therefore $M' = X - A' = X - A = M$. □

Exercises

12.1. [2, Exercise 5.2.3] Give an example of a submartingale $X = \{X_n\}$ such that $\{X_n^2\}$ is a supermartingale.

12.2. [2, Exercise 5.2.4] Give an example of a martingale $X = \{X_n\}$ such that $X_n \to -\infty$ a.s.

12.3. [2, Exercise 5.2.5] Let $\{\mathcal{F}_n\}$ be a filtration and let $B_n \in \mathcal{F}_n$ for each n. Define $X = \{X_n\}$ by $X_n = \sum_{m=0}^n 1_{B_m}$. Prove that X is an $\{\mathcal{F}_n\}$-submartingale and identify the Doob decomposition for X.

12.4. [2, Exercise 5.2.11] Let $\{\mathcal{F}_n\}$ be a filtration. Let $X = \{X_n\}$ and $Y = \{Y_n\}$ be integrable, positive, and adapted to $\{\mathcal{F}_n\}$. Assume that

$$E[X_{n+1} \mid \mathcal{F}_n] \leqslant (1 + Y_n)X_n$$

for all n. Also assume that $\sum_n Y_n < \infty$ a.s. Prove that there exists a real-valued random variable, X_∞, such that $X_n \to X_\infty$ a.s.

12.5. [2, Exercise 5.2.6] Let $\xi_1, \xi_2, \dots$ be independent with $E\xi_j = 0$ and $\sigma_j^2 := \operatorname{var}(\xi_j) < \infty$. Define $X_n = \xi_1 + \cdots + \xi_n$ and $s_n^2 = \sum_{j=1}^n \sigma_j^2$. Prove that $\{X_n^2 - s_n^2\}$ is a martingale.

12.2 Branching processes

This section corresponds to [2, Section 5.3.4].

Let $\{\xi_{n,j} : j, n \in \mathbb{N}\}$ be i.i.d., $\mathbb{N} \cup \{0\}$-valued random variables. Define the stochastic process $Z = \{Z_n : n \geqslant 0\}$ by $Z_0 = 1$ and

$$Z_n = \sum_{j=1}^{Z_{n-1}} \xi_{n,j} = \sum_{j=1}^{\infty} \xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}}.$$

The process Z is called the **Galton-Watson process**, or **Galton-Watson branching process**.

The Galton-Watson process can be used to model the size of a population evolving in discrete-time. In this case, Z_n represent the size of the population (in number of individuals) in the nth generation. The jth individual in the $(n-1)$th generation has $\xi_{n,j}$ offspring and then dies (i.e., is not part of the nth generation). Thus, the number of individuals in the nth generation is $\xi_{n,1} + \cdots + \xi_{n,Z_{n-1}}$, unless $Z_{n-1} = 0$, in which case $Z_n = 0$ also.

The random variable $\xi_{n,j}$ represents the number of offspring that the jth member of the $(n-1)$th generation contributes to the nth generation. Its distribution, which does not depend on j or n, is called the **offspring distribution** of the branching process Z, and is denoted by $p_k = P(\xi_{1,1} = k)$ for $k \in \mathbb{N} \cup \{0\}$.

Let us assume that $\xi_{1,1}$ is integrable and not identically zero. Let $\mu = E\xi_{1,1} \in (0, \infty)$ be the mean number of offspring per individual per generation. Let $\mathcal{F}_0 = \{\varnothing, \Omega\}$ and, for $n \in \mathbb{N}$, let $\mathcal{F}_n = \sigma(\{\xi_{m,j} : j \in \mathbb{N}, 1 \leqslant m \leqslant n\})$. Define $M = \{M_n : n \geqslant 0\}$ by $M_n = \mu^{-n} Z_n$.

Lemma 12.17. *The process M is an $\{\mathcal{F}_n\}$-martingale.*

Proof. Since $Z_0 = 1$ is constant, we have $Z_0 \in \mathcal{F}_0$.

Suppose $Z_{n-1} \in \mathcal{F}_{n-1}$ for some $n \in \mathbb{N}$. Then $1_{\{Z_{n-1} \geqslant j\}} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$ and $\xi_{n,j} \in \mathcal{F}_n$, for all $j \in \mathbb{N}$. Thus, $\xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}} \in \mathcal{F}_n$, and it follows that $Z_n \in \mathcal{F}_n$.

By induction, Z is adapted to $\{\mathcal{F}_n\}$, and consequently, M is adapted to $\mathcal{F}_n$.

Since
$$0 \leqslant \xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}} \leqslant \xi_{n,j},$$
it follows that $\xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}}$ is integrable, and so we have

$$\begin{aligned} E[\xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}}] &= E[E[\xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}} \mid \mathcal{F}_{n-1}]] \\ &= E[1_{\{Z_{n-1} \geqslant j\}} E[\xi_{n,j}]] = \mu P(Z_{n-1} \geqslant j). \end{aligned}$$

Thus,
$$EZ_n = \sum_{n=1}^{\infty} \mu P(Z_{n-1} \geqslant j) = \mu E Z_{n-1}.$$

Since $EZ_0 = 1$, we have $EZ_n = \mu^n$ for all n. In particular, each Z_n is integrable, and hence, each M_n in integrable.

Finally, as above, we have

$$\begin{aligned} E[M_n \mid \mathcal{F}_{n-1}] &= \mu^{-n} E[Z_n \mid \mathcal{F}_{n-1}] \\ &= \mu^{-n} E\left[\sum_{j=1}^{\infty} \xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}} \,\middle|\, \mathcal{F}_{n-1} \right] \\ &= \mu^{-n} \sum_{j=1}^{\infty} E[\xi_{n,j} 1_{\{Z_{n-1} \geqslant j\}} \mid \mathcal{F}_{n-1}] \\ &= \mu^{-n+1} \sum_{j=1}^{\infty} 1_{\{Z_{n-1} \geqslant j\}} \\ &= \mu^{-n+1} Z_{n-1} \\ &= M_{n-1}, \end{aligned}$$

where we have used the monotone convergence theorem for conditional expectations to reverse the sum and expectation. This shows that M is an $\{\mathcal{F}_n\}$-martingale. ☐

Since M is a nonnegative martingale, it follows from Theorem 12.14 that there exists an integrable, nonnegative random variable Y with $EY \leqslant EM_0 = 1$ and $M_n \to Y$ a.s.

Theorem 12.18. *If $\mu < 1$, then $Y = 0$ a.s., so that $Z_n = o(\mu^n)$. In fact, $P(Z_n > 0$ i.o.$) = 0$.*

Proof. Since $Z_n \geqslant 1$ on $\{Z_n > 0\}$, we have

$$P(Z_n > 0) = E[1_{\{Z_n>0\}}] \leqslant E[Z_n 1_{\{Z_n>0\}}] = EZ_n = \mu^n.$$

Thus,

$$\sum_{n=0}^{\infty} P(Z_n > 0) \leqslant \sum_{n=0}^{\infty} \mu^n < \infty.$$

By the Borel-Cantelli lemma, $P(Z_n > 0 \text{ i.o.}) = 0$. ☐

Theorem 12.19. *If $\mu = 1$ and $p_1 = P(\xi_{1,1} = 1) < 1$, then $Y = 0$ a.s., so that $Z_n = o(\mu^n)$. In fact, $P(Z_n > 0$ i.o.$) = 0$.*

Proof. Since $p_1 < 1$, we may choose $\ell \in \mathbb{N} \cup \{0\}$ such that $\ell \neq 1$ and $p_\ell = P(\xi_{1,1} = \ell) > 0$. Using Theorem 6.66, for any $k \in \mathbb{N}$, we have $P(Z_n \neq k \mid \mathcal{F}_{n-1}) = h_n(Z_{n-1})$, where

$$h_n(i) = P\left(\sum_{j=1}^{i} \xi_{n,j} \neq k\right).$$

Note that

$$h_n(k) = P\left(\sum_{j=1}^{k} \xi_{n,j} \neq k\right) \geqslant P\left(\bigcap_{j=1}^{k} \{\xi_{n,j} = \ell\}\right) = p_\ell^k.$$

We will now prove that for $N \leqslant n$ and $k \in \mathbb{N}$,

$$P(Z_N = Z_{N+1} = \cdots = Z_n = k) \leqslant P(Z_N = k)(1 - p_\ell^k)^{n-N}. \tag{12.1}$$

The result is trivial if $n = N$. Suppose it is true for some $n \geqslant N$. Then

$$\begin{aligned} P(Z_N =& Z_{N+1} = \cdots = Z_n = Z_{n+1} = k) \\ &= E[P(Z_N = Z_{N+1} = \cdots = Z_n = k, Z_{n+1} = k \mid \mathcal{F}_n)] \\ &= E[1_{\{Z_N = Z_{N+1} = \cdots = Z_n = k\}} P(Z_{n+1} = k \mid \mathcal{F}_n)] \\ &= E[1_{\{Z_N = Z_{N+1} = \cdots = Z_n = k\}} (1 - h_{n+1}(Z_n))] \\ &= (1 - h_{n+1}(k)) P(Z_N = Z_{N+1} = \cdots = Z_n = k) \\ &\leqslant (1 - p_\ell^k) P(Z_N = Z_{N+1} = \cdots = Z_n = k) \\ &\leqslant P(Z_N = k)(1 - p_\ell^k)^{n+1-N}, \end{aligned}$$

and (12.1) follows by induction.

Now, since $\mu = 1$, it follows that $Z_n \to Y$ a.s. But Z_n is $\mathbb{N} \cup \{0\}$-valued, so $Z_n \to Y$ a.s. implies that for P-a.e. ω, there exists $N(\omega)$ such that for all $n \geqslant N(\omega)$, we have $Z_n(\omega) = Y(\omega)$. Therefore, Y is $\mathbb{N} \cup \{0\}$-valued,

$$P\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{Z_n = Y\}\right) = 1,$$

and it will suffice to show that $Y = 0$ a.s.

Let $k \in \mathbb{N}$. Then

$$\begin{aligned} P(Y = k) &= P\left(\{Y = k\} \cap \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{Z_n = Y\}\right) \\ &= P\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{Y = k, Z_n = Y\}\right) \\ &= P\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{Y = k, Z_n = k\}\right) \\ &\leqslant P\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{Z_n = k\}\right) \\ &\leqslant \sum_{N=1}^{\infty} P\left(\bigcap_{n=N}^{\infty} \{Z_n = k\}\right). \end{aligned}$$

Using (12.1), we obtain

$$\begin{aligned} P\left(\bigcap_{n=N}^{\infty} \{Z_n = k\}\right) &= \lim_{n\to\infty} P(Z_N = Z_{N+1} = \cdots = Z_n = k) \\ &\leqslant \lim_{n\to\infty} P(Z_N = k)(1 - p_\ell^k)^{n-N} = 0. \end{aligned}$$

Thus, $P(Y = k) = 0$ for all $k \in \mathbb{N}$, and so $Y = 0$ a.s. □

For $s \in [0,1]$, let $\varphi(s) = E[s^{\xi_{1,1}}] = \sum_{k=0}^{\infty} p_k s^k$. The function φ is called the **generating function** of $\xi_{1,1}$ (or the generating function of the offspring distribution).

Theorem 12.20. *Let $\mu > 1$. Then φ has a unique fixed point $\rho \in [0,1)$ which satisfies $P(Z_n = 0$ for some $n) = \rho$.*

Proof. Note that

$$\varphi'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1},$$

and

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2}.$$

Also,

$$\varphi'(1) = \lim_{s \to 1^-} \varphi'(s) = \sum_{k=1}^{\infty} kp_k = E[\xi_{1,1}] = \mu.$$

Let $\psi(x) = \varphi(x) - x$, so that $\psi(0) = \varphi(0) = p_0 \geqslant 0$ and $\psi(1) = 0$. Since $\psi'(1) = \mu - 1 > 0$, it follows that for $\varepsilon > 0$ sufficiently small, we have $\psi(1-\varepsilon) < 0$. Thus, by the intermediate value theorem, there exists $\rho \in [0, 1-\varepsilon)$ such that $\psi(\rho) = 0$.

Suppose there exists $0 \leqslant \rho_1 < \rho_2 < 1$ such that $\psi(\rho_1) = \psi(\rho_2) = 0$. Since $\psi(1) = 0$ also, the mean value theorem implies that there exists $x_1 \in (\rho_1, \rho_2)$ and $x_2 \in (\rho_2, 1)$ such that $\psi'(x_1) = \psi'(x_2) = 0$. But $\psi'' = \varphi''$ is strictly positive, so ψ' is strictly increasing, which is a contradiction. Hence, there exists a unique $\rho \in [0, 1)$ such that $\psi(\rho) = 0$, that is φ has a unique fixed point $\rho \in [0, 1)$.

Note that $\varphi' > 0$, so φ is strictly increasing. In particular, if $x \leqslant \rho$, then $\varphi(x) \leqslant \varphi(\rho) = \rho$.

Now let $A_n = \{Z_n = 0\}$. Note that $A_n \subset A_{n+1}$ and

$$\bigcup A_n = \{Z_n = 0 \text{ for some } n\}.$$

For each $\ell \in \mathbb{N}$, let $Z^{(\ell)}$ be a branching process with the same offspring distribution as Z, constructed so that $Z^{(1)}, Z^{(2)}, \ldots$ are independent, and independent of $Z_1 = \xi_{1,1}$. Define $\widetilde{Z}_0 = 1$ and

$$\widetilde{Z}_n = \sum_{\ell=1}^{Z_1} Z_{n-1}^{(\ell)},$$

for $n \in \mathbb{N}$. It can be shown that $\widetilde{Z} = \{\widetilde{Z}_n\}$ and Z have the same distribution. Note that $\widetilde{Z}_1 = Z_1$. Thus, for $n, k \in \mathbb{N}$, we have

$$P(\widetilde{Z}_n = 0 \mid Z_1 = k) = P(Z_{n-1}^{(1)} = 0, \ldots, Z_{n-1}^{(k)} = 0) = P(Z_{n-1} = 0)^k.$$

Note that this also holds for $k = 0$. Hence, for $n \in \mathbb{N}$,

$$\begin{aligned}
P(A_n) &= P(\widetilde{Z}_n = 0) \\
&= \sum_{k=0}^{\infty} P(Z_1 = k) P(\widetilde{Z}_n = 0 \mid Z_1 = k) \\
&= \sum_{k=0}^{\infty} p_k P(A_{n-1})^k \\
&= \varphi(P(A_{n-1})).
\end{aligned}$$

Since φ is strictly increasing, it follows that $\{P(A_n)\}_{n=1}^{\infty}$ is a strictly increasing sequence. Also, since $x \leqslant \rho$ implies $\varphi(x) \leqslant \rho$, and $P(A_0) = 0 \leqslant \rho$, it follows that $P(A_n) \leqslant \rho$ for all n. Thus, there exists $\widetilde{\rho} \leqslant \rho$ such that $P(A_n) \uparrow \widetilde{\rho}$ as $n \to \infty$. Finally,

$$P(Z_n = 0 \text{ for some } n) = P\left(\bigcup_{n=0}^{\infty} A_n\right) = \lim_{n \to \infty} P(A_n) = \widetilde{\rho}.$$

On the other hand, since φ is continuous,

$$P(Z_n = 0 \text{ for some } n) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \varphi(P(A_{n-1})) = \varphi(\widetilde{\rho}).$$

Therefore $\widetilde{\rho} = \varphi(\widetilde{\rho})$. Since $\widetilde{\rho} \leqslant \rho < 1$, it follows that $\widetilde{\rho} = \rho$. □

Exercises

12.6. [2, Exercise 5.3.12] Let $Z = \{Z_n\}$ be a branching process whose offspring distribution has mean $\mu > 1$. Let ρ be the unique fixed point in $[0, 1)$ of the generating function, $\varphi(s) = E[s^{Z_1}]$. Prove that

$$P\left(\lim_{n\to\infty} \mu^{-n} Z_n = 0\right) \in \{\rho, 1\}.$$

Show that if the probability is ρ, then

$$P\left(\left\{\lim_{n\to\infty} \mu^{-n} Z_n > 0\right\} \triangle \{Z_n > 0 \text{ for all } n\}\right) = 0.$$

In other words, modulo a set of measure zero, $\mu^{-n} Z_n \to 0$ if and only if the population goes extinct in finite time.

12.3 Doob's inequality, convergence in L^p

This section corresponds to [2, Section 5.4].

Let $X = \{X_n\}$ be a submartingale. Recall that $EX_0 \leqslant EX_n$ for all n. The same is not true, in general, when n is replaced by a stopping time. For example, as we saw previously, if $Y_n = 1 + X_n$, where X is a simple random walk, and $N = \inf\{n : Y_n = 0\}$, then Y is a martingale (and therefore a submartingale, but $EY_0 = 1$ and $EY_N = 0$. In order to retain the inequality under a stopping time, we must add hypotheses. We will discuss this more when we cover optional stopping theorems. For now, here is a simple variant of such a result.

Theorem 12.21. *Let $X = \{X_n\}$ be an $\{\mathcal{F}_n\}$-submartingale and N an $\{\mathcal{F}_n\}$-stopping time. Suppose there exists $k \in \mathbb{N}$ such that $N \leqslant k$ a.s. Then*

$$EX_0 \leqslant EX_N \leqslant EX_k.$$

Proof. Since $\{X_{N\wedge n}\}$ is a submartingale, it follows that

$$EX_0 = EX_{N\wedge 0} \leqslant EX_{N\wedge k} = EX_N.$$

Let $H_n = 1_{\{N<n\}}$. Then $H = \{H_n\}$ is $\{\mathcal{F}_n\}$-predictable. By Remark 12.10,

$(H \cdot X)_n$ is a submartingale. Note that

$$\begin{aligned}(H \cdot X)_n &= \sum_{m=1}^{n} 1_{\{N<m\}}(X_m - X_{m-1}) \\ &= \sum_{m=(N\wedge n)+1}^{n} (X_m - X_{m-1}) \\ &= X_n - X_{N\wedge n}.\end{aligned}$$

Thus

$$0 = E(H \cdot X)_0 \leqslant E(H \cdot X)_k = EX_k - EX_{N\wedge k} = EX_k - EX_N,$$

so that $EX_N \leqslant EX_k$. □

Theorem 12.22 (Doob's inequality)**.** *Let $X = \{X_n\}$ be a submartingale. Define*

$$\overline{X}_n = \max_{0 \leqslant m \leqslant n} X_m^+.$$

Then

$$\lambda P(\overline{X}_n \geqslant \lambda) \leqslant E[X_n 1_{\{\overline{X}_n \geqslant \lambda\}}] \leqslant EX_n^+,$$

for all $\lambda > 0$ and all $n \in \mathbb{N}$.

Proof. Let $N = \inf\{m : X_m \geqslant \lambda\}$. If $\overline{X}_n(\omega) \geqslant \lambda > 0$, then there exists $m \in \{0, \ldots, n\}$ such that $X_m^+(\omega) = X_m(\omega) \geqslant \lambda$, which implies $N(\omega) \leqslant n$, so that

$$X_{N(\omega)\wedge n}(\omega) = X_{N(\omega)}(\omega) \geqslant \lambda.$$

In other words, $X_{N\wedge n} \geqslant \lambda$ on $\{\overline{X}_n \geqslant \lambda\}$. Therefore,

$$\lambda P(\overline{X}_n \geqslant \lambda) = E[\lambda 1_{\{\overline{X}_n \geqslant \lambda\}}] \leqslant E[X_{N\wedge n} 1_{\{\overline{X}_n \geqslant \lambda\}}]. \tag{12.2}$$

Next, if $\overline{X}_n(\omega) < \lambda$, then for all $m \leqslant n$, we

$$X_m(\omega) \leqslant X_m^+(\omega) \leqslant \overline{X}_n(\omega) < \lambda,$$

which implies $N(\omega) > n$. Thus, $X_{N\wedge n} = X_n$ on $\{\overline{X}_n < \lambda\}$. Since X is a submartingale and $N \wedge n$ is a stopping time with $N \wedge n \leqslant n$ a.s., Theorem 12.21 implies $EX_{N\wedge n} \leqslant EX_n$. Together, this gives

$$\begin{aligned}E[X_{N\wedge n} 1_{\{\overline{X}_n \geqslant \lambda\}}] &= EX_{N\wedge n} - E[X_{N\wedge n} 1_{\{\overline{X}_n < \lambda\}}] \\ &\leqslant EX_n - E[X_n 1_{\{\overline{X}_n < \lambda\}}] = E[X_n 1_{\{\overline{X}_n \geqslant \lambda\}}].\end{aligned}$$

Combined with (12.2), this gives the first inequality of the theorem. The second inequality follows from the fact that $Y 1_A \leqslant Y^+ 1_A \leqslant Y^+$ a.s., for any random variable Y and any event A. □

Example 12.23. Let $\xi_1, \xi_2, \ldots$ be independent with $E\xi_m = 0$ and $E\xi_m^2 = \sigma_m^2 \in (0, \infty)$. Let $X_0 = 0$ and $X_n = \xi_1 + \cdots + \xi_n$. Then $\{X_n^2\}$ is a submartingale. (Why?)

Let $x > 0$. By Doob's inequality,

$$x^2 P\left(\max_{0 \leqslant m \leqslant n} X_m^2 \geqslant x^2\right) \leqslant EX_n^2,$$

from which we obtain *Kolmogorov's maximal inequality:*

$$P\left(\max_{0 \leqslant m \leqslant n} |X_m| \geqslant x\right) \leqslant \frac{\operatorname{var}(X_n)}{x^2}.$$

Theorem 12.24 (L^p maximal inequality)**.** *Let $X = \{X_n\}$ be a submartingale. Define*

$$\overline{X}_n = \max_{0 \leqslant m \leqslant n} X_m^+.$$

Let $p \in (1, \infty)$. Then

$$E\overline{X}_n^p \leqslant \left(\frac{p}{p-1}\right)^p E(X_n^+)^p.$$

Proof. Fix $M > 0$. If $0 < \lambda \leqslant M$, then $\{\overline{X}_n \wedge M \geqslant \lambda\} = \{\overline{X}_n \geqslant \lambda\}$, so by Doob's inequality, we have

$$P(\overline{X}_n \wedge M \geqslant \lambda) \leqslant \lambda^{-1} E[X_n^+ 1_{\{\overline{X}_n \wedge M \geqslant \lambda\}}].$$

If $\lambda > M$, then $P(\overline{X}_n \wedge M \geqslant \lambda) = 0$, and the above is still true. Thus

$$\begin{aligned}
E(\overline{X}_n \wedge M)^p &= \int_0^\infty p\lambda^{p-1} P(\overline{X}_n \wedge M \geqslant \lambda)\, d\lambda \\
&\leqslant \int_0^\infty p\lambda^{p-2} E[X_n^+ 1_{\{\overline{X}_n \wedge M \geqslant \lambda\}}]\, d\lambda \\
&= E\left[\int_0^\infty p\lambda^{p-2} X_n^+ 1_{\{\overline{X}_n \wedge M \geqslant \lambda\}}\, d\lambda\right] \\
&= E\left[pX_n^+ \int_0^{\overline{X}_n \wedge M} \lambda^{p-2}\, d\lambda\right] \\
&= qE[X_n^+ (\overline{X}_n \wedge M)^{p-1}],
\end{aligned}$$

where $q = p/(p-1)$. Since p and q are conjugate exponents, Hölder's inequality gives

$$E(\overline{X}_n \wedge M)^p \leqslant q(E(X_n^+)^p)^{1/p} (E(\overline{X}_n \wedge M)^p)^{1/q}.$$

Since $\overline{X}_n \wedge M \leqslant M$ a.s., it follows that $E(\overline{X}_n \wedge M)^p \leqslant M^p < \infty$, so we may divide both sides by $(E(\overline{X}_n \wedge M)^p)^{1/q}$, obtaining

$$(E(\overline{X}_n \wedge M)^p)^{1/p} \leqslant q(E(X_n^+)^p)^{1/p}.$$

Note that this is still true even if $E(\overline{X}_n \wedge M)^p = 0$. Raising both sides to the p, it follows that

$$E(\overline{X}_n \wedge M)^p \leqslant \left(\frac{p}{p-1}\right)^p E(X_n^+)^p.$$

Letting $M \to \infty$ and applying the monotone convergence theorem finishes the proof. □

Corollary 12.25. *Let $X = \{X_n\}$ be a martingale. Define*

$$X_n^* = \max_{0 \leqslant m \leqslant n} |X_m|.$$

Then

$$E(X_n^*)^p \leqslant \left(\frac{p}{p-1}\right)^p E|X_n|^p.$$

Proof. By Theorem 12.5, $\{|X_n|\}$ is a submartingale and the result follows immediately by Theorem 12.24. □

Theorem 12.26 (L^p convergence theorem)**.** *Let $X = \{X_n\}$ be a martingale. Suppose there exists $p > 1$ such that $\sup_n E|X_n|^p < \infty$. Then there exists $X_\infty \in L^p(\Omega)$ such that $X_n \to X_\infty$ a.s. and in L^p.*

Proof. Since $EX_n^+ \leqslant E|X_n| \leqslant (E|X_n|^p)^{1/p}$, it follows that $\sup_n EX_n^+ < \infty$. By the martingale convergence theorem (Theorem 12.12), there exists $X_\infty \in L^1(\Omega)$ such that $X_n \to X_\infty$ a.s. Let $C = \sup_n E|X_n|^p < \infty$. By Corollary 12.25,

$$E(X_n^*)^p \leqslant \left(\frac{p}{p-1}\right)^p C.$$

Letting $n \to \infty$ and applying the monotone convergence theorem, we have that $\sup_n |X_n| \in L^p(\Omega)$. Since $|X_\infty| \leqslant \sup_n |X_n|$, it follows that $X_\infty \in L^p(\Omega)$. Moreover, using $|X_n - X_\infty|^p \leqslant (2 \sup |X_n|)^p$, it follows by the dominated convergence theorem that $X_n \to X_\infty$ in L^p. □

Theorem 12.27 (orthogonality of martingale increments)**.** *Let X be an L^2 martingale. That is, $X = \{X_n\}$ is a martingale and $X_n \in L^2(\Omega, \mathcal{F}, P)$ for all n. Let $m < n$. Then $X_n - X_m$ is orthogonal to the subspace $L^2(\Omega, \mathcal{F}_m, P)$. That is, if $Y \in L^2$ is $\mathcal{F}_m$-measurable, then $E[Y(X_n - X_m)] = 0$.*

Proof. Since

$$E[Y(X_n - X_m)] = E[E[Y(X_n - X_m) \mid \mathcal{F}_m]] = E[Y E[X_n - X_m \mid \mathcal{F}_m]],$$

and

$$E[X_n - X_m \mid \mathcal{F}_m] = E[X_n \mid \mathcal{F}_m] - X_m = X_m - X_m = 0,$$

it follows that $E[Y(X_n - X_m)] = 0$. □

The following result is the conditional analogue of the formula used to calculate variance: $E|X - EX|^2 = EX^2 - (EX)^2$.

Theorem 12.28. *Let X be an L^2 martingale. Then*

$$E[(X_n - X_m)^2 \mid \mathcal{F}_m] = E[X_n^2 \mid \mathcal{F}_m] - X_m^2,$$

for all $m < n$.

Proof. By the orthogonality of martingale increments, we have

$$\begin{aligned} E[X_n^2 \mid \mathcal{F}_m] &= E[(X_n - X_m + X_m)^2 \mid \mathcal{F}_m] \\ &= E[(X_n - X_m)^2 + 2X_m(X_n - X_m) + X_m^2 \mid \mathcal{F}_m] \\ &= E[(X_n - X_m)^2 \mid \mathcal{F}_m] + 2X_m E[X_n - X_m \mid \mathcal{F}_m] + X_m^2 \\ &= E[(X_n - X_m)^2 \mid \mathcal{F}_m] + X_m^2, \end{aligned}$$

and we are done. □

Exercises

12.7. [2, Exercise 5.4.4] Let $\xi_1, \xi_2, \ldots$ be independent with $E\xi_j = 0$ and $|\xi_j| \leqslant K$ a.s. for all j. Define $X_n = \xi_1 + \cdots + \xi_n$. Prove that

$$P\left(\max_{1 \leqslant m \leqslant n} |X_m| \leqslant x\right) \leqslant \frac{(x+K)^2}{\operatorname{var}(X_n)},$$

for all $x > 0$.

12.8. [2, Exercise 5.4.5] Let X be an L^2 martingale with $X_0 = 0$ a.s. Prove that

$$P\left(\max_{1 \leqslant m \leqslant n} X_m \geqslant \lambda\right) \leqslant \frac{EX_n^2}{EX_n^2 + \lambda^2},$$

for all $\lambda > 0$.

12.9. [2, Exercise 5.4.7] Let X and Y be L^2 martingales with respect to a common filtration $\{\mathcal{F}_n\}$. Prove that

$$EX_nY_n - EX_0Y_0 = \sum_{j=1}^n E[(X_j - X_{j-1})(Y_j - Y_{j-1})].$$

12.4 Uniform integrability, convergence in L^1

This section corresponds to [2, Section 5.5].

Lemma 12.29. *Let X be a real-valued random variable. Then X is integrable if and only if*

$$\lim_{M \to \infty} E[|X| 1_{\{|X| > M\}}] = 0.$$

Proof. First, suppose X is integrable. Since $|X|1_{\{|X|>M\}} \leqslant |X|$ for all M and $|X|1_{\{|X|>M\}} \to 0$ a.s. as $M \to \infty$, it follows from the dominated convergence theorem that $E[|X|1_{\{|X|>M\}}] \to 0$ as $M \to \infty$.

Now suppose $E[|X|1_{\{|X|>M\}}] \to 0$ as $M \to \infty$. Choose M such that $E[|X|1_{\{|X|>M\}}] < 1$. Then

$$E|X| = E[|X|1_{\{|X|\leqslant M\}}] + E[|X|1_{\{|X|>M\}}] < M + 1 < \infty,$$

and so X is integrable. □

A family of random variables, $\{X_\alpha\}_{\alpha\in A}$ is **uniformly integrable** if

$$\lim_{M\to\infty} \sup_{\alpha\in A} E[|X_\alpha|1_{\{|X_\alpha|>M\}}] = 0.$$

Note that if $\{X_\alpha\}$ is uniformly integrable, then any subset of $\{X_\alpha\}$ is also uniformly integrable.

Lemma 12.30. *If $\{X_\alpha\}_{\alpha\in A}$ is uniformly integrable, then $\sup_{\alpha\in A} E|X_\alpha| < \infty$.*

Proof. Choose M such that $E[|X_\alpha|1_{\{|X_\alpha|>M\}}] < 1$ for all $\alpha \in A$. Thus, as before, $E|X_\alpha| < M + 1$ for all $\alpha \in A$. □

Lemma 12.31. *Let $\{X_\alpha\}$ be a family of real-valued random variables. Suppose there exists an integrable random variable Y such that $|X_\alpha| \leqslant Y$ a.s. for all α. Then $\{X_\alpha\}$ is uniformly integrable.*

Proof. The result follows from the facts that $|X_\alpha|1_{\{|X_\alpha|>M\}} \leqslant |Y|1_{\{|Y|>M\}}$ a.s. and $E[|Y|1_{\{|Y|>M\}}] \to 0$ as $M \to \infty$. □

Theorem 12.32. *Let $X \in L^1(\Omega, \mathcal{F}, P)$. Let*

$$\mathbb{F} = \{\mathcal{G} \subset \mathcal{F} : \mathcal{G} \text{ is a } \sigma\text{-algebra}\}.$$

For each $\mathcal{G} \in \mathbb{F}$, let $X_\mathcal{G} = E[X \mid \mathcal{G}]$. Then $\{X_\mathcal{G}\}_{\mathcal{G}\in\mathbb{F}}$ is uniformly integrable.

Proof. Let $\varepsilon > 0$ be arbitrary. By Exercise 2.15, we may choose $\delta > 0$ such that $E[|X|1_A] < \varepsilon$ whenever $P(A) < \delta$. Define $M_0 = \delta^{-1}E|X|$. Let $M > M_0$ be arbitrary. Fix $\mathcal{G} \in \mathbb{F}$. Let $Z_\mathcal{G} = E[|X| \mid \mathcal{G}]$. By Jensen's inequality for conditional expectations, $|X_\mathcal{G}| \leqslant Z_\mathcal{G}$ a.s. Thus,

$$E[|X_\mathcal{G}|1_{\{|X_\mathcal{G}|>M\}}] \leqslant E[Z_\mathcal{G}1_{\{Z_\mathcal{G}>M\}}].$$

Since $\{Z_\mathcal{G} > M\} \in \mathcal{G}$, it follows from the definition of conditional expectation that

$$E[Z_\mathcal{G}1_{\{Z_\mathcal{G}>M\}}] = E[|X|1_{\{Z_\mathcal{G}>M\}}].$$

By Chebyshev's inequality,

$$P(Z_\mathcal{G} > M) \leqslant \frac{EZ_\mathcal{G}}{M} = \frac{E|X|}{M} < \frac{E|X|}{M_0} = \delta,$$

and it follows that

$$E[|X|1_{\{Z_\mathcal{G}>M\}}] < \varepsilon.$$

Putting it all together, we have $E[|X_\mathcal{G}|1_{\{|X_\mathcal{G}|>M\}}] < \varepsilon$. Since $\mathcal{G}$ was arbitrary, it follows that

$$\sup_{\mathcal{G}\in\mathbb{F}} E[|X_\mathcal{G}|1_{\{|X_\mathcal{G}|>M\}}] \leqslant \varepsilon,$$

for all $M > M_0$. Since ε was arbitrary, $\lim_{M\to\infty}\sup_{\mathcal{G}\in\mathbb{F}} E[|X_\mathcal{G}|1_{\{|X_\mathcal{G}|>M\}}] = 0$. □

Example 12.33. Let Y be an integrable random variable and $\{\mathcal{F}_n\}$ a filtration. Define $X = \{X_n\}$ by $X_n = E[Y \mid \mathcal{F}_n]$. Then X is an $\{\mathcal{F}_n\}$-martingale (check). Moreover, by the previous theorem, X is a uniformly integrable martingale. As we will see in this section, all uniformly integrable martingales can be written this way.

Proposition 12.34. *Let $\varphi : [0,\infty) \to [0,\infty)$ satisfy $x^{-1}\varphi(x) \to \infty$ as $x \to \infty$. If $\sup_{\alpha\in A} E\varphi(|X_\alpha|) < \infty$, then $\{X_\alpha\}_{\alpha\in A}$ is uniformly integrable.*

Proof. Exercise 12.10. □

Remark 12.35. A common choice for φ in Proposition 12.34 is $\varphi(x) = x^p$, where $p > 1$. Another possible choice is $\varphi(x) = x(\log x)^+$.

Theorem 12.36. *Suppose $X_n \to X$ in probability. Then the following are equivalent:*

(i) $\{X_n\}$ *is uniformly integrable,*

(ii) $X_n \to X$ *in* L^1,

(iii) $E|X_n| \to E|X|$

Proof. Uses truncation. See [2, Theorem 5.5.2] for details. □

Theorem 12.37. *Let $X = \{X_n\}$ be a submartingale. Then the following are equivalent:*

(i) X *is uniformly integrable,*

(ii) There exists X_∞ *such that* $X_n \to X_\infty$ *a.s. and in* L^1.

(iii) There exists X_∞ *such that* $X_n \to X_\infty$ *in* L^1.

Proof. Suppose X is uniformly integrable. Then $\sup_n E|X_n| < \infty$. Thus, by the martingale convergence theorem (Theorem 12.12), there exists $X_\infty \in L^1$ such that $X_n \to X_\infty$ a.s. This implies $X_n \to X_\infty$ in probability. Therefore, by Theorem 12.36, we have $X_n \to X_\infty$ in L^1 and we have proven that (i) implies (ii).

Trivially, (ii) implies (iii).

Now suppose (iii) holds. Since convergence in L^1 implies convergence in probability, we may apply Theorem 12.36, which gives us that X is uniformly integrable. Hence, (iii) implies (i). □

Lemma 12.38. *If $X_n \to X$ in L^1 and $A \in \mathcal{F}$, then $E[X_n 1_A] \to E[X 1_A]$.*

Proof. This follows since

$$|E[X_n 1_A] - E[X 1_A]| = |E[(X_n - X) 1_A]| \leqslant E|X_n - X| \to 0$$

as $n \to \infty$. □

Lemma 12.39. *Let $X = \{X_n\}$ be an $\{\mathcal{F}_n\}$-martingale. Suppose there exists X_∞ such that $X_n \to X_\infty$ in L^1. Then $X_n = E[X_\infty \mid \mathcal{F}_n]$.*

Proof. Fix $n \in \mathbb{N}$. Note that $X_n \in \mathcal{F}_n$. Let $A \in \mathcal{F}_n$. By the definition of conditional expectation, we need only show that $E[X_\infty 1_A] = E[X_n 1_A]$.

By the martingale property, if $k \in \mathbb{N}$, then $E[X_{n+k} \mid \mathcal{F}_n] = X_n$. Thus, by the definition of conditional expectation, $E[X_{n+k} 1_A] = E[X_n 1_A]$. Letting $k \to \infty$ and using the previous lemma, we get $E[X_\infty 1_A] = E[X_n 1_A]$. □

Theorem 12.40. *Let $X = \{X_n\}$ be an $\{\mathcal{F}_n\}$-martingale. Then the following are equivalent:*

(i) X is uniformly integrable,

(ii) There exists X_∞ such that $X_n \to X_\infty$ a.s. and in L^1.

(iii) There exists X_∞ such that $X_n \to X_\infty$ in L^1.

(iv) There exists $Y \in L^1$ such that $X_n = E[Y \mid \mathcal{F}_n]$ for all n.

Proof. Theorem 12.37 gives (i) implies (ii) implies (iii). Lemma 12.39 gives (iii) implies (iv). And Theorem 12.32 gives (iv) implies (i). □

Given a filtration $\{\mathcal{F}_n\}$, we will define $\mathcal{F}_\infty := \sigma(\bigcup_n \mathcal{F}_n)$.

Theorem 12.41. *Let $\{\mathcal{F}_n\}$ be a filtration and Y an integrable random variable. Then $E[Y \mid \mathcal{F}_n] \to E[Y \mid \mathcal{F}_\infty]$ a.s. and in L^1 as $n \to \infty$.*

Proof. Let $X_n = E[Y \mid \mathcal{F}_n]$ so that $X = \{X_n\}$ is a uniformly integrable martingale and there exists X_∞ such that $X_n \to X_\infty$ a.s. and in L^1. It therefore suffices to show that $X_\infty = E[Y \mid \mathcal{F}_\infty]$.

For each n, we have $X_n \in \mathcal{F}_\infty$. Thus, $\limsup_n X_n \in \mathcal{F}_\infty$. But $\limsup_n X_n = X_\infty$ a.s. So after changing X_∞ (if necessary) on a set of measure zero, we have $X_\infty \in \mathcal{F}_\infty$.

Let

$$\mathcal{L} = \{A \in \mathcal{F}_\infty : E[Y 1_A] = E[X_\infty 1_A]\}.$$

It remains only to show that $\mathcal{F}_\infty \subset \mathcal{L}$. By Lemma 12.39, we have $X_n = E[X_\infty \mid \mathcal{F}_n]$. Thus, $E[Y \mid \mathcal{F}_n] = E[X_\infty \mid \mathcal{F}_n]$ for all n. It follows from the definition of conditional expectation that $\mathcal{F}_n \subset \mathcal{L}$ for all n, and hence, $\bigcup_n \mathcal{F}_n \subset \mathcal{L}$. Since $\bigcup_n \mathcal{F}_n$ is a π-system and $\mathcal{L}$ is a λ-system, it follows that $\mathcal{F}_\infty \subset \mathcal{L}$. □

Theorem 12.42 (Lévy's 0-1 law)**.** *Let $\{\mathcal{F}_n\}$ be a filtration and $A \in \mathcal{F}_\infty$. Then $P(A \mid \mathcal{F}_n) \to 1_A$ a.s.*

Proof. Take $Y = 1_A$ in Theorem 12.41. □

Theorem 12.43. *Let $X_n \to X_\infty$ a.s. Suppose there exists integrable Y such that $|X_n| \leqslant Y$ a.s. for all $n \in \mathbb{N}$. If $\{\mathcal{F}_n\}$ is a filtration, then*

$$E[X_n \mid \mathcal{F}_n] \to E[X_\infty \mid \mathcal{F}_\infty] \text{ a.s.}$$

as $n \to \infty$.

Proof. Uses Theorem 12.41. See [2, Theorem 5.5.9] for details. □

Exercises

12.10. [2, Exercise 5.5.1] Prove Proposition 12.34.

12.11. [2, Exercise 5.5.2] Let θ be an integrable random variable. Let $Z = \{Z_n\}$ be an integrable stochastic process, independent of θ. Assume $Z_1, Z_2, ...$ are i.i.d. Define $Y_n = \theta + Z_n$. (For example, if $Z_1 \sim N(0,1)$, then, given θ, the sequence $Y_1, Y_2, \ldots$ is an i.i.d. sequence of $N(\theta, 1)$-distributed random variables.) Prove that $E[\theta \mid Y_1, \ldots, Y_n] \to \theta$ a.s. as $n \to \infty$.

12.12. [2, Exercise 5.5.7] Let $X = \{X_n\}$ be an $\{\mathcal{F}_n\}$-adapted, $[0,1]$ valued process. Let $\theta \in [0,1]$ and assume $X_0 = \theta$ a.s. Let $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and suppose

$$\begin{aligned} P(X_{n+1} = \alpha + \beta X_n \mid \mathcal{F}_n) &= X_n, \\ P(X_{n+1} = \beta X_n \mid \mathcal{F}_n) &= 1 - X_n. \end{aligned}$$

Prove that there exists $A \in \mathcal{F}$ such that $X_n \to 1_A$ a.s. and $P(A) = \theta$.

12.13. [2, Exercise 5.5.8] Prove that if $\{\mathcal{F}_n\}$ is a filtration and $X_n \to X_\infty$ in L^1, then $E[X_n \mid \mathcal{F}_n] \to E[X_\infty \mid \mathcal{F}_\infty]$ in L^1.

12.5 Optional Stopping Theorems

This section corresponds to [2, Section 5.7].

Our main theorems in this section will use the hypothesis that $\{X_{N \wedge n}\}$ is uniformly integrable. Our first two results provide tools for checking this hypothesis.

Theorem 12.44. *Let $X = \{X_n\}$ be a stochastic process and N an $\mathbb{N} \cup \{\infty\}$-valued random variable. If $X_N 1_{\{N<\infty\}}$ is integrable and $\{X_n 1_{\{N>n\}}\}$ is uniformly integrable, then $\{X_{N \wedge n}\}$ is uniformly integrable.*

Proof. Fix $n \in \mathbb{N}$ and $M > 0$. Then

$$\begin{aligned} &E[|X_{N \wedge n}| 1_{\{|X_{N \wedge n}| > M\}}] \\ &\quad = E[|X_N 1_{\{N<\infty\}}| 1_{\{|X_N 1_{\{N<\infty\}}| > M\}} 1_{\{N \leqslant n\}}] + E[|X_n| 1_{\{|X_n| > M\}} 1_{\{N>n\}}] \\ &\quad\quad \leqslant E[|X_N 1_{\{N<\infty\}}| 1_{\{|X_N 1_{\{N<\infty\}}| > M\}}] + E[|X_n 1_{\{N>n\}}| 1_{\{|X_n 1_{\{N>n\}}| > M\}}]. \end{aligned}$$

The result follows by taking the supremum and letting $M \to \infty$. □

Theorem 12.45. *Let X be a uniformly integrable $\{\mathcal{F}_n\}$-submartingale, and N an $\{\mathcal{F}_n\}$-stopping time. Then $\{X_{N\wedge n}\}$ is uniformly integrable.*

Proof. See [2, Theorem 5.7.1]. □

Theorem 12.46. *Let $\{X_n\}$ be a uniformly integrable $\{\mathcal{F}_n\}$-submartingale and N an $\{\mathcal{F}_n\}$ stopping time. By Theorem 12.37, there exists X_∞ such that $X_n \to X_\infty$ a.s. and in L^1. Hence, X_N is well-defined. We then have*

$$EX_0 \leqslant EX_N \leqslant EX_\infty.$$

Proof. By Theorem 12.21,

$$EX_0 \leqslant EX_{N\wedge n} \leqslant EX_n.$$

We have already established that $X_n \to X_\infty$ in L^1. By Theorem 12.45, we have $\{X_{N\wedge n}\}$ is uniformly integrable. Thus, by Theorem 12.37 and the fact that $X_{N\wedge n} \to X_N$ a.s., it follows that $X_{N\wedge n} \to X_N$ in L^1. Hence, the proof is completed by letting $n \to \infty$. □

Remark 12.47. It is instructive to compare this theorem with Theorem 12.21.

Theorem 12.48 (optional stopping theorem)**.** *Let $\{\mathcal{F}_n\}$ be a filtration. Let L and M be $\{\mathcal{F}_n\}$-stopping times with $L(\omega) \leqslant M(\omega)$ for all $\omega \in \Omega$. Let $\{Y_n\}$ be a stochastic process such that $\{Y_{M\wedge n}\}$ is a uniformly integrable $\{\mathcal{F}_n\}$-submartingale. Then $Y_n 1_{\{M=\infty\}}$ converges a.s. Define Y_∞ to be this limit, so that Y_L and Y_M are well-defined. Then $EY_L \leqslant EY_M$ and $Y_L \leqslant E[Y_M \mid \mathcal{F}_L]$ a.s.*

Proof. Let $X_n = Y_{M\wedge n}$, so that $X = \{X_n\}$ is a uniformly integrable submartingale. By Theorem 12.37, there exists X_∞ such that $X_n \to X_\infty$ a.s. and in L^1. In particular, this implies $Y_n 1_{\{M=\infty\}} = X_n 1_{\{M=\infty\}} \to X_\infty 1_{\{M=\infty\}} =: Y_\infty$ a.s.

By Theorem 12.46, $EX_L \leqslant EX_\infty$. But $X_L = Y_L$ and $X_\infty = Y_M$, so this proves the first inequality.

Now fix $A \in \mathcal{F}_L$. Let $N = L1_A + M1_{A^c}$. By Proposition 11.3, N is stopping time. Since $N \leqslant M$ pointwise, we have

$$E[Y_L 1_A] + E[Y_M 1_{A^c}] = EY_N \leqslant EY_M = E[Y_M 1_A] + E[Y_M 1_{A^c}].$$

Since $Y_M = X_\infty$ is integrable, we have $E[Y_M 1_{A^c}] < \infty$, so that

$$E[Y_L 1_A] \leqslant E[Y_M 1_A] = E[E[Y_M 1_A \mid \mathcal{F}_L]] = E[E[Y_M \mid \mathcal{F}_L] 1_A].$$

This holds for all $A \in \mathcal{F}_L$. By Exercise 11.6, $Y_L = Y_L 1_{\{L<\infty\}} + X_\infty 1_{\{L=\infty\}}$ is $\mathcal{F}_L$-measurable. Also, $E[Y_M \mid \mathcal{F}_L]$ is $\mathcal{F}_L$-measurable. Thus, by Lemma 6.50, we obtain the second inequality. □

For the proofs of the next two results, see [2, Theorems 5.7.5 and 5.7.6], respectively. The first is a generalization of Wald's equation.

Theorem 12.49. *Let X be an $\{\mathcal{F}_n\}$-submartingale. Suppose there exists $B > 0$ such that $E[|X_{n+1} - X_n| \mid \mathcal{F}_n] \leqslant B$ a.s. for all n. Let N be an $\{\mathcal{F}_n\}$-stopping time with $EN < \infty$. Then $\{X_{N \wedge n}\}$ is uniformly integrable and $EX_N \geqslant EX_0$. If X is a martingale, then $EX_N = EX_0$.*

Theorem 12.50. *Let X be a nonnegative $\{\mathcal{F}_n\}$-supermartingale and N an $\{\mathcal{F}_n\}$-stopping time. By Theorem 12.14, there exists $X_\infty \in L^1$ such that $X_n \to X_\infty$ a.s. and $EX_\infty \leqslant EX_0$. Hence, X_N is well-defined. It then follows that $EX_N \leqslant EX_0$.*

Theorem 12.51. *Let X be an asymmetric simple random walk with $p = P(\xi_1 = 1) > 1/2$. Let $\varphi(x) = (q/p)^x$ and $T_x = \inf\{n : X_n = x\}$. Fix $a < 0 < b$. Then*

(a) $\{\varphi(X_n)\}$ is a martingale,

(b) we have

$$P(T_a < T_b) = \frac{\varphi(b) - \varphi(0)}{\varphi(b) - \varphi(a)},$$

(c) $P(\inf_n X_n \leqslant a) = P(T_a < \infty) = \varphi(-a)$, and

(d) $P(T_b < \infty) = 1$ and $ET_b = b/(2p - 1)$.

Proof. Note that

$$E[\varphi(X_{n+1}) \mid \mathcal{F}_n^X] = E[\varphi(X_n + \xi_{n+1}) \mid \mathcal{F}_n^X] = h(X_n),$$

where

$$\begin{aligned} h(x) = E[\varphi(x + \xi_{n+1})] = p(q/p)^{x+1} + q(q/p)^{x-1} \\ = (q + p)(q/p)^x = (q/p)^x = \varphi(x). \end{aligned}$$

Thus, $E[\varphi(X_{n+1}) \mid \mathcal{F}_n^X] = \varphi(X_n)$, which proves (a).

Let $\theta = P(T_a < T_b)$. Let $N = T_a \wedge T_b$. By the same methods as in Example 11.12, it can be shown that $EN < \infty$, and therefore $N < \infty$ a.s. Since $|\varphi(X_{N \wedge n})| \leqslant |a| \vee b$ for all n, it follows that $\{\varphi(X_{N \wedge n})\}$ in uniformly integrable. By the optional stopping theorem,

$$\varphi(0) = E\varphi(X_0) = E\varphi(X_N) = \theta\varphi(a) + (1 - \theta)\varphi(b).$$

Solving for θ gives (b).

Since $\{T_a < \infty\} = \bigcup_{b=1}^\infty \{T_a < T_b\}$, and this is an increasing union, (c) follows by letting $b \to \infty$ in (b).

Similarly, letting $a \to -\infty$ gives $T_b < \infty$ a.s. Note that $\{X_n - (p - q)n\}$ is a mean zero random walk, and therefore a martingale. Since $T_b \wedge n$ is a bounded stopping time, we may apply Theorem 12.21 to obtain

$$0 = E[X_0 - (p - q)0] = E[X_{T_b \wedge n} - (p - q)(T_b \wedge n)].$$

Thus,

$$E[T_b \wedge n] = (p-q)^{-1} E[X_{T_b \wedge n}]. \tag{12.3}$$

Now,

$$|X_{T_b \wedge n}| \leqslant b + |\inf_m X_m|.$$

By (c),

$$\begin{aligned} E|\inf_m X_m| &= \sum_{k \in \mathbb{N}} k P(\inf_m X_m = -k) \\ &= \sum_{k \in \mathbb{N}} k(\varphi(k) - \varphi(k+1)) \\ &= \sum_{k \in \mathbb{N}} k(q/p)^k (1 - q/p) < \infty, \end{aligned}$$

since $q/p < 1$. Thus, we may apply dominated convergence to the right-hand side of (12.3). If we also apply monotone convergence to the left-hand side, we obtain

$$ET_b = (p-q)^{-1} E[X_{T_b}].$$

Since $X_{T_b} = b$ a.s., this proves (d). □

Exercises

12.14. Let X be a mean 0 random walk with $\sigma^2 = E\xi_1^2$. Prove that the process $\{X_n^2 - \sigma^2 n\}$ is a martingale.

12.15. [2, Exercise 5.7.2] Let X be an asymmetric random walk with $p > 1/2$. Prove that $\mathrm{var}(T_b) = 4bpq/(p-q)^3$.

12.16. [2, Exercise 5.7.8] Let X_n be the total bankroll of a poker players after his nth session, so that $\xi_n := X_n - X_{n-1}$ is his winnings (or losses, if negative) from the nth session. Assume that $\xi_1, \xi_2, \ldots$ are i.i.d. with $\xi_1 \sim N(\mu, \sigma^2)$, where $\mu > 0$. Let $X_0 = b \in (0, \infty)$ and consider

$$R = \{X_n \leqslant 0 \text{ for some } n \in \mathbb{N}\}.$$

Then R represents the event that the player eventually goes broke. Prove that

$$P(R) \leqslant e^{-2\mu b/\sigma^2}.$$

Chapter 13

Markov Chains

13.1 Definitions

This section corresponds to [2, Section 6.1].

Let $(S, \mathcal{S})$ be a measurable space. Recall from Section 6.4 that an S-valued stochastic process $X = \{X_n\}$ is a Markov chain with respect to a filtration $\{\mathcal{F}_n\}$ is X is adapted to $\{\mathcal{F}_n\}$ and satisfies the **Markov property**,

$$P(X_{n+1} \in B \mid \mathcal{F}_n) = P(X_{n+1} \in B \mid X_n), \tag{13.1}$$

for all $n \in \mathbb{N}$ and all $B \in \mathcal{S}$. The space S is called the *state space* of the Markov chain X.

If we say simply that X is a Markov chain, we mean that X is a Markov chain with respect to $\{\mathcal{F}_n^X\}$.

A **transition probability** on S is a probability kernel from S to S. Recall from Section 6.3.2 that a probability kernel from S to S is a function $p : S \to M_1(S)$ which is $(\mathcal{S}, \mathcal{M}(S))$-measurable. (Here, $M_1(S)$ is the set of probability measures on S.) Recall also that we write $p(x, A) = (p(x))(A)$, so that we may think of p as a mapping from $S \times \mathcal{S}$ to $\mathbb{R}$.

In Section 6.3.2, it was shows that $p : S \times \mathcal{S} \to \mathbb{R}$ is a transition probability if and only if $p(x, \cdot)$ is a probability measure for all $x \in S$ and $p(\cdot, A)$ is measurable for all $A \in \mathcal{S}$.

Lemma 13.1. *Let $(S, \mathcal{S})$ be a measurable space. Let X be an S-valued stochastic process, adapted to a filtration $\{\mathcal{F}_n\}$. If, for each $n \in \mathbb{N}$, there exists a transition probability p_n such that*

$$P(X_{n+1} \in B \mid \mathcal{F}_n) = p_n(X_n, B),$$

for all $B \in \mathcal{S}$, then X is an Markov chain with respect to $\{\mathcal{F}_n\}$.

Proof. Suppose there exists such a transition probability for each n. Fix $B \in \mathcal{S}$. Then $P(X_{n+1} \in B \mid \mathcal{F}_n)$ is $\sigma(X_n)$-measurable. Thus,

$$P(X_{n+1} \in B \mid X_n) = E[P(X_{n+1} \in B \mid \mathcal{F}_n) \mid X_n] = P(X_{n+1} \in B \mid \mathcal{F}_n),$$

and X is a Markov chain. □

Lemma 13.2. *Let $(S, \mathcal{S})$ be a measurable space. Let X be an S-valued Markov chain with respect to $\{\mathcal{F}_n\}$. If S is a standard Borel space, then, for each $n \in \mathbb{N}$, there exists a unique transition probability p_n such that*

$$P(X_{n+1} \in B \mid \mathcal{F}_n) = p_n(X_n, B),$$

for all $B \in \mathcal{S}$. (Uniqueness is in the sense that if, for fixed n, the transition probabilities p_n and $\widetilde{p}_n$ both satisfy the theorem, then for P-a.e. $\omega \in \Omega$, the measures $p_n(X_n(\omega), \cdot)$ and $\widetilde{p}_n(X_n(\omega), \cdot)$ are identical.)

Proof. This is just a special case of Theorem 6.64. □

For the remainder of this chapter, we shall assume S is a standard Borel space. In this case, we say that an $\{\mathcal{F}_n\}$-Markov chain is **time-homogeneous** if the transition probabilities p_n do not depend on n. That is, a Markov chain is time-homogeneous if there exists a single transition probability p such that

$$P(X_{n+1} \in B \mid \mathcal{F}_n) = p(X_n, B),$$

for all $n \in \mathbb{N}$ and all $B \in \mathcal{S}$.

For the remainder of this chapter, we shall also assume all Markov chains are time-homogeneous. As such, we will typically refer to time-homogeneous Markov chains simply as Markov chains.

If X is a Markov chain on S and $X_0 \sim \mu$, then we say μ is the **initial distribution** of X.

Theorem 13.3. *Let p be a transition probability on S and μ a probability measure on S. Then there exists a Markov chain X on S with transition probability p and initial distribution μ.*

Proof. Let $\Omega = S^{\mathbb{N} \cup \{0\}}$ and $\mathcal{F} = \mathcal{S}^{\mathbb{N} \cup \{0\}}$, so that a typical $\omega \in \Omega$ has the form $\omega = (\omega_0, \omega_1, \ldots)$. For $n \geqslant 0$, define $X_n : \Omega \to S$ by $X_n(\omega) = \omega_n$.

Let ν_n be the measure on $(S^{\{0,\ldots,n\}}, \mathcal{S}^{\{0,\ldots,n\}})$ determined by

$$\begin{aligned} &\nu_n(B_0 \times \cdots \times B_n) \\ &\qquad = \int_{B_0} \cdots \int_{B_{n-1}} p(x_{n-1}, B_n) p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1) \mu(dx_0) \end{aligned}$$

As in the proof of Theorem 6.26, we can use Kolmogorov's extension theorem (Theorem 2.52) to show that there exists a unique probability measure P_μ on $(\Omega, \mathcal{F})$ such that

$$P_\mu(X_0 \in B_0, \ldots, X_n \in B_n) = \nu_n(B_0 \times \cdots \times B_n)$$

for all $B_0, \ldots, B_n \in \mathcal{S}$. The details are left to the reader in Exercise 13.1.

We first note that

$$P_\mu(X_0 \in B_0) = \nu_0(B_0) = \int_{B_0} \mu(dx_0) = \mu(B_0).$$

It therefore suffices to show that X is a Markov chain with transition probability p.

Fix $n \in \mathbb{N}$ and $B \in \mathcal{S}$. We want to show that

$$P_\mu(X_{n+1} \in B \mid \mathcal{F}_n^X) = p(X_n, B).$$

Let

$$\mathcal{L} = \{A \in \mathcal{F}_n^X : E_\mu[1_{\{X_{n+1} \in B\}} 1_A] = E_\mu[p(X_n, B)1_A]\}.$$

Since $p(X_n, B)$ is $\mathcal{F}_n^X$-measurable, it suffices to show that $\mathcal{F}_n^X \subset \mathcal{L}$. Let

$$\mathcal{P} = \{\{X_0 \in B_0, \dots, X_n \in B_n\} : B_j \in \mathcal{S}\}.$$

Since $\mathcal{P}$ is a π-system and $\sigma(\mathcal{P}) = \mathcal{F}_n^X$, it suffices by the π-λ theorem to show that $\mathcal{P} \subset \mathcal{L}$.

Let $A = \{X_0 \in B_0, \dots, X_n \in B_n\} \in \mathcal{P}$. We will first show that

$$\int_{B_0} \cdots \int_{B_n} f(x_n) p(x_{n-1}, dx_n) \cdots p(x_0, dx_1) \mu(dx_0) = E_\mu[f(X_n)1_A], \qquad (13.2)$$

for all bounded, measurable $f : S \to \mathbb{R}$. To show this, first assume $f = 1_E$ for some $E \in \mathcal{S}$. Then the left-hand side of (13.2) becomes

$$\begin{aligned}
\int_{B_0} \cdots \int_{B_{n-1}} & p(x_{n-1}, E \cap B_n) p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1) \mu(dx_0) \\
&= \nu_n(B_0 \times \cdots \times B_{n-1} \times (E \cap B_n)) \\
&= P_\mu(X_0 \in B_0, \dots, X_{n-1} \in B_{n-1}, X_n \in E \cap B_n) \\
&= E_\mu[1_E(X_n)1_A],
\end{aligned}$$

and (13.2) is true for indicator functions. By linearity, it is true for simple functions, so by dominated convergence, it is true for all bounded, measurable functions.

We now show that $A \in \mathcal{L}$. For this, we note that

$$\begin{aligned}
E_\mu[1_{\{X_{n+1} \in B\}} 1_A] &= \nu_{n+1}(B_0 \times \ldots \times B_n \times B) \\
&= \int_{B_0} \cdots \int_{B_n} p(x_n, B) p(x_{n-1}, dx_n) \cdots p(x_0, dx_1) \mu(dx_0) \\
&= E_\mu[p(X_n, B)1_A],
\end{aligned}$$

where in the last line we have applied (13.2) to the bounded, measurable function $p(\cdot, B)$. □

Remark 13.4. Although the probability measure in the proof of Theorem 13.3 is denoted by P_μ, it actually depends on p as well. If we change p or μ, we will get a different probability measure.

Often, when we want to construct a variety of different stochastic processes (or even just a variety of different random variables), we fix one measurable space $(\Omega, \mathcal{F})$, one probability measure P, and then consider many different measurable functions.

Here we are taking a different approach. We fix one measurable space $(\Omega, \mathcal{F}) = (S^{\mathbb{N}\cup\{0\}}, \mathcal{S}^{\mathbb{N}\cup\{0\}})$, one measurable function $X(\omega) = \omega$, and then consider many different probability measures. We don't construct our Markov process by constructing X on top of a given probability space. Rather, we construct P_μ on top of the identity process on the canonical sequence space.

If $x \in S$, we will use P_x to denote P_{δ_x}. It can be show that

$$P_\mu(A) = \int_S P_x(A)\mu(dx),$$

so that for many purposes it suffices only to consider the measures P_x.

The Markov chain in Theorem 13.3 is unique in the following sense. If X and Y are Markov chains, both with transition probability p and initial distribution μ, then X and Y have the same finite-dimensional distributions. That is, for every $n \in \mathbb{N}$, we have $(X_0, \ldots, X_n) =_d (Y_0, \ldots, Y_n)$.

Theorem 13.5. *If X is a Markov chain on S with transition probability p and initial distribution μ, then*

$$\begin{aligned} &P(X_0 \in B_0, \ldots, X_n \in B_n) \\ &\qquad = \int_{B_0} \cdots \int_{B_{n-1}} p(x_{n-1}, B_n)p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1)\mu(dx_0), \end{aligned}$$

for all $B_0, \ldots, B_n \in \mathcal{S}$.

Proof. Note that $X_{n+1} \mid \mathcal{F}_n \sim p(X_n)$, that is, $p(X_n)$ is a regular conditional distribution for X_{n+1} given $\mathcal{F}_n$. By Corollary 6.67, we have

$$E[f(X_{n+1}) \mid \mathcal{F}_n] = \int_S f(x)p(X_n, dx) \quad \text{a.s.}$$

for all measurable $f : S \to \mathbb{R}$ satisfying $E|f(X_{n+1})| < \infty$. We will prove that

$$E\bigg[\prod_{m=0}^n f_m(X_m)\bigg] = \int_{S^{n+1}} \bigg(\prod_{m=0}^n f_m(x_m)\bigg)p(x_{n-1}, dx_n) \cdots p(x_0, dx_1)\mu(dx_0),$$

for all bounded, measurable f_m. Taking $f_m = 1_{B_m}$ will finish the proof.

The claim is true for $n = 0$, since $X_0 \sim \mu$. Suppose the claim is true for some n. Then

$$\begin{aligned} E\bigg[\prod_{m=0}^{n+1} f_m(X_m)\bigg] &= E\bigg[E\bigg[\prod_{m=0}^{n+1} f_m(X_m)\,\bigg|\,\mathcal{F}_n\bigg]\bigg] \\ &= E\bigg[\bigg(\prod_{m=0}^{n} f_m(X_m)\bigg)E[f_{n+1}(X_{n+1}) \mid \mathcal{F}_n]\bigg] \\ &= E\bigg[\bigg(\prod_{m=0}^{n} f_m(X_m)\bigg)\int_S f_{n+1}(x)p(X_n, dx)\bigg] \\ &= E\bigg[\bigg(\prod_{m=0}^{n-1} f_m(X_m)\bigg)\widetilde{f}_n(X_n)\bigg], \end{aligned}$$

where

$$\begin{aligned} \widetilde{f}_n(x_n) &= f_n(x_n)\int_S f_{n+1}(x)p(x_n, dx) \\ &= f_n(x_n)\int_S f_{n+1}(x_{n+1})p(x_n, dx_{n+1}), \end{aligned}$$

which is a bounded, measurable function. Thus, by the inductive hypothesis,

$$\begin{aligned} &E\bigg[\prod_{m=0}^{n+1} f_m(X_m)\bigg] \\ &\quad= \int_{S^{n+1}} \bigg(\prod_{m=0}^{n-1} f_m(x_m)\bigg)\widetilde{f}_n(x_n)p(x_{n-1}, dx_n)\cdots p(x_0, dx_1)\mu(dx_0) \\ &\quad= \int_{S^{n+2}} \bigg(\prod_{m=0}^{n+1} f_m(x_m)\bigg)p(x_n, dx_{n+1})p(x_{n-1}, dx_n)\cdots p(x_0, dx_1)\mu(dx_0), \end{aligned}$$

and this completes the proof. □

Exercises

13.1. Fill in the details in the proof of Theorem 13.3.

13.2 Examples

This section corresponds to [2, Section 6.2].

Example 13.6. Let $Y = \{Y_n\}$ be a random walk on $\mathbb{R}^d$, let $X_0 = x \in \mathbb{R}^d$, and define $X_n = X_0 + Y_n$. Note that since X_0 is not random, we have $\mathcal{F}_n^Y = \mathcal{F}_n^X = \mathcal{F}_n^\xi$, where $\xi_n = Y_n - Y_{n-1}$. Define $p : \mathbb{R}^d \times \mathcal{R}^d \to \mathbb{R}$ by

$$p(x, A) = \int_{\mathbb{R}^d} 1_A(x + z)\mu(dz) = \mu(A - x),$$

where $\xi_1 \sim \mu$ and
$$A - x = \{y - x \in \mathbb{R}^d : y \in A\}.$$
Note that p is a transition probability (check). Also note that
$$P(X_{n+1} \in A \mid \mathcal{F}_n) = E[1_A(X_n + \xi_{n+1}) \mid \mathcal{F}_n] = h(X_n),$$
where
$$h(x) = E[1_A(x + \xi_{n+1})] = p(x, A).$$
Thus, $X = \{X_n\}$ is a Markov chain with transition probability p and initial distribution δ_x.

Henceforth, we will extend our definition of a random walk to include any Markov chain with the above transition probability.

Let us now consider the special case where S is countable and $\mathcal{S} = 2^S$. For each $i, j \in S$, let $p(i,j) \geqslant 0$ and assume $\sum_{j \in S} p(i,j) = 1$ for all $i \in S$. The indexed collection $(p(i,j))_{i,j \in S}$ can be thought of as a (possibly infinite) stochastic matrix. (A so-called “stochastic” matrix is just a matrix whose rows sum to 1.)

With an abuse of notation, define $p : S \times \mathcal{S} \to \mathbb{R}$ by
$$p(i, A) = \sum_{j \in A} p(i,j).$$
It can be verified that p is a transition probability on S, and that all transition probabilities on S can be written this way.

Note that a stochastic process $X = \{X_n\}$ is a Markov chain with transition probability p is and only if
$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = p(i,j),$$
for all $n, i, j, i_0, \ldots, i_{n-1}$.

Theorem 13.7. *Let $Z = \{Z_n\}$ be the branching process defined in Section 12.2. Then Z is a Markov chain with transition probability*
$$p(i,j) = P\bigg(\sum_{m=1}^{i} \xi_{1,m} = j\bigg).$$

Proof. Exercise 13.2. □

Example 13.8. Consider a queue with one server. At time 0, the 0th customer is just beginning to receive service, and there are x customers in the queue. Each customer's service time is independent of other customers, and has a distribution ν, supported on $(0, \infty)$ and with finite moments of all orders. New customers arrive in the queue according to a Poisson process with rate $\lambda > 0$.

We wish to construct a stochastic process $X = \{X_n\}$ so that X_n represents the number of customers in the queue at the moment the nth customer begins receiving service. (In particular, we must have $X_0 = x$.)

We begin with a heuristic discussion. When the nth customer begins receiving service, there are X_n customers in the queue. The duration of the service will be $T \sim \nu$. During that time, N new customers will arrive. Since new customers arrive according to a Poisson process with rate $\lambda > 0$, we have

$$P(N = k \mid T = t) = e^{-\lambda} \frac{(\lambda t)^k}{k!}$$

Thus,

$$P(N = k) = E[P(N = k \mid T)] = E\left[e^{-\lambda} \frac{(\lambda T)^k}{k!}\right] = \int_0^\infty e^{-\lambda} \frac{(\lambda t)^k}{k!} \nu(dt) =: a_k,$$

for all $k \in \mathbb{N} \cup \{0\}$. Note that each $a_k > 0$. If $X_n + N > 0$, then we would have $X_{n+1} = X_n + N - 1$, since the customer at the front of the queue would step out and into service. On the other hand, if $X_n + N = 0$, this means the queue was empty when the nth customer began service, and no one arrived during that customer's service time. Thus, the size of the queue at the beginning of the $(n + 1)$th customer's service would be 0, that is, $X_{n+1} = 0$.

We turn this reasoning into a formal stochastic process as follows. Let $X_0 = x$. Let $\xi_1, \xi_2, \dots$ be i.i.d. $\{-1, 0, 1, 2, \dots\}$-valued random variables with $P(\xi_1 = k) = a_{k+1}$. For $n \in \mathbb{N}$, define $X_n = (X_{n-1} + \xi_n)^+$.

It can be shown (with details left to the reader) that X is a Markov chain with respect to $\{\mathcal{F}_n^\xi\}$ with transition probability

$$p(i, j) = \begin{cases} a_0 + a_1 & \text{if } i = 0,\ j = 0, \\ a_{j+1} & \text{if } i = 0,\ j \geqslant 1, \\ a_{j-(i-1)} 1_{\{j \geqslant i-1\}} & \text{if } i \geqslant 1. \end{cases}$$

In fact, the above is a transition probability for any sequence $\{a_k\}_{k=0}^\infty$ with $a_k > 0$ for all k and $\sum_k a_k = 1$.

This process is called an *M/G/1 queue*. The "M" refers to the fact that it is a Markov process, which comes from the assumption that the arrivals follow a Poisson process. The "G" (for "General") refers to the fact that the distribution of the service time, ν, is allowed to be any probability distribution. And the "1" refers to the fact that there is only one server.

Let X be a Markov chain on a countable state space S with transition probability $p(i, j)$ and initial distribution μ. We claim that

$$P_\mu(X_n = i_n, \dots, X_0 = i_0) = \mu(i_0) \prod_{m=1}^n p(i_{m-1}, i_m).$$

To see this, note that the equation holds for $n = 0$ since $X_0 \sim \mu$. Assume it holds for some n. Then

$$\begin{aligned}
P_\mu(X_{n+1} = i_{n+1}, \ldots, X_0 = i_0) &= E_\mu[P_\mu(X_{n+1} = i_{n+1}, \ldots, X_0 = i_0 \mid \mathcal{F}_n)] \\
&= E_\mu[1_{\{X_n = i_n, \ldots, X_0 = i_0\}} P_\mu(X_{n+1} = i_{n+1} \mid \mathcal{F}_n)] \\
&= E_\mu[1_{\{X_n = i_n, \ldots, X_0 = i_0\}} P_\mu(X_{n+1} = i_{n+1} \mid X_n)] \\
&= E_\mu[1_{\{X_n = i_n, \ldots, X_0 = i_0\}} p(X_n, i_{n+1})] \\
&= E_\mu[1_{\{X_n = i_n, \ldots, X_0 = i_0\}} p(i_n, i_{n+1})] \\
&= p(i_n, i_{n+1}) P_\mu(X_n = i_n, \ldots, X_0 = i_0),
\end{aligned}$$

and the result then follows from the inductive hypothesis.

Recall that $p : S^2 \to \mathbb{R}$ can be informally regarded as a (possibly infinite) matrix. Along these lines, let us define $p^0 = I$, where $I : S^2 \to \mathbb{R}$ is defined by $I(i,j) = 1_{\{i=j\}}$. We then define $p^n : S^2 \to \mathbb{R}$ by

$$p^n(i,j) = \sum_{k \in S} p^{n-1}(i,k) p(k,j).$$

The fact that this kind of "matrix" multiplication works more generally follows from Remark 13.12.

Similarly, $\mu \in M_1(S)$ can be considered a function from S to $\mathbb{R}$ where $\mu(j) = \mu(\{j\})$. Informally regarding μ as a (possibly infinite) row vector, we define $\mu p^n : S \to \mathbb{R}$ by

$$\mu p^n(j) = \sum_{k \in S} \mu(k) p^n(k,j).$$

Note that $\mu p^0 = \mu$. Also note that $\delta_x p^n(j) = p^n(x,j)$.

With this notation, we claim that

$$P_\mu(X_n = j) = \mu p^n(j).$$

To see this, note that the equation holds for $n = 0$ since $X_0 \sim \mu$. Assume it holds for some n. Then

$$\begin{aligned}
P_\mu(X_{n+1} = j) &= \sum_{k \in S} P_\mu(X_n = k) P_\mu(X_{n+1} = j \mid X_n = k) \\
&= \sum_{k \in S} \mu p^n(k) p(k,j) \\
&= \sum_{k \in S} \sum_{\ell \in S} \mu(\ell) p^n(\ell, k) p(k,j) \\
&= \sum_{\ell \in S} \mu(\ell) \sum_{k \in S} p^n(\ell, k) p(k,j) \\
&= \sum_{\ell \in S} \mu(\ell) p^{n+1}(\ell, j) = \mu p^{n+1}(j),
\end{aligned}$$

and the claim is proven by induction.

As a special case of this formula, we have

$$P_x(X_n = j) = p^n(x,j) \tag{13.3}$$

for all $x, j \in S$ and all $n \in \mathbb{N} \cup \{0\}$.

Exercises

13.2. Prove Theorem 13.7.

13.3. [2, Exercise 6.2.8] Let X_n be a simple random walk on $\mathbb{R}$ and define $Y_n = \max_{0 \leqslant m \leqslant n} X_m$. Prove that $\{Y_n\}$ is not a Markov chain.

13.3 Extensions of the Markov property

This section corresponds to [2, Section 6.3].

Recall that the general Markov property is given by (13.1). However, as stated previously, we are assuming that all our Markov chains are time homogeneous with a state space that is a standard Borel space. We further assume, as in the proof of Theorem 13.3, that our Markov process is built by constructing a probability measure P_μ on the canonical space $(\Omega, \mathcal{F}) = (S^{\mathbb{N}\cup\{0\}}, \mathcal{S}^{\mathbb{N}\cup\{0\}})$, so that the identity process $X = \{X_n\}$, where $X_n(\omega) = \omega_n$, is a Markov process with transition probability p and initial distribution μ. Because of this, we are free to use the shift operator θ, defined just prior to Example 11.9.

In this setting, the following theorem gives a useful reformulation of the Markov property.

Theorem 13.9 (Markov property)**.** *If Y is a bounded random variable, then for all $n \in \mathbb{N} \cup \{0\}$,*

$$E_\mu[Y \circ \theta^n \mid \mathcal{F}_n] = E_{X_n} Y,$$

where $E_{X_n} Y = \varphi(X_n)$, with $\varphi(x) = E_x Y$.

Proof. See [2, Theorem 6.3.1] □

Remark 13.10. Since Y is a random variable, we may write

$$Y = Y(\omega) = Y(\omega_0, \omega_1, \omega_2, \ldots).$$

But $X_n(\omega) = \omega_n$, so

$$Y(\omega) = Y(X_0(\omega), X_1(\omega), X_2(\omega), \ldots),$$

which is the same as

$$Y = Y(X_0, X_1, X_2 \ldots).$$

In other words, Y is just a function of the path of our Markov process. Also,

$$Y \circ \theta^n = Y(X_n, X_{n+1}, X_{n+2}, \ldots)$$

is that same function applied to the part of the path that starts at time n.

Hence, in words, the expected value of $Y(X_n, X_{n+1}, X_{n+2}, ...)$ given $\mathcal{F}_n$ is equal to the expected value of $Y(X_0, X_1, X_2, \ldots)$ starting at X_n.

Theorem 13.11 (Chapman-Kolmogorov equation)**.** *If S is countable, then*

$$P_x(X_{m+n} = z) = \sum_{y \in S} P_x(X_m = y) P_y(X_n = z),$$

for all $x, z \in S$ and all $m, n \in \mathbb{N} \cup \{0\}$.

Proof. We begin with

$$P_x(X_{m+n} = z) = E_x[P_x(X_{m+n} = z \mid \mathcal{F}_m)]$$

According to the Markov property,

$$\begin{aligned} P_x(X_{m+n} = z \mid \mathcal{F}_m) &= E_x[1_{\{X_{m+n}=z\}} \mid \mathcal{F}_m] \\ &= E_x[1_{\{X_n=z\}} \circ \theta^m \mid \mathcal{F}_m] = E_{X_m}[1_{\{X_n=z\}}] = P_{X_m}(X_n = z). \end{aligned}$$

This last quantity is to be understood as $\varphi(X_m)$, where $\varphi(x) = P_x(X_n = z)$. Thus,

$$\begin{aligned} P_x(X_{m+n} = z) &= E_x[\varphi(X_m)] \\ &= \sum_{y \in S} P_x(X_m = y)\varphi(y) = \sum_{y \in S} P_x(X_m = y) P_y(X_n = z), \end{aligned}$$

which is what we wanted to prove. □

Remark 13.12. By (13.3), the Chapman-Kolmogorov equation can be rewritten as

$$p^{m+n}(x, z) = \sum_{y \in S} p^m(x, y) p^n(y, z).$$

Theorem 13.13 (strong Markov property)**.** *Let $\{Y_n\}$ be a sequence of random variables. Suppose there exists $M > 0$ such that $|Y_n| \leqslant M$ a.s. for all n. Let N be an $\{\mathcal{F}_n\}$-stopping time, where $\mathcal{F}_n = \mathcal{F}_n^X$. Then*

$$E_\mu[Y_N \circ \theta^N \mid \mathcal{F}_N] 1_{\{N<\infty\}} = (E_{X_N} Y_N)\, 1_{\{N<\infty\}},$$

where $E_{X_N} Y_N = \varphi_N(X_N)$, with $\varphi_n(x) = E_x Y_n$.

Proof. Let $A \in \mathcal{F}_N$. Then

$$\begin{aligned} E_\mu[(Y_N \circ \theta^N) 1_{\{N<\infty\}} 1_A] &= \sum_{n=0}^\infty E_\mu[(Y_n \circ \theta^n) 1_{\{N=n\}} 1_A] \\ &= \sum_{n=0}^\infty E_\mu[E_\mu[(Y_n \circ \theta^n) 1_{\{N=n\}} 1_A \mid \mathcal{F}_n]]. \end{aligned}$$

By the definition of $\mathcal{F}_N$, we have $A \cap \{N = n\} \in \mathcal{F}_n$. Thus, by the Markov property,

$$\begin{aligned} E_\mu[(Y_n \circ \theta^n) 1_{\{N=n\}} 1_A \mid \mathcal{F}_n] &= E_\mu[Y_n \circ \theta^n \mid \mathcal{F}_n] 1_{\{N=n\}} 1_A \\ &= \varphi_n(X_n) 1_{\{N=n\}} 1_A, \end{aligned}$$

where $\varphi_n(x) = E_x[Y_n]$. Putting these together gives

$$\begin{aligned} E_\mu[(Y_N \circ \theta^N)1_{\{N<\infty\}}1_A] &= \sum_{n=0}^{\infty} E_\mu[\varphi_n(X_n)1_{\{N=n\}}1_A] \\ &= E_\mu[\varphi_N(X_N)1_{\{N<\infty\}}1_A]. \end{aligned}$$

Since $A \in \mathcal{F}_N$ was arbitrary and $\varphi_N(X_N) \in \mathcal{F}_N$, this shows that

$$\begin{aligned} \left(E_\mu[Y_N \circ \theta^N \mid \mathcal{F}_N]\right) 1_{\{N<\infty\}} &= E_\mu[(Y_N \circ \theta^N)1_{\{N<\infty\}} \mid \mathcal{F}_N] \\ &= \varphi_N(X_N)1_{\{N<\infty\}} \\ &= (E_{X_N} Y_N)\, 1_{\{N<\infty\}}, \end{aligned}$$

and we are done. □

Theorem 13.14 (reflection principle)**.** *Let X be a symmetric random walk on $\mathbb{R}$. Then*

$$P\left(\sup_{m\leqslant n} X_m > a\right) \leqslant 2P(X_n > a) \tag{13.4}$$

for all $a > 0$ and all $n \in \mathbb{N}$.

Remark 13.15. Equation (13.4) is equivalent to

$$P\left(\sup_{m\leqslant n} X_m \geqslant a\right) \leqslant 2P(X_n \geqslant a). \tag{13.5}$$

To see this, let $a_k \uparrow a$ in (13.4) and $a_k \downarrow a$ in (13.5).

Remark 13.16. The idea of the proof, and the origin of the name of the theorem is the following. Let $N = \inf\{n : X_n \geqslant a\}$. Then the left-hand side of (13.5) is simply $P(N \leqslant n)$. Let us suppose, for the purposes of this heuristic discussion, that $X_N = a$ a.s. and $P(X_n = a) = 0$.

Let us decompose the event $\{N \leqslant n\}$ into the two events,

$$U = \{N \leqslant n, X_n > a\} \quad \text{and} \quad L = \{N \leqslant n, X_n < a\}.$$

Note that $U = \{X_n > a\}$. Consider a particular sample path in U with steps $(\xi_1, \dots, \xi_N, \xi_{N+1}, \dots, \xi_n)$. From this we can create a corresponding "reflected" sample paths, with steps $(\xi_1, \dots, \xi_N, -\xi_{N+1}, -\xi_{N+2}, \dots, -\xi_n)$. The reflected sample path is in L and, in fact, this process of reflection creates a one-to-one correspondence between U and L. Moreover, because the random walk is symmetric, the original path and the reflected path have the same probability. Thus,

$$P(L) = P(U) = P(X_n > a) = P(X_n \geqslant a),$$

which gives

$$P\left(\sup_{m\leqslant n} X_n \geqslant a\right) = P(N \leqslant n) = P(U) + P(L) = 2P(U) = 2P(X_n \geqslant a).$$

Of course, this is not exactly correct, not only because the argument was purely heuristic, but also because of our simplifying assumption. In particular, we do not have $X_N = a$ a.s. Rather, the most we can say is that $X_N \geqslant a$ a.s.

Proof of Theorem 13.14. Fix $n \in \mathbb{N}$. Let $N = \inf\{m : X_m > a\}$. For $m \leqslant n$, define $Y_m = 1_{\{X_{n-m} > a\}}$. For $m > n$, let $Y_m = 0$. Note that $Y_m \circ \theta^m = 1_{\{X_n > a\}} 1_{\{m \leqslant n\}}$, so that

$$(Y_N \circ \theta^N) 1_{\{N < \infty\}} = 1_{\{X_n > a, N \leqslant n\}} = 1_{\{X_n > a\}}.$$

By the strong Markov property,

$$\begin{aligned} P(X_n > a \mid \mathcal{F}_N) &= E[(Y_N \circ \theta^N) 1_{\{N < \infty\}} \mid \mathcal{F}_N] \\ &= E[Y_N \circ \theta^N \mid \mathcal{F}_N] 1_{\{N < \infty\}} \\ &= \varphi_N(X_N) 1_{\{N < \infty\}}, \end{aligned}$$

where $\varphi_m(x) = E_x Y_m$. Thus, $\varphi_m(x) = 0$ for $m > n$. For $m \leqslant n$ and $x > a$, we have,

$$\varphi_m(x) = E_x Y_m = P_x(X_{n-m} > a) \geqslant P_x(X_{n-m} \geqslant x) \geqslant 1/2,$$

where the last inequality comes from the fact that X is a symmetric random walk. Since $X_N > a$ a.s. on the event $\{N < \infty\}$, we have

$$\begin{aligned} P(X_n > a) &= E[P(X_n > a \mid \mathcal{F}_N)] \\ &= E[\varphi_N(X_N) 1_{\{N < \infty\}}] \\ &= E[\varphi_N(X_N) 1_{\{N \leqslant n\}}] \\ &\geqslant \frac{1}{2} P(N \leqslant n) \end{aligned}$$

Since $\{N \leqslant n\} = \{\sup_{m \leqslant n} X_m > a\}$, this completes the proof. □

Exercises

13.4. Prove the following version of the strong Markov property: Let $X = \{X_n\}$ be a Markov chain with respect to $\{\mathcal{F}_n\}$ taking values in $(S, \mathcal{S})$. Let N be an $\{\mathcal{F}_n\}$-stopping time. Then, for all $m \in \mathbb{N}$ and all $B \in \mathcal{S}$,

$$P(X_{N+m} \in B \mid \mathcal{F}_N) = P(X_{N+m} \in B \mid X_N) \quad \text{a.s. on } \{N < \infty\}.$$

Note: When conditioning on X_N above, we are conditioning on the σ-algebra consisting of sets of the form $\{N < \infty\} \cap \{X_N \in A\}$ and $(\{N < \infty\} \cap \{X_N \in A\}) \cup \{N = \infty\}$, where $A \in \mathcal{S}$.

13.5. [2, Exercise 6.3.1] Let X be a (time-homogeneous) Markov chain on a (standard Borel) space S. Prove that the past and the future are conditionally independent given the present. More specifically, let $A \in \sigma(X_0, \dots, X_n)$ and $B \in \sigma(X_n, X_{n+1}, \dots)$. Prove that for any initial distribution μ, we have

$$P_\mu(A \cap B \mid X_n) = P_\mu(A \mid X_n) P_\mu(B \mid X_n).$$

13.4 Recurrence and transience

This section corresponds to [2, Section 6.4].

For the remainder of this chapter, we shall assume S is countable.

Let X be a Markov chain on S. For $y \in S$ and $k \in \mathbb{N}$, let $T_y^0 = 0$ and

$$T_y^k = \inf\{n > T_y^{k-1} : X_n = y\}.$$

For $k \in \mathbb{N}$, the stopping time T_y^k is the time of the kth return to y. (If it happens to be the case that $X_0 = y$, this does not count as a return to y.)

Let $T_y = T_y^1$ and, for $x, y \in S$, let $\rho_{xy} = P_x(T_y < \infty)$ be the probability, starting at x that X eventually visits/returns to y.

Theorem 13.17. *If $x, y \in S$ and $k \in \mathbb{N}$, then $P_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}$.*

Proof. The result is trivially true for $k = 1$. Suppose it is true for some $k \in \mathbb{N}$. Note that

$$(1_{\{T_y<\infty\}} \circ \theta^{T_y^k})1_{\{T_y^k<\infty\}} = 1_{\{T_y^{k+1}<\infty\}}.$$

By the strong Markov property,

$$\begin{aligned} P_x(T_y^{k+1} < \infty \mid \mathcal{F}_{T_y^k}) &= E_x[1_{\{T_y<\infty\}} \circ \theta^{T_y^k} \mid \mathcal{F}_{T_y^k}]1_{\{T_y^k<\infty\}} \\ &= P_{X_{T_y^k}}(T_y < \infty)1_{\{T_y^k<\infty\}} \\ &= P_y(T_y < \infty)1_{\{T_y^k<\infty\}} \\ &= \rho_{yy}1_{\{T_y^k<\infty\}}. \end{aligned}$$

Thus,

$$P_x(T_y^{k+1} < \infty) = E_x[P_x(T_y^{k+1} < \infty \mid \mathcal{F}_{T_y^k})] = \rho_{yy}P_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^k,$$

by the induction hypothesis. □

A state $y \in S$ is **recurrent** if $\rho_{yy} = 1$ and **transient** if $\rho_{yy} < 1$. If y is recurrent, then $P_y(T_y^k < \infty) = \rho_{yy}^k = 1$, which implies $P_y(X_n = y \text{ i.o.}) = 1$.

Theorem 13.18. *Let $y \in S$. Then y is recurrent if and only if $E_yN(y) = \infty$, where*

$$N(y) = \sum_{n=1}^{\infty} 1_{\{X_n=y\}}$$

is the number of returns to y.

Proof. Note that

$$E_xN(y) = \sum_{k=1}^{\infty} P_x(N(y) \geqslant k) = \sum_{k=1}^{\infty} P_x(T_y^k < \infty) = \sum_{k=1}^{\infty} \rho_{xy}\rho_{yy}^{k-1}.$$

If $x = y$ and y is recurrent, so that $\rho_{yy} = 1$, then this sum is infinite. If y is transient, then

$$E_xN(y) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty$$

for any $x \in S$. □

Theorem 13.19. *Let $x, y \in S$. Suppose x is recurrent and $\rho_{xy} > 0$. Then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.*

Proof. Let $x, y \in S$. Suppose x is recurrent and $\rho_{xy} > 0$. Assume $\rho_{yx} < 1$. Note that

$$\begin{aligned} p^k(x,y) &= P_x(X_k = y) \\ &= \sum_{y_1,\dots,y_{k-1}\in S} P_x(X_1 = y_1, \dots, X_{k-1} = y_{k-1}, X_k = y) \\ &= \sum_{y_1,\dots,y_{k-1}\in S} p(x,y_1)p(y_1,y_2)\cdots p(y_{k-1},y). \end{aligned}$$

Let $K = \inf\{k : p^k(x,y) > 0\}$. Since $\rho_{xy} > 0$, it follows that $K < \infty$. By the above, there exists $y_1, \dots, y_{K-1} \in S$ such that

$$p(x,y_1)p(y_1,y_2)\cdots p(y_{K-1},y) > 0.$$

Suppose $y_j = x$ for some $j \in \{1, \dots, K-1\}$. Then

$$p(x,y_{j+1})p(y_{j+2},y_{j+3})\cdots p(y_{K-1},y) > 0,$$

which implies $p^{K-j}(x,y) > 0$. But this contradicts the minimality of K. Therefore $y_j \neq x$ for all $j \in \{1, \dots, K-1\}$. Since x is recurrent,

$$\begin{aligned} 0 = P_x(T_x = \infty) &\geqslant P_x(X_1 = y_1, \dots, X_{K-1} = y_{K-1}, X_K = y, T_x = \infty) \\ &= P_x(X_1 = y_1, \dots, X_{K-1} = y_{K-1}, X_K = y, T_x \circ \theta^K = \infty) \\ &= E_x[1_{\{X_1=y_1,\dots,X_{K-1}=y_{K-1},X_K=y\}} P_x(T_x \circ \theta^K = \infty \mid \mathcal{F}_K)] \\ &= E_x[1_{\{X_1=y_1,\dots,X_{K-1}=y_{K-1},X_K=y\}} P_{X_K}(T_x = \infty)] \\ &= E_x[1_{\{X_1=y_1,\dots,X_{K-1}=y_{K-1},X_K=y\}} P_y(T_x = \infty)] \\ &= (1-\rho_{yx}) E_x[1_{\{X_1=y_1,\dots,X_{K-1}=y_{K-1},X_K=y\}}] \\ &= (1-\rho_{yx}) p(x,y_1)p(y_1,y_2)\cdots p(y_{K-1},y) > 0, \end{aligned}$$

a contradiction. Thus, $\rho_{yx} = 1$.

Now, since x is recurrent, we have

$$\infty = E_x N(x) = E_x \sum_{n=1}^{\infty} 1_{\{X_n = x\}} = \sum_{n=1}^{\infty} P_x(X_n = x) = \sum_{n=1}^{\infty} p^n(x,x).$$

Since $\rho_{yx} = 1$, we may choose $L \in \mathbb{N}$ such that $p^L(y,x) > 0$. For any $n \in \mathbb{N}$, we have

$$\begin{aligned} p^{L+n+K}(y,y) &= P_y(X_{L+n+K} = y) \\ &\geqslant P_y(X_L = x, X_{L+n} = x, X_{L+n+K} = y) \\ &= p^L(y,x)p^n(x,x)p^K(x,y). \end{aligned}$$

Thus,

$$E_y N(y) \geqslant \sum_{n=1}^{\infty} p^{L+n+K}(y,y) \geqslant p^L(y,x)p^K(x,y) \sum_{n=1}^{\infty} p^n(x,x).$$

Since $p^L(y,x)$ and $p^K(x,y)$ are strictly postive, this implies $E_y N(y) = \infty$. Thus, y is recurrent.

Finally, applying what we have proven so far to y and ρ_{yx} proves that $\rho_{xy} = 1$. □

Let $C \subset S$. Suppose that whenever $x \in C$ and $\rho_{xy} > 0$, we have $y \in C$. Then C is **closed**. Note that if $x \in C$, then $P_x(X_n \in C) = 1$ for all n (check). Let $D \subset S$. Suppose that, for all $x, y \in D$, we have $\rho_{xy} > 0$. Then D is **irreducible**.

Theorem 13.20. *If C is finite and closed, then there exists $x \in C$ such that x is recurrent. If C is also irreducible, then every $x \in C$ is recurrent.*

Proof. Let C be finite and closed. Suppose that for all $y \in C$, we have that y is transient. Fix $x \in C$. Then

$$\sum_{y \in C} E_x N(y) = \sum_{y \in C} \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty.$$

On the other hand,

$$\sum_{y \in C} E_x N(y) = \sum_{y \in C} \sum_{n=1}^{\infty} P_x(X_n = y) = \sum_{n=1}^{\infty} \sum_{y \in C} P_x(X_n = y) = \sum_{n=1}^{\infty} P_x(X_n \in C).$$

But $x \in C$ and C is closed, so $P_x(X_n \in C) = 1$ for all n, which is a contradiction. Hence, there exists $x \in C$ such that x is recurrent.

Suppose C is also irreducible. Let $y \in C$. Then $\rho_{xy} > 0$, so by Theorem 13.19, y is recurrent. □

Example 13.21. Let X be a Markov chain with state space $S = \{1, \dots, 7\}$ and transition probability

$$(p(i,j)) = \begin{pmatrix} 0.3 & 0 & 0 & 0 & 0.7 & 0 & 0 \\ 0.1 & 0.2 & 0.3 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

It can help to visualize the dynamics of X by considering a directed graph with an edge from i to j if $p(i,j) > 0$ (ignoring the cases where $i = j$). (See [2, Figure 6.4].)

Note that $\rho_{21} > 0$. If 2 were recurrent, this would imply that 1 is also recurrent and that $\rho_{12} = 1$. But $\rho_{12} = 0$, so it follows that 2 is transient. Similarly, $\rho_{34} > 0$, but $\rho_{43} = 0$, and we have that 3 is transient.

Since $\{1,5\}$ is an irreducible closed set, both these states are recurrent. Similarly for $\{4,6,7\}$.

Theorem 13.22. *Let X be a Markov chain and let $R \subset S$ be the recurrent states. Then R has a unique decomposition as $R = \biguplus_j R_j$, where each R_j is closed and irreducible.*

Proof. See [2, Theorem 6.4.5] □

Exercises

13.6. [2, Exercise 6.4.4] Prove that $\rho_{xz} \geqslant \rho_{xy}\rho_{yz}$ for all $x, y, z \in S$.

13.5 Stationary and limiting measures

This section corresponds to [2, Sections 6.5 and 6.6].

Let X be a Markov chain on S with transition probability $p(i,j)$. A measure μ on S is a **stationary measure** for X if

$$\sum_{i \in S} \mu(i)p(i,j) = \mu(j),$$

for all $j \in S$. A measure μ on S is a **stationary distribution** for X if it is a stationary measure and a probability distribution.

If μ is a stationary distribution for X, then

$$P_\mu(X_n = j) = \mu(j),$$

for all $j \in S$ and all $n \in \mathbb{N}$ (check).

Example 13.23. Let $f : \mathbb{Z}^d \to [0,1]$ satisfy $\sum_{z \in \mathbb{Z}^d} f(z) = 1$. Then $p(i,j) = f(j-i)$ is a transition probability on $\mathbb{Z}^d$ (check). Let μ be the measure on $\mathbb{Z}^d$ such that $\mu(j) = 1$ for all $j \in \mathbb{Z}^d$. Then

$$\sum_{i \in \mathbb{Z}^d} \mu(i)f(i-j) = \sum_{i \in \mathbb{Z}^d} f(i-j) = 1 = \mu(j).$$

Thus, μ is a stationary measure for the Markov chain with transition probability p.

Example 13.24. Let X be a random walk on $\mathbb{R}^d$ with $\xi_1 \in \mathbb{Z}^d$ a.s. Hence, X is actually a random walk on $\mathbb{Z}^d$. Let $f(z) = P(\xi_1 = z)$. Then X has transition probability $p(i,j) = f(j-i)$, so that $\mu \equiv 1$ is a stationary measure for X.

Example 13.25. Let X be an asymmetric simple random walk on $\mathbb{Z}$. By the above, $\mu \equiv 1$ is a stationary measure for X. In this example, we will construct another.

Let μ be the measure on $\mathbb{Z}$ satisfying $\mu(j) = (p/q)^j$ for all $j \in \mathbb{Z}$. Then

$$\begin{aligned}\sum_{i\in\mathbb{Z}} \mu(i)p(i,j) &= \mu(j-1)p(j-1,j) + \mu(j+1)p(j+1,j)\\ &= (p/q)^{j-1}p + (p/q)^{j+1}q\\ &= (p/q)^j(q+p)\\ &= (p/q)^j\\ &= \mu(j).\end{aligned}$$

Thus, μ is a stationary measure for X.

Example 13.26. Let S be countable. Let $a : S \times S \to \{0,1\}$ satisfy $a_{ij} = a_{ji}$ for all $i,j \in S$, and $a_{ii} = 0$ for all $i \in S$. Such a choice gives rise to an undirected graph $G = (S,E)$, where $\{i,j\} \in E$ if and only if $a_{ij} = 1$. The graph G has no loops.

Assume that each vertex belongs to only finitely many edges. That is, assume that

$$\mu(i) := \sum_{j\in S} a_{ij} < \infty,$$

for all $i \in S$. Note that $\mu(i)$ is the number of edges to which i belongs. Let us define

$$p(i,j) = \frac{a_{ij}}{\mu(i)},$$

for all $i,j \in S$. Then p is a transition probability on S (check). A Markov chain X with transition probability p is a random walk on G which, at each time n, selects uniformly from among the edges to which X_n belongs, and then moves along that edge to the next vertex.

Note that

$$\sum_{i\in S} \mu(i)p(i,j) = \sum_{i\in S} a_{ij} = \sum_{i\in S} a_{ji} = \mu(j),$$

so μ is a stationary measure for X.

Theorem 13.27. *Let X be a Markov chain on S. Suppose $x \in S$ is recurrent. Let $T = \inf\{n \geqslant 1 : X_n = x\}$. For $y \in S$, let*

$$\mu_x(y) = E_x\left[\sum_{n=0}^{T-1} 1_{\{X_n=y\}}\right]$$

be the expected number of visits to y before the first return to x. Then μ_x is a stationary measure for X.

Proof. See [2, Theorem 6.5.2]. □

Remark 13.28. Note that

$$\mu_x(y) = E_x\left[\sum_{n=0}^{\infty} 1_{\{X_n=y\}}1_{\{T>n\}}\right] = \sum_{n=0}^{\infty} P_x(X_n = y, T > n),$$

for all $x, y \in S$.

Theorem 13.29. *Let X be a Markov chain on S. Suppose that S is irreducible and every $x \in S$ is recurrent. Then the stationary measure is unique up to constant multiples.*

Proof. See [2, Theorem 6.5.3]. □

Theorem 13.30. *Let X be a Markov chain on S. Suppose there exists a stationary distribution π. Then, for all $y \in S$, if $\pi(y) > 0$, then y is recurrent.*

Proof. Suppose $\pi(y) > 0$ and y is transient, so that $\rho_{yy} < 1$. Then

$$\sum_{x\in S} \pi(x)E_xN(y) = \sum_{x\in S} \frac{\pi(x)\rho_{xy}}{1-\rho_{yy}} \leqslant \sum_{x\in S} \frac{\pi(x)}{1-\rho_{yy}} = \frac{1}{1-\rho_{yy}} < \infty.$$

On the other hand,

$$\begin{aligned}\sum_{x\in S} \pi(x)E_xN(y) &= \sum_{x\in S} \pi(x) \sum_{n=1}^{\infty} P_x(X_n = y)\\ &= \sum_{n=1}^{\infty}\sum_{x\in S} \pi(x)p^n(x, y)\\ &= \sum_{n=1}^{\infty} \pi p^n(y).\end{aligned}$$

By the definition of a stationary distribution, $\pi p = \pi$. By induction, then, $\pi p^n = \pi$ for all $n \in \mathbb{N}$. Thus,

$$\sum_{x\in S} \pi(x)E_xN(y) = \sum_{n=1}^{\infty} \pi(y) = \infty,$$

a contradiction. □

Theorem 13.31. *Let X be a Markov chain on S. Suppose that S is irreducible and that X has a stationary distribution π. Then*

$$\pi(x) = \frac{1}{E_xT_x}$$

for all $x \in S$, where $T_x = \inf\{n \geqslant 1 : X_n = x\}$.

Proof. Fix $x \in S$. Since π is a probability measure, there exists $y \in S$ such that $\pi(y) > 0$. Since S is irreducible, $\rho_{yx} > 0$. Thus, there exists $n \in \mathbb{N}$ such that $p^n(y,x) > 0$. We therefore have

$$\pi(x) = \pi p^n(x) = \sum_{z \in S} \pi(z) p^n(z,x) \geqslant \pi(y) p^n(y,x) > 0.$$

We have thus proved that $\pi(x) > 0$ for all $x \in S$. It follows that every $x \in S$ is recurrent.

Fix $x \in S$. By Theorem 13.27 and Remark 13.28,

$$\mu_x(y) = E_x\left[\sum_{n=0}^{T_x - 1} 1_{\{X_n = y\}}\right] = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$$

is a stationary measure for X with $\mu_x(x) = 1$ and

$$\sum_{y \in S} \mu_x(y) = \sum_{n=0}^{\infty} P_x(T_x > n) = E_x T_x.$$

Thus, $\mu_x / E_x T_x$ is a stationary distribution. By Theorem 13.29, the stationary measure is unique up to constant multiples. Thus, $\pi = \mu_x / E_x T_x$, and it follows that

$$\pi(x) = \frac{\mu_x(x)}{E_x T_x} = \frac{1}{E_x T_x},$$

and we are done. □

Let X be a Markov chain on S. Let $x \in S$ be recurrent and define $I_x = \{n \geqslant 1 : p^n(x,x) > 0\}$. Let d_x be the greatest common divisor of I_x. Then d_x is called the **period** of x.

For example, let X be a simple random walk on $\mathbb{Z}$ and $x = 0$. Then $I_0 = \{2, 4, 6, \ldots\}$ and the period of 0 is $d_0 = 2$.

By [2, Lemma 6.6.2], if $\rho_{xy} > 0$, then $d_x = d_y$. Thus, if S is irreducible, then $d_x = d_y$ for all $x, y \in S$. If this common value is 1, then X is said to be **aperiodic**.

Theorem 13.32. *Let X be a Markov chain on S. Assume S is irreducible, X is aperiodic, and there exists a stationary distribution π. Then $p^n(x,y) \to \pi(y)$ as $n \to \infty$ for all $x, y \in S$. More specifically,*

$$\sum_{y \in S} |p^n(x,y) - \pi(y)| \to 0,$$

as $n \to \infty$ for all $x \in S$.

Proof. See [2, Theorem 6.6.4] and its proof. □

Corollary 13.33. *Let X be a Markov chain on S. Assume S is irreducible, X is aperiodic, and there exists a stationary distribution π. Let Y be an S-valued random variable with distribution π. Then $X_n \Rightarrow Y$ as $n \to \infty$ under P_μ, for any initial distribution μ.*

Proof. Let $f : S \to \mathbb{R}$ be bounded with $|f| \leqslant M$. Then

$$\begin{aligned}
|E[f(X_n)] - E[f(Y)]| &= \left| \sum_{y \in S} f(y) P_\mu(X_n = y) - \sum_{y \in S} f(y)\pi(y) \right| \\
&= \left| \sum_{y \in S} f(y) \sum_{x \in S} \mu(x) p^n(x,y) - \sum_{y \in S} f(y) \sum_{x \in S} \mu(x)\pi(y) \right| \\
&= \left| \sum_{y \in S} f(y) \sum_{x \in S} \mu(x)(p^n(x,y) - \pi(y)) \right| \\
&\leqslant M \sum_{x \in S} \mu(x) \sum_{y \in S} |p^n(x,y) - \pi(y)|.
\end{aligned}$$

Note that

$$\mu(x) \sum_{y \in S} |p^n(x,y) - \pi(y)| \leqslant \mu(x) \sum_{y \in S} (p^n(x,y) + \pi(y)) = 2\mu(x),$$

and $\sum_{x \in S} \mu(x) = 1$. Also, by Theorem 13.32,

$$\mu(x) \sum_{y \in S} |p^n(x,y) - \pi(y)| \to 0$$

as $n \to \infty$ for each $x \in S$. Thus, by dominated convergence,

$$|E[f(X_n)] - E[f(Y)]| \to 0,$$

showing that $X_n \Rightarrow Y$. □

Exercises

13.7. [2, Exercise 6.5.8] Compute the expected number of moves it takes a knight to return to its initial position if it starts in a corner of the chessboard, assuming there are no other pieces on the board, and each time it chooses a move uniformly from among its legal moves.

Part V

Continuous-time Stochastic Processes

Chapter 14

Continuous-time Martingales

14.1 Continuous-time stochastic processes

This section corresponds to [8, Section 1.1].

Recall the following from Section 6.4. A stochastic process is a collection of random variables $\{X(t) : t \in T\}$ indexed by some set T, defined on a common probability space, $(\Omega, \mathcal{F}, P)$, and taking values in a common measurable space, $(S, \mathcal{S})$. We may occasionally use the notation X_t instead of $X(t)$. Recall also that we adopt the notation $X(t, \omega) = (X(t))(\omega)$.

For fixed $\omega \in \Omega$, the function $X(\cdot, \omega)$ is called a **sample path** and is an an element of S^T. The set S^T coincides with the product space $\prod_{t \in T} S$. As such, we can endow it with product σ-algebra,

$$\mathcal{S}^T = \bigotimes_{t \in T} \mathcal{S} = \sigma(\{\pi_t^{-1}(E) : E \in \mathcal{S}, t \in T\}),$$

where $\pi_t : S^T \to S$ is the projection defined by $\pi_t(f) = f(t)$. Recall that $\mathcal{S}^T$ is the smallest σ-algebra on S^T under which all the projections are measurable. Recall also that, for any σ-algebra $\mathcal{G}$ on Ω, a function $X : \Omega \to S^T$ is $(\mathcal{G}, \mathcal{S}^T)$-measurable if and only if $\pi_t \circ X$ is $(\mathcal{G}, \mathcal{S})$-measurable for all $t \in T$.

Lemma 14.1. *With notation as above, we have the following.*

(i) If $\{X(t) : t \in T\}$ is an S-valued stochastic process, and $X : \Omega \to S^T$ is defined by $X(\omega) = X(\cdot, \omega)$, then X is an S^T-valued random variable.

(ii) If $X : \Omega \to S^T$ is an S^T-valued random variable, and $X(t) : \Omega \to S$ is defined by $X(t) = \pi_t \circ X$, then $\{X(t) : t \in T\}$ is an S-valued stochastic process.

(iii) In either case, $\sigma(X) = \sigma(\{X(t) : t \in T\})$.

Proof. By the definition of $\mathcal{S}^T$, for any σ-algebra $\mathcal{G}$ on Ω, we have that X is $(\mathcal{G}, \mathcal{S}^T)$-measurable if and only if $\pi_t \circ X$ is $(\mathcal{G}, \mathcal{S})$-measurable for $t \in T$. Taking $\mathcal{G} = \mathcal{F}$ gives us (i) and (ii). Taking $\mathcal{G} = \sigma(X)$ gives us $\sigma(X) \supset \sigma(\{X(t) : t \in T\})$. And taking $\mathcal{G} = \sigma(\{X(t) : t \in T\})$ gives us $\sigma(X) \subset \sigma(\{X(t) : t \in T\})$. □

From now on, we identify an S-valued stochastic process $\{X(t) : t \in T\}$ with the corresponding S^T-valued random variable X, writing $X = \{X(t) : t \in T\}$. By Lemma 6.33, if Y is an extended real-valued random variable that is measurable with respect to $\sigma(X) = \sigma(\{X(t) : t \in T\})$, then there exists an $\mathcal{S}^T$-measurable function $h : S^T \to \mathbb{R}^*$ such that $Y = h(X)$. The following proposition gives us a practical way to make use of this fact.

Proposition 14.2. *Let $Y : \Omega \to \mathbb{R}^*$. Then Y is $\sigma(X)$-measurable if and only if there exists a sequence $\{t_n\}_{n=1}^\infty$ in T and a function $h : S^\infty \to \mathbb{R}^*$ which is $(\mathcal{S}^\infty, \mathcal{R}^*)$-measurable such that*

$$Y = h(X(t_1), X(t_2), \ldots).$$

Proof. See [1, Proposition II.4.6]. □

If X and Y are S-valued stochastic processes, then X and Y are **indistinguishable** if

$$P(X(t) = Y(t) \text{ for all } t \in T) = 1.$$

We say that Y is a **modification** of X if

$$P(X(t) = Y(t)) = 1 \text{ for all } t \in T.$$

If T is countable, then these concepts coincide.

Let $n \in \mathbb{N}$ and $\mathbf{t} = (t_1, \ldots, t_n) \in T^n$, where the t_j's are distinct. Define the measure $Q_{\mathbf{t}}$ on $(S^n, \mathcal{S}^n)$ by

$$Q_{\mathbf{t}}(B_1 \times \cdots \times B_n) = P(X(t_1) \in B_1, \ldots, X(t_n) \in B_n).$$

The family of measures $\{Q_{\mathbf{t}}\}$ are called the **finite-dimensional distributions** of X. Two processes X and Y have the same finite-dimensional distributions if

$$(X(t_1), \ldots, X(t_n)) \overset{d}{=} (Y(t_1), \ldots, Y(t_n)),$$

whenever $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in T$.

Note that X and Y must be defined on the same probability space in order to be indistinguishable or to be modifications of one another. They may, however, be defined on different probability spaces and still have the same finite-dimensional distributions.

A **continuous-time stochastic process** is one in which $T = I$, where $I \subset \mathbb{R}$ is an interval. Typically, we will have $I = [0, \infty)$, although we may sometimes have $I = [0, 1]$ or $I = [0, T]$ for some $T > 0$. We may even occasionally take $I = \mathbb{R}$, utilizing the notion of "negative time".

It has already been remarked that a continuous-time stochastic process, which is a family of random variables,

$$X(t) : \Omega \to S, \text{ for all } t \in [0, \infty),$$

can be regarded as a single function mapping ω to the sample path,

$$X : \Omega \to S^{[0,\infty)},$$

and that X is $(\mathcal{F}, \mathcal{S}^{[0,\infty)})$-measurable if and only if each $X(t)$ is measurable. In addition to this, though, we could also regard X as a function of two variables,

$$X : [0, \infty) \times \Omega \to S,$$

where $X(t, \omega) = (X(t))(\omega)$. We say that the stochastic process X is **measurable** if X (regarded in this way) is $(\mathcal{B}_{[0,\infty)} \otimes \mathcal{F}, \mathcal{S})$-measurable.

By Fubini's theorem, if X is a measurable stochastic process taking values in a standard Borel space S, then the sample paths, $t \mapsto X(t, \omega)$ are measurable functions from $[0, \infty)$ to S. As a consequence, it follows that not every stochastic process is measurable. For example, let $f : [0, \infty) \to \mathbb{R}$ be any non-measurable functions. For each $t \geqslant 0$, define $X(t, \omega) = f(t)$ for all ω. Then $X(t, \cdot)$ is a constant, so it is $(\mathcal{F}, \mathcal{R})$-measurable. Thus, $X = \{X(t) : t \geqslant 0\}$ is a stochastic process. But the sample paths (each of which is f) are non-measurable.

From this point forward, unless otherwise specified, we will take $S = \mathbb{R}^d$ and $\mathcal{S} = \mathcal{R}^d$.

A **(continuous-time) filtration** is a collection, $\{\mathcal{F}_t\}_{t \geqslant 0}$, of σ-algebras on Ω such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ whenever $0 \leqslant s < t$. Given a filtration, we define $\mathcal{F}_\infty = \sigma(\bigcup_{t \geqslant 0} \mathcal{F}_t)$. The filtration generated by a stochastic process X is $\{\mathcal{F}_t^X\}$, where

$$\mathcal{F}_t^X = \sigma(\{X(s) : 0 \leqslant s \leqslant t\}).$$

We define $\mathcal{F}_{t-} = \sigma(\bigcup_{s<t} \mathcal{F}_s)$ and $\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$. We adopt the convention that $\mathcal{F}_{0-} = \mathcal{F}_0$. A filtration is **right-continuous** if $\mathcal{F}_{t+} = \mathcal{F}_t$ for all $t \geqslant 0$. It is **left-continuous** if $\mathcal{F}_{t-} = \mathcal{F}_t$ for all $t \geqslant 0$. And it is **continuous** if it is both right- and left-continuous.

Lemma 14.3. *Let $\{\mathcal{F}_t\}$ be a filtration. Then $\{\mathcal{F}_{t+}\}$ is a right-continuous filtration and $\{\mathcal{F}_{t-}\}$ is a left-continuous filtration.*

Proof. Exercise 14.1. □

A stochastic process X is **adapted** to a filtration $\{\mathcal{F}_t\}$ if $X(t) \in \mathcal{F}_t$ for all $t \geqslant 0$. The process X is **progressively measurable** with respect to $\{\mathcal{F}_t\}$ if, for all $T > 0$,

$$X : [0, T] \times \Omega \to \mathbb{R}^d$$

is $(\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T, \mathcal{R}^d)$-measurable.

Proposition 14.4. *If X is progressively measurable, then X is measurable and adapted. Conversely, if X is measurable and adapted, then X has a progressively measurable modification.*

Proof. See [8, Proposition 1.1.12] and the references therein. □

Proposition 14.5. *Let X be a right-continuous stochastic process. That is, each sample path is right-continuous. If X is adapted to $\{\mathcal{F}_t\}$, then X is progressively measurable with respect to $\{\mathcal{F}_t\}$.*

Proof. Fix $T > 0$. For $n \in \mathbb{N}$, define

$$X_n(t) = X(0)1_{\{0\}}(t) + \sum_{k=1}^{2^n} X(k2^{-n}T)1_{((k-1)2^{-n}T,k2^{-n}T]}(t).$$

Since every sample-path is right-continuous, it follows that $X_n(t,\omega) \to X(t,\omega)$ as $n \to \infty$, for all t and ω. Since $X_n : [0,T] \times \Omega \to \mathbb{R}^d$ is $(\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T, \mathcal{R}^d)$-measurable for each n, it follows that $X : [0,T] \times \Omega \to \mathbb{R}^d$ is $(\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T, \mathcal{R}^d)$-measurable. □

A similar proof establishes the following proposition.

Proposition 14.6. *Let X be a right-continuous stochastic process. That is, each sample path is right-continuous. Then X is measurable.*

Both of these propositions also hold if right-continuity is replaced by left-continuity.

A **random time** is a $[0,\infty]$-valued random variable. If $X = \{X(t) : t \in [0,\infty)\}$ is a stochastic process and T is a random time, then $X(T)$ denotes the function

$$X(T) : \{T < \infty\} \to \mathbb{R}^d$$

given by $(X(T))(\omega) = X(T(\omega),\omega)$.

Lemma 14.7. *If X is a measurable stochastic process and T is a random time, then the function $X(T)$ is $(\mathcal{F}|_{\{T<\infty\}}, \mathcal{R}^d)$-measurable.*

Proof. Exercise 14.2. □

Strictly speaking, $\sigma(X(T))$ should be a σ-algebra on $\{T < \infty\}$. However, with an abuse of notation, when X is a measurable process, we will define

$$\sigma(X(T)) = \{\{X(T) \in B\} : B \in \mathcal{R}^d\} \cup \{\{X(T) \in B\} \cup \{T = \infty\} : B \in \mathcal{R}^d\}.$$

It can be shown (see Exercise 14.6) that, with this definition, $\sigma(X(T))$ is a σ-algebra on Ω, and in fact, $\sigma(X(T)) \subset \mathcal{F}$.

Exercises

14.1. Prove Lemma 14.3.

14.2. Prove Lemma 14.7.

14.3. [8, Problem 1.1.5] Let X and Y be stochastic processes. Suppose that almost every sample path of X is right-continuous, and almost every sample path of Y is right-continuous. Prove that if X and Y are modifications of one another, then they are indistinguishable.

14.4. [8, Exercise 1.1.7] Let X be a cadlag process. That is, every sample path is cadlag. Let

$$A = \{\omega \in \Omega : X(\cdot, \omega) \text{ is continuous on } [0, t_0)\}. \tag{14.1}$$

Prove that $A \in \mathcal{F}^X_{t_0}$.

14.5. [8, Exercise 1.1.8] Give an example of a process X such that almost every sample path is cadlag, but the event A in (14.1) is not in $\mathcal{F}^X_{t_0}$.

14.6. [8, Problem 1.1.17] Let X be a measurable process and T a random time. Prove that $\sigma(X(T))$ is a sub-σ-algebra of $\mathcal{F}$.

14.2 Stopping times

This section corresponds to [8, Section 1.2].

Let T be a random time and $\{\mathcal{F}_t\}$ a filtration. If $\{T \leqslant t\} \in \mathcal{F}_t$ for all $t \geqslant 0$, then T is a **stopping time** with respect to $\{\mathcal{F}_t\}$. If $\{T < t\} \in \mathcal{F}_t$ for all $t \geqslant 0$, then T is an **optional time** with respect to $\{\mathcal{F}_t\}$.

Note that if T is a constant, then T is both a stopping time and an optional time.

If X is a stochastic process and T is a stopping time with respect to $\{\mathcal{F}^X_t\}$, then the value of T depends only on the values of $X(t)$ for $t \in [0, T]$. The following proposition makes this notion rigorous.

Proposition 14.8. *Let X be a stochastic process and T a stopping time with respect to $\{\mathcal{F}^X_t\}$. Suppose $\omega_0, \omega_1 \in \Omega$ satisfy*

$$X(t, \omega_0) = X(t, \omega_1) \text{ for all } t \in [0, T(\omega_0)] \cap [0, \infty).$$

Then $T(\omega_0) = T(\omega_1)$.

Proof. Exercise 14.7. □

Proposition 14.9. *Every stopping time is an optional time.*

Proof. Let T be an $\{\mathcal{F}_t\}$-stopping time. Fix $t > 0$. Choose n_0 such that $t - 1/n_0 > 0$. Then

$$\{T < t\} = \bigcup_{n=n_0}^{\infty} \left\{T \leqslant t - \frac{1}{n}\right\}.$$

Since T is a stopping time, we have, for each fixed n,

$$\left\{T \leqslant t - \frac{1}{n}\right\} \in \mathcal{F}_{t-1/n} \subset \mathcal{F}_t.$$

Thus, $\{T < t\} \in \mathcal{F}_t$, and T is an optional time. □

Proposition 14.10. *If $\{\mathcal{F}_t\}$ is right-continuous, then every $\{\mathcal{F}_t\}$-optional time is an $\{\mathcal{F}_t\}$-stopping time.*

Proof. Let T be an $\{\mathcal{F}_t\}$-optional time. Fix $m \in \mathbb{N}$ and note that

$$\{T \leqslant t\} = \bigcap_{n=m}^{\infty} \left\{T < t + \frac{1}{n}\right\}.$$

Since T is an optional time, for every $n \geqslant m$, we have

$$\left\{T < t + \frac{1}{n}\right\} \in \mathcal{F}_{t+1/n} \subset \mathcal{F}_{t+1/m}.$$

It follows that $\{T \leqslant t\} \in \mathcal{F}_{t+1/m}$ for every $m \in \mathbb{N}$. Thus,

$$\{T \leqslant t\} \in \bigcap_{m=1}^{\infty} \mathcal{F}_{t+1/m} = \mathcal{F}_{t+} = \mathcal{F}_t,$$

and so T is an $\{\mathcal{F}_t\}$-stopping time. □

Proposition 14.11. *Let $\{\mathcal{F}_t\}$ be a filtration and T a random time. Then T is an $\{\mathcal{F}_t\}$-optional time if and only if T is an $\{\mathcal{F}_{t+}\}$-stopping time.*

Proof. Let T be an $\{\mathcal{F}_t\}$-optional time. Since $\mathcal{F}_t \subset \mathcal{F}_{t+}$, it follows that T is an $\{\mathcal{F}_{t+}\}$-optional time. But $\{\mathcal{F}_{t+}\}$ is right-continuous, so T is an $\{\mathcal{F}_{t+}\}$-stopping time.

Conversely, suppose T is an $\{\mathcal{F}_{t+}\}$-stopping time. Then

$$\{T < t\} = \bigcup_{n=n_0}^{\infty} \left\{T \leqslant t - \frac{1}{n}\right\}.$$

Since

$$\left\{T \leqslant t - \frac{1}{n}\right\} \in \mathcal{F}_{(t-1/n)+} = \bigcap_{s>t-1/n} \mathcal{F}_s \subset \mathcal{F}_t,$$

it follows that $\{T < t\} \in \mathcal{F}_t$, and so T is an $\{\mathcal{F}_t\}$-optional time. □

Let X be a right-continuous process and $B \in \mathcal{R}^d$. The **hitting time** of B is defined as

$$H_B = \inf\{t \geqslant 0 : X(t) \in B\}.$$

Proposition 14.12. *If B is open, then H_B is an optional time. If B is closed and X is a continuous process, then H_B is a stopping time.*

Proof. Exercise 14.8. □

Lemma 14.13. *Let S and T be stopping times. Then $S \wedge T$, $S \vee T$, and $S + T$ are also stopping times.*

Proof. Since

$$\{S \wedge T \leqslant t\} = \{S \leqslant t\} \cup \{T \leqslant t\}$$

and

$$\{S \vee T \leqslant t\} = \{S \leqslant t\} \cap \{T \leqslant t\},$$

it follows that $S \wedge T$ and $S \vee T$ are stopping time.

For $S+T$, first note that $\{S+T \leqslant 0\} = \{S \leqslant 0\} \cap \{T \leqslant 0\} \in \mathcal{F}_0$. Now fix $t > 0$. Then $\{S+T \leqslant t\} = \{S+T > t\}^c$ and

$$\{S+T > t\} = \{S = 0, T > t\} \cup \{0 < S < t, S+T > t\} \cup \{S = t, T > 0\} \cup \{S > t\}.$$

We have

$$\{S = 0, T > t\} = \{S \leqslant 0\} \cap \{T \leqslant t\}^c \in \mathcal{F}_t$$

and

$$\{S = t, T > 0\} = \{S \leqslant t\} \cap \{S < t\}^c \cap \{T \leqslant 0\}^c \in \mathcal{F}_t.$$

Here, we have used the fact that a stopping time is an optional time, and that $\mathcal{F}_0 \subset \mathcal{F}_t$. We also have $\{S > t\} = \{S \leqslant t\}^c \in \mathcal{F}_t$. For the final set, observe that if $S(\omega) + T(\omega) > t$, then there exists a positive rational r such that

$$S(\omega) + T(\omega) > r + T(\omega) > t.$$

Thus,

$$\{0 < S < t, S+T > t\} = \bigcup_{r \in \mathbb{Q} \cap [0,t)} \{r < S < t, T > t - r\}.$$

Since this is a countable union, and

$$\{r < S < t, T > t - r\} = \{S \leqslant r\}^c \cap \{S < t\} \cap \{T \leqslant t - r\}^c \in \mathcal{F}_t,$$

we are done. □

Lemma 14.14. *If $\{T_n\}_{n=1}^\infty$ is a sequence of optional times, then $\sup_n T_n$, $\inf_n T_n$, $\limsup_{n\to\infty} T_n$, and $\liminf_{n\to\infty} T_n$ are all optional times. If $\{T_n\}_{n=1}^\infty$ is a sequence of stopping times, then $\sup_n T_n$ is a stopping times.*

Proof. Exercise 14.9. □

If T is an $\{\mathcal{F}_t\}$-stopping time, we define

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leqslant t\} \in \mathcal{F}_t \text{ for all } t \geqslant 0\}.$$

Lemma 14.15. *If T is a $\{\mathcal{F}_t\}$-stopping time, then $\mathcal{F}_T$ is a σ-algebra and $T \in \mathcal{F}_T$. Moreover, if $T \equiv t$ for some $t \geqslant 0$, then $\mathcal{F}_T = \mathcal{F}_t$.*

Proof. Exercise 14.10. □

Lemma 14.16. *Let S and T be $\{\mathcal{F}_t\}$-stopping times. If $A \in \mathcal{F}_S$, then $A \cap \{S \leqslant T\} \in \mathcal{F}_T$. If $S \leqslant T$ pointwise, then $\mathcal{F}_S \subset \mathcal{F}_T$.*

Proof. See [8, Lemma 1.2.15]. □

Lemma 14.17. *Let S and T be $\{\mathcal{F}_t\}$-stopping times. Then $\mathcal{F}_{S \wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$ and*

$$\{S < T\}, \{T < S\}, \{S \leqslant T\}, \{T \leqslant S\}, \{S = T\}$$

are all elements of $\mathcal{F}_S \cap \mathcal{F}_T$.

Proof. See [8, Lemma 1.2.16]. □

Proposition 14.18. *Let S and T be $\{\mathcal{F}_t\}$-stopping times and $Z \in L^1(\Omega)$. Then*

(i) $E[Z \mid \mathcal{F}_S]1_{\{S \leqslant T\}} = E[Z \mid \mathcal{F}_{S \wedge T}]1_{\{S \leqslant T\}}$ *a.s., and*

(ii) $E[E[Z \mid \mathcal{F}_S] \mid \mathcal{F}_T] = E[Z \mid \mathcal{F}_{S \wedge T}]$ *a.s.*

Proof. Exercise 14.11. □

Proposition 14.19. *Let X be $\{\mathcal{F}_t\}$-progressively measurable and T an $\{\mathcal{F}_t\}$-stopping time. Then $X(T)$ is $\mathcal{F}_T|_{\{T<\infty\}}$-measurable and the stopped process $\{X(T \wedge t)\}$ is $\{\mathcal{F}_t\}$-progressively measurable.*

Proof. See [8, Proposition 1.2.18]. □

If T is an $\{\mathcal{F}_t\}$-optional time, we define

$$\mathcal{F}_{T+} = \{A \in \mathcal{F} : A \cap \{T \leqslant t\} \in \mathcal{F}_{t+} \text{ for all } t \geqslant 0\}.$$

Lemma 14.20. *If T is an $\{\mathcal{F}_t\}$-optional time, then $\mathcal{F}_{T+}$ is a σ-algebra, $T \in \mathcal{F}_T+$, and*

$$\mathcal{F}_{T+} = \{A \in \mathcal{F} : A \cap \{T < t\} \in \mathcal{F}_t \textit{ for all } t \geqslant 0\}.$$

Moreover, if T is an $\{\mathcal{F}_t\}$-stopping time, then $\mathcal{F}_T \subset \mathcal{F}_{T+}$.

Proof. Exercise 14.12. □

Given a probability space $(\Omega, \mathcal{F}, P)$, recall that a null set (or a P-null set) is an event $N \in \mathcal{F}$ such that $P(N) = 0$. Recall also that $A \subset \Omega$ is called negligible if $A \subset N$ for some null set N.

Lemma 14.21. *Let Y be an $\mathbb{R}^*$-valued random variable and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra containing all null sets. If N is a null set, then $Y1_N \in \mathcal{G}$.*

Proof. First assume $Y = 1_A$ for some some $A \in \mathcal{F}$. Then $Y1_N = 1_{A \cap N}$. Since $A \cap N$ is a null set, we have $A \cap N \in \mathcal{G}$, and so $Y1_N \in \mathcal{G}$. By linearity, the claim is true for simple Y. Since any nonnegative Y is the pointwise limit of simple functions, Proposition 2.11 implies the claim is true for nonnegative Y. For general Y, we obtain the result by considering the positive and negative parts of Y. □

Lemma 14.22. *Let X be a $\mathcal{G}$-measurable random variable, where $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra containing all null sets. If Y is a random variable with $Y = X$ a.s., then $Y \in \mathcal{G}$.*

Proof. Since $Y = X$ a.s., there exists $N \in \mathcal{F}$ with $P(N) = 0$ such that

$$Y = X1_{N^c} + Y1_N.$$

Since N is a null set, we have $N \in \mathcal{G}$. Since $X \in \mathcal{G}$, we also have $X1_{N^c} \in \mathcal{G}$. The preceding lemma shows that $Y1_N \in \mathcal{G}$. □

A filtration $\{\mathcal{F}_t\}$ satisfies the **usual conditions** if it is right-continuous and $\mathcal{F}_0$ contains all negligible sets. Note that this implies $\mathcal{F}_t$ contains all negligible sets, for all $t \geqslant 0$.

Exercises

14.7. [8, Problem 1.2.2] Prove Proposition 14.8.

14.8. [8, Problems 1.2.6] Prove Proposition 14.12 in the case that B is open.

14.9. Prove Lemma 14.14.

14.10. [8, Problem 1.2.13] Prove Lemma 14.15.

14.11. [8, Problem 1.2.17] Prove Proposition 14.18.

14.12. [8, Problem 1.2.21] Prove Lemma 14.20.

14.3 Martingales: definition and properties

This section corresponds to [8, Section 1.3].

Fix a filtration $\{\mathcal{F}_t\}$. In this section, we consider only processes X which are real-valued, $\{\mathcal{F}_t\}$-adapted, and integrable (that is, $E|X(t)| < \infty$ for all $t \geqslant 0$).

The process X is a **submartingale** with respect to $\{\mathcal{F}_t\}$ if $E[X(t) \mid \mathcal{F}_s] \geqslant X(s)$ a.s. whenever $s < t$. It is a **supermartingale** with respect to $\{\mathcal{F}_t\}$ if $E[X(t) \mid \mathcal{F}_s] \leqslant X(s)$ a.s. whenever $s < t$. It is a **martingale** with respect to $\{\mathcal{F}_t\}$ if it is both a submartingale and a supermartingale. If the filtration is not mentioned, then it is taken to be $\{\mathcal{F}_t^X\}$.

A process N is a Poisson process with respect to $\{\mathcal{F}_t\}$ is N is a Poisson process and $N(t) - N(s)$ is independent of $\mathcal{F}_s$ for all $s < t$. Note that an $\{\mathcal{F}_t\}$-Poisson process is an $\{\mathcal{F}_t\}$-submartingale (check).

Let N be an $\{\mathcal{F}_t\}$-Poisson process with rate $\lambda > 0$. The **compensated Poisson process** is the process

$$M(t) = N(t) - \lambda t.$$

Lemma 14.23. *The compensated Poisson process is an $\{\mathcal{F}_t\}$-martingale.*

Proof. Exercise 14.13. □

If X is an $\{\mathcal{F}_t\}$-submartingale and $X(\infty) \in L^1(\Omega)$, then $\{X(t) : t \in [0, \infty]\}$ is an $\{\mathcal{F}_t\}$**-submartingale (with last element** $X(\infty)$**)** if $X(\infty) \in \mathcal{F}_\infty$ and

$$E[X(\infty) \mid \mathcal{F}_s] \geqslant X(s) \quad \text{a.s.},$$

for all $s \geqslant 0$. We make a similar definition for supermartingales and martingales.

Theorems 12.5 and 12.7 and their respective corollaries are still true in the continuous-time case, and their proofs are similar.

Doob's inequality also holds in continuous time and its proof also uses "upcrossings". The continuous-time formulation of Doob's inequality is contained in the following theorem.

Theorem 14.24. *Let X be a right-continuous submartingale. If $a < b$ and $\lambda > 0$, then we have the following:*

(i) $\lambda P\left(\sup_{t \in [a,b]} X(t) \geqslant \lambda\right) \leqslant E[X(b)^+]$.

(ii) $\lambda P\left(\inf_{t \in [a,b]} X(t) \leqslant -\lambda\right) \leqslant E[X(b)^+] - E[X(a)]$.

(iii) If $X(t) \geqslant 0$ a.s. for all $t \geqslant 0$ and $p > 1$, then

$$E\left|\sup_{t \in [a,b]} X(t)\right|^p \leqslant \left(\frac{p}{p-1}\right)^p E|X(b)|^p.$$

(iv) Almost every sample path of X is cadlag.

Proof. See [8, Theorem 1.3.8]. □

Remark 14.25. For the definition of "cadlag", refer back to Section 3.5.2.

Theorem 14.26. *Let X be an $\{\mathcal{F}_t\}$-submartingale, where $\{\mathcal{F}_t\}$ satisfies the usual conditions. Then X has a right-continuous modification if and only if $t \mapsto EX(t)$ is right-continuous. Moreover, in this case, the modification can be chosen so that it is cadlag and $\{\mathcal{F}_t\}$-adapted, and therefore also an $\{\mathcal{F}_t\}$-submartingale.*

Proof. See [8, Theorem 1.3.13]. □

Theorem 14.27 (martingale convergence theorem)**.** *Let $X = \{X(t) : t \geqslant 0\}$ be a right-continuous submartingale with $\sup_t E[X(t)^+] < \infty$. Then there exists an integrable random variable $X(\infty)$ such that $X(t) \to X(\infty)$ a.s. as $t \to \infty$.*

Proof. Uses "upcrossings". See [8, Theorem 1.3.15]. □

Theorem 14.28. *Let $X = \{X(t) : t \geqslant 0\}$ be a right-continuous, nonnegative $\{\mathcal{F}_t\}$-supermartingale (that is, $X(t) \geqslant 0$ a.s. for all $t \geqslant 0$). Then there exists an integrable, nonnegative random variable $X(\infty)$ such that $X(t) \to X(\infty)$ a.s. Moreover, $\{X(t) : t \in [0, \infty]\}$ is an $\{\mathcal{F}_t\}$-supermartingale.*

Proof. Exercise 14.14. □

Theorem 14.29. *Let $X = \{X(t) : t \geqslant 0\}$ be a non-negative, right-continuous $\{\mathcal{F}_t\}$-submartingale. Then the following are equivalent:*

(i) X is uniformly integrable,

(ii) There exists $X(\infty) \in L^1(\Omega)$ such that $X(t) \to X(\infty)$ a.s., and $\{X(t) : t \in [0, \infty]\}$ is an $\{\mathcal{F}_t\}$-submartingale.

(iii) There exists $X(\infty)$ such that $X(t) \to X(\infty)$ in $L^1(\Omega)$.

Moreover, (i) implies (iii) implies (ii) without the assumption of nonnegativity.

Proof. Exercise 14.15. □

Theorem 14.30. *Let X be a right-continuous $\{\mathcal{F}_t\}$-martingale. Then the following are equivalent:*

(i) X is uniformly integrable,

(ii) There exists $X(\infty) \in L^1(\Omega)$ such that $X(t) \to X(\infty)$ a.s., and $\{X(t) : t \in [0, \infty]\}$ is an $\{\mathcal{F}_t\}$-martingale.

(iii) There exists $X(\infty)$ such that $X(t) \to X(\infty)$ in $L^1(\Omega)$.

(iv) There exists $Y \in L^1(\Omega)$ such that $X(t) = E[Y \mid \mathcal{F}_t]$ a.s., for all $t \geqslant 0$.

Moreover, if (iv) holds and $X(\infty)$ is the random variable appearing in (ii), then $E[Y \mid \mathcal{F}_\infty] = X(\infty)$ a.s.

Proof. Exercise 14.16. □

Theorem 14.31 (optional sampling theorem)**.** *Let $X = \{X(t) : t \in [0, \infty)\}$ be a right-continuous $\{\mathcal{F}_t\}$-submartingale. Let S and T be $\{\mathcal{F}_t\}$-optional times with $S \leqslant T$ pointwise. Assume at least one of the following conditions hold:*

(i) T is bounded. That is, there exists $a > 0$ such that $T \leqslant a$ pointwise.

(ii) There exists $Y \in L^1(\Omega)$ such that $X(t) \leqslant E[Y \mid \mathcal{F}_t]$ a.s. for all $t \geqslant 0$.

Then

$$E[X(T) \mid \mathcal{F}_{S+}] \geqslant X(S) \quad \text{a.s.}$$

If S is a stopping time, then

$$E[X(T) \mid \mathcal{F}_S] \geqslant X(S) \quad \text{a.s.}$$

Consequently, $EX(T) \geqslant EX(0)$ and, if X is a martingale with last element, then $EX(T) = EX(0)$.

Proof. If (ii) holds and $Y \in \mathcal{F}_\infty$, then, defining $X(\infty) = Y$, we have that $\{X(t) : t \in [0, \infty]\}$ is an $\{\mathcal{F}_t\}$-submartingale with last element $X(\infty)$. In this case, the proof uses "backward" martingales (see [8, Theorem 1.3.22]). The proofs of the other cases are Exercises 14.17 and 14.18. □

Exercises

14.13. [8, Problem 1.3.4] Prove Lemma 14.23.

14.14. [8, Problem 1.3.16] Prove Theorem 14.28.

14.15. [8, Problem 1.3.19] Prove Theorem 14.29.

14.16. [8, Problem 1.3.20] Prove Theorem 14.30.

14.17. [8, Problem 3.23(i)] Prove Theorem 14.31 under condition (i). You may use the fact that it is true under condition (ii) when $Y \in \mathcal{F}_\infty$.

14.18. [8, Problem 3.23(ii)] Prove Theorem 14.31 under condition (ii). You may use the fact that it is true under condition (ii) when $Y \in \mathcal{F}_\infty$.

14.4 The Doob-Meyer decomposition

This section corresponds to [8, Section 1.4].

In this section, we extend Doob's decomposition (Theorem 12.16) to the cases of continuous-time submartingales. We must first establish an additional fact in the discrete setting. Let $\{\mathcal{F}_n\}_{n=0}^\infty$ be a discrete-time filtration, and let $A = \{A_n\}$ be an $\{\mathcal{F}_n\}$-adapted, increasing process with $A_0 = 0$ a.s. and $E|A_n| < \infty$ a.s. for all n. Then A is **natural** if

$$E[A_n M_n] = E \sum_{m=1}^{n} M_{m-1}(A_m - A_{m-1}), \tag{14.2}$$

for all $n \geqslant 1$, whenever $M = \{M_n\}$ is an $\{\mathcal{F}_n\}$-martingale which is bounded. (Here, "bounded" has the same meaning as in Theorem 12.9. That is, for all $n \in \mathbb{N}$, there exists $C_n > 0$ such that $|M_n| \leqslant C_n$ a.s.)

Remark 14.32. Note that

$$\begin{aligned}
(A \cdot M)_n &= \sum_{m=0}^{n} A_m(M_m - M_{m-1}) \\
&= \sum_{m=0}^{n} A_m M_m - \sum_{m=0}^{n} A_m M_{m-1} \\
&= \sum_{m=1}^{n+1} A_{m-1} M_{m-1} - \sum_{m=0}^{n} A_m M_{m-1} \\
&= A_n M_n - \sum_{m=1}^{n} M_{m-1}(A_m - A_{m-1})
\end{aligned}$$

Thus, (14.2) is equivalent to $E(A \cdot M)_n = 0$.

Proposition 14.33. *Let $\{\mathcal{F}_n\}_{n=0}^\infty$ be a filtration. Assume that $\mathcal{F}_0$ contains all null sets. Let $A = \{A_n\}$ be an $\{\mathcal{F}_n\}$-adapted, increasing process with $A_0 = 0$ a.s. and $E|A_n| < \infty$ a.s. for all n. Then A is natural if and only if A is predictable.*

Proof. Assume A is predictable. Let M be a bounded $\{\mathcal{F}_n\}$-martingale. By Theorem 12.9 and Remark 12.10, we have that $A \cdot M$ is an $\{\mathcal{F}_n\}$-martingale, so that $E(A \cdot M)_n = E(A \cdot M)_0 = 0$. Thus, by Remark 14.32, A is natural.

Now assume A is natural. We will first show that if M is any bounded $\{\mathcal{F}_n\}$-martingale, then

$$E[M_n(A_n - E[A_n \mid \mathcal{F}_{n-1}])] = 0,$$

for all $n \in \mathbb{N}$. To see this, note that

$$\begin{aligned} E[M_n(A_n - E[A_n \mid \mathcal{F}_{n-1}])] = E[(M_n - M_{n-1})A_n] \\ + E[M_{n-1}(A_n - E[A_n \mid \mathcal{F}_{n-1}])] \\ - E[(M_n - M_{n-1})E[A_n \mid \mathcal{F}_{n-1}]]. \end{aligned}$$

For the first term, Remark 14.32 implies

$$E[(M_n - M_{n-1})A_n] = E[(A \cdot M)_n - (A \cdot M)_{n-1}] = 0.$$

The second and third terms are both zero, which can be seen by conditioning on the inside by $\mathcal{F}_{n-1}$.

Now fix $k \in \mathbb{N}$. We wish to show that $A_k \in \mathcal{F}_{k-1}$. Define the random variable $Y = \operatorname{sgn}(A_k - E[A_k \mid \mathcal{F}_{k-1}])$ and then define

$$M_n = \begin{cases} E[Y \mid \mathcal{F}_n] & \text{if } 0 \leqslant n < k, \\ Y & \text{if } n \geqslant k. \end{cases}$$

Then M_n is integrable and $M = \{M_n\}$ is $\{\mathcal{F}_n\}$-adapted. If $n < k$, then

$$E[M_n \mid \mathcal{F}_{n-1}] = E[E[Y \mid \mathcal{F}_n] \mid \mathcal{F}_{n-1}] = E[Y \mid \mathcal{F}_{n-1}] = M_{n-1}.$$

If $n = k$, then

$$E[M_n \mid \mathcal{F}_{n-1}] = E[Y \mid \mathcal{F}_{n-1}] = M_{n-1}.$$

And if $n > k$, then $n - 1 \geqslant k$, which implies $\mathcal{F}_k \subset \mathcal{F}_{n-1}$. Hence,

$$E[M_n \mid \mathcal{F}_{n-1}] = E[Y \mid \mathcal{F}_{n-1}] = Y = M_{n-1}.$$

We have therefore shown that M is a bounded $\{\mathcal{F}_n\}$-martingale. From what we proved initially, it now follows that

$$E[M_n(A_n - E[A_n \mid \mathcal{F}_{n-1}])] = 0,$$

for all n. In particular,

$$\begin{aligned} 0 &= E[M_k(A_k - E[A_k \mid \mathcal{F}_{k-1}])] \\ &= E[Y(A_k - E[A_k \mid \mathcal{F}_{k-1}])] \\ &= E|A_k - E[A_k \mid \mathcal{F}_{k-1}]|, \end{aligned}$$

from which it follows that $A_k = E[A_k \mid \mathcal{F}_{k-1}]$ a.s. Or, more specifically, if Z is a version of $E[A_k \mid \mathcal{F}_{k-1}]$, then $A_k = Z$ a.s. Since $Z \in \mathcal{F}_{k-1}$ and $\mathcal{F}_{k-1}$ contains all null sets, it follows from Lemma 14.22 that $A_k \in \mathcal{F}_{k-1}$. □

We now turn our attention back to the continuous-time setting.

Definition 14.34. An $\{\mathcal{F}_t\}$-adapted process $A = \{A(t) : t \geqslant 0\}$ is **natural** if

(a) $A(0) = 0$ a.s.,

(b) almost every sample path is increasing and right-continuous,

(c) $EA(t) < \infty$ for all $t \geqslant 0$, and

(d) whenever M is a bounded, right-continuous $\{\mathcal{F}_t\}$-martingale, we have

$$E\int_{(0,t]} M(s)\,dA(s) = E\int_{(0,t]} M_-(s)\,dA(s).$$

Remark 14.35. Recall the notation M_- from Section 3.5.2.

Remark 14.36. The term "bounded" in (d) means there exists $C > 0$ such that $P(|M(t)| \leqslant C \text{ for all } t \geqslant 0) = 1$.

Remark 14.37. If M is a right-continuous $\{\mathcal{F}_t\}$-martingale, then by Theorem 14.24, almost every sample path is cadlag. Therefore, as in Section 3.5.3, the integrals in (d) are almost surely well-defined Lebesgue-Stieltjes integrals.

Moreover, Lebesgue-Stieltjes integrals involving cadlag function can be written as limits of Riemann sums (see Theorem 3.31). This fact can be used to show that the function

$$\omega \mapsto \int_{(0,t]} M(s,\omega)\,dA(s,\omega)$$

is a random variable.

If M is bounded, then

$$\left|\int_{(0,t]} M(s)\,dA(s)\right| \leqslant CA(t).$$

Since $A(t)$ is integrable by (c), it follows that the expectation on the left-hand side in (d) is well-defined. A similar argument shows the same for the right-hand side.

Remark 14.38. If an $\{\mathcal{F}_t\}$-adapted process satisfies (a)-(c) and is continuous, then it automatically satisfies (d). By Theorem 3.26, a cadlag function has at most countably many discontinuities. Since the Lebesgue-Stieltjes measure corresponding to a continuous function assigns 0 mass to singletons, it follows that

$$\int_{(0,t]} (M(s) - M_-(s))\, dA(s) = 0 \quad \text{a.s.}$$

whenever A is continuous.

Remark 14.39. If $\{\mathcal{F}_t\}$ satisfies the usual conditions and A is $\{\mathcal{F}_t\}$-natural, then A is adapted to $\{\mathcal{F}_{t-}\}$. See [8, Remark 1.4.6(iii)] and the references therein.

Lemma 14.40. *Let A be an $\{\mathcal{F}_t\}$-adapted process with $A(0) = 0$ a.s. Suppose that almost every sample path of A is cadlag and of bounded variation on compact intervals. Further suppose that $E[T_A(t)] < \infty$ for all $t \geqslant 0$. Let M be a bounded, right-continuous $\{\mathcal{F}_t\}$-martingale. Then*

$$E[M(t)A(t)] = E\int_{(0,t]} M(s)\, dA(s),$$

for all $t \geqslant 0$.

Remark 14.41. The notation $T_A(t)$ refers to the total variation of the sample path on the interval $[0,t]$, which is defined in Section 3.4.

Remark 14.42. Note that Definition 14.34 implies that almost every sample path is cadlag and of bounded variation on compact intervals. Thus, Lemma 14.40 shows that Definition 14.34(d) is equivalent to

$$E[M(t)A(t)] = E\int_{(0,t]} M_-(s)\, dA(s),$$

and this is analogous to the definition of natural in the discrete-time setting.

Proof of Lemma 14.40. Choose $C > 0$ such that $|M(t)| \leqslant C$ for all $t \geqslant 0$ a.s. Fix $t > 0$. As in Theorem 3.31, let $\{\mathcal{P}_m\}$ be a sequence of partitions of $[0,t]$ with $\|\mathcal{P}_m\| \to 0$. Let

$$M_m(s) = \sum_{k=1}^{n} M(t_k) 1_{(t_{k-1},t_k]}(s).$$

Since $t_n = t$, $t_0 = 0$, and $A(0) = 0$, we have

$$\begin{aligned} E\int_{(0,t]} M_m(s)\, dA(s) &= E\sum_{k=1}^{n} M(t_k)(A(t_k) - A(t_{k-1})) \\ &= E\sum_{k=1}^{n} M(t_k)A(t_k) - E\sum_{k=0}^{n-1} M(t_{k+1})A(t_k) \\ &= EM(t)A(t) - E\sum_{k=1}^{n-1} (M(t_{k+1}) - M(t_k))A(t_k). \end{aligned}$$

Note that

$$\begin{aligned} E[(M(t_{k+1}) - M(t_k))A(t_k)] &= E[E[(M(t_{k+1}) - M(t_k))A(t_k) \mid \mathcal{F}_{t_k}]] \\ &= E[A(t_k)E[M(t_{k+1}) - M(t_k) \mid \mathcal{F}_{t_k}]] = 0. \end{aligned}$$

Thus,

$$E\int_{(0,t]} M_m(s)\,dA(s) = EM(t)A(t), \tag{14.3}$$

for all m. On the other hand, since almost every sample path of M and A are cadlag, Theorem 3.31 implies

$$\int_{(0,t]} M_m(s)\,dA(s) \to \int_{(0,t]} M(s)\,dA(s) \quad \text{a.s.}$$

Since

$$\left|\int_{(0,t]} M_m(s)\,dA(s)\right| \leqslant CT_A(t)$$

and $T_A(t)$ is integrable, it follows by dominated convergence that

$$E\int_{(0,t]} M_m(s)\,dA(s) \to E\int_{(0,t]} M(s)\,dA(s).$$

Letting $m \to \infty$ in (14.3) finishes the proof. □

Lemma 14.43. *Let $\{\mathcal{F}_t\}$ be a filtration satisfying the usual conditions. Let B be a right-continuous $\{\mathcal{F}_t\}$-martingale with $B(0) = 0$ a.s. Suppose that $B = A - A'$, where A and A' are natural. Then*

$$P(B(t) = 0 \textit{ for all } t \geqslant 0) = 1.$$

Proof. Fix $t \geqslant 0$. Let $M(s) = E[\operatorname{sgn}(B(t)) \mid \mathcal{F}_s]$, so that M is a bounded martingale. By Theorem 14.26, M has a cadlag modification which is also an $\{\mathcal{F}_t\}$-martingale. By Lemma 14.40,

$$E[M(t)B(t)] = E\int_{(0,t]} M_-(s)\,dB(s).$$

As in Theorem 3.31, let $\{\mathcal{P}_m\}$ be a sequence of partitions of $[0,t]$ with $\|\mathcal{P}_m\| \to 0$. By Theorem 3.31,

$$\sum_{k=1}^n M(t_{k-1})(B(t_k) - B(t_{k-1})) \to \int_{(0,t]} M_-(s)\,dB(s) \quad \text{a.s.}$$

Since

$$\left|\sum_{k=1}^n M(t_{k-1})(B(t_k) - B(t_{k-1}))\right| \leqslant A(t) + A'(t),$$

dominated convergence gives

$$E\sum_{k=1}^{n} M(t_{k-1})(B(t_k) - B(t_{k-1})) \to E\int_{(0,t]} M_-(s)\,dB(s).$$

But M and B are both martingales, $E[M(t_{k-1})(B(t_k) - B(t_{k-1}))] = 0$. It therefore follows that

$$0 = E\int_{(0,t]} M_-(s)\,dB(s) = EM(t)B(t) = E|B(t)|.$$

Hence, $B(t) = 0$. Since t was arbitrary, it follows that B is a modification of the 0 process. By Exercise 14.3, B is indistinguishable from the 0 process. □

Let $\{\mathcal{F}_t\}$ be a filtration. Let $\mathcal{S}$ be the set of all $\{\mathcal{F}_t\}$-stopping times that are finite almost surely. For $a > 0$, let $\mathcal{S}_a$ be the set of all $\{\mathcal{F}_t\}$-stopping times T such that $T \leqslant a$ a.s.

Let X be a right-continuous $\{\mathcal{F}_t\}$-adapted process. Then X is *of class D* if the family of random variables $\{X(T) : T \in \mathcal{S}\}$ is uniformly integrable. We say X is *of class DL* if $\{X(T) : T \in \mathcal{S}_a\}$ is uniformly integrable for all $a > 0$.

Note that $\mathcal{S}_a \subset \mathcal{S}$. Thus, if X is of class D, then X is of class DL.

Proposition 14.44. *Let X be a right-continuous $\{\mathcal{F}_t\}$-submartingale. Suppose that $X(t) \geqslant 0$ a.s. for all $t \geqslant 0$. Then X is of class DL.*

Proof. Exercise 14.19. □

Proposition 14.45. *Let X be a right-continuous $\{\mathcal{F}_t\}$-submartingale. Suppose $X = M + A$, where M is an $\{\mathcal{F}_t\}$-martingale and A is an $\{\mathcal{F}_t\}$-adapted process satisfying (a)-(c) of Definition 14.34. Then X is of class DL.*

Proof. Exercise 14.19. □

Proposition 14.46. *If X is a uniformly integrable, right-continuous $\{\mathcal{F}_t\}$-martingale, then X is of class D.*

Proof. Exercise 14.19. □

Theorem 14.47 (Doob-Meyer decomposition)**.** *Let X be a right-continuous $\{\mathcal{F}_t\}$-submartingale, where $\{\mathcal{F}_t\}$ satisfies the usual conditions. Suppose X is of class DL. Then $X = M + A$, where M is a right-continuous $\{\mathcal{F}_t\}$-martingale and A is $\{\mathcal{F}_t\}$-natural. If $X = M' + A'$ is another such decomposition, then M and M' are indistinguishable, as are A and A'. Moreover, if X is of class D, then M is uniformly integrable.*

Proof. We prove only uniqueness here. For the rest of the proof, see [8, Theorem 1.4.10].

Suppose $X = M + A = M' + A'$ are two Doob-Meyer decompositions of X. Let $B = A - A' = M' - M$. Then B is the difference of two natural processes, and is also a martingale. By Lemma 14.43, B is indistinguishable from the 0 process. Thus, A and A' are indistinguishable, and M and M' are indistinguishable. □

Suppose we wish to apply Theorem 14.47 to a submartingale X which is continuous. The theorem does not guarantee that the resulting processes M and A are continuous. To ensure this, we need the next result.

Let X be an $\{\mathcal{F}_t\}$-submartingale. Then X is **regular** if for all $a > 0$ and all sequences $\{T_n\}_{n=1}^{\infty}$ in $\mathcal{S}_a$ with $T_n \leqslant T_{n+1}$ pointwise, we have $EX(T_n) \to EX(T)$, where $T = \lim_n T_n$.

Theorem 14.48. *Let X be a right-continuous $\{\mathcal{F}_t\}$-submartingale, where $\{\mathcal{F}_t\}$ satisfies the usual conditions. Suppose X is of class DL. Let A be the natural process in the Doob-Meyer decomposition. Then A is continuous if and only if X is regular.*

Proof. See [8, Theorem 1.4.14]. □

Exercises

14.19. [8, Problem 1.4.9] Prove Propositions 14.44, 14.45, and 14.46.

14.20. [8, Problem 4.13] Prove that a continuous, nonnegative submartingale is regular.

14.5 Continuous L^2 martingales

This section corresponds to [8, Section 1.5].

Throughout this section, we fix a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_t\}$ that satisfies the usual conditions.

Let X be a right-continuous $\{\mathcal{F}_t\}$-martingale. If $EX(t)^2 < \infty$ for all $t \geqslant 0$, then X is **square-integrable** (or an L^2 **martingale**). Let $\mathcal{M}_2$ denote the set of square-integrable $\{\mathcal{F}_t\}$-martingales with $X(0) = 0$ a.s. Let $\mathcal{M}_2^c$ denote the set of all $X \in \mathcal{M}_2$ that have continuous sample paths. Note that $\mathcal{M}_2$ and $\mathcal{M}_2^c$ are both vectors spaces over the reals.

Suppose $X \in \mathcal{M}_2$. By the continuous version of Theorem 12.5, we have that $X^2 = \{X(t)^2 : t \geqslant 0\}$ is a submartingale, and so, by Theorem 14.44, is of class DL. Thus, by the Doob-Meyer decomposition, there exist unique (up to indistinguishability) processes M and A, such that M is a martingale, A is natural, and $X^2 = M + A$. We define the **angle bracket (process) of** X to be $\langle X \rangle := A$. For the value of the process at a specific time, we will typically adopt the notation $\langle X \rangle_t$, but sometimes also $\langle X \rangle (t)$.

Suppose $X \in \mathcal{M}_2^c$. By Exercise 14.20, we have that X^2 is regular. Thus, by Theorem 14.48, $\langle X \rangle$ is continuous.

Note that $\langle X \rangle$ is the unique (up to indistinguishability) $\{\mathcal{F}_t\}$-natural process such that $X^2 - \langle X \rangle$ is an $\{\mathcal{F}_t\}$-martingale. In particular, this implies that, for all $X \in \mathcal{M}_2$, we have $E[X(t)^2] = E\langle X \rangle_t$.

Let $X, Y \in \mathcal{M}_2$. The **angle bracket (process) of** X **and** Y is

$$\langle X, Y \rangle := \frac{1}{4}(\langle X + Y \rangle - \langle X - Y \rangle).$$

Theorem 14.49. *Let $X, Y \in \mathcal{M}_2$. There is a unique (up to indistinguishability) process B which is the difference of two $\{\mathcal{F}_t\}$-natural processes, satisfies $B(0) = 0$ a.s., and makes $XY - B$ an $\{\mathcal{F}_t\}$-martingale. This process is given by the angle bracket $\langle X, Y \rangle$.*

Proof. By the definition of the angle bracket, $\langle X, Y \rangle$ is the difference of two natural processes, and

$$M_1 = (X + Y)^2 - \langle X + Y \rangle$$

and

$$M_2 = (X - Y)^2 - \langle X - Y \rangle$$

are both martingales. Thus,

$$\frac{1}{4}(M_1 - M_2) = XY - \langle X, Y \rangle$$

is also a martingale.

Suppose B is the difference of two natural processes and $XY - B$ is a martingale. Then

$$(XY - \langle X, Y \rangle) - (XY - B) = B - \langle X, Y \rangle$$

is a martingale that is the difference of two natural processes. By Lemma 14.43, B and $\langle X, Y \rangle$ are indistinguishable. □

Two martingales $X, Y \in \mathcal{M}_2$ are **orthogonal** if $\langle X, Y \rangle_t = 0$ a.s. for all $t \geqslant 0$. Note that $\langle X, X \rangle = \langle X \rangle$.

Recall the total variation function T_F from Section 3.4.

Proposition 14.50. *For any $X, Y, Z \in \mathcal{M}_2$ and any $\alpha, \beta \in \mathbb{R}$, we have the following.*

(i) $\langle \alpha X + \beta Y, Z \rangle = \alpha \langle X, Z \rangle + \beta \langle Y, Z \rangle$.

(ii) $\langle X, Y \rangle = \langle Y, X \rangle$.

(iii) $|\langle X, Y \rangle_t|^2 \leqslant \langle X \rangle_t \langle Y \rangle_t$ *for all* $t \geqslant 0$ *a.s.*

(iv) We have

$$T_{\langle X,Y \rangle}(t) - T_{\langle X,Y \rangle}(s) \leqslant \frac{1}{2}(\langle X \rangle_t - \langle X \rangle_s + \langle Y \rangle_t - \langle Y \rangle_t)$$

for all $s < t$ *a.s.*

Proof. Exercise 14.21 □

In the same way that we defined the total variation of a (deterministic) function in Section 3.4, we could define the quadratic variation of a (deterministic) function by using the squares of the increments instead of the absolute values of the increments.

Defining the quadratic variation of a stochastic process is a little more involved. We will not present such a definition here, but the usual notation for the quadratic variation of $X \in \mathcal{M}_2$ is $[X]$. The angle bracket process $\langle X \rangle$ is sometimes called the **predictable quadratic variation** of X. It turns out that if X is continuous, then $[X] = \langle X \rangle$. Since we will focus primarily on continuous processes, we will follow the convention in [8] and refer to $\langle X \rangle$ as the **quadratic variation of** X. We will also refer to $\langle X, Y \rangle$ as the **cross-variation (or co-variation) of** X **and** Y. When reading other sources, however, one should recognize that this is an abuse of standard terminology and notation.

The following theorem illustrates the connection between the angle bracket process and the usual notion of quadratic variation. Before stating the theorem, we need some notation. If X is a stochastic process, $\mathcal{P}$ is a partition of $[0, t]$ and $p > 0$, we define

$$V_t^p(\mathcal{P}) = \sum_{k=1}^n |X(t_k) - X(t_{k-1})|^p.$$

If L is a random variable, we say that $V_t^p(\mathcal{P}) \to L$ in probability (as $\|\mathcal{P}\| \to 0$) if for all $\varepsilon > 0$ and all $\eta > 0$, there exists $\delta > 0$ such that whenever $\|\mathcal{P}\| < \delta$, we have $P(|L - V_t^p(\mathcal{P})| \geqslant \varepsilon) < \eta$. We say that $V_t^p(\mathcal{P}) \to \infty$ in probability (as $\|\mathcal{P}\| \to 0$) if for all $K > 0$ and all $\eta > 0$, there exists $\delta > 0$ such that whenever $\|\mathcal{P}\| < \delta$, we have $P(|V_t^p(\mathcal{P})| \leqslant K) < \eta$.

Theorem 14.51. *Let $X \in \mathcal{M}_2^c$. Fix $t > 0$. Then $V_t^2(\mathcal{P}) \to \langle X \rangle_t$ in probability.*

Proof. See [8, Theorem 1.5.8] □

Proposition 14.52. *Let X be a stochastic process with continuous sample paths. Let L be a $(0, \infty)$-valued stochastic process. Let $p > 0$. Suppose that for all $t > 0$, $V_t^p(\mathcal{P}) \to L_t$ in probability. Then $V_t^q(\mathcal{P}) \to 0$ in probability whenever $t > 0$ and $q > p$, and $V_t^q(\mathcal{P}) \to \infty$ in probability whenever $t > 0$ and $q < p$.*

Proof. Exercise 14.22. □

Proposition 14.53. *Let $X \in \mathcal{M}_2^c$ and let T be an $\{\mathcal{F}_t\}$-stopping time. If $\langle X \rangle_T = 0$ a.s., then $X(T \wedge t) = 0$ for all $t \geqslant 0$ a.s.*

Proof. Exercise 14.23. □

Corollary 14.54. *Let X be a continuous $\{\mathcal{F}_t\}$-martingale with $X(0) = 0$ a.s. Suppose the sample paths of X have bounded variation on compact intervals. Then $X(t) = 0$ for all $t \geqslant 0$ a.s.*

Proof. Define

$$T_n = \inf\{t \geqslant 0 : |X(t)| = n\}.$$

Let $X_n(t) = X(t \wedge T_n)$. Then $X_n \in \mathcal{M}_2^c$. By Theorem 14.51 and Proposition 14.52, $\langle X_n \rangle_t = 0$ a.s. for all $t \geqslant 0$. By Proposition 14.53, it follows that $X_n(t) = 0$ for all $t \geqslant 0$ a.s., which implies that $X(t) = 0$ for all $t \geqslant 0$ a.s. □

Theorem 14.55. *Let $X, Y \in \mathcal{M}_2^c$. There is a unique (up to indistinguishability) continuous, $\{\mathcal{F}_t\}$-adapted process B whose sample paths have bounded variation on compact intervals, satisfies $B(0) = 0$ a.s., and makes $XY - B$ a continuous $\{\mathcal{F}_t\}$-martingale. This process is given by the cross-variation $\langle X, Y\rangle$.*

Proof. The existence of B is given by Theorem 14.49. Suppose B' is another such process. Then

$$(XY - B) - (XY - B') = B' - B$$

is a continuous $\{\mathcal{F}_t\}$-martingale with sample paths that are of bounded variation on compact intervals. By Corollary 14.54, B and B' are indistinguishable. □

Proposition 14.56. *If $X, Y \in \mathcal{M}_2^c$, then for all $t \geqslant 0$,*

$$\sum_{k=1}^{n} (X(t_k) - X(t_{k-1}))(Y(t_k) - Y(t_{k-1})) \to \langle X, Y\rangle_t$$

in probability as $\|\mathcal{P}\| \to 0$.

Proof. Exercise 14.24. □

Let X be an $\{\mathcal{F}_t\}$-adapted stochastic process. Suppose there exists a sequence $\{T_n\}$ of $\{\mathcal{F}_t\}$-stopping times such that

(i) $T_n \leqslant T_{n+1}$ pointwise,

(ii) $T_n \uparrow \infty$ a.s., and

(iii) each X_n is an $\{\mathcal{F}_t\}$-martingale, where $X_n(t) = X(t \wedge T_n)$.

Then X is **local martingale with respect to** $\{\mathcal{F}_t\}$.

The set of $\{\mathcal{F}_t\}$-local martingales X satisfying $X(0) = 0$ a.s. is denoted by $\mathcal{M}^{\mathrm{loc}}$. The set of $X \in \mathcal{M}^{\mathrm{loc}}$ that are continuous is denoted by $\mathcal{M}^{\mathrm{c,loc}}$.

Remark 14.57. Every martingale is a local martingale, but not every local martingale is a martingale. In [8, Chapter 3], there are examples of uniformly integrable local martingales that are not martingales.

Theorem 14.58. *Let $X, Y \in \mathcal{M}^{\mathrm{c,loc}}$. There is a unique (up to indistinguishability) continuous, $\{\mathcal{F}_t\}$-adapted process B whose sample paths have bounded variation on compact intervals, satisfies $B(0) = 0$ a.s., and makes $XY - B$ a continuous $\{\mathcal{F}_t\}$-local martingale.*

Proof. This result is [8, Problem 1.5.17]. The text includes a worked solution. □

The process B in Theorem 14.58 is denoted by $\langle X, Y\rangle$ and, in keeping with our established abuse of terminology, is called the **cross-variation** (or **co-variation**) of X and Y. If $X \in \mathcal{M}^{c,\mathrm{loc}}$, then $\langle X\rangle = \langle X, X\rangle$ is the **quadratic variation** of X.

Remark 14.59. The proof of Theorem 14.58 shows that a single sequence of stopping times, $\{T_n\}$, can be used for both X and Y in the definition of a local martingale, and in this case, $\langle X, Y\rangle_{t\wedge T_n} = \langle X_n, Y_n\rangle_t$. In particular, $\langle X\rangle_{t\wedge T_n} = \langle X_n\rangle_t$, and this shows that $\langle X\rangle$ is an increasing process.

Proposition 14.60. *A local martingale of class DL is a martingale.*

Proof. Exercise 14.25. □

Proposition 14.61. *A nonnegative local martingale is a supermartingale.*

Proof. Exercise 14.25. □

Proposition 14.62. *Let $M \in \mathcal{M}^{c,\mathrm{loc}}$ and define $M(\infty) = \liminf_{t\to\infty} M(t)$. Let S be an $\{\mathcal{F}_t\}$-stopping time. Then $E[M(S)^2] \leqslant E\langle M\rangle_S$.*

Proof. Exercise 14.25. □

For $X \in \mathcal{M}_2$, define

$$\|X\| := \sum_{n=1}^{\infty} \frac{\|X(n)\|_{L^2(\Omega)} \wedge 1}{2^n}.$$

If we identify processes that are indistinguishable, then $\|\cdot\|$ is a norm on $\mathcal{M}_2$. Under the induced metric $(X, Y) \mapsto \|X - Y\|$, the space $\mathcal{M}_2$ is a complete metric space. (That is, $\mathcal{M}_2$ is a Banach space.) Moreover, $\mathcal{M}_2^c$ is a closed subspace of $\mathcal{M}_2$. See [8, Proposition 1.5.23] for details.

Exercises

14.21. [8, Problem 1.5.7] Prove Proposition 14.50.

14.22. [8, Problem 1.5.11] Prove Proposition 14.52.

14.23. [8, Problem 1.5.12] Prove Proposition 14.53.

14.24. [8, Problem 1.5.14] Prove Proposition 14.56.

14.25. [8, Problem 1.5.19] Prove Propositions 14.60, 14.61, and 14.62.

Chapter 15

Brownian Motion

15.1 Introduction

This section is inspired by the beginning of [11, Section 3.1].

Consider the following simple example. Let $Y(t)$ denote the amount of money you have invested in a savings account at time t (measured in years). Let r denote the annual interest rate your receive. Without any further deposits or withdrawals, the function (or process) Y satisfies

$$\frac{dY}{dt} = rY(t).$$

Now suppose that the interest rate changes with time. At any given moment, the interest rate is random, but with mean r, and the randomness is independent from moment to moment. Then we might be led to write

$$\frac{dY}{dt} = (r + W(t))Y(t), \tag{15.1}$$

where W is a stochastic process that represents this ongoing random perturbation of the interest rate. Such a model is often used as an elementary model for your wealth if you have invested in a risky asset such as a stock.

The process W is referred to as "white noise", and ideally, we would like it to have the following properties:

(i) If $s \neq t$, then $W(s)$ and $W(t)$ are independent.

(ii) The process W is stationary. That is,

$$(W(t_1), \ldots, W(t_k)) \stackrel{d}{=} (W(t_1 + t), \ldots, W(t_k + t)).$$

(iii) For all $t \geqslant 0$, $EW(t) = 0$.

Unfortunately, no such process exists, at least not in any reasonable sense. For instance, there is no continuous process with these properties. And there is no measurable process with these properties which also satisfies $E[W(t)^2] = 1$ for all $t \geqslant 0$. (See [11, Section 3.1] and the references therein.)

Nevertheless, let us proceed heuristically as if a nice white noise process exists, and let us consider what type of properties its integral would have. That is, let us "define"

$$B(t) = \int_0^t W(s)\,ds.$$

Then B ought to satisfy the following:

(i) $B(0) = 0$.

(ii) If $0 \leqslant t_1 < \cdots < t_k$, then

$$B(t_1), B(t_2) - B(t_1), \ldots, B(t_k) - B(t_{k-1})$$

are independent.

(iii) $B(t) - B(s)$ is normally distributed with mean 0 and variance proportional to $t - s$.

(iv) B is continuous.

Condition (iii) follows from a heuristic application of the central limit theorem.

It turns out that there *is* a stochastic process which satisfies all of these conditions, and it is Brownian motion, which we will formally define shortly. White noise, then, ought to be the derivative of Brownian motion. Unfortunately, Brownian motion is almost surely nowhere differentiable. (This fact will also be established later.)

Returning to our heuristically derived differential equation (15.1), we find that it now becomes

$$\frac{dY}{dt} = \left(r + \frac{dB}{dt}\right) Y(t).$$

If we write this an integral equation, it becomes

$$Y(t) = Y(0) + r\int_0^t Y(s)\,ds + \int_0^t Y(s)\frac{dB}{ds}\,ds,$$

or

$$Y(t) = Y(0) + r\int_0^t Y(s)\,ds + \int_0^t Y(s)\,dB(s).$$

The nonexistence of white noise manifests itself in this integral equation through the nonexistence of the final integral. It will later be established that Brownian motion is of unbounded variation on all intervals. Hence, the final integral above cannot be understood as an ordinary Lebesgue-Stieltjes integral. A new theory must be developed in order to make sense of and work with integrals of this type. This new theory is theory of Itô integration, which is the starting point in the study of stochastic differential equations.

Remark 15.1. Although white noise cannot be defined as an ordinary stochastic process, it can be made rigorous as an $\mathcal{S}'$-valued random variables, where $\mathcal{S}'$ is the space of what are called "generalized functions" on $[0, \infty)$. In this sense, white noise is analogous to the delta function. The delta function is not a function in the ordinary sense, but it can be rigorously defined as a generalized function. Likewise, white noise is not a stochastic process in the ordinary sense, but it can be rigorously defined as a random generalized function. All of this, however, is beyond the scope of these notes. We will not work formally with white noise in these notes, only with Brownian motion.

15.2 Definition

This section corresponds to [8, Sections 2.1 and 2.2].

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t\}$ a filtration. Let B be a continuous, $\{\mathcal{F}_t\}$-adapted stochastic process such that

(i) $B(0) = 0$ a.s.,

(ii) If $0 \leqslant s < t$, then $B(t) - B(s) \sim N(0, t - s)$, and

(iii) If $0 \leqslant s < t$, then $B(t) - B(s)$ and $\mathcal{F}_s$ are independent.

Then B is a **(standard, one-dimensional) Brownian motion** with respect to $\{\mathcal{F}_t\}$. The word "standard" refers to the fact that $B(0) = 0$ a.s.

Proposition 15.2. *Let B be a continuous stochastic process such that*

(i) $B(0) = 0$ a.s., and

(ii) If $0 \leqslant s < t$, then $B(t) - B(s) \sim N(0, t - s)$.

Then B is an $\{\mathcal{F}_t^B\}$-Brownian motion if and only if

(iii)′ $B(t_1), B(t_2) - B(t_1), \ldots, B(t_k) - B(t_{k-1})$ are independent,

for all $0 \leqslant t_1 < \cdots < t_k$.

Proof. Exercise 15.1. □

In general, if we say that B is a Brownian motion, without reference to a filtration, then we mean that B is an $\{\mathcal{F}_t^B\}$-Brownian motion.

The first issue that must be resolved is whether or not Brownian motion exists. That is, does there exist a stochastic process that satisfies the definition of Brownian motion? The answer, of course, is yes. To prove this, we begin with the following continuous-time version of Kolmogorov's extension theorem.

Theorem 15.3. *For each $n \in \mathbb{N}$ and $\mathbf{t} = (t_1, \ldots, t_n) \in [0, \infty)^n$ where the t_j's are distinct, let $Q_{\mathbf{t}}$ be a probability measure on $(\mathbb{R}^n, \mathcal{R}^n)$. Assume that $\{Q_{\mathbf{t}}\}$ is consistent. That is, assume:*

(i) If $\mathbf{t} = (t_1, \ldots, t_n)$, $\mathbf{s} = (t_1, \ldots, t_{n-1})$, *and* $A \in \mathcal{R}^{n-1}$, *then*

$$Q_{\mathbf{t}}(A \times \mathbb{R}) = Q_{\mathbf{s}}(A).$$

(ii) If $\mathbf{s} = (t_{i_1}, \ldots, t_{i_n})$ *is a permutation of* $\mathbf{t} = (t_1, \ldots, t_n)$ *and* $A_i \in \mathcal{R}$, *then*

$$Q_{\mathbf{t}}(A_1 \times \cdots \times A_n) = Q_{\mathbf{s}}(A_{i_1} \times \cdots \times A_{i_n}).$$

Then there exists a probability measure P *on* $(\mathbb{R}^{[0,\infty)}, \mathcal{R}^{[0,\infty)})$ *such that*

$$Q_{\mathbf{t}}(A) = P(\{\omega \in \mathbb{R}^{[0,\infty)} : (\omega(t_1), \ldots, \omega(t_n)) \in A\}).$$

Proof. See [8, Theorem 2.2.2]. □

Remark 15.4. The σ-algebra $\mathcal{R}^{[0,\infty)}$ is just $\bigotimes_{t\in[0,\infty)} \mathcal{R}$, the same σ-algebra we reviewed at the beginning of Section 14.1.

For $t > 0$ and $x, y \in \mathbb{R}$, let p be the Gaussian kernel given by

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t}.$$

That is, $p(t, x, \cdot)$ is the density of the $N(x, t)$ distribution.

Let $\mathbf{t} = (t_1, \ldots, t_n)$. First assume $t_j \neq 0$ for all j. Let $\mathbf{s} = (t_{i_1}, \ldots, t_{i_n})$ be a permutation of $\mathbf{t}$ such that $0 < s_1 < \cdots < s_n$. Let $U = (U_1, \ldots, U_n)$ be an $\mathbb{R}^n$-valued random variable with density

$$(x_1, \ldots, x_n) \mapsto p(s_1, 0, x_1)p(s_2 - s_1, x_1, x_2) \cdots p(s_n - s_{n-1}, x_{n-1}, x_n),$$

and define $Q_{\mathbf{t}}$ via $Q_{\mathbf{t}}(A_1 \times \cdots \times A_n) = P(U \in A_{i_1} \times \cdots \times A_{i_n})$.

Now assume $t_j = 0$ for some j. Let $\mathbf{s} = (t_{i_1}, \ldots, t_{i_n})$ be a permutation of $\mathbf{t}$ such that $0 = s_1 < s_2 < \cdots < s_n$. Define $Q_{\mathbf{t}}$ via

$$Q_{\mathbf{t}}(A_1 \times \cdots \times A_n) = \delta_0(A_{i_1}) Q_{(s_2,\ldots,s_n)}(A_{i_2} \times \cdots \times A_{i_n}).$$

By showing that $\{Q_{\mathbf{t}}\}$ is consistent and using Theorem 15.3, we can obtain the following.

Corollary 15.5. *Let* $\Omega = \mathbb{R}^{[0,\infty)}$ *and* $\mathcal{F} = \mathcal{R}^{[0,\infty)}$. *For each* $t \geqslant 0$, *define* $B(t) : \Omega \to \mathbb{R}$ *by* $B(t, \omega) = \omega(t)$. *Let* $B = \{B(t) : t \geqslant 0\}$. *Then there exists a probability measure* P *on* $(\Omega, \mathcal{F})$ *such that* B *is a stochastic process that satisfies (i), (ii), and (iii)′ of Proposition 15.2.*

Proof. Exercise 15.2. □

According to Proposition 15.2, the process B in Corollary 15.5 would be a Brownian motion if it were continuous. Even if it were continuous almost surely, that would be enough, since we could change it to the zero process on the corresponding null set. Unfortunately, the set

$$C = \{\omega \in \Omega : B(\cdot, \omega) \text{ is continuous}\}$$

is not in $\mathcal{F}$. So it is not an event, and $P(C)$ is undefined. In fact, it can be shown that if $A \subset C$ and $A \in \mathcal{F}$, then $A = \varnothing$. It follows from this fact that B is not continuous almost surely. To see this, suppose B is continuous almost surely. Then there exists $N \in \mathcal{F}$ such that $P(N) = 0$ and B is continuous on N^c. That is $N^c \subset C$. But this implies $N^c = \varnothing$, and so $N = \Omega$. But $P(\Omega) = 1$, and we have a contradiction.

So the process B in Corollary 15.5 is not a Brownian motion. Our goal is to create a modification of it which is.

Theorem 15.6 (Kolmogorov-Čentsov theorem)**.** *Let $X = \{X(t) : t \in [0,T]\}$ be a stochastic process. Suppose there exists $\alpha, \beta, C > 0$ such that for all $\varepsilon > 0$,*

$$P(|X(t) - X(s)| \geqslant \varepsilon) \leqslant C\varepsilon^{-\alpha}|t-s|^{1+\beta}, \tag{15.2}$$

whenever $s, t \in [0,T]$. Then X has a continuous modification $\widetilde{X}$ which does not depend on α, β, or C, and which satisfies the following: for all $\gamma \in (0, \beta/\alpha)$, there exists a random variable δ with $\delta > 0$ a.s. and

$$\sup_{\substack{0<t-s<\delta \\ s,t\in[0,T]}} \frac{|\widetilde{X}(t) - \widetilde{X}(s)|}{|t-s|^\gamma} \leqslant \frac{2}{1-2^{-\gamma}} \quad a.s.$$

Proof. See [8, Theorem 2.2.8] □

Remark 15.7. Suppose

$$E|X(t) - X(s)|^\alpha \leqslant C|t-s|^{1+\beta}. \tag{15.3}$$

Then, by Chebyshev,

$$P(|X(t) - X(s)| \geqslant \varepsilon) \leqslant \frac{E|X(t) - X(s)|^\alpha}{\varepsilon^\alpha} \leqslant C\varepsilon^{-\alpha}|t-s|^{1+\beta}.$$

Thus, (15.2) can be replaced by (15.3). In fact, the version of the Kolmogorov-Čentsov theorem in [8] uses (15.3), but the first step in the proof is to apply Chebyshev in order to obtain (15.2).

A continuous function $f : [0,T] \to \mathbb{R}$ is said to be **Hölder-continuous with exponent** $\gamma > 0$ (or γ**-Hölder**) if

$$\sup_{\substack{s,t\in[0,T] \\ s\neq t}} \frac{|f(t) - f(s)|}{|t-s|^\gamma} < \infty.$$

Lemma 15.8. *Let $f : [0,T] \to \mathbb{R}$ be continuous and $\gamma > 0$. If there exists $\delta > 0$ such that*

$$\sup_{\substack{s,t\in[0,T] \\ 0<t-s<\delta}} \frac{|f(t) - f(s)|}{|t-s|^\gamma} < \infty,$$

then f is γ-Hölder.

Proof. Note that

$$\sup_{\substack{s,t\in[0,T]\\ t-s>\delta}} \frac{|f(t)-f(s)|}{|t-s|^\gamma} \leqslant \delta^{-\gamma} \sup_{\substack{s,t\in[0,T]\\ t-s>\delta}} |f(t)-f(s)| \leqslant 2\delta^{-\gamma} \sup_{t\in[0,T]} |f(t)| < \infty.$$

Thus,

$$\sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{|f(t)-f(s)|}{|t-s|^\gamma} \leqslant \sup_{\substack{s,t\in[0,T]\\ 0<t-s<\delta}} \frac{|f(t)-f(s)|}{|t-s|^\gamma} + \sup_{\substack{s,t\in[0,T]\\ t-s>\delta}} \frac{|f(t)-f(s)|}{|t-s|^\gamma} < \infty,$$

so f is γ-Hölder. □

Lemma 15.9. *If $f : [0,T] \to \mathbb{R}$ is γ-Hölder and $\lambda < \gamma$, then f is λ-Hölder.*

Proof. This follows from

$$\begin{aligned}
\sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{|f(t)-f(s)|}{|t-s|^\lambda} &\leqslant \sup_{\substack{s,t\in[0,T]\\ 0<t-s<1}} \frac{|f(t)-f(s)|}{|t-s|^\lambda} + \sup_{\substack{s,t\in[0,T]\\ t-s\geqslant 1}} \frac{|f(t)-f(s)|}{|t-s|^\lambda} \\
&\leqslant \sup_{\substack{s,t\in[0,T]\\ 0<t-s<1}} \frac{|f(t)-f(s)|}{|t-s|^\gamma} + \sup_{\substack{s,t\in[0,T]\\ t-s\geqslant 1}} |f(t)-f(s)| \\
&\leqslant \sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{|f(t)-f(s)|}{|t-s|^\gamma} + 2 \sup_{t\in[0,T]} |f(t)|,
\end{aligned}$$

which is finite. □

A continuous function $f : [0,\infty) \to \mathbb{R}$ is said to be **locally Hölder-continuous with exponent** $\gamma > 0$ (or **locally γ-Hölder**) if $f|_{[0,T]}$ if γ-Hölder for all $T > 0$.

Theorem 15.10. *Let $\Omega = \mathbb{R}^{[0,\infty)}$ and $\mathcal{F} = \mathcal{R}^{[0,\infty)}$. There exists a probability measure P on $(\Omega, \mathcal{F})$ and a stochastic process $W = \{W(t) : t \geqslant 0\}$ defined on $(\Omega, \mathcal{F}, P)$ such that W is a Brownian motion. Moreover, almost every sample path of W is locally Hölder-continuous with exponent γ, for every $\gamma \in (0, 1/2)$.*

Proof. Let B be the process in Corollary 15.5. Since $B(t) - B(s) \sim N(0, t-s)$, we have

$$E|B(t) - B(s)|^{2n} = (2n-1)!!|t-s|^n.$$

Taking $n = 2$, we may apply Theorem 15.6 with $\alpha = 4$ and $\beta = 1$. Doing this for each $T \in \mathbb{N}$, we obtain continuous processes $W^T = \{W^T(t) : t \in [0,T]\}$, which are modifications of B on $[0,T]$.

Let

$$\Omega_1 = \bigcap_{T=1}^{\infty} \bigcap_{t\in\mathbb{Q}\cap[0,T]} \{W^T(t) = B(t)\},$$

so that $P(\Omega_1) = 1$. Fix $\omega \in \Omega_1$ and $T \in \mathbb{N}$. Then for any $t \in \mathbb{Q} \cap [0, T]$, we have

$$W^T(t, \omega) = B(t, \omega) = W^{T+1}(t, \omega).$$

Since W^T and W^{T+1} are continuous, we have that $W^T(t, \omega) = W^{T+1}(t, \omega)$ for all $t \in [0, T]$.

For $t \geqslant 0$, choose $T \geqslant t$ and define $W(t) = W^T(t) 1_{\Omega_1}$. Then W is a continuous modification of B, and so W is a Brownian motion.

Now fix $T \in \mathbb{N}$. Fix $j \in \mathbb{N}$. Choose $n \in \mathbb{N}$ such that

$$\frac{1}{2} - \frac{1}{j} < \frac{n-1}{2n}.$$

By Theorem 15.6 with $\alpha = 2n$, $\beta = n - 1$, and $\gamma_j = 1/2 - 1/j$, we may choose a positive random variable δ_j and an event $A_{j,T}$ with $P(A_{j,T}) = 1$ such that for all $\omega \in A_{j,T}$,

$$\sup_{\substack{0 < t-s < \delta(\omega) \\ s,t \in [0,T]}} \frac{|W^T(t, \omega) - W^T(s, \omega)|}{|t - s|^{\gamma_j}} < \infty.$$

By Lemma 15.8, $W^T(\cdot, \omega)$ is γ_j-Hölder on $[0, T]$.

Now let

$$\Omega_2 = \bigcap_{T=1}^{\infty} \bigcap_{j=1}^{\infty} A_{j,T},$$

so that $P(\Omega_2) = 1$. Let $\omega \in \Omega_2$, $T \in (0, \infty)$, and $\gamma \in (0, 1/2)$. Choose $\widetilde{T} \in \mathbb{N}$ with $T < \widetilde{T}$ and $j \in \mathbb{N}$ with $\gamma < \gamma_j$. Then $\omega \in A_{j,\widetilde{T}}$ implies $W(\cdot, \omega)$ is γ_j-Hölder on $[0, \widetilde{T}]$, and therefore on $[0, T]$. By Lemma 15.9, $W(\cdot, \omega)$ is γ-Hölder on $[0, T]$. Since this is true for all $T > 0$, the sample path $W(\cdot, \omega)$ is locally γ-Hölder. Since γ was arbitrary, this completes the proof. □

Exercises

15.1. [8, Problem 2.1.4] Prove Proposition 15.2.

15.2. [8, Problem 2.2.5] Prove Corollary 15.5.

15.3 Donsker's invariance principle

Roughly speaking, Donsker's invariance principle states that a sequence of random walks, in which the sizes of the temporal and spatial steps go to zero at appropriate rates, will converge in distribution to a Brownian motion. This theorem provides an alternative proof of the existence of Brownian motion. But it also provides an intuitive understanding of Brownian motion itself. Brownian motion is the continuous-time analog of a random walk.

Donsker's invariance principle is a theorem which asserts the convergence in distribution of a sequence of random variables taking values in the space of continuous functions. To properly understand the theorem, we must generalize some of the notions from Section 7.3. In particular, we will need to generalize the notion of tightness and use the proof method described in Remark 7.28.

Let $C[0,\infty)$ denote the set of continuous functions from $[0,\infty)$ to $\mathbb{R}$. Note that $C[0,\infty)$ is a vector space over $\mathbb{R}$.

Proposition 15.11. *For $f,g \in C[0,\infty)$, define*

$$\rho(f,g) = \sum_{n=1}^{\infty} 2^{-n} \sup_{t\in[0,n]} (|f(t)-g(t)| \wedge 1).$$

Then ρ is a metric on $C[0,\infty)$ which makes $C[0,\infty)$ a complete and separable metric space. Moreover, $f_n \to f$ in this metric if and only if $f_n \to f$ uniformly on compact intervals.

Proof. Exercise 15.3. □

A **cylinder set** is a set $C \subset C[0,\infty)$ of the form

$$C = \{f \in C[0,\infty) : (f(t_1),\ldots,f(t_n)) \in A\}, \tag{15.4}$$

for some $n \in \mathbb{N}$, $t_j \in [0,\infty)$, and $A \in \mathcal{R}^n$.

Proposition 15.12. *Let $\mathcal{C}$ denote the collection of all cylinder sets in $C[0,\infty)$. Then $\mathcal{B}_{C[0,\infty)} = \sigma(\mathcal{C})$.*

Proof. Exercise 15.4. □

Lemma 15.13. *Let $X : \Omega \to C[0,\infty)$ and let $\mathcal{G}$ be a σ-algebra on Ω. Then X is $(\mathcal{G},\mathcal{B}_{C[0,\infty)})$-measurable if and only if $\pi_t \circ X$ is $(\mathcal{G},\mathcal{R})$-measurable for all $t \geqslant 0$.*

Proof. Suppose X is $(\mathcal{G},\mathcal{B}_{C[0,\infty)})$. Since $\pi_t : C[0,\infty) \to \mathbb{R}$ is continuous, it follows that π_t is $(\mathcal{B}_{C[0,\infty)},\mathcal{R})$-measurable. Thus, $\pi_t \circ X$ is $(\mathcal{G},\mathcal{R})$-measurable.

Now suppose $\pi_t \circ X$ is $(\mathcal{G},\mathcal{R})$-measurable for all $t \geqslant 0$. Let $C \in \mathcal{C}$ be given by (15.4). Then

$$X^{-1}(C) = \{(\pi_{t_1} \circ X, \ldots, \pi_{t_n} \circ X) \in A\}.$$

By Corollary 2.7, $(\pi_{t_1} \circ X, \ldots, \pi_{t_n} \circ X)$ is $(\mathcal{F},\mathcal{R}^n)$-measurable. It therefore follows that $X^{-1}(C) \in \mathcal{F}$. Since $\mathcal{B}_{C[0,\infty)} = \sigma(\mathcal{C})$, Proposition 2.2 implies that X is $(\mathcal{G},\mathcal{B}_{C[0,\infty)})$-measurable. □

We can now prove the following variation on Lemma 14.1.

Lemma 15.14. *With notation as above, we have the following.*

(i) If $\{X(t) : t \geqslant 0\}$ is a real-valued stochastic process with continuous sample paths, and $X : \Omega \to C[0,\infty)$ is defined by $X(\omega) = X(\cdot,\omega)$, then X is a $C[0,\infty)$-valued random variable.

(ii) If $X : \Omega \to C[0,\infty)$ is a $C[0,\infty)$-valued random variable, and $X(t) : \Omega \to \mathbb{R}$ is defined by $X(t) = \pi_t \circ X$, then $\{X(t) : t \geqslant 0\}$ is a real-valued stochastic process with continuous sample paths.

(iii) In either case, $\sigma(X) = \sigma(\{X(t) : t \geqslant 0\})$.

Proof. By Lemma 15.13, for any σ-algebra $\mathcal{G}$ on Ω, we have that the function X is $(\mathcal{G}, \mathcal{B}_{C[0,\infty)})$-measurable if and only if $\pi_t \circ X$ is $(\mathcal{G}, \mathcal{R})$-measurable for all $t \geqslant 0$. Taking $\mathcal{G} = \mathcal{F}$ gives us (i) and (ii). Taking $\mathcal{G} = \sigma(X)$ gives us $\sigma(X) \supset \sigma(\{X(t) : t \geqslant 0\})$. And taking $\mathcal{G} = \sigma(\{X(t) : t \geqslant 0\})$ gives us $\sigma(X) \subset \sigma(\{X(t) : t \geqslant 0\})$. □

We can now identify continuous stochastic processes with $C[0,\infty)$-valued random variables. If X and Y are continuous stochastic processes, then they can both be regarded as $C[0,\infty)$-valued random variables. To say that $X =_d Y$ in $C[0,\infty)$ is to say that $P(X \in A) = P(Y \in A)$ for all $A \in \mathcal{B}_{C[0,\infty)}$.

Lemma 15.15. *Let X and Y be continuous stochastic processes. Then $X =_d Y$ in $C[0,\infty)$ if and only if X and Y have the same finite-dimensional distributions.*

Proof. Suppose $X =_d Y$ in $C[0,\infty)$. Fix $d \in \mathbb{N}$ and $t_1, \ldots, t_d \in [0,\infty)$. Define

$$\pi_{t_1,\ldots,t_d} : C[0,\infty) \to \mathbb{R}^d$$

by $\pi_{t_1,\ldots,t_d}(f) = (f(t_1), \ldots, f(t_d))$. Note that $\pi_{t_1,\ldots,t_d}$ is continuous and therefore measurable. Thus, for any $A \in \mathcal{R}^d$,

$$\begin{aligned} P((X(t_1), \ldots, X(t_d)) \in A) &= P(\pi_{t_1,\ldots,t_d} \circ X \in A) \\ &= P(X \in \pi_{t_1,\ldots,t_d}^{-1}(A)) \\ &= P(Y \in \pi_{t_1,\ldots,t_d}^{-1}(A)) \\ &= P((Y(t_1), \ldots, Y(t_d)) \in A). \end{aligned}$$

Thus, X and Y have the same finite-dimensional distributions.

Now assume X and Y have the same finite-dimensional distributions. Let

$$\mathcal{L} = \{A \in \mathcal{B}_{C[0,\infty)} : P(X \in A) = P(Y \in A)\}.$$

Since X and Y have the same finite-dimensional distributions, it follows that $\mathcal{C} \subset \mathcal{L}$, where $\mathcal{C}$ is the collection of cylinder sets. Since $\mathcal{C}$ is a π-system and $\mathcal{L}$ is a λ-system, it follows that $X =_d Y$ in $C[0,\infty)$. □

According to the definition in Section 7.3, to say that $X_n \Rightarrow X_\infty$ in $C[0,\infty)$ is to say that $E[G(X_n)] \to E[G(X_\infty)]$ for all bounded, continuous functions $G : C[0,\infty) \to \mathbb{R}$. The function G is continuous if $G(f_n) \to G(f_\infty)$ whenever $f_n \to f_\infty$ locally uniformly.

Another way to think about $X_n \Rightarrow X_\infty$ in $C[0,\infty)$ is given by Remark 7.16. By the Skorohod representation theorem, $X_n \Rightarrow X_\infty$ in $C[0,\infty)$ if and only if there are continuous stochastic processes Y_n such that X_n and Y_n have the same finite-dimensional distributions for each n, and $Y_n \to Y_\infty$ locally uniformly, a.s.

We will say that $X_n \Rightarrow X_\infty$ **in the fdd sense** if the finite-dimensional distributions of X_n converge in distribution to those of X_∞. That is

$$(X_n(t_1), \ldots, X_n(t_d)) \Rightarrow (X_\infty(t_1), \ldots, X_\infty(t_d))$$

as $n \to \infty$, for any $d \in \mathbb{N}$ and any $t_j \in [0, \infty)$.

Lemma 15.16. *If $X_n \Rightarrow X_\infty$ in $C[0,\infty)$, then $X_n \Rightarrow X_\infty$ in the fdd sense.*

Proof. Suppose $X_n \Rightarrow X_\infty$ in $C[0,\infty)$. By Remark 7.20 and the fact that $\pi_{t_1,\ldots,t_d}$ is continuous, we have

$$\begin{aligned}(X_n(t_1), \ldots, X_n(t_d)) &= \pi_{t_1,\ldots,t_d}(X_n)\\ &\Rightarrow \pi_{t_1,\ldots,t_d}(X_\infty) = (X_n(t_\infty), \ldots, X_n(t_\infty)).\end{aligned}$$

Thus, $X_n \Rightarrow X_\infty$ in the fdd sense. □

The converse of Lemma 15.16 is not true. Convergence in the fdd sense is not sufficient to give convergence in $C[0,\infty)$. To obtain convergence in $C[0,\infty)$, we need both convergence in the fdd sense and the additional property of "tightness".

In Section 7.3, we defined what it means for a sequence of real-valued random variables to be tight. Here, we extend that definition to $C[0,\infty)$-valued random variables. In fact, we will extend it to M-valued random variables, where M is any metric space.

Let (M, ρ) be a metric space. Let $\{\mu_\alpha\}_{\alpha \in A}$ be a family of Borel probability measures on M. We say that $\{\mu_\alpha\}$ is **tight** if, for all $\varepsilon > 0$, there exists a compact $K \subset M$ such that

$$\mu_\alpha(K) \geqslant 1 - \varepsilon,$$

for all $\alpha \in A$. A sequence of M-valued random variables, $\{X_n\}_{n=1}^\infty$, is **tight** if $\{\mu_n\}_{n=1}^\infty$ is tight, where $X_n \sim \mu_n$. That is, if, for all $\varepsilon > 0$, there exists a compact $K \subset M$ such that

$$P(X_n \in K) \geqslant 1 - \varepsilon,$$

for all n.

The following is a generalization of Theorem 7.27.

Theorem 15.17. *Let M be a complete and separable metric space. Let $\{\mu_\alpha\}_{\alpha \in A}$ be a family of Borel probability measures on M. Then $\{\mu_\alpha\}$ is tight if and only if it is relatively compact, that is, every sequence $\{\mu_{\alpha(n)}\}_{n=1}^\infty$ has a subsequence that converges weakly.*

In particular, a sequence of random variables taking values in a complete and separable metric space is tight if and only if every subsequence has a further subsequence that converges in distribution.

Proof. See [8, Theorem 2.4.7] and the references therein. □

To apply this theorem to $M = C[0,\infty)$, we must first characterize the compact subsets of $C[0,\infty)$. If $f \in C[0,\infty)$, $T > 0$, and $\delta > 0$, define

$$m^T(f,\delta) = \sup_{\substack{|s-t|\leqslant\delta \\ 0\leqslant s,t\leqslant T}} |f(s) - f(t)|,$$

which we call the **modulus of continuity of f on** $[0,T]$.

Proposition 15.18. *The function $m^T(\cdot,\delta)$ is continuous, the function $m^T(f,\cdot)$ is increasing, and $m^T(f,\delta) \downarrow 0$ as $\delta \downarrow 0$.*

Proof. Exercise 15.5. □

Theorem 15.19 (Arzelà-Ascoli theorem)**.** *Let $A \subset C[0,\infty)$. Then the closure of A is compact if and only if the following conditions hold:*

(i) $\sup\{|f(0)| : f \in A\} < \infty$, *and*

(ii) for all $T > 0$, we have $\sup\{m^T(f,\delta) : f \in A\} \to 0$ *as $\delta \downarrow 0$.*

Proof. See [8, Theorem 2.4.9]. □

Combining Theorems 15.19 and 15.17, we obtain the following necessary and sufficient conditions for tightness in $C[0,\infty)$.

Theorem 15.20. *Let $\{X_n\}$ be a sequence of $C[0,\infty)$-valued random variables. Then $\{X_n\}$ is tight if and only if the following conditions hold:*

(i) $\sup_n P(|X_n(0)| > M) \to 0$ *as $M \to \infty$, and*

(ii) $\sup_n P(m^T(X_n,\delta) > \varepsilon) \to 0$ *as $\delta \downarrow 0$ for all $T > 0$ and all $\varepsilon > 0$.*

Proof. See [8, Theorem 2.4.10]. □

Remark 15.21. Condition (i) in Theorem 15.20 is equivalent to the assertion that the sequence of real-valued random variables $\{X_n(0)\}_{n=1}^\infty$ is tight.

Remark 15.22. For simplicity, Theorem 15.20 is stated under the assumption that the X_n's are defined on the same probability space, and so P does not depend on n. The theorem is still true if each X_n is defined on its own probability space.

Although Theorem 15.20 provides necessary and sufficient conditions for tightness, it is often not a very practical tool for verifying tightness in specific applications. Using methods similar to the proof of the Kolmogorov-Čentsov theorem (Theorem 15.6), we obtain the following sufficient conditions for tightness.

Proposition 15.23. *Let $\{X_n\}$ be a sequence of $C[0,\infty)$-valued random variables. Suppose that there exists $\alpha,\beta,\nu > 0$ and a family of positive constants $\{C_T\}_{T>0}$ such that* *(i)* $\sup_n E|X_n(0)|^\nu < \infty$, *and*

(ii) $\sup_n E|X_n(t) - X_n(s)|^\alpha \leqslant C_T|t-s|^{1+\beta}$ *for all* $0 \leqslant s, t \leqslant T$.

Then $\{X_n\}$ *is tight.*

Proof. See [8, Problem 2.4.11]. □

Finally, the proof method described in Remark 7.28, as it applies to the current situation, is encoded in the following theorem.

Theorem 15.24. *Let* $\{X_n\}$ *be a sequence of continuous stochastic processes. Suppose* $\{X_n\}$ *is tight and that the finite-dimensional distributions converge in distribution. Then there exists a continuous process* X_∞ *such that* $X_n \Rightarrow X_\infty$ *in* $C[0,\infty)$ *as* $n \to \infty$.

Remark 15.25. In Theorem 15.24, each X_n might be defined on its own probability space. In particular, even if $\{X_n : n \in \mathbb{N}\}$ are all defined on a common probability space, the theorem does not guarantee that X_∞ can be defined on that space.

Remark 15.26. To say that the finite-dimensional distributions of the sequence of processes $\{X_n\}$ converge in distribution is to say that for all $d \in \mathbb{N}$ and all $t_1, \ldots, t_d \in [0,\infty)$, there exists an $\mathbb{R}^d$-valued random vector U such that

$$(X_n(t_1), \ldots, X_n(t_d)) \Rightarrow U,$$

as $n \to \infty$.

Proof of Theorem 15.24. By Remark 7.26, we may employ a subsequential argument (using Theorem 7.2).

By Theorem 15.17, there exists a subsequence $\{X_{n(m)}\}$ and a continuous process X_∞ such that $X_{n(m)} \Rightarrow X_\infty$ in $C[0,\infty)$ as $m \to \infty$.

Let $\{X_{\tilde{n}(m)}\}$ be an arbitrary subsequence. By Theorem 15.17, there exists a further subsequence $\{X_{\tilde{n}(m_k)}\}$ and a continuous process $\widetilde{X}_\infty$ such that $X_{\tilde{n}(m_k)} \Rightarrow \widetilde{X}_\infty$ in $C[0,\infty)$ as $k \to \infty$. Since the finite-dimensional distributions of $\{X_n\}$ converge, it follows that X_∞ and $\widetilde{X}_\infty$ have the same finite-dimensional distributions. By Lemma 15.15, $X_\infty =_d \widetilde{X}_\infty$ in $C[0,\infty)$. Thus, $X_{\tilde{n}(m_k)} \Rightarrow X_\infty$ in $C[0,\infty)$ as $k \to \infty$. Since $\{X_{\tilde{n}(m)}\}$ was arbitrary, it follows from Theorem 7.2 that $X_n \Rightarrow X_\infty$ in $C[0,\infty)$ as $n \to \infty$. □

Corollary 15.27. *Let* $\{X_n\}$ *and* X *be continuous stochastic processes. Suppose* $\{X_n\}$ *is tight and* $X_n \Rightarrow X$ *in the fdd sense. Then* $X_n \Rightarrow X$ *in* $C[0,\infty)$.

Proof. By Theorem 15.24, there exists a continuous process $\widetilde{X}$ such that $X_n \Rightarrow \widetilde{X}$ in $C[0,\infty)$. But then X and $\widetilde{X}$ have the same finite-dimensional distributions. By Lemma 15.15, $X =_d \widetilde{X}$ in $C[0,\infty)$. Thus, $X_n \Rightarrow X$ in $C[0,\infty)$. □

The following result is a generalization of Exercise 7.10, which is often useful when working with convergence in distribution in $C[0,\infty)$.

Proposition 15.28. *Let (M, ρ) be a separable metric space. Let X_n, Y_n, X be M-valued random variables. Suppose $X_n \Rightarrow X$ and $\rho(X_n, Y_n) \to 0$ in probability. Then $Y_n \Rightarrow X$.*

Proof. Exercise 15.6. □

Let $\{\xi_j\}$ be i.i.d. real-valued random variables with mean 0 and variance 1. Let $S_0 = 0$ and $S_n = \xi_1 + \cdots + \xi_n$. For $t \geqslant 0$, define

$$Y(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\xi_{\lfloor t \rfloor + 1}.$$

Then $Y = \{Y(t) : t \geqslant 0\}$ is a continuous stochastic process which is a linear interpolation of a mean 0 random walk. For $n \in \mathbb{N}$ and $t \geqslant 0$, let

$$X_n(t) = n^{-1/2} Y(nt).$$

Then $X_n = \{X_n(t) : t \geqslant 0\}$ is the same continuous process, but with time and space scaled by a factor of n^{-1} and $n^{-1/2}$, respectively.

Theorem 15.29 (Donsker's invariance principle)**.** *There exists a Brownian motion W such that $X_n \Rightarrow W$.*

Proof idea. The substantial portion of the proof, which requires multiple lemmas in the text, is to prove that $\{X_n\}$ is tight. In addition to this, one must use the central limit theorem to prove that the finite-dimensional distributions of X_n converge. Theorem 15.24 then implies there exists a continuous process W such that $X_n \Rightarrow W$. One can then use the earlier calculations from the central limit theorem to verify that W has the necessary finite-dimensional distributions to be a Brownian motion. For details, see [8, Theorem 2.4.20]. □

Remark 15.30. Donsker's invariance principle can be regarded as a constructive proof of the existence of Brownian motion. It constructs Brownian motion W as a $C[0, \infty)$-valued random variable. As stated here in these notes, the underlying probability space on which W is built is left unspecified. However, it can always be built in the following canonical fashion.

Suppose W is a $C[0, \infty)$-valued random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $\widetilde{\Omega} = C[0, \infty)$, $\widetilde{\mathcal{F}} = \mathcal{B}_{C[0,\infty)}$, and $\widetilde{W}(t, \omega) = \omega(t)$. Define $\widetilde{P}$ on $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ by $\widetilde{P}(A) = P(W \in A)$. Then $\widetilde{W} =_d W$. In particular, if W is a Brownian motion, then $\widetilde{W}$ is a Brownian motion. In this case, the canonical measure $\widetilde{P}$ is called **Wiener measure** and the probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ is the **canonical probability space** for Brownian motion.

Exercises

15.3. [8, Problem 2.4.1] Prove Proposition 15.11.

15.4. [8, Problem 2.4.2] Prove Proposition 15.12.

15.5. [8, Problem 2.4.8] Prove Proposition 15.18.

15.6. [8, Problem 2.4.16] Prove Proposition 15.28.

15.4 Properties of Brownian motion

Brownian motion is a continuous, square-integrable martingale (see Exercise 15.7). We would like to say that $B \in \mathcal{M}_2^c$, so that we can use all the results of Section 14.5. But to do that, we need to have a filtration that satisfies the usual conditions. Unfortunately, $\{\mathcal{F}_t^B\}$ does not. To this end, we introduce a new concept.

Let $\{\mathcal{G}_t\}$ be a filtration.

Exercises

15.7. Prove that Brownian motion is a continuous, square-integrable martingale.

Bibliography

[1] Erhan Çınlar. *Probability and stochastics*, volume 261 of *Graduate Texts in Mathematics*. Springer, New York, 2011.

[2] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[3] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[4] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[5] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[6] Jean Jacod and Philip Protter. *Probability essentials*. Universitext. Springer-Verlag, Berlin, second edition, 2003.

[7] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, 1997.

[8] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[9] John M. Lee. *Introduction to topological manifolds*, volume 202 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2011.

[10] James R. Munkres. *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. Second edition of [MR0464128].

[11] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, 1985. An introduction with applications.

[12] Sheldon Ross. *A first course in probability.* Pearson, ninth edition, 2012.

[13] Stephen Willard. *General topology.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1970.