Probability Theory

Lectures and Problems

Brando Magnani

Contents

1 Measures and Generated Sigma-Algebras **1**

1.1 Measurable spaces and measures 1

1.2 Additivity and continuity 2

1.3 Discrete and nonatomic probabilities 3

1.4 Generated sigma-algebras 4

1.5 Extension of measures 6

1.6 Dynkin's theorem and uniqueness 8

Problems 10

2 Random Variables, Distributions, and Expectation **11**

2.1 Measurable functions and simple approximation 11

2.2 Generated sigma-algebras and almost-sure statements 13

2.3 Laws and distribution functions 13

2.4 Integration and expectation 16

2.5 Transformation and change of measure 17

2.6 Densities and integration against a distribution function 18

2.7 Three useful inequalities 20

Problems 21

3 Modes of Convergence **23**

3.1 Almost-sure convergence and convergence in probability 23

3.2 Convergence in L^p 24

3.3 Passing to the limit in expectations 25

3.4 General measure spaces and almost-uniform convergence 29

Problems 33

4 Product Measures, Characteristic Functions, and Weak Convergence 34
4.1 Product spaces and independence . . . 34
4.2 Integration on product spaces . . . 36
4.3 Characteristic functions and moments . . . 36
4.4 Fourier inversion and uniqueness . . . 39
4.5 Convolution and sums of independent variables . . . 41
4.6 Weak convergence and the continuity theorem . . . 42
Problems . . . 46

5 Laws of Large Numbers and Convergent Series 49
5.1 Weak laws of large numbers . . . 49
5.2 The Borel–Cantelli lemmas . . . 50
5.3 A first strong law . . . 51
5.4 A maximal inequality and the one-series theorem . . . 52
5.5 The strong law under a first-moment assumption . . . 53
5.6 The strong law with an extended expectation . . . 56
5.7 The two-series theorem . . . 57
Problems . . . 58

6 Central Limit Theorems 61
6.1 The iid central limit theorem . . . 61
6.2 Lindeberg's condition for triangular arrays . . . 62
6.3 Gaussian replacement . . . 63
6.4 Lyapunov's condition . . . 66
6.5 An application: the three-series theorem . . . 67
Problems . . . 69

7 Conditional Expectation and Conditional Distributions 73
7.1 Conditioning on finitely many atoms . . . 73
7.2 Definition, uniqueness, and existence . . . 74
7.3 Properties of conditional expectation . . . 76
7.4 Conditional expectation as an orthogonal projection . . . 78
7.5 Conditional densities . . . 81
7.6 Regular conditional distributions . . . 83
Problems . . . 85

8 Markov Chains and Recurrence **87**
8.1 Filtrations and transition probabilities 87
8.2 The law on path space 89
8.3 Stopping times and stopped sigma-fields 90
8.4 The ordinary and strong Markov properties 91
8.5 Countable chains and communication 93
8.6 Recurrence, visits, and the Green function 94
8.7 Types of communicating classes 96
8.8 Martingales at bounded stopping times 98
Problems 99

A Integration Against Distribution Functions **103**

Chapter 1

Measures and Generated Sigma-Algebras

1.1 Measurable spaces and measures

Definition 1.1. Let Ω be a set. A collection $\mathcal{F} \subseteq 2^\Omega$ is a *σ-algebra* if it satisfies the following conditions:

(a) $\Omega \in \mathcal{F}$.

(b) If $A \in \mathcal{F}$, then $A^c = \Omega \setminus A \in \mathcal{F}$.

(c) If $A_n \in \mathcal{F}$ for every $n \geq 1$, then $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is a *measurable space*; the members of $\mathcal{F}$ are its measurable sets.

In particular, $\varnothing \in \mathcal{F}$. A σ-algebra is also closed under countable intersections, by De Morgan's law:

$$\bigcap_{n=1}^\infty A_n = \left(\bigcup_{n=1}^\infty A_n^c \right)^c \in \mathcal{F}.$$

If we require closure under finite unions instead of countable unions, we obtain an *algebra* of sets. An algebra is closed under finite intersections and set differences.

Definition 1.2. A *measure* on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ such that $\mu(\varnothing) = 0$ and

$$\mu\left(\bigcup_{n=1}^\infty A_n \right) = \sum_{n=1}^\infty \mu(A_n) \tag{1.1}$$

for every countable collection of pairwise disjoint sets $A_n \in \mathcal{F}$. This property is called *countable additivity*.

If additivity is required only for finite disjoint collections, the set function is *finitely additive*. Every measure is finitely additive, by applying (1.1) with all but finitely many sets empty.

A measure μ is *finite* if $\mu(\Omega) < \infty$, and is *σ-finite* if there are sets $B_n \in \mathcal{F}$ such that

$$\Omega = \bigcup_{n=1}^\infty B_n, \qquad \mu(B_n) < \infty \quad \text{for every } n.$$

If $\mu(\Omega) = 1$, we call μ a *probability measure* and usually write P for it. The triple $(\Omega, \mathcal{F}, \mu)$ is a *measure space*; $(\Omega, \mathcal{F}, P)$ is a *probability space*.

1.2 Additivity and continuity

We write $A_n \uparrow A$ when $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_n A_n$. Similarly, $A_n \downarrow A$ means $A_1 \supseteq A_2 \supseteq \cdots$ and $A = \bigcap_n A_n$.

The following properties are elementary exercises in applying countable additivity.

Proposition 1.3. *Let μ be a measure on $(\Omega, \mathcal{F})$.*

(i) If $A \subseteq B$, then $\mu(A) \le \mu(B)$.

(ii) If $A \subseteq \bigcup_{n=1}^{\infty} A_n$, then $\mu(A) \le \sum_{n=1}^{\infty} \mu(A_n)$.

(iii) If $A_n \uparrow A$, then $\mu(A_n) \uparrow \mu(A)$ (continuity from below).

(iv) If $A_n \downarrow A$ and $\mu(A_m) < \infty$ for some m, then $\mu(A_n) \downarrow \mu(A)$ (continuity from above).

All the sets appearing here are assumed measurable.

The first three properties hold for every measure, without a finiteness assumption. In a probability space the fourth also always applies.

Example 1.4. The finiteness qualification in continuity from above cannot be omitted. With Lebesgue measure on $\mathbb{R}$, the sets $A_n = [n, \infty)$ decrease to $\varnothing$, while $\mu(A_n) = \infty$ for every n. The same phenomenon occurs with counting measure: on $\mathbb{N}$, take $A_n = \{n, n+1, \ldots\}$ and let $\mu(A)$ be the number of elements of A, with value ∞ for infinite sets.

Continuity can also be used to establish countable additivity of a finitely additive set function. When its domain is only an algebra, the limit set or union in the assertion must still belong to that algebra.

Proposition 1.5. *Let $\mathcal{A}$ be an algebra and let $\mu : \mathcal{A} \to [0, \infty]$ be finitely additive, with $\mu(\varnothing) = 0$.*

(i) Countable additivity for disjoint sequences whose union belongs to $\mathcal{A}$ is equivalent to continuity from below whenever the increasing union belongs to $\mathcal{A}$.

(ii) The condition

$$B_n \in \mathcal{A}, \quad B_n \downarrow \varnothing \quad \Longrightarrow \quad \mu(B_n) \downarrow 0 \tag{1.2}$$

is sufficient for countable additivity, even if $\mu(\Omega) = \infty$. If $\mu(\Omega) < \infty$, it is also necessary.

Proof. For (i), suppose $C_n \uparrow C \in \mathcal{A}$. The disjoint increments $D_1 = C_1$ and $D_n = C_n \setminus C_{n-1}$, $n \ge 2$, give

$$\mu(C) = \sum_{n=1}^{\infty} \mu(D_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu(D_n) = \lim_{N\to\infty} \mu(C_N).$$

Conversely, if $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ is a disjoint union, then finite additivity and continuity from below give

$$\mu(A) = \lim_{N\to\infty} \mu\left(\bigcup_{n=1}^{N} A_n\right) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu(A_n).$$

For (ii), set $B_N = A \setminus \bigcup_{n=1}^{N} A_n$. These sets belong to $\mathcal{A}$ and decrease to $\varnothing$. Finite additivity gives

$$\mu(A) = \sum_{n=1}^{N} \mu(A_n) + \mu(B_N).$$

Letting $N \to \infty$ proves countable additivity under (1.2), without subtracting infinite quantities. If $\mu(\Omega) < \infty$, necessity follows from continuity from above; its proof uses only sets and unions that belong to $\mathcal{A}$. □

Requiring (1.2) only for sequences with a finite-mass member is a weaker condition. For an infinite set function, that weaker condition need not imply countable additivity.

Example 1.6. On $\Omega = \mathbb{N}$, with $\mathcal{F} = 2^{\mathbb{N}}$, define

$$\mu(A) = \begin{cases} 0, & |A| < \infty, \\ \infty, & |A| = \infty. \end{cases}$$

This set function is finitely additive: a finite union of finite sets is finite, whereas a union containing an infinite set is infinite. It also satisfies continuity from above for every decreasing sequence having a finite-mass member, since all later sets then have measure zero. Nevertheless, for $A_i = \{i\}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu(\mathbb{N}) = \infty, \qquad \sum_{i=1}^{\infty} \mu(A_i) = 0.$$

Thus it is not countably additive. Indeed, even though $\mathbb{N}$ is covered by its zero-mass singletons, this does not repair the failure. For the decreasing tails $B_n = \{n, n+1, \ldots\}$, $\mu(B_n) = \infty$ for all n, so the unrestricted condition (1.2) fails.

1.3 Discrete and nonatomic probabilities

If Ω is countable, we may take $\mathcal{F} = 2^{\Omega}$ and assign a number $p_\omega \geq 0$ to each $\omega \in \Omega$, with $\sum_{\omega \in \Omega} p_\omega = 1$. Then

$$P(A) = \sum_{\omega \in A} p_\omega, \qquad A \subseteq \Omega,$$

defines a probability measure. Conversely, every probability measure on this space has this form, with $p_\omega = P(\{\omega\})$. Countable additivity follows by regrouping a nonnegative series over disjoint sets.

For an uncountable Ω, assigning masses to individual points does not describe every probability measure. Interpret the sum of a nonnegative family as the supremum of its finite subsums. A family of nonnegative point masses with total sum one has at most countably many positive terms: for each $k \geq 1$, only finitely many can be at least $1/k$. Such an assignment still defines a probability, but it is concentrated on a countable set. In particular, it cannot produce a nonatomic probability.

Definition 1.7. A probability measure P is *nonatomic* if, for every $A \in \mathcal{F}$ with $P(A) > 0$, there is a measurable $B \subseteq A$ such that

$$0 < P(B) < P(A).$$

There are two useful consequences of this definition.

Proposition 1.8. *Suppose P is nonatomic and $A \in \mathcal{F}$ has $P(A) > 0$.*

(i) For every $\varepsilon > 0$, there is a measurable $B \subseteq A$ with $0 < P(B) < \varepsilon$.

(ii) For every $0 < \delta < P(A)$, there is a measurable $B \subseteq A$ with $P(B) = \delta$.

Proof. For (i), split A into two measurable sets of strictly positive probability and retain one with the smaller probability. Repeating this operation gives a positive-mass subset of probability at most $2^{-n}P(A)$ after n steps. Choose n so that this is less than ε.

For (ii), begin with $B_0 = \varnothing$. If $P(B_n) = \delta$, the construction is complete. Otherwise, put $r_n = \delta - P(B_n) > 0$ and

$$s_n = \sup\{P(C) : C \in \mathcal{F},\ C \subseteq A \setminus B_n,\ P(C) \le r_n\}.$$

Since $P(A \setminus B_n) > 0$, part (i) shows that $s_n > 0$. Choose such a set C_n with $P(C_n) > s_n/2$, and set $B_{n+1} = B_n \cup C_n$. Then $B_n \uparrow B \subseteq A$ and $P(B) = \lim_n P(B_n) \le \delta$.

If $r = \delta - P(B) > 0$, part (i) gives a measurable $C \subseteq A \setminus B$ with $0 < P(C) < r$. This same C is admissible in the definition of every s_n, so $P(C_n) > s_n/2 \ge P(C)/2$ for every n. That contradicts the convergence of $P(B_n)$, whose successive increments are $P(C_n)$. Hence $P(B) = \delta$. □

1.4 Generated sigma-algebras

As an elementary exercise, verify that an intersection of σ-algebras on the same set Ω is again a σ-algebra, even when the intersection is over an uncountable family. This property permits the following definition.

Definition 1.9. For a family $\{A_\alpha : \alpha \in I\}$ of subsets of Ω, the *generated σ-algebra* is the smallest σ-algebra containing every A_α:

$$\sigma(\{A_\alpha : \alpha \in I\}) = \bigcap \{\mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-algebra on } \Omega,\ A_\alpha \in \mathcal{G} \text{ for all } \alpha \in I\}.$$

The family being intersected is nonempty, since it contains 2^Ω.

The index set I may be uncountable. More briefly, for any collection $\mathcal{A} \subseteq 2^\Omega$, we write $\sigma(\mathcal{A})$ for the σ-algebra it generates. If $\mathcal{A} \subseteq \mathcal{F}$ and $\mathcal{F}$ is a σ-algebra, then $\sigma(\mathcal{A}) \subseteq \mathcal{F}$. Consequently, to show $\sigma(\mathcal{A}) = \sigma(\mathcal{A}')$, it suffices to prove the two inclusions

$$\mathcal{A} \subseteq \sigma(\mathcal{A}'), \qquad \mathcal{A}' \subseteq \sigma(\mathcal{A}).$$

1.4.1 Borel sets

Definition 1.10. If S is a topological space, its *Borel σ-algebra* is

$$\mathcal{B}_S = \sigma(\{O \subseteq S : O \text{ is open}\}).$$

On $\mathbb{R}$, we simply write $\mathcal{B}$.

Proposition 1.11. *Each of the following families generates $\mathcal{B}$:*

$$\{(a,b) : a,b \in \mathbb{Q},\ a < b\},$$
$$\{(a,b] : a,b \in \mathbb{R},\ a < b\},$$
$$\{[a,b) : a,b \in \mathbb{R},\ a < b\},$$
$$\{(-\infty, b] : b \in \mathbb{Q}\}.$$

In the two half-open interval families, rational endpoints also suffice.

Proof. All the displayed intervals are Borel sets. Every open set in $\mathbb{R}$ is the union of the rational-endpoint open intervals it contains; this is a countable union. Thus the first family generates $\mathcal{B}$.

For rational $a < b$, choose rational $b_n \uparrow b$ with $a < b_n < b$. Then $(a,b) = \bigcup_n (a, b_n]$. Likewise, choosing rational $a_n \downarrow a$ with $a < a_n < b$ gives $(a,b) = \bigcup_n [a_n, b)$. These identities prove the assertions for both half-open interval families, even with rational endpoints.

Finally, if $a < b$ are rational, then

$$(a,b) = \left(\bigcup_{\substack{q \in \mathbb{Q} \\ q < b}} (-\infty, q] \right) \setminus (-\infty, a].$$

Hence the rational-endpoint closed half-lines also generate all rational open intervals, and therefore $\mathcal{B}$. □

1.4.2 A countable subcollection for each set

Proposition 1.12. *Let $\mathcal{F} = \sigma(\{A_\alpha : \alpha \in I\})$, with I possibly uncountable. For every $B \in \mathcal{F}$, there is an at most countable $J \subseteq I$ such that*

$$B \in \sigma(\{A_\alpha : \alpha \in J\}).$$

Proof. Let $\mathcal{M}$ be the collection of subsets $B \subseteq \Omega$ with the desired property. Clearly $\mathcal{M} \subseteq \mathcal{F}$. We show that $\mathcal{M}$ is itself a σ-algebra.

First, $\Omega \in \mathcal{M}$, since it belongs even to the σ-algebra generated by the empty collection. If $B \in \sigma(\{A_\alpha : \alpha \in J\})$, then B^c belongs to the same σ-algebra, so $B^c \in \mathcal{M}$. Now suppose $B_n \in \mathcal{M}$, and choose an at most countable $J_n \subseteq I$ for each B_n. The set $J = \bigcup_{n=1}^\infty J_n$ is at most countable. Every B_n belongs to $\sigma(\{A_\alpha : \alpha \in J\})$, and therefore so does $\bigcup_{n=1}^\infty B_n$.

Thus $\mathcal{M}$ is a σ-algebra containing every A_α. It follows that

$$\mathcal{F} = \sigma(\{A_\alpha : \alpha \in I\}) \subseteq \mathcal{M} \subseteq \mathcal{F},$$

which proves the claim. □

1.5 Extension of measures

It is often easier to prescribe a measure on an algebra than on an entire σ-algebra. Countable additivity on an algebra means that the additivity identity is required whenever the disjoint union still belongs to that algebra.

Definition 1.13. A *premeasure* on an algebra $\mathcal{A}$ is a function $\mu_0 : \mathcal{A} \to [0, \infty]$ with $\mu_0(\varnothing) = 0$ such that

$$\mu_0(A) = \sum_{n=1}^{\infty} \mu_0(A_n)$$

whenever $A_n \in \mathcal{A}$ are pairwise disjoint and $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

For a finite, finitely additive μ_0, this is equivalent to $\mu_0(A_n) \downarrow 0$ whenever $A_n \in \mathcal{A}$ and $A_n \downarrow \varnothing$, by Proposition 1.5. For an infinite premeasure, continuity from above is asserted only when one member of the decreasing sequence has finite mass.

We use the following extension theorem.

Theorem 1.14 (Carathéodory's extension theorem)**.** *Let $\mathcal{A}$ be an algebra on Ω, and let μ_0 be a premeasure on $\mathcal{A}$. There is a measure μ on $(\Omega, \sigma(\mathcal{A}))$ such that $\mu|_{\mathcal{A}} = \mu_0$. If μ_0 is σ-finite on $\mathcal{A}$, meaning that*

$$\Omega = \bigcup_{n=1}^{\infty} B_n, \qquad B_n \in \mathcal{A}, \qquad \mu_0(B_n) < \infty,$$

then this extension is unique. In particular, uniqueness holds when $\mu_0(\Omega) < \infty$.

Thus every countably additive probability on an algebra extends uniquely as a countably additive probability to the σ-algebra it generates. The existence assertion is used here without proof. The uniqueness assertion follows from Dynkin's theorem, developed below.

1.5.1 Lebesgue measure and the uniform probability

Consider first a finite disjoint union of bounded half-open intervals,

$$A = \bigcup_{k=1}^{n} (a_k, b_k], \qquad a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_n < b_n.$$

Its length is

$$\mu_0(A) = \sum_{k=1}^{n} (b_k - a_k). \tag{1.3}$$

Subdividing at all endpoints shows that this value does not depend on the disjoint representation and that length is finitely additive.

To obtain an algebra $\mathcal{A}$ on $\mathbb{R}$, allow also unbounded intervals: $(-\infty, b]$, (a, ∞), and $\mathbb{R}$, together with the empty set and finite disjoint unions. Equivalently, we allow infinite

endpoints in $(a, b]$, interpreting $(a, \infty]$ as (a, ∞). Give every unbounded interval length ∞ and use (1.3) for disjoint unions. Complements, finite unions, and intersections remain in $\mathcal{A}$, and $\sigma(\mathcal{A}) = \mathcal{B}$.

Countable additivity of length on this algebra uses the Heine–Borel theorem. Here is the compactness argument. Suppose $A = \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$ is a disjoint union with A bounded. Finite additivity and monotonicity give

$$\sum_{j=1}^{N} \mu_0(A_j) \le \mu_0(A) \quad \text{for every } N.$$

For the reverse inequality, fix $\varepsilon > 0$. Shorten each component of A at its left endpoint to obtain a compact union of intervals $K \subseteq A$ whose total length is at least $\mu_0(A) - \varepsilon$. Enlarge each component of each A_j slightly at its right endpoint to an open interval, distributing the extra lengths so that their total is at most ε. These open intervals cover K. Heine–Borel supplies a finite subcover. For finite interval covers, subdivision at the endpoints gives the elementary inequality that the total covered length is at most the sum of the covering lengths. Consequently,

$$\mu_0(A) - \varepsilon \le \text{length}(K) \le \sum_{j=1}^{\infty} \mu_0(A_j) + \varepsilon.$$

Let $\varepsilon \downarrow 0$. If A is unbounded, apply the bounded case to $A \cap (-M, M]$ and the disjoint sets $A_j \cap (-M, M]$. The length of $A \cap (-M, M]$ tends to ∞ as $M \to \infty$, so the same additivity identity follows. Thus μ_0 is a premeasure.

Since $\mathbb{R} = \bigcup_{n \ge 1} (-n, n]$ and $\mu_0((-n, n]) = 2n$, the premeasure is σ-finite. The extension theorem therefore gives a unique measure λ on $(\mathbb{R}, \mathcal{B})$ satisfying

$$\lambda\left(\bigcup_{k=1}^{n} (a_k, b_k]\right) = \sum_{k=1}^{n} (b_k - a_k)$$

for finite disjoint unions of bounded half-open intervals. This is *Lebesgue measure on the Borel sets*.

The *uniform probability on* $[0, 1]$ is its restriction to the measurable space $([0, 1], \mathcal{B}_{[0,1]})$, where

$$\mathcal{B}_{[0,1]} = \{B \cap [0, 1] : B \in \mathcal{B}\}, \qquad P(A) = \lambda(A) \quad (A \in \mathcal{B}_{[0,1]}).$$

Here $\lambda(\{0\}) = 0$, since $\{0\} \subseteq (-1/n, 0]$ for every n, and hence $\lambda([0, 1]) = 1$. This probability is nonatomic. Indeed, divide $[0, 1]$ into finitely many measurable intervals of length less than $P(A)$ for a given A of positive probability. At least one intersection with A has positive probability, and each has probability strictly less than $P(A)$.

1.6 Dynkin's theorem and uniqueness

Definition 1.15. A collection $\mathcal{P}$ of subsets of Ω is a *π-system* if $I, J \in \mathcal{P}$ implies $I \cap J \in \mathcal{P}$. A collection $\mathcal{L}$ is a *λ-system* if

(i) $\Omega \in \mathcal{L}$;

(ii) $A, B \in \mathcal{L}$ and $A \subseteq B$ imply $B \setminus A \in \mathcal{L}$;

(iii) $A_n \in \mathcal{L}$ and $A_n \uparrow A$ imply $A \in \mathcal{L}$.

Every algebra is a π-system. A λ-system is closed under complements, by taking $B = \Omega$ in (ii), and hence also under decreasing intersections, by taking complements in (iii). It is also closed under finite disjoint unions, since $A \cup B = (A^c \setminus B)^c$ when $A \cap B = \varnothing$. The increasing-limit condition can equivalently be replaced by closure under countable disjoint unions: differences of successive increasing sets give one implication, and finite disjoint unions followed by an increasing limit give the other.

Proposition 1.16. *A collection is both a π-system and a λ-system if and only if it is a σ-algebra.*

Proof. A σ-algebra has both properties. Conversely, the λ-system property gives Ω and complements. Together with the π-system property, De Morgan's law then gives finite unions. If A_n belong to the collection, their finite unions $\bigcup_{n=1}^{N} A_n$ increase to $\bigcup_{n=1}^{\infty} A_n$. The λ-system property puts this countable union in the collection. □

Theorem 1.17 (Dynkin's π–λ theorem)**.** *If $\mathcal{P}$ is a π-system, $\mathcal{L}$ is a λ-system, and $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

Proof. Let $\mathcal{D}$ be the smallest λ-system containing $\mathcal{P}$, obtained by intersecting all such λ-systems. Then $\mathcal{D} \subseteq \mathcal{L}$. The essential point is that $\mathcal{D}$ is also a π-system.

For $I \in \mathcal{D}$, define

$$\mathcal{D}_I = \{J \in \mathcal{D} : I \cap J \in \mathcal{D}\}.$$

This is a λ-system. It contains Ω because $I \in \mathcal{D}$; if $J \subseteq K$ and $J, K \in \mathcal{D}_I$, then

$$I \cap (K \setminus J) = (I \cap K) \setminus (I \cap J) \in \mathcal{D};$$

and increasing unions are preserved by intersection with I.

First fix $I \in \mathcal{P}$. Since $\mathcal{P}$ is a π-system, $\mathcal{P} \subseteq \mathcal{D}_I$. Minimality of $\mathcal{D}$ gives $\mathcal{D} \subseteq \mathcal{D}_I$, so $I \cap J \in \mathcal{D}$ for every $I \in \mathcal{P}$ and $J \in \mathcal{D}$. Now fix any $J \in \mathcal{D}$. The conclusion just proved shows that $\mathcal{P} \subseteq \mathcal{D}_J$. Applying minimality once more gives $\mathcal{D} \subseteq \mathcal{D}_J$. Thus $I \cap J \in \mathcal{D}$ for all $I, J \in \mathcal{D}$.

By Proposition 1.16, $\mathcal{D}$ is a σ-algebra. Since it contains $\mathcal{P}$, we have $\sigma(\mathcal{P}) \subseteq \mathcal{D} \subseteq \mathcal{L}$. □

Corollary 1.18 (Uniqueness of finite measures)**.** *Let μ_1 and μ_2 be measures on $(\Omega, \mathcal{F})$, and suppose*

$$\mathcal{F} = \sigma(\mathcal{P}), \qquad \mu_1 = \mu_2 \text{ on } \mathcal{P}, \qquad \mu_1(\Omega) = \mu_2(\Omega) < \infty,$$

where $\mathcal{P}$ is a π-system. Then $\mu_1 = \mu_2$ on $\mathcal{F}$.

Proof. Set

$$\mathcal{L} = \{A \in \sigma(\mathcal{P}) : \mu_1(A) = \mu_2(A)\}.$$

By assumption, $\mathcal{P} \subseteq \mathcal{L}$ and $\Omega \in \mathcal{L}$. If $A \subseteq B$ and $A, B \in \mathcal{L}$, then

$$\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A).$$

The measures are finite, so every difference is well defined. If $A_n \uparrow A$ with $A_n \in \mathcal{L}$, continuity from below gives

$$\mu_1(A) = \lim_n \mu_1(A_n) = \lim_n \mu_2(A_n) = \mu_2(A).$$

Thus $\mathcal{L}$ is a λ-system. Dynkin's theorem yields $\sigma(\mathcal{P}) \subseteq \mathcal{L}$, as required. □

In the extension theorem, the generating family is an algebra $\mathcal{A}$, which is in particular a π-system. The corollary therefore proves uniqueness for finite premeasures, including probabilities. For the σ-finite case, suppose μ_1 and μ_2 extend the same premeasure μ_0 and let $B_n \in \mathcal{A}$ cover Ω with $\mu_0(B_n) < \infty$. Set $C_n = \bigcup_{k=1}^{n} B_k \in \mathcal{A}$. For each fixed n, the finite measures

$$A \longmapsto \mu_1(A \cap C_n), \qquad A \longmapsto \mu_2(A \cap C_n)$$

have the same total mass and agree on $\mathcal{A}$, since $A \cap C_n \in \mathcal{A}$ when $A \in \mathcal{A}$. They therefore agree on $\sigma(\mathcal{A})$. As $C_n \uparrow \Omega$, continuity from below gives $\mu_1(A) = \mu_2(A)$ for every $A \in \sigma(\mathcal{A})$.

1.6.1 Monotone classes

For an algebra of generators, an alternative to Dynkin's theorem is the monotone-class theorem. We recall its set form.

Definition 1.19. A collection $\mathcal{M}$ of subsets of Ω is a *monotone class* if it is closed under increasing countable unions and decreasing countable intersections. That is,

$$A_n \in \mathcal{M},\ A_n \uparrow A \implies A \in \mathcal{M}, \qquad A_n \in \mathcal{M},\ A_n \downarrow A \implies A \in \mathcal{M}.$$

Theorem 1.20 (Halmos's monotone-class theorem)**.** *If $\mathcal{A}$ is an algebra, $\mathcal{M}$ is a monotone class, and $\mathcal{A} \subseteq \mathcal{M}$, then $\sigma(\mathcal{A}) \subseteq \mathcal{M}$.*

Problems

Problem 1.1 (Properties of probability measures)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $A, B, A_k \in \mathcal{F}$ for $k \geq 1$. Prove the following properties.

(a) *Monotonicity.* If $A \subseteq B$, then $P(A) \leq P(B)$.

(b) *Subadditivity.* If $A \subseteq \bigcup_{k=1}^{\infty} A_k$, then

$$P(A) \leq \sum_{k=1}^{\infty} P(A_k).$$

(c) *Continuity from below.* If $A_k \uparrow A$, that is, $A_k \subseteq A_{k+1}$ for all k and $\bigcup_{k=1}^{\infty} A_k = A$, then $P(A_k) \uparrow P(A)$.

(d) *Continuity from above.* If $A_k \downarrow A$, that is, $A_k \supseteq A_{k+1}$ for all k and $\bigcap_{k=1}^{\infty} A_k = A$, then $P(A_k) \downarrow P(A)$.

Problem 1.2 (Sigma-algebra constructions and monotone classes)**.** Let $\mathcal{F}$ be a field (an algebra) of subsets of Ω.

(a) Show that, if $\{\mathcal{G}_\alpha : \alpha \in I\}$ is a possibly uncountable family of σ-algebras on Ω, then $\bigcap_{\alpha \in I} \mathcal{G}_\alpha$ is also a σ-algebra. Conclude that

$$\sigma(\mathcal{F}) = \bigcap \{\mathcal{G} \supseteq \mathcal{F} : \mathcal{G} \text{ is a } \sigma\text{-algebra on } \Omega\}.$$

(b) Verify that, for any σ-algebras $\mathcal{H} \subseteq \mathcal{G}$ on Ω and any $H \in \mathcal{H}$, the collection

$$\mathcal{H}^H = \{A \in \mathcal{G} : A \cap H \in \mathcal{H}\}$$

is a σ-algebra on Ω.

(c) With $\mathcal{H} \subseteq \mathcal{G}$ as in part (b), show that $H \mapsto \mathcal{H}^H$ is nonincreasing with respect to set inclusion:

$$H \subseteq H' \quad \Longrightarrow \quad \mathcal{H}^{H'} \subseteq \mathcal{H}^H, \qquad H, H' \in \mathcal{H}.$$

Show also that

$$\mathcal{H}^\Omega = \mathcal{H}, \qquad \mathcal{H}^\varnothing = \mathcal{G}.$$

Deduce that, for every $H, H' \in \mathcal{H}$,

$$\mathcal{H}^{H \cup H'} = \mathcal{H}^H \cap \mathcal{H}^{H'}.$$

(d) Prove that, if $\mathcal{M}$ is a monotone class and $\mathcal{F} \subseteq \mathcal{M}$, then $\sigma(\mathcal{F}) \subseteq \mathcal{M}$. Conclude that

$$\sigma(\mathcal{F}) = m(\mathcal{F}) := \bigcap \{\mathcal{M} \supseteq \mathcal{F} : \mathcal{M} \text{ is a monotone class on } \Omega\}.$$

Chapter 2

Random Variables, Distributions, and Expectation

2.1 Measurable functions and simple approximation

Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be measurable spaces. A map $X : \Omega \to S$ is *measurable* if

$$X^{-1}(B) \in \mathcal{F} \qquad \text{for every } B \in \mathcal{S}.$$

On a probability space $(\Omega, \mathcal{F}, P)$, such a map is called a *random variable*. Unless another range is specified, we take $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra. We denote the class of real-valued measurable functions by $m\mathcal{F}$. When extended values are allowed, we say so explicitly and use the Borel σ-algebra of $\overline{\mathbb{R}} = [-\infty, \infty]$. The nonnegative measurable functions, with $+\infty$ admitted, form the class $m\mathcal{F}^+$.

It is enough to check measurability on generators. Indeed, if $\mathcal{S} = \sigma(\mathcal{A})$ and $X^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{A}$, then

$$\{B \subseteq S : X^{-1}(B) \in \mathcal{F}\}$$

is a σ-algebra containing $\mathcal{A}$, and therefore contains $\mathcal{S}$. In particular, a real-valued X is measurable if and only if $\{X \le x\} \in \mathcal{F}$ for every $x \in \mathbb{R}$.

For $A \in \mathcal{F}$, the indicator $\mathbf{1}_A$ is measurable. More generally,

$$\varphi = \sum_{i=1}^{m} c_i \mathbf{1}_{A_i}, \qquad A_i \in \mathcal{F}, \quad c_i \in \mathbb{R}, \quad m < \infty,$$

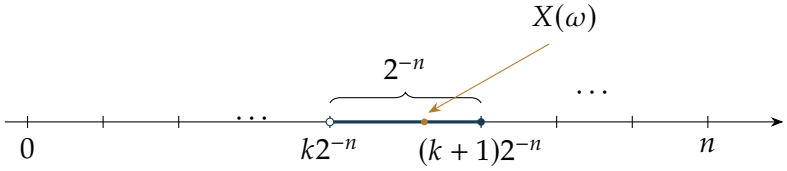
is measurable. These are the *simple functions*; we denote their class by SF and its nonnegative members by SF^+. The sets A_i need not initially be disjoint. By taking a common finite partition, any simple function can be written using disjoint measurable sets.

Proposition 2.1 (Measurable approximation)**.** *Every real-valued measurable function X is a pointwise limit of simple functions. If $X \ge 0$, the approximating functions can be chosen so that $0 \le X_n \uparrow X$, also when X takes the value $+\infty$. For bounded X, the approximation can be made uniform.*

Proof. For $X \ge 0$ and $n \ge 1$, set

$$X_n = n\mathbf{1}_{\{X > n\}} + \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mathbf{1}_{\{k2^{-n} < X \le (k+1)2^{-n}\}}. \tag{2.1}$$

Thus we truncate the range at n, divide it into intervals of length 2^{-n}, and take the left endpoint of the interval containing X. The intervals are open on the left and closed on the right; at $X = 0$ the formula gives zero.



take the left endpoint for $X_n(\omega)$

Every X_n has finitely many values and is measurable. Each dyadic interval is split into two at the next stage, and the truncation level increases; hence $0 \le X_n \le X_{n+1} \le X$. If $0 < X(\omega) \le n$, then

$$0 \le X(\omega) - X_n(\omega) \le 2^{-n}.$$

This proves convergence at every finite value of X; if $X(\omega) = \infty$, then $X_n(\omega) = n$. The same estimate proves uniform convergence when X is bounded and nonnegative.

For general X, write

$$X = X^+ - X^-, \qquad X^+ = \max(X, 0), \qquad X^- = \max(-X, 0),$$

and approximate the two nonnegative parts separately. Their simple approximations have a simple difference converging to X. If X is bounded, both approximations are uniform. □

Theorem 2.2 (Functional monotone-class theorem)**.** *Let $\mathcal{H}$ be a vector space of real-valued functions on Ω containing the constant function 1. Suppose that whenever $0 \le h_n \uparrow h$, with every $h_n \in \mathcal{H}$ and h finite-valued, we have $h \in \mathcal{H}$. If $\mathcal{P}$ is a π-system and $\mathbf{1}_A \in \mathcal{H}$ for every $A \in \mathcal{P}$, then $\mathcal{H}$ contains every real-valued $\sigma(\mathcal{P})$-measurable function.*

There is also a bounded version: if closure under increasing nonnegative limits is required only when the limit is bounded, the conclusion is that $\mathcal{H}$ contains every bounded $\sigma(\mathcal{P})$-measurable function.

Proof. Set

$$\mathcal{L} = \{A \subseteq \Omega : \mathbf{1}_A \in \mathcal{H}\}.$$

Because $1 \in \mathcal{H}$, we have $\Omega \in \mathcal{L}$. Closure under linear combinations gives $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A \in \mathcal{H}$ when $A \in \mathcal{L}$. For pairwise disjoint $A_j \in \mathcal{L}$,

$$\sum_{j=1}^{n} \mathbf{1}_{A_j} \uparrow \mathbf{1}_{\bigcup_{j=1}^{\infty} A_j}.$$

The limit is bounded by one, so either version of the closure hypothesis shows that $\mathcal{L}$ is a λ-system. Dynkin's theorem (Theorem 1.17) now gives $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.

Thus all measurable indicators, and then all measurable simple functions, belong to $\mathcal{H}$. Proposition 2.1 and monotone closure give all nonnegative finite-valued measurable functions, or all bounded nonnegative ones in the bounded version. Taking the difference of the positive and negative parts proves the respective conclusions. □

2.2 Generated sigma-algebras and almost-sure statements

For a measurable map $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$, define

$$\sigma(X) = \mathcal{F}^X = \{X^{-1}(B) : B \in \mathcal{S}\}.$$

Inverse images commute with complements and countable unions, so this is a σ-algebra. It is the smallest σ-algebra on Ω with respect to which X is measurable, and it is contained in $\mathcal{F}$.

For real or extended-real random variables X, Y on the same probability space, we write $X = Y$ *almost surely*, or $X \overset{\text{a.s.}}{=} Y$, if

$$P\{\omega : X(\omega) \neq Y(\omega)\} = 0.$$

Countable extrema and limits preserve measurability when extended values are admitted. In particular, for extended-real random variables X_n,

$$\inf_n X_n, \qquad \sup_n X_n, \qquad \liminf_{n\to\infty} X_n, \qquad \limsup_{n\to\infty} X_n$$

are extended-real random variables. For example, $\{\sup_n X_n > a\} = \bigcup_n \{X_n > a\}$ and $\{\inf_n X_n < a\} = \bigcup_n \{X_n < a\}$; the limits then follow from

$$\liminf_n X_n = \sup_m \inf_{n\geq m} X_n, \qquad \limsup_n X_n = \inf_m \sup_{n\geq m} X_n.$$

We write $X_n \xrightarrow{\text{a.s.}} X_\infty$ if

$$P\{\omega : \lim_{n\to\infty} X_n(\omega) = X_\infty(\omega)\} = 1.$$

In particular, if $\liminf_n X_n = \limsup_n X_n$ almost surely, their common value defines an extended-real limit almost surely. One may take $X_\infty = \limsup_n X_n$ everywhere; a real-valued limit requires this value to be finite almost surely and a measurable choice on the exceptional null set.

2.3 Laws and distribution functions

Definition 2.3. The *law*, or *distribution*, of a real-valued random variable X on $(\Omega, \mathcal{F}, P)$ is the probability measure P_X on $(\mathbb{R}, \mathcal{B})$ defined by

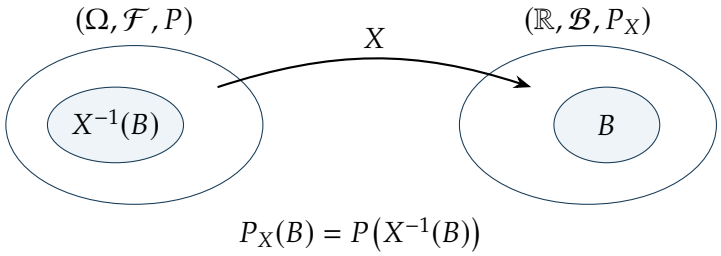
$$P_X(B) = P\big(X^{-1}(B)\big) = P\{X \in B\}, \qquad B \in \mathcal{B}.$$

It is also called the *pushforward* of P by X and is written PX^{-1}. We shall also use the local notation $\alpha = PX^{-1}$.

To check that this is a probability measure, observe that $X^{-1}(\mathbb{R}) = \Omega$ and $X^{-1}(\varnothing) = \varnothing$. Inverse images of pairwise disjoint sets are pairwise disjoint, and

$$P_X\left(\bigcup_{j=1}^{\infty} B_j\right) = P\left(\bigcup_{j=1}^{\infty} X^{-1}(B_j)\right) = \sum_{j=1}^{\infty} P_X(B_j).$$

The same construction works for a random variable with any measurable range $(S, \mathcal{S})$.



Two random variables have the *same law*, written $X \stackrel{d}{=} Y$, if $P_X = P_Y$. They may be defined on different probability spaces. The *distribution function* of X is

$$F_X(x) = P\{X \le x\} = P_X((-\infty, x]), \qquad x \in \mathbb{R}.$$

Thus P_X or α is a measure on Borel sets, whereas F_X is a function of a real number. If α is fixed, we often write simply $F(x) = \alpha((-\infty, x])$.

Theorem 2.4 (Characterization of distribution functions)**.** *A function $F : \mathbb{R} \to [0, 1]$ is the distribution function of a real-valued random variable if and only if it is nondecreasing, right-continuous, and*

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$$

It determines a unique probability measure on $(\mathbb{R}, \mathcal{B})$.

Proof. If $F = F_X$, monotonicity follows from the inclusion $\{X \le x\} \subseteq \{X \le y\}$ for $x \le y$. When $x_n \downarrow x$, the events $\{X \le x_n\}$ decrease to $\{X \le x\}$, so continuity of a finite measure from above gives $F(x_n) \downarrow F(x)$. The events $\{X \le n\}$ increase to Ω, whereas $\{X \le -n\}$ decrease to the empty set. These give the two tail limits.

Conversely, take Lebesgue probability measure λ on $(0, 1]$. For $0 < \omega < 1$, define the generalized inverse

$$X(\omega) = \inf\{y \in \mathbb{R} : F(y) \ge \omega\}, \tag{2.2}$$

and set $X(1) = 0$. For $0 < \omega < 1$, the set in (2.2) is nonempty by the upper tail limit and is bounded below by the lower tail limit. Thus $X(\omega)$ is finite. If q is this infimum, then $F(q + 1/m) \ge \omega$ for every m; right continuity gives $F(q) \ge \omega$. Consequently,

$$X(\omega) \le x \quad \Longleftrightarrow \quad \omega \le F(x), \qquad 0 < \omega < 1.$$

This proves measurability, including the separately defined value at 1, and gives

$$\lambda\{\omega : X(\omega) \le x\} = F(x).$$

The choice at $\omega = 1$ does not affect the law. This is the generalized-inverse (Skorokhod) representation of a single prescribed distribution.

Finally, any two probability measures with distribution function F agree on the generating π-system $\{(-\infty, x] : x \in \mathbb{R}\}$, so measure uniqueness (Corollary 1.18) makes them equal on $\mathcal{B}$. □

2.3.1 Construction by extension from intervals

There is a second construction of the measure corresponding to F. The extension theorem reduces the construction of a measure to the construction of a countably additive set function on an algebra. Let

$$\mathcal{I} = \{(a, b] \cap \mathbb{R} : -\infty \le a < b \le +\infty\},$$

and let $\mathcal{A}$ consist of the empty set and finite disjoint unions of members of $\mathcal{I}$. The intersection with $\mathbb{R}$ means, in particular, that the interval with right endpoint $+\infty$ is (a, ∞). This is an algebra and $\sigma(\mathcal{A}) = \mathcal{B}$. Put $F(-\infty) = 0$, $F(+\infty) = 1$, and define

$$P((a, b] \cap \mathbb{R}) = F(b) - F(a).$$

For a disjoint union of intervals, define P by summing these increments. Refining two finite interval decompositions by all their endpoints and telescoping the increments shows that the definition is independent of the decomposition. It is nonnegative, finitely additive, and $P(\mathbb{R}) = 1$.

For a nondecreasing F with these tail limits, this set function is countably additive on $\mathcal{A}$ if and only if F is right-continuous. Here is the continuity argument that supplies the needed premeasure. If P is countably additive, apply continuity from above to $(-\infty, x_n]$ with $x_n \downarrow x$ to obtain right continuity of F.

For the converse, right continuity and the tail limits imply that for every $A \in \mathcal{A}$ and $\eta > 0$ there is $B \in \mathcal{A}$ with

$$B \subseteq \overline{B} \subseteq A, \qquad \overline{B} \text{ compact}, \qquad P(A \setminus B) < \eta.$$

For a bounded component $(a, b]$, replace it by $(a + u, b]$ with $u > 0$ small; the lost mass is $F(a + u) - F(a)$. Unbounded components can first be truncated using the tail limits. Applying these choices to the finitely many components gives the assertion.

Suppose now that $A_n \downarrow \varnothing$, with $A_n \in \mathcal{A}$. Choose B_n as above so that $P(A_n \setminus B_n) < \varepsilon 2^{-n}$. The compact sets $\overline{B}_n$ have empty total intersection, since $\overline{B}_n \subseteq A_n$. By compactness, some finite intersection $\bigcap_{n=1}^{N} \overline{B}_n$ is empty. Because $A_N \subseteq A_n$ for $n \le N$,

$$A_N \subseteq \bigcup_{n=1}^{N} (A_n \setminus B_n), \qquad P(A_N) < \varepsilon.$$

Thus $P(A_n) \downarrow 0$. A finite finitely additive set function with this continuity property is countably additive: for disjoint C_j whose union C belongs to $\mathcal{A}$, apply it to $C \setminus \bigcup_{j=1}^{n} C_j \downarrow \varnothing$.

The extension theorem (Theorem 1.14) therefore extends P uniquely to $\mathcal{B}$. Its distribution function is F. This gives the one-to-one correspondence between distribution functions and Borel probability measures directly through interval increments.

2.4 Integration and expectation

Let $(S, \mathcal{F}, \mu)$ be a measure space. We construct the integral in four steps. First, for an indicator, set

$$\mu_0(\mathbf{1}_A) = \mu(A), \qquad A \in \mathcal{F}.$$

Second, if $\varphi = \sum_{i=1}^{m} c_i \mathbf{1}_{A_i} \in \mathrm{SF}^+$ with $c_i \geq 0$, define

$$\mu_0(\varphi) = \sum_{i=1}^{m} c_i \mu(A_i),$$

with $0 \cdot \infty = 0$. A common disjoint refinement and finite additivity of μ show that this does not depend on the representation. They also give additivity, nonnegative homogeneity, and monotonicity for nonnegative simple functions.

Third, for a nonnegative measurable $f : S \to [0, \infty]$, define

$$\mu(f) = \int_S f \, \mathrm{d}\mu = \sup\{\mu_0(\varphi) : \varphi \in \mathrm{SF}^+,\ \varphi \leq f\}. \tag{2.3}$$

This may equal $+\infty$. It agrees with μ_0 on nonnegative simple functions and is monotone directly from the definition.

Fourth, for a real or extended-real measurable f, write $f = f^+ - f^-$ and define

$$\mu(f) = \int_S f \, \mathrm{d}\mu = \mu(f^+) - \mu(f^-)$$

provided at least one of $\mu(f^+)$, $\mu(f^-)$ is finite. If both are infinite, the integral is undefined. The function is *integrable* if $\mu(|f|) < \infty$, equivalently if both parts have finite integral. We write $f \in L^1(\mu)$ for this condition, identifying functions that agree μ-almost everywhere when regarding $L^1(\mu)$ as a space.

On a probability space, this construction is called *expectation* and is denoted by E:

$$\begin{aligned} E[\mathbf{1}_A] &= P(A), \\ E\left[\sum_{i=1}^{m} c_i \mathbf{1}_{A_i}\right] &= \sum_{i=1}^{m} c_i P(A_i) \quad (c_i \geq 0), \\ E[X] &= \sup\{E[Y] : Y \in \mathrm{SF}^+,\ Y \leq X\} \quad (X \geq 0), \\ E[X] &= E[X^+] - E[X^-] \quad \text{when at least one part is finite.} \end{aligned}$$

On a finite measure space, linearity for bounded functions can already be obtained from uniform simple approximation. For bounded $f \geq 0$, choose simple $s \leq f \leq s + \eta$. The supremum definition gives

$$\mu_0(s) \leq \mu(f) \leq \mu_0(s) + \eta\mu(S).$$

Apply the same bounds to two bounded nonnegative functions and the sum of their simple approximations, and let $\eta \downarrow 0$. This proves additivity for bounded nonnegative functions; scalar multiplication and the positive and negative parts give linearity for bounded signed functions. This argument uses only simple functions and the definition of the integral.

Theorem 2.5 (Monotone convergence)**.** *If $h_n : S \to [0, \infty]$ are measurable and $h_n(s) \uparrow h(s)$ for every $s \in S$, then*

$$\mu(h_n) \uparrow \mu(h) \leq \infty.$$

We use monotone convergence here as an integration theorem. Its proof through Fatou's lemma is given in Theorem 3.11.

2.4.1 The standard machine

Many integration identities are established by the following sequence. First prove the identity for $h = \mathbf{1}_A$. Use linearity for finite sums to pass to SF^+. Next use increasing simple approximation and monotone convergence to pass to $m\mathcal{F}^+$, the nonnegative measurable functions. Finally write $h = h^+ - h^-$ and subtract, provided the integrals involved are defined and no subtraction of two infinities occurs. This is the *standard machine*.

For example, nonnegative simple approximations $f_n \uparrow f$ and $g_n \uparrow g$ also satisfy $f_n + g_n \uparrow f + g$. Monotone convergence and simple-function additivity give

$$\mu(f + g) = \lim_n \mu_0(f_n + g_n) = \lim_n \big(\mu_0(f_n) + \mu_0(g_n)\big) = \mu(f) + \mu(g)$$

for $f, g \geq 0$. Nonnegative homogeneity follows in the same way. For integrable signed functions the decomposition into positive and negative parts gives ordinary linearity. Monotonicity then gives

$$\left| \int_S f \, \mathrm{d}\mu \right| \leq \int_S |f| \, \mathrm{d}\mu.$$

Measurable changes on a null set do not change an integral: every nonnegative simple function supported on a null set has integral zero, and the supremum definition extends this to every nonnegative measurable function supported there. These facts allow almost-everywhere versions of the preceding statements.

2.5 Transformation and change of measure

Theorem 2.6 (Change of measure under a measurable map)**.** *Let $X : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ be measurable, let P be a probability measure on $(\Omega_1, \mathcal{F}_1)$, and put $Q = PX^{-1} = P_X$. For measurable f on Ω_2, the composition $g = f \circ X$ is measurable on Ω_1. If $f \geq 0$, then*

$$\int_{\Omega_2} f(\omega_2) \, \mathrm{d}Q(\omega_2) = \int_{\Omega_1} f(X(\omega_1)) \, \mathrm{d}P(\omega_1), \tag{2.4}$$

with $+\infty$ allowed. For signed f, the identity holds whenever one of the positive and negative integrals is finite. Moreover,

$$f \in L^1(Q) \quad \Longleftrightarrow \quad f \circ X \in L^1(P).$$

Proof. For a Borel set B in the range of f,

$$(f \circ X)^{-1}(B) = X^{-1}\big(f^{-1}(B)\big) \in \mathcal{F}_1,$$

so the composition is measurable. For $f = \mathbf{1}_A$, $A \in \mathcal{F}_2$,

$$\int_{\Omega_2} \mathbf{1}_A \, \mathrm{d}Q = Q(A) = P(X^{-1}(A)) = \int_{\Omega_1} \mathbf{1}_A(X(\omega_1)) \, \mathrm{d}P(\omega_1).$$

Linearity proves the identity for nonnegative simple f. For general nonnegative f, choose simple $f_n \uparrow f$. Then $f_n \circ X \uparrow f \circ X$, and monotone convergence on both spaces proves (2.4). Applying this result to $|f|$ proves the equivalence of integrability. Applying it separately to f^+ and f^- proves the signed identity with its stated qualification. □

For real-valued X, choosing the identity function $f(x) = x$ gives

$$E[X] = \int_\Omega X(\omega) \, \mathrm{d}P(\omega) = \int_{\mathbb{R}} x \, \mathrm{d}Q(x), \qquad Q = PX^{-1},$$

whenever the expectation is defined.

More generally, let $\alpha = PX^{-1}$ and let $g : \mathbb{R} \to \mathbb{R}$ be Borel measurable. Then $Y = g(X)$ is a random variable and its law is

$$\beta = PY^{-1} = PX^{-1}g^{-1} = \alpha g^{-1}.$$

For each $B \in \mathcal{B}$, this notation means $\beta(B) = \alpha(g^{-1}(B))$. The transformation theorem gives

$$E[g(X)] = \int_\Omega g(X(\omega)) \, \mathrm{d}P(\omega) = \int_{\mathbb{R}} g(x) \, \mathrm{d}\alpha(x). \tag{2.5}$$

The formula holds for nonnegative g, for integrable g, and more generally when one of the two sign parts has finite integral.

The same observation applies to a random vector $X = (X_1, \dots, X_m)$: its joint distribution is $\alpha = PX^{-1}$ on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, and for Borel measurable g with the same integral qualifications,

$$E[g(X_1, \dots, X_m)] = \int_{\mathbb{R}^m} g(x_1, \dots, x_m) \, \mathrm{d}\alpha(x).$$

2.6 Densities and integration against a distribution function

Let $f : \mathbb{R} \to [0, \infty)$ be Lebesgue measurable with $\int_{\mathbb{R}} f(x) \, \mathrm{d}x = 1$. The function

$$F(x) = \int_{-\infty}^{x} f(s) \, \mathrm{d}s$$

is a continuous distribution function. In fact, it is absolutely continuous, and the fundamental theorem for the Lebesgue integral gives

$$F'(x) = f(x) \qquad \text{for Lebesgue-almost every } x.$$

The function f is called a *density* of the corresponding probability measure P_F. Densities are determined only up to Lebesgue-null sets; continuity of a distribution function alone does not imply that it has a density.

Theorem 2.7 (Expectation from a density)**.** *Suppose that a real-valued random variable X has distribution function*

$$F_X(b) = \int_{-\infty}^{b} f_X(x)\,\mathrm{d}x, \qquad f_X \geq 0, \qquad \int_{\mathbb{R}} f_X(x)\,\mathrm{d}x = 1.$$

For every nonnegative Borel measurable h,

$$E[h(X)] = \int_{\mathbb{R}} h(x) f_X(x)\,\mathrm{d}x. \tag{2.6}$$

For signed Borel measurable h, the two sides are defined under exactly the same condition: at least one of $\int h^+ f_X\,\mathrm{d}x$, $\int h^- f_X\,\mathrm{d}x$ is finite. Under that condition they are equal. In particular,

$$h(X) \in L^1(P) \quad \Longleftrightarrow \quad \int_{\mathbb{R}} |h(x)| f_X(x)\,\mathrm{d}x < \infty.$$

Proof. By the transformation theorem, $E[h(X)] = \int h\,\mathrm{d}P_X$ whenever the integrals are defined. For $h = \mathbf{1}_{(-\infty,b]}$,

$$\begin{aligned} E[h(X)] &= P_X((-\infty, b]) = F_X(b) \\ &= \int_{-\infty}^{b} f_X(x)\,\mathrm{d}x = \int_{\mathbb{R}} h(x) f_X(x)\,\mathrm{d}x. \end{aligned}$$

Define, for $B \in \mathcal{B}$,

$$\mu(B) = \int_{\mathbb{R}} \mathbf{1}_B(x) f_X(x)\,\mathrm{d}x.$$

This is a probability measure: its total mass is one, and countable additivity follows by applying monotone convergence to the partial sums of indicators of disjoint sets, multiplied by f_X. The measures μ and P_X agree on

$$\mathcal{P} = \{(-\infty, b] : b \in \mathbb{R}\},$$

a π-system generating $\mathcal{B}$. By the π–λ uniqueness theorem, $\mu = P_X$ on $\mathcal{B}$.

Therefore (2.6) holds for every Borel indicator $h = \mathbf{1}_B$, and linearity gives it for every nonnegative simple h. For a nonnegative Borel h, take nonnegative simple $h_n \uparrow h$. Monotone convergence for P_X and for Lebesgue measure gives

$$\int_{\mathbb{R}} h\,\mathrm{d}P_X = \lim_n \int_{\mathbb{R}} h_n\,\mathrm{d}P_X = \lim_n \int_{\mathbb{R}} h_n f_X\,\mathrm{d}x = \int_{\mathbb{R}} h f_X\,\mathrm{d}x.$$

Applying this identity to $|h|$ gives the integrability criterion. Applying it to h^+ and h^-, and subtracting when permitted, proves the assertion for signed functions. □

For complex-valued h, define the integral by its real and imaginary parts when $\int |h| f_X \, \mathrm{d}x < \infty$. Applying the theorem to those two parts gives the same expectation formula. Likewise the transformation theorem extends to integrable complex-valued functions.

If X takes its values in a Borel set $S \subseteq \mathbb{R}$ and has a density, then $f_X = 0$ Lebesgue-almost everywhere on S^c. We may choose a version that is identically zero there, and all the preceding integrals can be restricted to S.

A density is not needed to integrate using a distribution function. For bounded continuous g, the Riemann–Stieltjes integral satisfies

$$E[g(X)] = \int_{\mathbb{R}} g(x) \, \mathrm{d}P_X(x) = \int_{-\infty}^{\infty} g(x) \, \mathrm{d}F_X(x). \tag{2.7}$$

The last integral is the limit of sums

$$\sum_{j=1}^{N} g(x_j)\big[F_X(a_{j+1}) - F_X(a_j)\big], \qquad x_j \in (a_j, a_{j+1}],$$

as the finite intervals expand to $\mathbb{R}$ and their subdivisions become fine enough that the oscillation of g on each subinterval tends uniformly to zero. Appendix A gives the precise construction and the full approximation argument.

2.7 Three useful inequalities

We recall three basic inequalities for expectation. The passage from simple to general functions in integration arguments is supplied by the standard machine.

Proposition 2.8 (Markov's inequality)**.** *If $X \geq 0$ and $a > 0$, then*

$$P\{X \geq a\} \leq \frac{E[X]}{a}.$$

The bound is valid also when $E[X] = \infty$.

Proposition 2.9 (Jensen's inequality)**.** *If X is integrable and $\varphi : \mathbb{R} \to \mathbb{R}$ is a finite convex function, then*

$$\varphi(E[X]) \leq E[\varphi(X)].$$

Here the expectation on the right is well-defined and may equal $+\infty$.

Proposition 2.10 (Hölder's inequality)**.** *Let $p, q > 1$ satisfy $1/p + 1/q = 1$. If $E[|X|^p] < \infty$ and $E[|Y|^q] < \infty$, then XY is integrable and*

$$E[|XY|] \leq \big(E[|X|^p]\big)^{1/p} \big(E[|Y|^q]\big)^{1/q}.$$

The endpoint version is $E[|XY|] \leq E[|X|] \operatorname{ess\,sup} |Y|$ when X is integrable and Y is essentially bounded.

Problems

Problem 2.1. Prove that if $f : \mathbb{R}^n \to [-\infty, \infty]$ is lower semicontinuous, that is,

$$\liminf_{\|x-x_0\|\downarrow 0} f(x) \geq f(x_0) \qquad \text{for every } x_0 \in \mathbb{R}^n,$$

then it is a Borel function. Conclude that continuous functions are Borel measurable.

Hint. First show that every set of the form $\{x : f(x) \leq a\}$, $a \in \mathbb{R}$, is closed.

Problem 2.2. Let $m\mathcal{F}$ be the set of measurable functions from $(\Omega, \mathcal{F})$ to $([-\infty, \infty], \mathcal{B}_{[-\infty,\infty]})$, where

$$\mathcal{B}_{[-\infty,\infty]} = \sigma\big(\{[-\infty, a] : a \in \mathbb{R}\}\big).$$

Prove that:

(a) Every simple function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ belongs to $m\mathcal{F}$.

(b) If $X_n \in m\mathcal{F}$ for $n \geq 1$, then $\liminf_n X_n$ and $\limsup_n X_n$ also belong to $m\mathcal{F}$.

Conclude that $m\mathcal{F}$ is the smallest class of functions satisfying (a) and (b).

Problem 2.3. Let X be a nonnegative random variable with $E[X^2] < \infty$, and set $m_i = E[X^i]$ for $i = 1, 2$.

(i) Prove that for every $0 \leq x < m_1$,

$$P(X > x) \geq \frac{(m_1 - x)^2}{m_2}.$$

(ii) Prove that

$$\big(E[|X^2 - m_2|]\big)^2 \leq 4m_2(m_2 - m_1^2).$$

(iii) For measurable events $A_1, \ldots, A_n$, show that

$$P\left(\bigcup_{k=1}^n A_k\right) \geq \sum_{k=1}^n P(A_k) - \sum_{1\leq k<\ell\leq n} P(A_k \cap A_\ell),$$

and compare this inequality with the bound in part (i) for $X = \sum_{k=1}^n \mathbf{1}_{A_k}$.

Problem 2.4. Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be a random variable with $E[|X|] < \infty$.

(i) If $A_n \in \mathcal{F}$, $n \geq 1$, are pairwise disjoint and $A = \bigcup_{n\geq 1} A_n$, prove that

$$\sum_{n=1}^{\infty} E[X\mathbf{1}_{A_n}]$$

converges absolutely and has sum $E[X\mathbf{1}_A]$.

(ii) Deduce that if $X \geq 0$ and $0 < E[X] < \infty$, then

$$Q(A) = \frac{E[X\mathbf{1}_A]}{E[X]}, \qquad A \in \mathcal{F},$$

defines a probability measure on $(\Omega, \mathcal{F})$.

(iii) Suppose that X and Y are nonnegative random variables on the same probability space $(\Omega, \mathcal{F}, P)$, with $E[X] = E[Y] < \infty$. Deduce from the preceding parts that if

$$E[X\mathbf{1}_A] = E[Y\mathbf{1}_A] \qquad \text{for every } A \in \mathcal{A},$$

where $\mathcal{A}$ is a π-system with $\sigma(\mathcal{A}) = \mathcal{F}$, then $X = Y$ almost surely.

Problem 2.5. Let $Y = \sum_{k=1}^{n} \mathbf{1}_{A_k}$, where $A_1, \ldots, A_n$ are measurable events. Express the variance

$$\mathrm{Var}(Y) = E\big[(Y - E[Y])^2\big]$$

in terms of $P(A_k)$ and $P(A_k \cap A_\ell)$.

Use your expression to calculate $\mathrm{Var}(Y)$ in the following setting. Each of m players independently selects a number uniformly from $\{1, \ldots, n\}$, and A_k is the event that number k is not selected by any player. Here $m, n \geq 1$ are integers; explicitly, the finite sample space is $\Omega = \{1, \ldots, n\}^m$, with all subsets measurable and each of its n^m tuples assigned probability n^{-m}.

Problem 2.6. Let X be a random variable, and let $f, g : \mathbb{R} \to \mathbb{R}$ be bounded nondecreasing functions. Prove that

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)].$$

Chapter 3

Modes of Convergence

3.1 Almost-sure convergence and convergence in probability

Let X_n and X_∞ be real-valued random variables on the same probability space $(\Omega, \mathcal{F}, P)$. Recall that $X_n \xrightarrow{\text{a.s.}} X_\infty$ means

$$P\{\omega : X_n(\omega) \longrightarrow X_\infty(\omega)\} = 1.$$

Thus almost-sure convergence is pointwise convergence outside a single null set.

Definition 3.1 (Convergence in probability)**.** We write $X_n \xrightarrow{P} X_\infty$ if, for every $\varepsilon > 0$,

$$P\{|X_n - X_\infty| \geq \varepsilon\} \longrightarrow 0.$$

Equivalently, $|X_n - X_\infty| \xrightarrow{P} 0$. Replacing $\geq \varepsilon$ by $> \varepsilon$ gives the same definition.

Proposition 3.2. *If $X_n \xrightarrow{\text{a.s.}} X_\infty$, then $X_n \xrightarrow{P} X_\infty$.*

Proof. Fix $\varepsilon > 0$ and set

$$A_n = \{|X_n - X_\infty| \geq \varepsilon\}, \qquad S = \{X_n \text{ does not converge to } X_\infty\}.$$

If the events A_n occur infinitely often, convergence fails, so

$$\{A_n \text{ infinitely often}\} = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n \subseteq S.$$

The tail unions decrease, and $P(S) = 0$. Continuity from above gives

$$0 = P\left(\bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n\right) = \lim_{N \to \infty} P\left(\bigcup_{n \geq N} A_n\right).$$

Since $P(A_N) \leq P(\bigcup_{n \geq N} A_n)$, we obtain $P(A_N) \to 0$. □

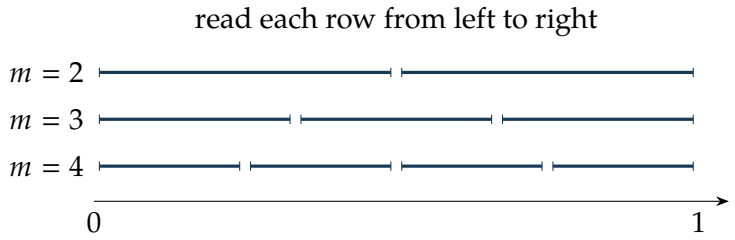
Example 3.3 (The typewriter sequence)**.** The converse is false. On $(0, 1]$ with Lebesgue probability measure, list the indicators of the intervals

$$(0, \tfrac{1}{2}],\ (\tfrac{1}{2}, 1], \quad (0, \tfrac{1}{3}],\ (\tfrac{1}{3}, \tfrac{2}{3}],\ (\tfrac{2}{3}, 1], \quad (0, \tfrac{1}{4}], \ldots.$$

More precisely, for $m \geq 2$, list in order

$$I_{m,j} = ((j-1)/m, j/m], \qquad j = 1, \ldots, m,$$

and let f_n be the resulting sequence of indicators. Each interval in the mth row has probability $1/m$. Hence $P\{f_n \neq 0\} \to 0$, and $f_n \xrightarrow{P} 0$. Every $\omega \in (0,1]$ belongs to exactly one interval in each row and lies outside the other intervals. Thus $f_n(\omega)$ takes both values zero and one infinitely often: it has no limit at any ω.



3.2 Convergence in L^p

A random variable X is *integrable* if $E|X| < \infty$, or, equivalently, if both $E[X^+]$ and $E[X^-]$ are finite. Fix $1 \leq p < \infty$. We write $X \in L^p(\Omega, \mathcal{F}, P)$ when

$$\|X\|_p = \left(E|X|^p\right)^{1/p} < \infty.$$

As usual, random variables equal almost surely represent the same member of L^p.

Definition 3.4. We write $X_n \xrightarrow{L^p} X_\infty$ if $X_n, X_\infty \in L^p$ and

$$\|X_n - X_\infty\|_p \longrightarrow 0.$$

Proposition 3.5. *If $X_n \xrightarrow{L^p} X_\infty$, then $X_n \xrightarrow{P} X_\infty$.*

Proof. Markov's inequality gives

$$P\{|X_n - X_\infty| \geq \varepsilon\} \leq \frac{E|X_n - X_\infty|^p}{\varepsilon^p} \longrightarrow 0.$$

Here the inequality follows directly by integrating the pointwise bound $\varepsilon^p \mathbf{1}_{\{|X_n - X_\infty| \geq \varepsilon\}} \leq |X_n - X_\infty|^p$. □

Remark 3.6 (Moments and agreement of limits). If $X_n \xrightarrow{L^p} X_\infty$, then

$$E|X_n|^p \longrightarrow E|X_\infty|^p.$$

Verify this using Minkowski's inequality, in the form $|\|X_n\|_p - \|X_\infty\|_p| \leq \|X_n - X_\infty\|_p$. For an integer $p \geq 1$, also verify that $E[X_n^p] \to E[X_\infty^p]$.

There is a further useful consistency check: if $X_n \xrightarrow{\text{a.s.}} X_\infty$ and $X_n \xrightarrow{L^p} X'_\infty$, then $X_\infty = X'_\infty$ almost surely.

Example 3.7 (A concentrating spike)**.** Almost-sure convergence does not imply L^p convergence. On $(0,1]$ with Lebesgue probability measure, take

$$X_n(\omega) = n^p \mathbf{1}_{(0,1/n]}(\omega).$$

For every $\omega > 0$, eventually $1/n < \omega$, so $X_n(\omega) \to 0$. But

$$E|X_n|^p = n^{p^2}\frac{1}{n} = n^{p^2-1} \geq 1.$$

Thus $X_n \xrightarrow{\text{a.s.}} 0$ while $X_n \overset{L^p}{\nrightarrow} 0$. In particular, when $p = 1$, every spike has integral one.

Conversely, the typewriter sequence in Example 3.3 satisfies $E|f_n|^p \to 0$ for every finite p, so it converges to zero in L^p and fails to converge almost surely. Almost-sure and L^p convergence therefore do not imply one another, although both imply convergence in probability. The broader question of when convergence in probability yields L^1 convergence leads to uniform integrability and Vitali's convergence theorem. We first consider boundedness and domination.

3.3 Passing to the limit in expectations

Theorem 3.8 (Bounded convergence)**.** *Suppose $X_n \xrightarrow{P} X_\infty$ and there is a constant $M < \infty$ such that $|X_n| \leq M$ almost surely for every n. Then*

$$E|X_n - X_\infty| \longrightarrow 0, \qquad E[X_n] \longrightarrow E[X_\infty].$$

Proof. First, $|X_\infty| \leq M$ almost surely. Indeed, for every $\eta > 0$,

$$P\{|X_\infty| > M + \eta\} \leq P\{|X_n - X_\infty| > \eta\} \longrightarrow 0.$$

Taking $\eta = 1/k$ proves the assertion. Subtracting X_∞, it is enough to consider a sequence converging in probability to zero and bounded by a common constant C; here we may take $C = 2M$. For every $\varepsilon > 0$,

$$\begin{aligned} E|X_n - X_\infty| &= E\big[|X_n - X_\infty|\mathbf{1}_{\{|X_n - X_\infty| \leq \varepsilon\}}\big] \\ &\quad + E\big[|X_n - X_\infty|\mathbf{1}_{\{|X_n - X_\infty| > \varepsilon\}}\big] \\ &\leq \varepsilon + C\,P\{|X_n - X_\infty| > \varepsilon\}. \end{aligned}$$

The probability tends to zero. Taking the upper limit and then letting $\varepsilon \downarrow 0$ proves L^1 convergence. Finally,

$$|E[X_n] - E[X_\infty]| \leq E|X_n - X_\infty| \longrightarrow 0.$$

All the functions integrated in this proof are bounded. Their linearity and the integral triangle inequality follow from uniform simple approximation on a finite measure space. □

Theorem 3.9 (Fatou's lemma). *If $X_n \geq 0$ and $X_n \xrightarrow{P} X$, then $X \geq 0$ almost surely and*

$$E[X] \leq \liminf_{n\to\infty} E[X_n]. \tag{3.1}$$

The expectations may be infinite. On an arbitrary measure space $(S, \mathcal{F}, \mu)$, nonnegative measurable functions f_n satisfy

$$\int_S \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_S f_n \, \mathrm{d}\mu. \tag{3.2}$$

Extended nonnegative values are allowed in (3.2).

Proof. For the probability statement, $P\{X < -\eta\} \leq P\{|X - X_n| > \eta\} \to 0$, so $X \geq 0$ almost surely. Changing X on a measurable null set, take $X \geq 0$ everywhere. Recall the definition

$$E[X] = \sup\{E[Y] : Y \in \mathrm{SF}^+,\ 0 \leq Y \leq X\}.$$

Take such a simple Y. The variables $X_n \wedge Y = \min(X_n, Y)$ are uniformly bounded and converge to $X \wedge Y = Y$ in probability, since

$$|(X_n \wedge Y) - (X \wedge Y)| \leq |X_n - X|.$$

Bounded convergence and monotonicity of the integral give

$$E[Y] = \lim_n E[X_n \wedge Y] \leq \liminf_n E[X_n].$$

Taking the supremum over Y proves (3.1).

For completeness, the simple-minorant argument also gives the arbitrary-measure assertion without a finiteness assumption on μ. Put $f = \liminf_n f_n$ and take a nonnegative simple minorant $s = \sum_{j=1}^{r} a_j \mathbf{1}_{A_j} \leq f$, where the A_j are disjoint and $a_j > 0$. Fix $0 < c < 1$ and set

$$B_{N,j} = A_j \cap \bigcap_{n \geq N} \{f_n \geq c a_j\}.$$

Then $B_{N,j} \uparrow A_j$: at each point of A_j, the lower limit is at least $a_j > c a_j$. For $n \geq N$, integral monotonicity gives

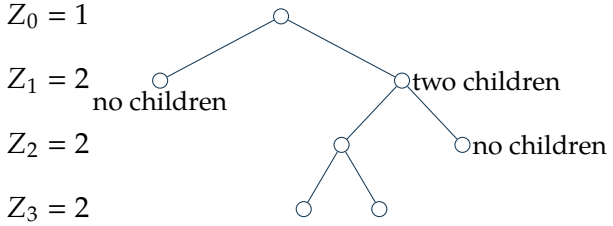
$$\int_S f_n \, \mathrm{d}\mu \geq c \sum_{j=1}^{r} a_j \mu(B_{N,j}).$$

First take the lower limit in n, then let $N \to \infty$ and use continuity from below of the measure. This yields $\liminf_n \int f_n \, \mathrm{d}\mu \geq c \int s \, \mathrm{d}\mu$, including when the latter integral is infinite. Let $c \uparrow 1$ and take the supremum over s. The same conclusion holds if $s \leq f$ only almost everywhere, by removing a measurable null set from the A_j. □

Example 3.10 (A critical branching process)**.** Start with one individual, $Z_0 = 1$. Each individual has either no children or two children, with probability $1/2$ for each possibility. Let $\xi_{n,i}$ be the number of children of individual i in generation $n-1$. All these offspring variables are independent and have the same law. The population in generation n is

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{n,i}, \qquad P\{\xi_{n,i} = 0\} = P\{\xi_{n,i} = 2\} = \tfrac{1}{2},$$

where an empty sum is zero. Once the population is zero, it stays zero. Independence means that probabilities for the separate offspring choices multiply; see also the general definition in 4.2.



Since $Z_{n-1} \le 2^{n-1}$, we can write the random sum as a finite sum,

$$Z_n = \sum_{i=1}^{2^{n-1}} \xi_{n,i} \mathbf{1}_{\{Z_{n-1} \ge i\}}.$$

The offspring in generation n are independent of all earlier choices, and $E[\xi_{n,i}] = 1$. Therefore

$$E[Z_n] = \sum_{i=1}^{2^{n-1}} P\{Z_{n-1} \ge i\} = E[Z_{n-1}] = 1.$$

Only expectations of bounded variables and finite sums are used here.

To determine the limit, let $q_n = P\{Z_n = 0\}$, with $q_0 = 0$. If the first individual has no children, extinction is immediate. If it has two children, their two descendant trees are independent copies of the original tree. Thus

$$q_{n+1} = \tfrac{1}{2} + \tfrac{1}{2} q_n^2.$$

The events $\{Z_n = 0\}$ increase, so $q_n \uparrow q$, where q is the probability of eventual extinction. Passing to the limit gives

$$q = \tfrac{1}{2} + \tfrac{1}{2} q^2, \qquad (q-1)^2 = 0.$$

Hence $q = 1$: almost surely the process reaches zero in a finite generation. In particular, $Z_n \to Z_\infty = 0$ almost surely. Fatou's lemma gives a strict inequality,

$$E[Z_\infty] = 0 < 1 = \liminf_n E[Z_n].$$

This is another example of almost-sure convergence without L^1 convergence.

For a general offspring law on the nonnegative integers, write $m = E[\xi]$ and

$$g(s) = E[s^\xi] = \sum_{k=0}^{\infty} P\{\xi = k\} s^k, \qquad 0 \le s \le 1,$$

with $s^0 = 1$ also at $s = 0$. The same tree argument gives $q_{n+1} = g(q_n)$, $q_0 = 0$. Thus the extinction probability is the smallest fixed point of g in $[0, 1]$: iteration stays below every fixed point, and continuity of g shows that its limit is a fixed point. The mean m, which may be infinite, is the left derivative $g'(1-)$. Convexity of g shows that $q = 1$ if $m < 1$, and also if $m = 1$ unless $P\{\xi = 1\} = 1$, in which case $Z_n = 1$ for all n. If $m > 1$, then $q < 1$; when $P\{\xi = 0\} > 0$, we also have $q > 0$. For example, for offspring zero or two with $P\{\xi = 2\} = p > 1/2$, the smaller fixed point is $q = (1 - p)/p$.

In the supercritical case the population tends to infinity on survival, while it tends to zero on extinction. To see the first assertion when $p_0 = P\{\xi = 0\} > 0$, fix K. At each visit to a population between one and K, the fresh offspring choices give a probability at least p_0^K of extinction in the next generation. The probability of surviving r such visits is at most $(1 - p_0^K)^r$. Survival therefore allows only finitely many visits to $\{1, \dots, K\}$. This holds for every K. If $p_0 = 0$, the population is nondecreasing, and $P\{\xi = 1\} < 1$ when $m > 1$; the probability of remaining forever at any fixed positive population is zero. Thus supercriticality does not imply certain divergence when extinction has positive probability.

Theorem 3.11 (Monotone convergence)**.** *Let $(S, \mathcal{F}, \mu)$ be a measure space. If $f_n \ge 0$ are measurable, f is measurable, and $f_n \uparrow f$ almost everywhere, then*

$$\int_S f_n \, \mathrm{d}\mu \uparrow \int_S f \, \mathrm{d}\mu.$$

The integrals and the limit may be infinite. In particular, on a probability space, $0 \le X_n \uparrow X$ almost surely implies $E[X_n] \uparrow E[X]$.

Proof. Choose a measurable null set N outside which $f_n \uparrow f$. Replacing every f_n and f by zero on N preserves measurability and all the integrals, so we may assume $f_n \uparrow f$ everywhere. Integral monotonicity, which follows from the definition by simple minorants, gives

$$\int_S f_n \, \mathrm{d}\mu \le \int_S f_{n+1} \, \mathrm{d}\mu \le \int_S f \, \mathrm{d}\mu.$$

Thus the increasing limit of the integrals is at most $\int f \, \mathrm{d}\mu$. The reverse inequality is Fatou's lemma:

$$\int_S f \, \mathrm{d}\mu \le \liminf_n \int_S f_n \, \mathrm{d}\mu.$$

Together the two inequalities prove the assertion. □

Theorem 3.12 (Dominated convergence)**.** *Suppose* $X_n \xrightarrow{P} X_\infty$ *and, for some* $Y \in L^1$,

$$|X_n| \le Y \quad \text{almost surely for every } n.$$

Then $X_\infty \in L^1$ *and*

$$E[X_n] \longrightarrow E[X_\infty].$$

In fact, $X_n \xrightarrow{L^1} X_\infty$.

Proof. The bound passes to the limit. For every $\eta > 0$,

$$P\{|X_\infty| > Y + \eta\} \le P\{|X_n - X_\infty| > \eta\} \longrightarrow 0.$$

Taking a countable sequence $\eta \downarrow 0$ gives $|X_\infty| \le Y$ almost surely, so X_∞ is integrable.

Now $Y + X_n \ge 0$ and $Y + X_n \xrightarrow{P} Y + X_\infty$. Fatou's lemma therefore gives

$$E[Y + X_\infty] \le \liminf_n E[Y + X_n].$$

Subtracting the finite number $E[Y]$, we obtain

$$E[X_\infty] \le \liminf_n E[X_n].$$

Apply the same argument to $Y - X_n \ge 0$, which converges in probability to $Y - X_\infty$. It yields

$$E[Y] - E[X_\infty] \le E[Y] - \limsup_n E[X_n], \qquad E[X_\infty] \ge \limsup_n E[X_n].$$

These two inequalities prove convergence of the expectations. Finally, $|X_n - X_\infty| \xrightarrow{P} 0$ and $|X_n - X_\infty| \le 2Y$. The assertion just proved, applied to these nonnegative differences, gives $E|X_n - X_\infty| \to 0$. □

3.4 General measure spaces and almost-uniform convergence

Now let $(S, \mathcal{F}, \mu)$ be an arbitrary measure space, and let f_n, f be real-valued measurable functions. Convergence *almost everywhere*, written $f_n \xrightarrow{\text{a.e.}} f$, means pointwise convergence outside a measurable set of measure zero. Convergence *in measure*, written $f_n \xrightarrow{\mu} f$, means

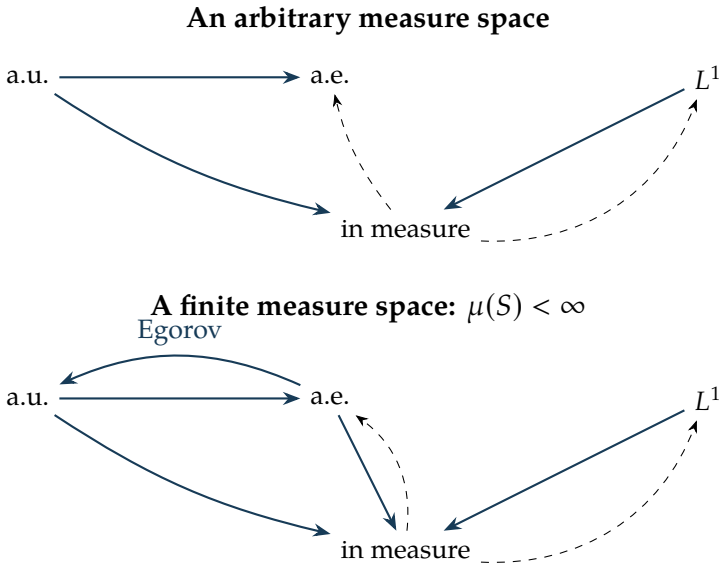
$$\mu\{|f_n - f| \ge \varepsilon\} \longrightarrow 0 \qquad \text{for every } \varepsilon > 0.$$

For a probability measure this is precisely convergence in probability. We write $f_n \xrightarrow{L^1(\mu)} f$ when $f_n, f \in L^1(\mu)$ and $\int_S |f_n - f|\,\mathrm{d}\mu \to 0$.

Definition 3.13 (Almost-uniform convergence)**.** We say that f_n converges to f *almost uniformly*, written $f_n \xrightarrow{\text{a.u.}} f$, if for every $\delta > 0$ there is a measurable set E_δ with $\mu(E_\delta) < \delta$ such that $f_n \to f$ uniformly on E_δ^c. Explicitly, after choosing this exceptional set, for every $\varepsilon > 0$ there is a single N such that

$$n \ge N, \quad x \in E_\delta^c \quad \Longrightarrow \quad |f_n(x) - f(x)| < \varepsilon.$$

The following two diagrams compare the modes. A solid arrow is a valid implication. A dashed arrow marks a direction that fails in general; it does not assert an implication.



The L^1 implication follows from

$$\mu\{|f_n - f| \geq \varepsilon\} \leq \frac{1}{\varepsilon} \int_S |f_n - f| \, \mathrm{d}\mu.$$

The arrow labelled Egorov is the following named theorem.

Theorem 3.14 (Egorov)**.** *On a finite measure space, a sequence of real-valued measurable functions converging almost everywhere to a real-valued measurable function converges almost uniformly to that function.*

3.4.1 Four counterexamples

The first three examples use Lebesgue measure, on the indicated interval or on $\mathbb{R}$. Write $\chi_A = \mathbf{1}_A$ for an indicator.

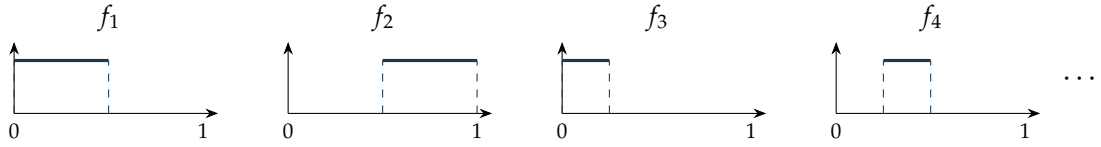
(i) On $\mathbb{R}$, let $f_n = n^{-1}\chi_{(0,n)}$. Then $\sup_x |f_n(x)| = 1/n \to 0$, so convergence to zero is uniform, almost uniform, almost everywhere, and in measure. Nevertheless, $\int_{\mathbb{R}} |f_n| \, \mathrm{d}x = 1$, so there is no L^1 convergence.

(ii) On $[0, 1]$, let $f_n = n\chi_{[0,1/n)}$. For every $x > 0$, eventually $f_n(x) = 0$, and the support has measure $1/n$. Thus $f_n \to 0$ almost everywhere and in measure, while $\int_0^1 |f_n| \, \mathrm{d}x = 1$. It also converges almost uniformly: remove $[0, a)$ with $0 < a < \delta$; the functions are eventually zero on its complement. This is the $p = 1$ concentrating spike, with a different, immaterial choice of endpoints.

(iii) On $\mathbb{R}$, let $f_n = \chi_{(n,n+1)}$. Every fixed point eventually lies outside the moving interval, so $f_n \to 0$ everywhere. But $\mu\{|f_n| \geq 1/2\} = 1$ for every n. Hence almost-everywhere

convergence need not imply convergence in measure on an infinite measure space. It cannot imply almost-uniform convergence there either, since almost-uniform convergence implies convergence in measure.

(iv) On $[0,1]$, move an interval first through the two halves, then through the four quarters, then through the eight eighths, and so on. For $k \geq 1$ and $0 \leq j < 2^k$, take, in this order, $\chi_{[j2^{-k},(j+1)2^{-k})}$, assigning the value zero at $x = 1$. The integrals are 2^{-k} and tend to zero, so the sequence converges in L^1 and in measure. At every $x \in [0,1)$, however, it is one once in each row and zero at the other positions. It does not converge almost everywhere, and therefore does not converge almost uniformly. This is the dyadic version of the typewriter construction.



3.4.2 The almost-uniform implications

Proposition 3.15. *On any measure space, almost-uniform convergence implies almost-everywhere convergence.*

Proof. For each integer $m \geq 1$, choose a measurable set $E_{1/m}$ such that $\mu(E_{1/m}) < 1/m$ and $f_n \to f$ uniformly on $E_{1/m}^c$. Put

$$E = \bigcap_{m=1}^{\infty} E_{1/m}.$$

If $x \in E^c = \bigcup_m E_{1/m}^c$, then $x \in E_{1/m}^c$ for at least one m, and consequently $f_n(x) \to f(x)$. Moreover,

$$0 \leq \mu(E) \leq \mu(E_{1/m}) < \frac{1}{m} \qquad \text{for every } m \geq 1.$$

Thus $\mu(E) = 0$, which proves almost-everywhere convergence. The exceptional sets do not have to be nested. □

Proposition 3.16. *On any measure space, almost-uniform convergence implies convergence in measure.*

Proof. Suppose convergence in measure fails. There are $\varepsilon, \delta > 0$ and a subsequence f_{n_k} such that

$$\mu\{|f_{n_k} - f| \geq \varepsilon\} \geq \delta \qquad \text{for every } k.$$

Almost-uniform convergence supplies a measurable E_δ with $\mu(E_\delta) < \delta$ and uniform convergence on E_δ^c. Choose N so that

$$|f_n(x) - f(x)| < \varepsilon \qquad \text{for all } n \geq N \text{ and all } x \in E_\delta^c.$$

For every k with $n_k \geq N$, we have $\{|f_{n_k} - f| \geq \varepsilon\} \subseteq E_\delta$. Therefore

$$0 < \delta \leq \mu\{|f_{n_k} - f| \geq \varepsilon\} \leq \mu(E_\delta) < \delta,$$

a contradiction. Hence $f_n \xrightarrow{\mu} f$. □

Proposition 3.17. *If $\mu(S) < \infty$ and $f_n \to f$ almost everywhere, then $f_n \xrightarrow{\mu} f$.*

Proof. Fix $\varepsilon > 0$ and define the tail sets

$$B_N = \bigcup_{n \geq N} \{|f_n - f| \geq \varepsilon\}.$$

The sets B_N decrease, and their intersection is contained in the null set where convergence fails. Since $\mu(B_1) \leq \mu(S) < \infty$, continuity from above gives

$$\lim_{N \to \infty} \mu(B_N) = \mu\left(\bigcap_{N=1}^{\infty} B_N\right) = 0.$$

Finally, $\{|f_N - f| \geq \varepsilon\} \subseteq B_N$, so its measure also tends to zero. The finiteness of the measure is exactly what justifies continuity from above here. □

Problems

Problem 3.1 (Exponential transform and logarithmic limits)**.** Let X be a real-valued random variable.

(a) Prove that the function
$$f(x) = E\big[\exp(-|X - x|)\big]$$
is continuous on $\mathbb{R}$. This function is sometimes called the *bilateral exponential transform*.

(b) Suppose further that $X \geq 0$ and $E[X^{p_0}] < \infty$ for some fixed $p_0 > 0$. Use the convention $\log 0 = -\infty$, including when $E[X^p] = 0$. Interpret $E[\log X]$ and the following limits in the extended real sense; the positive part of $\log X$ is integrable under the stated moment assumption.

(b.1) Show that
$$\lim_{p \downarrow 0} \frac{E[X^p] - 1}{p} = E[\log X].$$

(b.2) Conclude that
$$\lim_{p \downarrow 0} \frac{\log(E[X^p])}{p} = E[\log X].$$

Hint. For fixed $x \geq 0$, use convexity of $p \mapsto x^p$ to deduce that
$$\frac{x^p - 1}{p} \downarrow \log x \qquad \text{as } p \downarrow 0.$$

Problem 3.2 (Random indices and almost-sure convergence)**.** Let $(X_n)_{n \geq 1}$ be real-valued random variables on $(\Omega, \mathcal{F}, P)$. Prove that $X_n \xrightarrow{\text{a.s.}} 0$ if and only if, for every $\varepsilon > 0$, there is a deterministic integer $n \geq 1$ such that for every measurable random integer
$$N : \Omega \longrightarrow \{n, n+1, n+2, \ldots\},$$
we have
$$P\big\{\omega : |X_{N(\omega)}(\omega)| > \varepsilon\big\} < \varepsilon.$$

Chapter 4

Product Measures, Characteristic Functions, and Weak Convergence

4.1 Product spaces and independence

Given measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, their product space is

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_i \in \Omega_i\}, \qquad \mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\{A_1 \times A_2 : A_i \in \mathcal{F}_i\}.$$

The sets $A_1 \times A_2$ are called *measurable rectangles*. The family

$$\mathcal{G} = \left\{ \bigsqcup_{j=1}^{m} A_1^{(j)} \times A_2^{(j)} : A_i^{(j)} \in \mathcal{F}_i,\ m < \infty \right\}$$

is an algebra, and $\sigma(\mathcal{G}) = \mathcal{F}_1 \otimes \mathcal{F}_2$. Indeed, intersections of rectangles are rectangles, and the complement of a rectangle is the disjoint union

$$(A_1 \times A_2)^c = (A_1^c \times \Omega_2) \sqcup (A_1 \times A_2^c).$$

Finite refinements give the algebra properties. In general $\mathcal{G}$ is not a sigma-algebra. For example, on $\mathbb{N} \times \mathbb{N}$ with all subsets measurable, the diagonal is a countable union of singleton rectangles, but not a finite union of rectangles: a nonempty rectangle contained in the diagonal has at most one point.

Theorem 4.1 (Product measure)**.** *Let $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, 2$, be probability spaces. There is a unique probability measure $P = P_1 \times P_2$ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ such that*

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2).$$

Consequently, on every finite disjoint union of rectangles,

$$P\left(\bigsqcup_{j=1}^{m} A_1^{(j)} \times A_2^{(j)} \right) = \sum_{j=1}^{m} P_1(A_1^{(j)})P_2(A_2^{(j)}).$$

The same construction gives the unique product measure of two sigma-finite measures, with the usual convention $0 \cdot \infty = 0$ on rectangles.

Proof. First consider probability measures. For $A \in \mathcal{G}$, let $A_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in A\}$. The proposed value is

$$p(A) = \int_{\Omega_1} P_2(A_{\omega_1}) \, \mathrm{d}P_1(\omega_1).$$

For a disjoint rectangle representation, its integrand is the measurable simple function $\sum_j \mathbf{1}_{A_1^{(j)}}(\omega_1) P_2(A_2^{(j)})$. Thus the value does not depend on the representation. If $A = \bigsqcup_{k\geq 1} A_k$ with $A, A_k \in \mathcal{G}$, countable additivity on each section and monotone convergence give $p(A) = \sum_k p(A_k)$. The extension theorem (Theorem 1.14) therefore applies to this premeasure on $\mathcal{G}$. Its finite total mass also gives uniqueness.

For sigma-finite measures, partition each space into countably many measurable pieces of finite measure. On each pair of pieces use the finite construction, obtained by scaling the probability construction (zero-mass pieces give the zero measure). Sum these measures over the disjoint product pieces. This gives the rectangle formula. Any other measure with that formula agrees on each finite product piece by finite-measure uniqueness, and hence agrees everywhere. □

Definition 4.2 (Independence)**.** Real random variables X, Y on the same probability space are *independent* if their joint law is the product of their laws:

$$P_{(X,Y)} = P_X \times P_Y.$$

More generally, $X_1, \ldots, X_m$ are independent if $P_{(X_1,\ldots,X_m)} = P_{X_1} \times \cdots \times P_{X_m}$. Equivalently,

$$P\left(\bigcap_{j=1}^{m} \{X_j \in A_j\}\right) = \prod_{j=1}^{m} P(X_j \in A_j), \qquad A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R}). \tag{4.1}$$

A sequence $(X_n)_{n\geq 1}$ is independent if every finite subcollection is independent. Independent variables with the same law are called *independent and identically distributed*, abbreviated iid.

The equivalence follows because measurable rectangles form a generating pi-system for $\mathcal{B}(\mathbb{R}^m)$, so their probabilities determine the joint law. Independence of events means the corresponding factorization of probabilities of all finite intersections; it is equivalently independence of their indicator variables. Independence is preserved under measurable functions of the separate coordinates, since their inverse images may be substituted for the A_j in (4.1).

Example 4.3 (Independent copies)**.** Ten independent tosses of a coin with head probability p are represented on $\{0,1\}^{10}$ by the product of ten Bernoulli measures. A particular sequence with k heads has probability $p^k(1-p)^{10-k}$. More generally, if X is defined on $(\Omega, \mathcal{F}, P)$, the variables

$$X_j(\omega_1, \ldots, \omega_m) = X(\omega_j), \qquad 1 \leq j \leq m,$$

on $(\Omega^m, \mathcal{F}^{\otimes m}, P^{\otimes m})$ are iid copies of X. This works for every finite m. For an infinite sequence of real random variables, the consistent finite product laws extend to a law on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}})$ by the Kolmogorov extension theorem; its coordinate variables provide countably many independent copies.

4.2 Integration on product spaces

Theorem 4.4 (Tonelli and Fubini)**.** *Let μ_1 and μ_2 be sigma-finite measures on $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$, and let $\mu = \mu_1 \times \mu_2$. Suppose h is $\mathcal{X} \otimes \mathcal{Y}$-measurable. If $h \geq 0$, then the section integrals are measurable, may equal $+\infty$, and*

$$\begin{aligned}\int_{\mathsf{X}\times\mathsf{Y}} h \,\mathrm{d}\mu &= \int_{\mathsf{X}} \left[\int_{\mathsf{Y}} h(x,y) \,\mathrm{d}\mu_2(y) \right] \mathrm{d}\mu_1(x) \\ &= \int_{\mathsf{Y}} \left[\int_{\mathsf{X}} h(x,y) \,\mathrm{d}\mu_1(x) \right] \mathrm{d}\mu_2(y).\end{aligned}$$

If h is real or complex valued and $\int |h| \,\mathrm{d}\mu < \infty$, the same equalities hold with finite integrals. In this case the sections are integrable for almost every outer variable. Each inner integral in the display is set equal to zero on the measurable null set where its section is not integrable.

Proof. Sections of measurable sets are measurable: the class of sets with this property is a sigma-algebra containing the rectangles. Thus sections of measurable functions are measurable as well.

For finite measures, the integral identities and measurability of the section integrals hold first for indicators of rectangles. The class of measurable sets for whose indicators they hold is a lambda-system: complements follow by subtracting from the finite total masses, and disjoint countable unions follow by monotone convergence. The pi-lambda theorem extends the identities to all measurable indicators. For sigma-finite measures, apply this finite argument on increasing product pieces of finite measure, and use monotone convergence again.

The standard integration procedure now applies: linearity gives the result for nonnegative simple functions; increasing simple approximation and monotone convergence give it for nonnegative h. Apply that result to $|h|$ in the integrable case. Its iterated integrals are finite, so almost every section is integrable. Subtract the positive and negative parts of the real and imaginary parts to finish. □

In particular, if X, Y are independent and f, g are integrable with respect to their respective laws, Tonelli applied to $|f(x)g(y)|$ and then Fubini give

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}f(X)\,\mathbb{E}g(Y). \tag{4.2}$$

This also holds for nonnegative functions with the corresponding extended-integral convention, and for bounded complex functions.

4.3 Characteristic functions and moments

Definition 4.5. The *characteristic function* of a real random variable X is

$$\Phi_X : \mathbb{R} \longrightarrow \mathbb{C}, \qquad \Phi_X(t) = \mathbb{E}\mathrm{e}^{\mathrm{i}tX} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \,\mathrm{d}P_X(x).$$

It is defined for every real t, because $|\mathrm{e}^{\mathrm{i}tX}| = 1$.

Proposition 4.6. *Every characteristic function satisfies* $\Phi_X(0) = 1$, $|\Phi_X(t)| \le 1$, *and* $\Phi_X(-t) = \overline{\Phi_X(t)}$. *It is uniformly continuous on* $\mathbb{R}$ *and positive semidefinite: for every* $m \ge 1$, $t_1, \dots, t_m \in \mathbb{R}$, *and* $\xi_1, \dots, \xi_m \in \mathbb{C}$,

$$\sum_{j,k=1}^{m} \Phi_X(t_j - t_k)\xi_j\overline{\xi_k} \ge 0.$$

Equivalently, the matrix $(\Phi_X(t_j - t_k))_{j,k=1}^m$ *is positive semidefinite.*

Proof. Normalization, the bound, and conjugate symmetry follow directly from the definition. For uniform continuity, observe that

$$\begin{aligned} |\Phi_X(t+h) - \Phi_X(t)| &= \left|\mathbb{E}\left[e^{itX}(e^{ihX} - 1)\right]\right| \\ &\le \mathbb{E}|e^{ihX} - 1| \longrightarrow 0 \qquad (h \longrightarrow 0). \end{aligned}$$

Bounded convergence applies because $|e^{ihX} - 1| \le 2$, and the bound does not depend on t. Finally,

$$\sum_{j,k} \Phi_X(t_j - t_k)\xi_j\overline{\xi_k} = \mathbb{E}\sum_{j,k} \xi_j\overline{\xi_k}e^{i(t_j - t_k)X} = \mathbb{E}\left|\sum_j \xi_j e^{it_jX}\right|^2 \ge 0.$$

□

Proposition 4.7 (Moments and derivatives)**.** *If* $\mathbb{E}|X|^m < \infty$ *for a positive integer* m, *then* $\Phi_X \in C^m(\mathbb{R})$, *and for* $1 \le k \le m$,

$$\Phi_X^{(k)}(t) = \mathbb{E}[(iX)^k e^{itX}], \qquad \Phi_X^{(k)}(0) = i^k\mathbb{E}X^k.$$

Conversely, if $m = 2k$ *is even and the finite derivative* $\Phi_X^{(2k)}(0)$ *exists, then* $\mathbb{E}|X|^{2k} < \infty$. *In particular,*

$$\Phi_X'(0) = i\mathbb{E}X \quad \text{when } \mathbb{E}|X| < \infty, \qquad \Phi_X''(0) = -\mathbb{E}X^2 \quad \text{when } \mathbb{E}X^2 < \infty.$$

Proof. Since $|e^{iu} - 1| \le |u|$,

$$\left|\frac{(iX)^{j-1}e^{i(t+h)X} - (iX)^{j-1}e^{itX}}{h}\right| \le |X|^j.$$

Differentiating successively under the expectation is justified by dominated convergence. The same theorem gives continuity of the derivatives.

For the even-order converse, the nonnegative functions

$$\frac{|e^{ihx} - 1|^{2k}}{h^{2k}} \longrightarrow |x|^{2k} \qquad (h \longrightarrow 0)$$

give, by Fatou's lemma,

$$\mathbb{E}|X|^{2k} \le \liminf_{h\to 0} \frac{\mathbb{E}|e^{ihX}-1|^{2k}}{h^{2k}} = \lim_{h\to 0} \frac{(-1)^k}{h^{2k}} \sum_{j=0}^{2k} (-1)^j \binom{2k}{j} \Phi_X((j-k)h) = (-1)^k \Phi_X^{(2k)}(0) < \infty.$$

The last equality follows by the Taylor expansion at zero: the finite difference annihilates every power of degree less than $2k$, and its coefficient at degree $2k$ is $(2k)!$. Once the moment is finite, the forward implication also gives equality with that moment. □

Example 4.8 (Cauchy law)**.** The standard Cauchy density and characteristic function are

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \qquad \Phi_X(t) = e^{-|t|}.$$

The characteristic function is not differentiable at zero. Consequently $\mathbb{E}|X| = \infty$, also seen directly by integrating $|x|/[\pi(1+x^2)]$. Symmetry does not give a finite mean: both $\mathbb{E}X^+$ and $\mathbb{E}X^-$ are infinite, so $\mathbb{E}X$ is undefined.

Example 4.9 (A differentiable characteristic function without an absolute mean)**.** Choose $C > 0$ so that

$$P(X = m) = P(X = -m) = \frac{C}{m^2 \log m}, \qquad m \ge 2, \qquad 2C \sum_{m=2}^{\infty} \frac{1}{m^2 \log m} = 1.$$

Then

$$\Phi_X(t) = 2C \sum_{m=2}^{\infty} \frac{\cos(mt)}{m^2 \log m}, \qquad \mathbb{E}|X| = 2C \sum_{m=2}^{\infty} \frac{1}{m \log m} = \infty.$$

Again $\mathbb{E}X$ is undefined, whereas $\Phi_X'(0) = 0$. In fact $\Phi_X \in C^1(\mathbb{R})$.

To see the derivative at zero directly, set $N = \lfloor 1/|t| \rfloor$ for small nonzero t. The bounds $1 - \cos u \le u^2/2$ and $1 - \cos u \le 2$ yield

$$\frac{|\Phi_X(t) - 1|}{|t|} \le C|t| \sum_{m=2}^{N} \frac{1}{\log m} + \frac{4C}{|t|} \sum_{m>N} \frac{1}{m^2 \log m} \longrightarrow 0.$$

Indeed, splitting the first sum at $\sqrt{N}$ bounds it by $\sqrt{N}/\log 2 + 2N/\log N$, while the second sum is at most $1/(N \log N)$.

For completeness, away from $2\pi\mathbb{Z}$, summation by parts gives locally uniform convergence of

$$\Phi_X'(t) = -2C \sum_{m=2}^{\infty} \frac{\sin(mt)}{m \log m},$$

since the partial sums of $\sin(mt)$ are bounded by a constant times $1/|\sin(t/2)|$. Near zero, the sum up to N is bounded in modulus by $|t| \sum_{m=2}^{N} 1/\log m$, and the remaining sum is

bounded by a constant times $1/[N \log N \, |\sin(t/2)|]$. Both tend to zero. Periodicity gives continuity of the derivative at every point of $2\pi\mathbb{Z}$.

The failure of the converse also occurs at higher odd orders. For example, the symmetric law

$$P(Y = m) = P(Y = -m) = \frac{C_3}{m^4 \log m}, \qquad m \geq 2,$$

has $\mathbb{E}|Y|^3 = \infty$ but $\Phi_Y \in C^3(\mathbb{R})$. Its second derivative is $-2C_3 \sum_{m\geq 2} \cos(mt)/(m^2 \log m)$, a constant multiple of the preceding C^1 function, so $\Phi_Y'''(0) = 0$. The nearby logarithmic variant with masses $\widetilde{C}/[m^4(\log m)^2]$ has a finite absolute third moment and an infinite fourth moment; it therefore has three continuous derivatives but no finite fourth derivative at zero.

4.4 Fourier inversion and uniqueness

The distribution function and the law of X are related by $F_X(x) = P_X((-\infty, x])$. If X has density f_X, then

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u, \qquad \Phi_X(t) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} f_X(x)\,\mathrm{d}x.$$

Thus the characteristic function is a Fourier transform of the density. Fourier inversion suggests recovering the density, or more generally the law, from the characteristic function.

There are two useful normalizations. With the Fourier-analysis convention

$$\widehat{f}(t) = \int_{\mathbb{R}} \mathrm{e}^{-2\pi\mathrm{i}tx} f(x)\,\mathrm{d}x,$$

the inversion theorem for $f, \widehat{f} \in L^1(\mathbb{R})$ says

$$f(x) = \int_{\mathbb{R}} \mathrm{e}^{2\pi\mathrm{i}tx} \widehat{f}(t)\,\mathrm{d}t \quad \text{for almost every } x.$$

The right side is a continuous representative, and equality holds at every point where f is continuous. For a probability density the exact bridge is

$$\widehat{f_X}(t) = \Phi_X(-2\pi t), \qquad \Phi_X(t) = \widehat{f_X}\left(-\frac{t}{2\pi}\right).$$

In particular, when $\Phi_X \in L^1(\mathbb{R})$, the probability normalization gives

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx} \Phi_X(t)\,\mathrm{d}t$$

for the continuous representative of the density. The following inversion formula does not require a density or integrability of Φ_X.

Theorem 4.10 (Lévy's inversion formula)**.** *Let F and Φ be the distribution function and characteristic function of a probability measure P. If $a < b$ are continuity points of F, then*

$$F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \Phi(t)\,\mathrm{d}t.$$

At $t = 0$ the quotient is given its continuous value $b - a$. For arbitrary $a < b$, the same limit equals

$$\frac{F(b) + F(b-)}{2} - \frac{F(a) + F(a-)}{2} = P((a,b)) + \frac{P(\{a\}) + P(\{b\})}{2}.$$

Proof. Substitute $\Phi(t) = \int \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}P(x)$ into the truncated integral. For fixed T the interchange is justified by Fubini, since

$$\left| \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \mathrm{e}^{\mathrm{i}tx} \right| \le b - a, \qquad |t| \le T,$$

and the product measure of $[-T, T] \times \mathbb{R}$ is finite. The result is $\int K_T(x)\,\mathrm{d}P(x)$, where

$$K_T(x) = \frac{1}{\pi} \left[\int_0^T \frac{\sin(t(x-a))}{t}\,\mathrm{d}t - \int_0^T \frac{\sin(t(x-b))}{t}\,\mathrm{d}t \right].$$

The Dirichlet integral gives

$$\int_0^\infty \frac{\sin u}{u}\,\mathrm{d}u = \frac{\pi}{2}, \qquad \sup_{v\ge 0} \left| \int_0^v \frac{\sin u}{u}\,\mathrm{d}u \right| < \infty.$$

One way to verify the value is to insert $\mathrm{e}^{-\varepsilon u}$: differentiate $\int_0^\infty \mathrm{e}^{-\varepsilon u} \sin(bu)/u\,\mathrm{d}u$ in b, then integrate $\varepsilon/(\varepsilon^2 + b^2)$ from 0 to 1. This gives $\arctan(1/\varepsilon) \to \pi/2$. Integration by parts bounds both the damped and undamped tails beyond A by $2/A$, uniformly in $\varepsilon > 0$, so the damped limit equals the improper integral. The same tail bound gives boundedness of its partial integrals.

It follows that K_T is bounded uniformly in T and x, and tends to 1 on (a, b), to $1/2$ at a, b, and to 0 outside $[a, b]$. Bounded convergence proves the formula, including its endpoint terms. □

Corollary 4.11 (Uniqueness)**.** *If $\Phi_X(t) = \Phi_Y(t)$ for every $t \in \mathbb{R}$, then X and Y have the same law, written $X \overset{d}{=} Y$.*

Proof. Let C be the set of common continuity points of F_X, F_Y. Each distribution function has at most countably many discontinuities, so C is dense. Inversion gives $F_X(b) - F_X(a) = F_Y(b) - F_Y(a)$ whenever $a < b$ lie in C. Letting $a \to -\infty$ through C gives equality at each $b \in C$. Right continuity then gives equality everywhere, since

$$F_X(x) = \inf\{F_X(z) : z > x,\ z \in C\},$$

and the same holds for F_Y. □

4.5 Convolution and sums of independent variables

For probability measures P, Q on $\mathbb{R}$, their *convolution* is the probability measure

$$\begin{aligned}(P * Q)(A) &= \int_{\mathbb{R}} P(A - y)\,\mathrm{d}Q(y) = \int_{\mathbb{R}} Q(A - x)\,\mathrm{d}P(x) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_A(x + y)\,\mathrm{d}(P \times Q)(x, y), \qquad A \in \mathcal{B}(\mathbb{R}),\end{aligned}$$

where $A - y = \{x : x + y \in A\}$. It is the pushforward of $P \times Q$ under addition; the two iterated forms follow from Tonelli.

Proposition 4.12. *If X, Y are independent, then $P_{X+Y} = P_X * P_Y$ and*

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t).$$

In particular, if $X_1, \dots, X_n$ are iid and $S_n = \sum_{j=1}^n X_j$, then $\Phi_{S_n}(t) = \Phi_{X_1}(t)^n$.

Proof. The joint law is the product law, so its pushforward under addition is the convolution. Fubini, applied to a function of modulus one, gives

$$\begin{aligned}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tz}\,\mathrm{d}(P_X * P_Y)(z) &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}t(x+y)}\,\mathrm{d}P_X(x)\,\mathrm{d}P_Y(y) \\ &= \left(\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}P_X(x)\right)\left(\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}ty}\,\mathrm{d}P_Y(y)\right).\end{aligned}$$

Iteration proves the iid assertion. □

Example 4.13 (Factorization for the sum does not imply independence)**.** Let X be standard Cauchy and put $Y = X$. Then X, Y are dependent, but

$$\Phi_{X+Y}(t) = \Phi_X(2t) = \mathrm{e}^{-2|t|} = \Phi_X(t)\Phi_Y(t).$$

For instance $P(X \le 0, Y \le 0) = 1/2$, whereas $P(X \le 0)P(Y \le 0) = 1/4$. The characteristic function of the full joint law is $\mathbb{E}\mathrm{e}^{\mathrm{i}(sX+tY)} = \mathrm{e}^{-|s+t|}$; its factorization would have to hold for *every pair* (s, t), and fails at $(1, -1)$.

Example 4.14 (Rare events and a Poisson limit)**.** Let $\lambda > 0$, and for each sufficiently large n let $X_{n1}, \dots, X_{nn}$ be independent Bernoulli variables of parameter λ/n. Then $S_n = \sum_{j=1}^n X_{nj}$ is binomial with parameters $(n, \lambda/n)$, and

$$\Phi_{S_n}(t) = \left(1 + \frac{\lambda}{n}(\mathrm{e}^{\mathrm{i}t} - 1)\right)^n \longrightarrow \exp\{\lambda(\mathrm{e}^{\mathrm{i}t} - 1)\}.$$

The limit is the characteristic function of a Poisson variable W of parameter λ, since

$$\mathbb{E}\mathrm{e}^{\mathrm{i}tW} = \sum_{k=0}^{\infty} \mathrm{e}^{\mathrm{i}tk}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = \exp\{\lambda(\mathrm{e}^{\mathrm{i}t} - 1)\}.$$

In particular, success probability $1/n$ gives the parameter-one limit. Uniqueness identifies distributions whose characteristic functions are equal. To pass from the displayed convergence of characteristic functions to convergence of distributions, we need a continuity theorem.

4.6 Weak convergence and the continuity theorem

Definition 4.15 (Weak convergence)**.** Let P_n, P be probability measures on $\mathbb{R}$ with distribution functions F_n, F. We write $F_n \Rightarrow F$, or $P_n \Rightarrow P$, if

$$F_n(x) \longrightarrow F(x) \quad \text{at every continuity point } x \text{ of } F.$$

For random variables with these laws this is *convergence in distribution*, written $X_n \xrightarrow{d} X$. The variables need not be defined on one probability space. Let $C_b(\mathbb{R})$ denote the bounded continuous real-valued functions on $\mathbb{R}$.

Theorem 4.16 (Lévy–Cramér continuity theorem)**.** *Let P_n and P be probability measures on $\mathbb{R}$, with distribution functions F_n, F and characteristic functions Φ_n, Φ. The following are equivalent:*

(i) $F_n \Rightarrow F$.

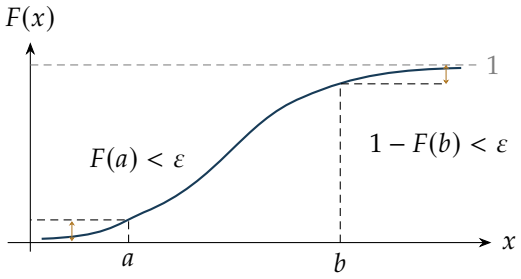
(ii) $\displaystyle\int f \, \mathrm{d}P_n \longrightarrow \int f \, \mathrm{d}P$ *for every* $f \in C_b(\mathbb{R})$.

(iii) $\Phi_n(t) \longrightarrow \Phi(t)$ *for every* $t \in \mathbb{R}$.

Equivalently, in (i) *one may require* $P_n([a,b]) \to P([a,b])$ *for every finite* $a < b$ *such that* $P(\{a\}) = P(\{b\}) = 0$.

We prove the first two implications directly, then establish the selection and tightness facts needed for the converse.

Proof of (i) $\Rightarrow$ (ii) *and* (ii) $\Rightarrow$ (iii). Fix $f \in C_b(\mathbb{R})$ and put $M = \|f\|_\infty$. Given $\varepsilon > 0$, choose continuity points $a < b$ of F with $F(a) < \varepsilon$ and $1 - F(b) < \varepsilon$. For all sufficiently large n, $F_n(a) < 2\varepsilon$ and $1 - F_n(b) < 2\varepsilon$. Thus restricting the integrals to $(a, b]$ incurs errors at most $2M\varepsilon$ for P and $4M\varepsilon$ for P_n.



On $[a, b]$, f is uniformly continuous. Choose a finite partition $a = a_0 < a_1 < \cdots < a_N = b$ consisting of continuity points of F, sufficiently fine that the oscillation of f on each $[a_{j-1}, a_j]$ is less than ε. Set

$$T_n = \sum_{j=1}^{N} f(a_j)[F_n(a_j) - F_n(a_{j-1})], \qquad T = \sum_{j=1}^{N} f(a_j)[F(a_j) - F(a_{j-1})].$$

The mesh error is at most ε under either measure. Hence

$$\left|\int f\,\mathrm{d}P_n - T_n\right| \le \varepsilon + 4M\varepsilon, \qquad \left|\int f\,\mathrm{d}P - T\right| \le \varepsilon + 2M\varepsilon.$$

Moreover,

$$|T_n - T| \le M \sum_{j=1}^{N} \bigl(|F_n(a_j) - F(a_j)| + |F_n(a_{j-1}) - F(a_{j-1})|\bigr) \to 0.$$

Taking the upper limit and then letting $\varepsilon \downarrow 0$ proves (ii). Applying (ii) to $\cos(tx)$ and $\sin(tx)$ proves (iii). □

The interval formulation has the same endpoint qualification. From (i), probabilities of $(a, b]$ converge by subtraction. For a continuity point a, $P_n(\{a\}) \le F_n(a) - F_n(a - \delta)$, where $a - \delta$ may be chosen a continuity point; the upper limit tends to zero as $\delta \downarrow 0$. This gives the assertion for $[a, b]$. Conversely, choose atom-free endpoints $a < x < b$ so that $P([a, b]) > 1 - \varepsilon$. The assumed interval convergence controls the mass outside $[a, b]$, and convergence on $[a, x]$ bounds $F_n(x)$ within that tail error of $F(x)$. Letting $\varepsilon \downarrow 0$ gives (i) at every continuity point x.

Lemma 4.17 (Helly selection)**.** *Every sequence of distribution functions (F_n) has a subsequence (F_{n_k}) and a nondecreasing right-continuous function $G : \mathbb{R} \to [0, 1]$ such that*

$$F_{n_k}(x) \longrightarrow G(x) \quad \textit{at every continuity point of } G.$$

The function G need not be a distribution function.

Proof. Enumerate the rationals as $r_1, r_2, \ldots$. The sequence $F_n(r_1)$ lies in $[0, 1]$, so choose a convergent subsequence. From it choose a subsequence converging at r_2, and continue. The diagonal subsequence converges at every rational. Write

$$G_0(r) = \lim_{k\to\infty} F_{n_k}(r), \qquad r \in \mathbb{Q}, \qquad G(x) = \inf\{G_0(r) : r > x,\ r \in \mathbb{Q}\}.$$

The rational limit G_0 is nondecreasing, and so is G. To check right continuity, choose a rational $r > x$ with $G_0(r) < G(x) + \eta$. For $x < y < r$ we have $G(x) \le G(y) \le G_0(r)$; let $y \downarrow x$ and then $\eta \downarrow 0$.

For rationals $u < x < v$, monotonicity gives

$$G_0(u) \le \liminf_k F_{n_k}(x) \le \limsup_k F_{n_k}(x) \le G_0(v).$$

The infimum of the right bounds is $G(x)$. The supremum of the left bounds is $G(x-)$: $G_0(u) \le G(u) \le G(x-)$, whereas for any $y < x$ a rational $u \in (y, x)$ gives $G(y) \le G_0(u)$. At a continuity point these bounds coincide, proving the assertion. □

Example 4.18 (Loss of mass at infinity)**.** For the deterministic variables $X_n = n$,

$$F_n(x) = \begin{cases} 0, & x < n, \\ 1, & x \geq n, \end{cases} \qquad F_n(x) \longrightarrow 0 \quad \text{for every finite } x.$$

The limit $G \equiv 0$ is nondecreasing and right-continuous, but is not a distribution function. A Helly limit is a distribution function precisely when $G(-\infty) = 0$ and $G(+\infty) = 1$.

Definition 4.19. A sequence (P_n) of probability measures on $\mathbb{R}$ is *tight* if for every $\varepsilon > 0$ there exists $K < \infty$ such that

$$P_n([-K, K]) > 1 - \varepsilon \qquad \text{for every } n.$$

Lemma 4.20. *If $\Phi_n(t) \to \Phi(t)$ for every real t, where Φ is the characteristic function of a probability measure, then (P_n) is tight.*

Proof. For $\delta > 0$, Fubini on a finite interval gives

$$\begin{aligned} A_n(\delta) &:= \frac{1}{2\delta} \int_{-\delta}^{\delta} (1 - \operatorname{Re} \Phi_n(t)) \, \mathrm{d}t \\ &= \int_{\mathbb{R}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right) \mathrm{d}P_n(x), \end{aligned}$$

with the quotient equal to 1 at $x = 0$. The integrand is nonnegative; if $|x| \geq 2/\delta$, it is at least $1/2$. Therefore

$$P_n(|x| \geq 2/\delta) \leq 2A_n(\delta).$$

For fixed δ, bounded convergence in t gives

$$A_n(\delta) \longrightarrow A(\delta) = \frac{1}{2\delta} \int_{-\delta}^{\delta} (1 - \operatorname{Re} \Phi(t)) \, \mathrm{d}t.$$

Continuity of Φ at zero and $\Phi(0) = 1$ imply $A(\delta) \to 0$ as $\delta \downarrow 0$. Choose δ so that $A(\delta) < \varepsilon/8$. For all sufficiently large n, $A_n(\delta) < \varepsilon/4$, giving the required tail bound. Enlarge $2/\delta$ to control the finitely many remaining probability measures as well. □

Completion of the continuity theorem. Assume (iii). Apply Helly selection to any subsequence to obtain a further subsequence converging to G at its continuity points. By the preceding lemma the original laws, and hence this subsequence, are tight. Given $\varepsilon > 0$, choose K so that the mass of $[-K, K]$ is greater than $1 - \varepsilon$ for every member. At continuity points $a < -K$ and $b > K$ of G, pass to the limit to get $G(a) \leq \varepsilon$ and $G(b) \geq 1 - \varepsilon$. Monotonicity and arbitrary ε show $G(-\infty) = 0$ and $G(+\infty) = 1$. Thus G is a distribution function.

The already proved implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) show that along this further subsequence the characteristic functions tend to Φ_G. They also tend to Φ, so uniqueness gives $G = F$.

Finally, if the whole sequence failed to converge at some continuity point x of F, there would be $\eta > 0$ and a subsequence (F_{n_k}) with $|F_{n_k}(x) - F(x)| \geq \eta$ for all k. The argument just given produces a further subsequence converging weakly to F, hence converging at x, a contradiction. This proves (i). $\square$

In Example 4.14 the target is the Poisson law with parameter λ. The continuity theorem therefore proves $S_n \xrightarrow{d} W$.

Lemma 4.21 (Converging together; Slutsky)**.** *Suppose $X_n \xrightarrow{d} X$ and $Y_n - X_n \xrightarrow{P} 0$, with X_n, Y_n defined on the same probability space for each n. Then $Y_n \xrightarrow{d} X$. In particular, if $Z_n \xrightarrow{P} c$ for a constant c, then $X_n + Z_n \xrightarrow{d} X + c$.*

Proof. For each $t \in \mathbb{R}$ and $\eta > 0$,

$$\begin{aligned} |\Phi_{Y_n}(t) - \Phi_{X_n}(t)| &\leq \mathbb{E}|\mathrm{e}^{\mathrm{i}t(Y_n - X_n)} - 1| \\ &\leq |t|\eta + 2P(|Y_n - X_n| > \eta). \end{aligned}$$

Let $n \to \infty$, then $\eta \downarrow 0$. The characteristic functions of Y_n consequently have the same limit Φ_X as those of X_n; the continuity theorem proves the first assertion. For the second, $\Phi_{X_n + c}(t) = \mathrm{e}^{\mathrm{i}tc}\Phi_{X_n}(t) \to \Phi_{X+c}(t)$, and $(X_n + Z_n) - (X_n + c) = Z_n - c \xrightarrow{P} 0$. $\square$

Problems

In Problems 4.3–4.9, you may cite the statements of other problems without proving them.

Problem 4.1. Give an example of a random variable X that has a bounded probability density function and yet its characteristic function satisfies

$$\int_{\mathbb{R}} |\Phi_X(t)| \, \mathrm{d}t = \infty.$$

Give another example of a random variable X such that Φ_X is not differentiable at $t = 0$.

Problem 4.2. Let X and X' be iid random variables, and let $Z = X - X'$.

(i) Show that the characteristic function of Z is nonnegative and real valued.

(ii) Show that there do not exist $a < b$ such that $Z \sim \text{Uniform}(a, b)$.

Problem 4.3. Let $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$, and X be random variables on a probability space $(\Omega, \mathcal{F}, P)$, and let $c \in \mathbb{R}$. Suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$.

(i) Show that $\overline{C} = \{x : F_X \text{ is discontinuous at } x\}$ is countable.

(ii) Deduce that $X_n + Y_n \xrightarrow{d} X + c$. Further, show that if $(Z_n)_{n\geq 1}$ and Z are random variables on the same probability space and $Z_n - X_n \xrightarrow{d} 0$, then

$$Z_n \xrightarrow{d} Z \quad \Longleftrightarrow \quad X_n \xrightarrow{d} Z.$$

(iii) Prove that $X_n Y_n \xrightarrow{d} cX$.

Problem 4.4. For $n \geq 2$, let $U_1, \ldots, U_n$ be iid Uniform$(-1, 1)$ random variables, and set $S_n = \sum_{j=1}^{n} U_j$. Prove that S_n has probability density function

$$f_{S_n}(x) = \frac{1}{\pi} \int_0^\infty \cos(tx) \left(\frac{\sin t}{t}\right)^n \mathrm{d}t,$$

and deduce that

$$\int_0^\infty \cos(tx) \left(\frac{\sin t}{t}\right)^n \mathrm{d}t = 0 \qquad \text{for every } x > n \geq 2.$$

Hint. Use the characteristic function of a uniform variable: $\Phi_U(t) = \sin(t)/t$ for $U \sim$ Uniform$(-1, 1)$, with $\Phi_U(0) = 1$.

Problem 4.5. Let $(X_j)_{j\geq 1}$ be iid random variables and let $S_n = \sum_{j=1}^n X_j$.

(i) Suppose X_j has Cauchy density

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Show that its characteristic function is $\Phi(t) = \mathrm{e}^{-|t|}$, and use this characteristic function to deduce that $S_n/n \stackrel{d}{=} X_1$ for every $n \geq 1$.

(ii) Suppose X_j is standard normal. Prove that $S_n/\sqrt{n} \stackrel{d}{=} X_1$, and that for every $x > 0$,

$$(x^{-1} - x^{-3})\mathrm{e}^{-x^2/2} \leq \int_x^\infty \mathrm{e}^{-t^2/2}\,\mathrm{d}t \leq x^{-1}\mathrm{e}^{-x^2/2}.$$

Deduce that, for some absolute constant $C > 0$ and every sufficiently large n,

$$P\left(|S_n| \geq 2\sqrt{n\log n}\right) \leq Cn^{-2}.$$

Problem 4.6. Let $(X_j)_{j\geq 1}$ be iid standard normal random variables, let $M_n = \max_{1\leq j\leq n} X_j$, and for $n \geq 2$ choose a_n so that $F_{X_1}(a_n) = (n-1)/n$.

(a) Show that for every $x \in \mathbb{R}$,

$$\lim_{t\to\infty} \frac{1 - F_{X_1}(t + x/t)}{1 - F_{X_1}(t)} = \mathrm{e}^{-x}.$$

You may use the Gaussian tail bounds in Problem 4.5(ii).

(b) Prove that $a_n(M_n - a_n) \xrightarrow{d} M$, where $F_M(x) = \exp(-\mathrm{e}^{-x})$, and that

$$\lim_{n\to\infty} \frac{a_n}{\sqrt{2\log n}} = 1.$$

Deduce that $M_n/\sqrt{2\log n} \xrightarrow{P} 1$.

Problem 4.7. Let $(X_k)_{k\geq 1}$ be iid random variables with distribution function F_X, and let $M_n = \max_{1\leq k\leq n} X_k$. Establish that $(M_n - a_n)/b_n \xrightarrow{d} M$, with the specified distribution function F_M, in each case below.

(a) $F_X(x) = 1 - \mathrm{e}^{-x}$ for $x \geq 0$, and $F_X(x) = 0$ for $x < 0$. Here $a_n = \log n$, $b_n = 1$, and

$$F_M(x) = \exp(-\mathrm{e}^{-x}), \qquad x \in \mathbb{R}.$$

(b) $F_X(x) = 1 - x^{-\alpha}$ for $x \geq 1$, where $\alpha > 0$, and $F_X(x) = 0$ for $x < 1$. Here $a_n = 0$, $b_n = n^{1/\alpha}$, and

$$F_M(x) = \begin{cases} \exp(-x^{-\alpha}), & x > 0, \\ 0, & x \leq 0. \end{cases}$$

(c) $F_X(x) = 1 - |x|^\alpha$ for $-1 \le x \le 0$, where $\alpha > 0$, with $F_X(x) = 0$ for $x < -1$ and $F_X(x) = 1$ for $x > 0$. Here $a_n = 0$, $b_n = n^{-1/\alpha}$, and

$$F_M(x) = \begin{cases} \exp(-|x|^\alpha), & x \le 0, \\ 1, & x > 0. \end{cases}$$

Problem 4.8. (i) Let X_n, Y_n be a pair of independent random variables for each $n \ge 1$, and let X, Y be independent random variables such that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$. Prove that $X_n + Y_n \xrightarrow{d} X + Y$.

(ii) Let X, Y be $[0,1]$-valued random variables such that $\mathbb{E}X^n = \mathbb{E}Y^n$ for every integer $n \ge 0$. Show first that $\mathbb{E}p(X) = \mathbb{E}p(Y)$ for every polynomial p. Then show that $\mathbb{E}f(X) = \mathbb{E}f(Y)$ for every continuous function $f : [0,1] \to \mathbb{R}$, and conclude that $X \overset{d}{=} Y$.
Hint. Use the Weierstrass approximation theorem: for every continuous $f : [0,1] \to \mathbb{R}$, there are polynomials p_n such that $\sup_{x\in[0,1]} |f(x) - p_n(x)| \to 0$.

Problem 4.9. (i) Show that if $X \ge 0$ and $Y \ge 0$ satisfy $\mathbb{E}e^{-tX} = \mathbb{E}e^{-tY}$ for every $t > 0$, then $X \overset{d}{=} Y$. You may use Problem 4.8(ii).

(ii) Suppose $X_n \ge 0$ and

$$g(t) := \lim_{n\to\infty} \mathbb{E}e^{-tX_n}$$

exists for every $t > 0$, with $\lim_{t\downarrow 0} g(t) = 1$. Show that the laws of the X_n are uniformly tight and that there exists a random variable $X \ge 0$ such that $X_n \xrightarrow{d} X$ and $g(t) = \mathbb{E}e^{-tX}$ for every $t > 0$.

(iii) Let $X_n = n^{-1} \sum_{j=1}^n jI_j$, where the $\{0,1\}$-valued variables I_j are independent and $P(I_j = 1) = 1/j$. Show that $X_n \xrightarrow{d} X$ for some $X \ge 0$ with

$$\mathbb{E}e^{-tX} = \exp\left\{ \int_0^1 \frac{e^{-xt} - 1}{x}\, dx \right\}, \qquad t > 0.$$

Chapter 5

Laws of Large Numbers and Convergent Series

5.1 Weak laws of large numbers

Write $S_n = \sum_{i=1}^n X_i$. A weak law of large numbers asserts that S_n/n converges in probability to a constant. Under a second-moment assumption, a variance calculation gives the stronger convergence in L^2.

Theorem 5.1 (L^2 weak law)**.** *Let $(X_i)_{i\geq 1}$ be uncorrelated random variables such that, for all i,*

$$\mathbb{E}X_i = \mu, \qquad \operatorname{Var}(X_i) \leq \sigma^2 < \infty.$$

Then

$$\frac{S_n}{n} \xrightarrow{L^2} \mu, \qquad \text{and hence} \qquad \frac{S_n}{n} \xrightarrow{P} \mu.$$

Proof. Uncorrelatedness means that $\operatorname{Cov}(X_i, X_j) = 0$ whenever $i \neq j$. Consequently,

$$\mathbb{E}\left|\frac{S_n}{n} - \mu\right|^2 = \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}(X_i) \leq \frac{\sigma^2}{n} \longrightarrow 0.$$

Chebyshev's inequality gives convergence in probability. □

Are second moments needed for convergence in probability? For independent, identically distributed variables, a finite first absolute moment is enough. In fact, the convergence then holds in L^1.

Theorem 5.2 (L^1 weak law)**.** *Let $(X_i)_{i\geq 1}$ be iid with $\mathbb{E}|X_1| < \infty$, and put $\mu = \mathbb{E}X_1$. Then $S_n/n \xrightarrow{L^1} \mu$.*

Proof. Fix $M > 0$ and truncate at this fixed level:

$$X_i' = X_i \mathbf{1}_{\{|X_i| \leq M\}}, \qquad Y_i = X_i - X_i' = X_i \mathbf{1}_{\{|X_i| > M\}}.$$

Set $\mu' = \mathbb{E}X_i'$ and $\eta = \mathbb{E}Y_i$, so that $\mu = \mu' + \eta$. The triangle inequality and Cauchy–Schwarz give

$$\begin{aligned}\mathbb{E}\left|\frac{S_n}{n} - \mu\right| &\leq \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n X_i' - \mu'\right| + \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n Y_i - \eta\right| \\ &\leq \left(\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n X_i' - \mu'\right|^2\right)^{1/2} + 2\mathbb{E}|Y_1|.\end{aligned}$$

For fixed M, the bounded iid variables X_i' satisfy Theorem 5.1, so the first term tends to zero as $n \to \infty$. Moreover,

$$|X_1|\mathbf{1}_{\{|X_1|\le M\}} \uparrow |X_1| \quad \text{as } M \uparrow \infty.$$

Monotone convergence therefore yields

$$\mathbb{E}|Y_1| = \mathbb{E}|X_1| - \mathbb{E}\big[|X_1|\mathbf{1}_{\{|X_1|\le M\}}\big] \longrightarrow 0.$$

Taking first the limit superior in n and then the limit in M proves $\mathbb{E}|S_n/n - \mu| \to 0$. □

5.2 The Borel–Cantelli lemmas

For events $(A_n)_{n\ge1}$, the event that infinitely many of them occur is

$$\{A_n \text{ infinitely often}\} = \limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n.$$

We abbreviate "infinitely often" by "i.o.".

Lemma 5.3 (Borel–Cantelli I)**.** *If $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$.*

Proof. The nonnegative random variable

$$X(\omega) = \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega)$$

counts the number of events that occur. By monotone convergence,

$$\mathbb{E}X = \sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty.$$

Thus $X < \infty$ almost surely. Since $\{A_n \text{ i.o.}\} = \{X = \infty\}$, the conclusion follows. □

Example 5.4. On $[0,1]$ with the uniform probability measure, take $A_n = (0, 1/n]$. Then $\mathbb{P}(A_n) = 1/n$ and $\sum_n \mathbb{P}(A_n) = \infty$, but the decreasing events have empty intersection. Hence $\mathbb{P}(A_n \text{ i.o.}) = 0$. The converse to the first lemma needs an additional assumption.

Lemma 5.5 (Borel–Cantelli II)**.** *Suppose that $(A_n)_{n\ge1}$ are mutually independent events and*

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

Then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. For fixed $m \le n$, independence and $1 - x \le \mathrm{e}^{-x}$ give

$$\mathbb{P}\left(\bigcap_{k=m}^{n} A_k^c\right) = \prod_{k=m}^{n} \bigl(1 - \mathbb{P}(A_k)\bigr) \le \exp\left(-\sum_{k=m}^{n} \mathbb{P}(A_k)\right).$$

Letting $n \to \infty$ and using continuity from above,

$$\mathbb{P}\left(\bigcap_{k=m}^{\infty} A_k^c\right) = 0, \qquad \mathbb{P}\left(\bigcup_{k=m}^{\infty} A_k\right) = 1.$$

Intersecting these probability-one tail unions over all positive integers m proves the assertion. □

5.3 A first strong law

Recall the role of Kolmogorov's consistency, or extension, theorem: consistent finite-dimensional probability laws determine a unique law on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}})$. In particular, the finite product laws of independent coordinates give the infinite product law. This permits an infinite iid sequence with a prescribed common distribution.

A strong law asks for almost-sure convergence of S_n/n. A fourth moment makes a direct proof possible.

Theorem 5.6 (Strong law under a fourth-moment assumption)**.** *Let $(X_i)_{i\ge1}$ be iid with $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^4 \le K < \infty$. Then $S_n/n \to 0$ almost surely.*

Proof. Expand the fourth power of S_n. Every term in which some index appears exactly once has expectation zero, by centering and independence. The remaining terms give

$$\begin{aligned}\mathbb{E}S_n^4 &= \sum_{i=1}^{n} \mathbb{E}X_i^4 + 6 \sum_{1\le i<j\le n} \mathbb{E}X_i^2\,\mathbb{E}X_j^2 \\ &\le nK + 3n(n-1)K \le 3n^2K.\end{aligned}$$

Here $(\mathbb{E}X_i^2)^2 \le \mathbb{E}X_i^4 \le K$ by Cauchy–Schwarz. For every $\varepsilon > 0$, Markov's inequality now gives

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) = \mathbb{P}\bigl(S_n^4 > \varepsilon^4 n^4\bigr) \le \frac{3K}{\varepsilon^4 n^2}.$$

These probabilities are summable in n. By Borel–Cantelli I, for each $\varepsilon > 0$ the inequality $|S_n/n| > \varepsilon$ holds only finitely often, almost surely. Apply this simultaneously to $\varepsilon = 1/j$, $j \ge 1$, by taking a countable intersection of probability-one events. On that intersection $S_n/n \to 0$. □

5.4 A maximal inequality and the one-series theorem

Chebyshev's inequality controls a single partial sum. Kolmogorov's maximal inequality controls all partial sums up to a given time with the same variance bound.

Lemma 5.7 (Kolmogorov's maximal inequality)**.** *Let $X_1, \dots, X_n$ be independent, centered random variables with finite variances, and let $S_k = \sum_{i=1}^{k} X_i$. For every $a > 0$,*

$$\mathbb{P}\left(\max_{1 \le k \le n} |S_k| \ge a \right) \le \frac{\operatorname{Var}(S_n)}{a^2} = \frac{1}{a^2} \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

Proof. Separate the event according to the first crossing of the level a:

$$E_k = \{|S_k| \ge a\} \cap \bigcap_{i=1}^{k-1} \{|S_i| < a\}, \qquad 1 \le k \le n.$$

The empty intersection for $k = 1$ is the whole sample space. These events are disjoint, and

$$\left\{ \max_{1 \le k \le n} |S_k| \ge a \right\} = \bigcup_{k=1}^{n} E_k.$$

On E_k, we have $S_k^2 \ge a^2$, hence $a^2 \mathbb{P}(E_k) \le \mathbb{E}[S_k^2 \mathbf{1}_{E_k}]$. The variable $S_n - S_k$ depends only on $X_{k+1}, \dots, X_n$, whereas $S_k \mathbf{1}_{E_k}$ is measurable with respect to $\sigma(X_1, \dots, X_k)$. The two are independent and $\mathbb{E}(S_n - S_k) = 0$. Thus

$$\mathbb{E}\big[(S_n - S_k) S_k \mathbf{1}_{E_k}\big] = \mathbb{E}(S_n - S_k)\, \mathbb{E}[S_k \mathbf{1}_{E_k}] = 0.$$

Expanding the square, we obtain

$$\begin{aligned} \mathbb{E}[S_n^2 \mathbf{1}_{E_k}] &= \mathbb{E}[S_k^2 \mathbf{1}_{E_k}] + \mathbb{E}[(S_n - S_k)^2 \mathbf{1}_{E_k}] + 2\mathbb{E}[(S_n - S_k) S_k \mathbf{1}_{E_k}] \\ &\ge \mathbb{E}[S_k^2 \mathbf{1}_{E_k}] \ge a^2 \mathbb{P}(E_k). \end{aligned}$$

Finally, disjointness gives $\sum_{k=1}^{n} \mathbf{1}_{E_k} \le 1$, so

$$a^2 \mathbb{P}\left(\max_{1 \le k \le n} |S_k| \ge a \right) = a^2 \sum_{k=1}^{n} \mathbb{P}(E_k) \le \mathbb{E}\left[S_n^2 \sum_{k=1}^{n} \mathbf{1}_{E_k} \right] \le \mathbb{E} S_n^2 = \operatorname{Var}(S_n).$$

□

Theorem 5.8 (Kolmogorov's one-series theorem)**.** *Let $(X_n)_{n \ge 1}$ be independent, centered random variables. If*

$$\sum_{n=1}^{\infty} \operatorname{Var}(X_n) < \infty,$$

then $\sum_{n=1}^{\infty} X_n$ converges almost surely to a finite limit.

Proof. For integers $N > n$, apply the maximal inequality to $X_{n+1}, \dots, X_N$. For every $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{n \le k \le N} |S_k - S_n| > \varepsilon \right) \le \frac{1}{\varepsilon^2} \sum_{i=n+1}^{N} \operatorname{Var}(X_i).$$

Letting $N \to \infty$ by continuity from below gives

$$\mathbb{P}\left(\sup_{k \ge n} |S_k - S_n| > \varepsilon \right) \le \frac{1}{\varepsilon^2} \sum_{i=n+1}^{\infty} \operatorname{Var}(X_i) \longrightarrow 0 \qquad (n \to \infty). \tag{5.1}$$

To obtain almost-sure convergence, consider the tail oscillation

$$R_n = \sup_{m,k \ge n} |S_m - S_k|.$$

It is a measurable, possibly extended-valued random variable, and the triangle inequality gives

$$R_n \le 2 \sup_{k \ge n} |S_k - S_n|.$$

Therefore $\mathbb{P}(R_n > 2\varepsilon) \to 0$ by (5.1). For every ω, the sequence $R_n(\omega)$ is nonincreasing, so its limit $R_\infty(\omega)$ exists in $[0, \infty]$. For every n,

$$\mathbb{P}(R_\infty > 2\varepsilon) \le \mathbb{P}(R_n > 2\varepsilon) \longrightarrow 0.$$

Taking $\varepsilon = 1/j$ for all positive integers j shows that $R_\infty = 0$ almost surely. Thus $(S_n(\omega))$ is a Cauchy sequence for almost every ω. Completeness of $\mathbb{R}$ gives a finite limit, which is measurable as a limit of measurable partial sums on their convergence event. □

5.5 The strong law under a first-moment assumption

Theorem 5.9 (Strong law of large numbers)**.** *Let $(X_n)_{n \ge 1}$ be iid with $\mathbb{E}|X_1| < \infty$. Then*

$$\frac{1}{n} \sum_{k=1}^{n} X_k \longrightarrow \mathbb{E}X_1 \qquad \textit{almost surely.}$$

In particular, for centered variables $S_n/n \to 0$ almost surely.

Two deterministic lemmas connect convergence of a series to convergence of averages.

Lemma 5.10 (Weighted Cesàro lemma)**.** *Let $\beta_n > 0$ be nondecreasing with $\beta_n \to \infty$, and set $\beta_0 = 0$. If $\alpha_n \to \alpha$ in $\mathbb{R}$, then*

$$\frac{1}{\beta_n} \sum_{k=1}^{n} (\beta_k - \beta_{k-1}) \alpha_k \longrightarrow \alpha.$$

Proof. Fix $\varepsilon > 0$. Choose N so that $\alpha - \varepsilon \le \alpha_k \le \alpha + \varepsilon$ for $k > N$. All weights $\beta_k - \beta_{k-1}$ are nonnegative. Put $C_N = \sum_{k=1}^{N} (\beta_k - \beta_{k-1})\alpha_k$. For $n > N$, splitting the sum at N gives both bounds

$$\frac{C_N}{\beta_n} + \frac{\beta_n - \beta_N}{\beta_n}(\alpha - \varepsilon) \le \frac{1}{\beta_n} \sum_{k=1}^{n} (\beta_k - \beta_{k-1})\alpha_k,$$

$$\frac{1}{\beta_n} \sum_{k=1}^{n} (\beta_k - \beta_{k-1})\alpha_k \le \frac{C_N}{\beta_n} + \frac{\beta_n - \beta_N}{\beta_n}(\alpha + \varepsilon).$$

Since $C_N/\beta_n \to 0$ and $(\beta_n - \beta_N)/\beta_n \to 1$, the limit inferior is at least $\alpha - \varepsilon$ and the limit superior is at most $\alpha + \varepsilon$. Let $\varepsilon \downarrow 0$. The ordinary Cesàro lemma is the case $\beta_n = n$. □

Lemma 5.11 (Kronecker's lemma)**.** *Let $(x_n)_{n\ge1}$ be real numbers, and let $\beta_n > 0$ be nondecreasing with $\beta_n \to \infty$. If the signed series $\sum_{n=1}^{\infty} x_n/\beta_n$ converges to a finite real number, then*

$$\frac{1}{\beta_n} \sum_{k=1}^{n} x_k \longrightarrow 0.$$

Absolute convergence of the series is not required.

Proof. Set $\alpha_0 = \beta_0 = 0$ and

$$\alpha_n = \sum_{k=1}^{n} \frac{x_k}{\beta_k}, \qquad \alpha = \lim_{n\to\infty} \alpha_n.$$

Then $\alpha_k - \alpha_{k-1} = x_k/\beta_k$. Summation by parts gives, with $S_n = \sum_{k=1}^{n} x_k$,

$$\begin{aligned} S_n &= \sum_{k=1}^{n} \beta_k(\alpha_k - \alpha_{k-1}) \\ &= \sum_{k=1}^{n} \beta_k\alpha_k - \sum_{k=1}^{n} \beta_k\alpha_{k-1} \\ &= \beta_n\alpha_n - \sum_{k=1}^{n} (\beta_k - \beta_{k-1})\alpha_{k-1}. \end{aligned}$$

Consequently,

$$\frac{S_n}{\beta_n} = \alpha_n - \frac{1}{\beta_n} \sum_{k=1}^{n} (\beta_k - \beta_{k-1})\alpha_{k-1}.$$

Apply the weighted Cesàro lemma to the sequence (α_{k-1}), which also tends to α. The right-hand side tends to $\alpha - \alpha = 0$. □

Proof of Theorem 5.9. Replacing X_n by $X_n - \mathbb{E}X_1$, we may assume that $\mathbb{E}X_1 = 0$. Define the moving truncations and their centered, scaled versions by

$$Y_n = X_n \mathbf{1}_{\{|X_n| \le n\}}, \qquad Z_n = \frac{Y_n - \mathbb{E}Y_n}{n}.$$

The variables Z_n are independent and centered. We shall prove that

$$\sum_{n=1}^{\infty} \operatorname{Var}(Z_n) = \sum_{n=1}^{\infty} \frac{\operatorname{Var}(Y_n)}{n^2} \le 2\mathbb{E}|X_1| < \infty. \tag{5.2}$$

Assuming this bound for the moment, the one-series theorem shows that $\sum_n Z_n$ converges almost surely.

We next compare the original variables with their truncations. For any nonnegative random variable U, the pointwise inequalities

$$\sum_{n=1}^{\infty} \mathbf{1}_{\{U > n\}} \le U \le 1 + \sum_{n=1}^{\infty} \mathbf{1}_{\{U > n\}}$$

imply, by monotone convergence,

$$\sum_{n=1}^{\infty} \mathbb{P}(U > n) \le \mathbb{E}U \le 1 + \sum_{n=1}^{\infty} \mathbb{P}(U > n). \tag{5.3}$$

Thus, using the common law of the X_n,

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n) \le \mathbb{E}|X_1| < \infty.$$

Borel–Cantelli I gives $X_n = Y_n$ eventually almost surely. On the probability-one event where this holds and $\sum_n Z_n$ converges,

$$\sum_{n=1}^{\infty} \frac{X_n - \mathbb{E}Y_n}{n}$$

also converges: it differs from $\sum_n Z_n$ in only finitely many terms. Kronecker's lemma, with $\beta_n = n$, therefore gives

$$\frac{1}{n} \sum_{k=1}^{n} (X_k - \mathbb{E}Y_k) \longrightarrow 0 \qquad \text{almost surely.}$$

On the other hand, identical distributions and dominated convergence yield

$$\mathbb{E}Y_k = \mathbb{E}\big[X_1 \mathbf{1}_{\{|X_1| \le k\}}\big] \longrightarrow \mathbb{E}X_1 = 0.$$

The ordinary Cesàro lemma then gives $n^{-1} \sum_{k=1}^{n} \mathbb{E}Y_k \to 0$. Adding the two limits proves the strong law once (5.2) is established.

To verify that bound, use $\operatorname{Var}(Y_n) \le \mathbb{E}Y_n^2$ and interchange the nonnegative sum and expectation:

$$\sum_{n=1}^{\infty} \frac{\operatorname{Var}(Y_n)}{n^2} \le \sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E}\left[X_1^2 \mathbf{1}_{\{|X_1| \le n\}}\right]$$
$$= \mathbb{E}\left[X_1^2 \sum_{\substack{n \ge 1 \\ n \ge |X_1|}} \frac{1}{n^2}\right].$$

For $0 \le x \le 1$,

$$x^2 \sum_{n=1}^{\infty} \frac{1}{n^2} \le x^2 \left(1 + \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n}\right)\right) = 2x^2 \le 2x.$$

For $x > 1$, let $k = \lceil x \rceil \ge 2$. The reciprocal-square tail satisfies

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \le \sum_{n=k}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n}\right) = \frac{1}{k-1} \le \frac{2}{x},$$

because $x \le k \le 2(k-1)$. It follows in both cases that

$$x^2 \sum_{\substack{n \ge 1 \\ n \ge x}} \frac{1}{n^2} \le 2x.$$

Apply this pointwise with $x = |X_1|$ inside the expectation to obtain (5.2). □

5.6 The strong law with an extended expectation

Theorem 5.12 (Strong law with an extended expectation)**.** *Let $(X_n)_{n \ge 1}$ be iid random variables such that at least one of $\mathbb{E}(X_1)_+$ and $\mathbb{E}(X_1)_-$ is finite. Then the extended expectation*

$$\mathbb{E}X_1 = \mathbb{E}(X_1)_+ - \mathbb{E}(X_1)_-$$

is well defined, and

$$\frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow \mathbb{E}X_1 \qquad \text{almost surely},$$

where the limit may be $+\infty$ or $-\infty$.

Proof. First suppose that $X_1 \ge 0$ and $\mathbb{E}X_1 = \infty$. For each positive integer M, put

$$X_i^M = \min\{X_i, M\}, \qquad S_n^M = \sum_{i=1}^{n} X_i^M.$$

The variables $(X_i^M)_{i \ge 1}$ are iid and bounded, so the finite-mean strong law gives

$$\frac{S_n^M}{n} \longrightarrow \mathbb{E}X_1^M \qquad \text{almost surely.}$$

Take the intersection of these probability-one events over integer M. On the resulting common event, $S_n^M \le S_n$ for every M, n, and hence

$$\liminf_{n\to\infty} \frac{S_n}{n} \ge \lim_{n\to\infty} \frac{S_n^M}{n} = \mathbb{E}X_1^M \qquad \text{for every positive integer } M.$$

As $M \to \infty$, $X_1^M \uparrow X_1$ and monotone convergence gives $\mathbb{E}X_1^M \uparrow \infty$. Thus $S_n/n \to +\infty$ almost surely. Together with Theorem 5.9, this proves the result for every nonnegative iid sequence, whether its mean is finite or infinite.

For the general case, write

$$X_i = (X_i)_+ - (X_i)_-, \qquad \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n (X_i)_+ - \frac{1}{n}\sum_{i=1}^n (X_i)_-.$$

Apply the nonnegative result to each of the two iid sequences. On the intersection of their probability-one convergence events,

$$\frac{1}{n}\sum_{i=1}^n (X_i)_+ \longrightarrow \mathbb{E}(X_1)_+, \qquad \frac{1}{n}\sum_{i=1}^n (X_i)_- \longrightarrow \mathbb{E}(X_1)_-.$$

At least one of these limits is finite, so their difference converges to the well-defined extended value $\mathbb{E}X_1$. □

5.7 The two-series theorem

Centering allows the one-series theorem to handle variables with nonzero means.

Theorem 5.13 (Kolmogorov's two-series theorem)**.** *Let $(X_n)_{n\ge1}$ be independent random variables with*

$$\mu_n = \mathbb{E}X_n, \qquad \sigma_n^2 = \operatorname{Var}(X_n) < \infty.$$

If the signed series $\sum_{n=1}^\infty \mu_n$ converges to a finite real number and $\sum_{n=1}^\infty \sigma_n^2 < \infty$, then $\sum_{n=1}^\infty X_n$ converges almost surely to a finite limit.

Proof. Set $\overline{X}_n = X_n - \mu_n$. These variables are independent and centered, with $\operatorname{Var}(\overline{X}_n) = \sigma_n^2$. The one-series theorem therefore gives almost-sure convergence of $\sum_n \overline{X}_n$. Since $\sum_n \mu_n$ converges, the identity

$$\sum_{k=1}^n X_k = \sum_{k=1}^n \overline{X}_k + \sum_{k=1}^n \mu_k$$

proves the result. □

The series of means may converge conditionally; absolute convergence is not assumed. A full criterion using truncated variables is given by the three-series theorem, Theorem 6.8. Its proof will follow the central limit theorem.

Problems

When solving these problems, you may cite the results of other problems without proving them.

Problem 5.1 (Three weak-law formulations)**.** For $q > 0$, write $X_n \xrightarrow{L^q} X$ if $X_n, X \in L^q$ and $\mathbb{E}|X_n - X|^q \to 0$, where

$$L^q(\Omega, \mathcal{F}, \mathbb{P}) = \{Y : Y \text{ is measurable and } \|Y\|_q < \infty\}, \qquad \|Y\|_q = (\mathbb{E}|Y|^q)^{1/q}.$$

(i) Establish the following L^2 weak law. Suppose $(X_i)_{i\ge1}$ are in L^2, $\mathbb{E}X_i = \mu$, and

$$\operatorname{Cov}(X_i, X_j) \le a_{|i-j|} \qquad \text{for all } i, j,$$

where $(a_k)_{k\ge0}$ is bounded and $a_k \to 0$ as $k \to \infty$. Show that $n^{-1}\sum_{i=1}^n X_i \xrightarrow{L^2} \mu$.

(ii) Establish the following weak law. If $(X_i)_{i\ge1}$ are iid and $k\mathbb{P}(|X_1| > k) \to 0$ as $k \to \infty$, then

$$\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}\big[X_1 \mathbf{1}_{\{|X_1|\le n\}}\big] \xrightarrow{P} 0.$$

Hint. Establish a weak law for the truncated variables $X_i' = X_i\mathbf{1}_{\{|X_i|\le n\}}$, $1 \le i \le n$, using $\operatorname{Var}(X_i')/n \to 0$, and then compare $\sum_{i=1}^n X_i$ with $\sum_{i=1}^n X_i'$.

(iii) Let $(X_i)_{i\ge1}$ be iid with law

$$\mathbb{P}\big(X_1 = (-1)^k k\big) = \frac{1}{c_0 k^2 \log k}, \qquad k = 2, 3, \ldots, \qquad c_0 = \sum_{k=2}^{\infty} \frac{1}{k^2 \log k}.$$

Prove that $\mathbb{E}|X_1| = \infty$, yet there is a finite real constant μ such that $n^{-1}\sum_{i=1}^n X_i \xrightarrow{P} \mu$.

Problem 5.2 (Pairwise independent event counts)**.** Let $(A_n)_{n\ge1}$ be pairwise independent events with $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$, and let $S_n = \sum_{k=1}^n \mathbf{1}_{A_k}$. Ratios involving $\mathbb{E}S_n$ are considered for sufficiently large n, when this denominator is positive.

(a) Show that $\operatorname{Var}(S_n) \le \mathbb{E}S_n$, and deduce that $S_n/\mathbb{E}S_n \xrightarrow{P} 1$.

(b) For $k \ge 1$, set

$$n_k = \inf\{n \ge 1 : \mathbb{E}S_n \ge k^2\}.$$

Show that $S_{n_k}/\mathbb{E}S_{n_k} \to 1$ almost surely as $k \to \infty$. *Hint.* Use Borel–Cantelli I.

(c) Prove that $\mathbb{E}S_{n_{k+1}}/\mathbb{E}S_{n_k} \to 1$ and, using the fact that S_n is nondecreasing, deduce that $S_n/\mathbb{E}S_n \to 1$ almost surely.

Problem 5.3 (Infinite moments and strong-law obstructions)**.**

(a) Let X be a nonnegative random variable. Show that its integer part $Y = \lfloor X \rfloor$ satisfies

$$Y = \sum_{n=1}^{\infty} \mathbf{1}_{\{X \geq n\}},$$

and deduce both bounds

$$\mathbb{E}X - 1 \leq \sum_{n=1}^{\infty} \mathbb{P}(X \geq n) \leq \mathbb{E}X.$$

(b) Let $(X_n)_{n\geq 1}$ be iid real random variables with $\mathbb{E}|X_1|^\alpha = \infty$ for some $\alpha > 0$. Show that, for every $\beta > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\big(|X_n| > \beta n^{1/\alpha}\big) = \infty,$$

and deduce that $\limsup_{n\to\infty} n^{-1/\alpha}|X_n| = \infty$ almost surely.

(c) Conclude that $S_n = \sum_{k=1}^{n} X_k$ satisfies

$$\limsup_{n\to\infty} n^{-1/\alpha}|S_n| = \infty \qquad \text{almost surely.}$$

Problem 5.4 (Integrability, single terms, and maxima)**.** Let $(X_k)_{k\geq 1}$ be iid random variables taking values in $\overline{\mathbb{R}} = [-\infty, \infty]$, and let $M_n = \max_{1\leq k\leq n} X_k$.

(a) Show that

$$\mathbb{P}\big(\{|X_n| > n\} \text{ i.o.}\big) = 0 \quad \Longleftrightarrow \quad \mathbb{E}|X_1| < \infty.$$

(b) Show that $X_n/n \to 0$ almost surely if and only if $\mathbb{E}|X_1| < \infty$.

(c) Show that $M_n/n \to 0$ almost surely if and only if

$$\mathbb{E}(X_1)_+ < \infty \quad \text{and} \quad \mathbb{P}(X_1 > -\infty) > 0.$$

Further show that $M_n/n \xrightarrow{P} 0$ if and only if

$$n\mathbb{P}(X_1 > n) \longrightarrow 0 \quad \text{and} \quad \mathbb{P}(X_1 > -\infty) > 0.$$

(d) Show that $X_n/n \xrightarrow{P} 0$ if and only if $\mathbb{P}(|X_1| < \infty) = 1$.

Problem 5.5 (Kronecker's lemma and a weighted strong law)**.** Let $(X_k)_{k\ge 1}$ be integrable iid random variables with $\mathbb{E}X_k = 0$.

(a) Let (a_n) and (b_n) be real sequences with $b_n > 0$ and $b_n \uparrow \infty$. Show that convergence of the signed series $\sum_{n=1}^{\infty} a_n/b_n$ implies $b_n^{-1}\sum_{k=1}^{n} a_k \to 0$.

(b) Show that
$$\sum_{k=1}^{\infty} k^{-2}\operatorname{Var}\bigl(X_k \mathbf{1}_{\{|X_k|\le k\}}\bigr) \le 2\mathbb{E}|X_1|.$$

(c) Conclude from parts (a), (b), and Kolmogorov's one-series theorem that, for every bounded sequence (c_n) of nonrandom real constants,
$$\frac{1}{n}\sum_{k=1}^{n} c_k X_k \longrightarrow 0 \qquad \text{almost surely.}$$

Problem 5.6 (Converse strong law)**.** Let $(X_k)_{k\ge 1}$ be iid real random variables, and put $S_n = \sum_{k=1}^{n} X_k$. Show that if S_n/n converges almost surely to a finite limit as $n \to \infty$, then X_1 is necessarily integrable.

Problem 5.7 (Differentiability at zero and the weak law)**.** Let $(X_k)_{k\ge 1}$ be iid real random variables, put $S_n = \sum_{k=1}^{n} X_k$, and write $\Phi_{X_1}(t) = \mathbb{E}e^{itX_1}$.

(a) Prove that if $\Phi'_{X_1}(0) = a + ib \in \mathbb{C}$, with $a, b \in \mathbb{R}$, then $a = 0$ and $S_n/n \xrightarrow{P} b$ as $n \to \infty$.

(b) Suppose that $S_n/n \xrightarrow{P} b \in \mathbb{R}$. For every $t > 0$ and every positive sequence $h_k \downarrow 0$, set $n_k = \lfloor t/h_k \rfloor$. Prove both one-sided limits
$$\Phi_{X_1}(h_k)^{n_k} \longrightarrow e^{ibt}, \qquad \Phi_{X_1}(-h_k)^{n_k} \longrightarrow e^{-ibt}.$$
Deduce that $\Phi'_{X_1}(0) = ib$, and conclude that S_n/n converges in probability to a real constant if and only if Φ_{X_1} is differentiable at zero.

(c) Using the distribution in Problem 5.1(iii), give an example of a random variable X for which Φ_X is differentiable at zero yet $\mathbb{E}|X| = \infty$.

Chapter 6

Central Limit Theorems

6.1 The iid central limit theorem

We begin with independent, identically distributed random variables. The characteristic function of a sum then reduces to a power of a single characteristic function.

Theorem 6.1 (Central limit theorem for iid variables)**.** *Let $X_1, X_2, \dots$ be iid, with $\mathbb{E}X_1 = 0$ and $0 < \operatorname{Var}(X_1) = \sigma^2 < \infty$. Then*

$$\widehat{S}_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} X_k \xrightarrow{d} G, \qquad G \sim N(0,1).$$

Proof. With the convention $\Phi_X(t) = \mathbb{E}e^{itX}$, independence gives

$$\Phi_{\widehat{S}_n}(t) = \left[\Phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

The finite second moment permits differentiation under the expectation twice, and dominated convergence makes the second derivative continuous:

$$\Phi'_{X_1}(0) = i\mathbb{E}X_1 = 0, \qquad \Phi''_{X_1}(0) = -\mathbb{E}X_1^2 = -\sigma^2.$$

Consequently, the second-order expansion at zero is

$$\Phi_{X_1}(u) = 1 - \frac{\sigma^2 u^2}{2} + o(u^2).$$

For each fixed $t \in \mathbb{R}$ this yields

$$\Phi_{\widehat{S}_n}(t) = \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^n \longrightarrow e^{-t^2/2} = \Phi_G(t).$$

The continuity theorem, Theorem 4.16, proves the claim. □

For iid variables with mean μ and variance $\sigma^2 \in (0, \infty)$, apply the theorem to $X_k - \mu$. Thus $(\sum_{k=1}^n X_k - n\mu)/(\sigma\sqrt{n}) \xrightarrow{d} N(0,1)$.

6.2 Lindeberg's condition for triangular arrays

In a triangular array the distributions of the summands may change with both the row and the position in that row. For each $n \geq 1$, let $\widehat{X}_{n,1}, \ldots, \widehat{X}_{n,n}$ be mutually independent random variables on a probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Different rows may live on different probability spaces; no independence between rows is required. The hats indicate the normalization used for the row sums.

$$\begin{array}{ll} \text{row } 1 & \widehat{X}_{1,1} \\ \text{row } 2 & \widehat{X}_{2,1} \quad \widehat{X}_{2,2} \\ \text{row } 3 & \widehat{X}_{3,1} \quad \widehat{X}_{3,2} \quad \widehat{X}_{3,3} \\ \vdots & \qquad\qquad \ddots \\ \text{row } n & \widehat{X}_{n,1} \quad \widehat{X}_{n,2} \quad \cdots \quad \widehat{X}_{n,n} \end{array}$$

Mutual independence within each row

Assume throughout this section that the variables are centered and have finite variances. Put

$$\sigma_{n,k}^2 = \mathbb{E}\widehat{X}_{n,k}^2, \qquad \widehat{S}_n = \sum_{k=1}^{n} \widehat{X}_{n,k}, \qquad \widehat{V}_n = \sum_{k=1}^{n} \sigma_{n,k}^2 \longrightarrow 1. \tag{6.1}$$

Expectations refer to the probability space of the row in question. The convergence of the total variance in (6.1) is a separate assumption.

Theorem 6.2 (Lindeberg's central limit theorem)**.** *Under the row independence, centering and variance assumptions above, suppose that, for every $\varepsilon > 0$,*

$$g_n(\varepsilon) := \sum_{k=1}^{n} \mathbb{E}\left[\widehat{X}_{n,k}^2 \mathbf{1}_{\{|\widehat{X}_{n,k}| > \varepsilon\}}\right] \longrightarrow 0. \tag{6.2}$$

Then $\widehat{S}_n \xrightarrow{d} N(0,1)$.

Example 6.3 (The iid case)**.** For the centered iid variables of Theorem 6.1, take $\widehat{X}_{n,k} = X_k/(\sigma\sqrt{n})$. Then $\widehat{V}_n = 1$ and

$$\begin{aligned} g_n(\varepsilon) &= n\mathbb{E}\left[\widehat{X}_{n,1}^2 \mathbf{1}_{\{|\widehat{X}_{n,1}| > \varepsilon\}}\right] \\ &= \frac{1}{\sigma^2}\mathbb{E}\left[X_1^2 \mathbf{1}_{\{|X_1| > \varepsilon\sigma\sqrt{n}\}}\right] \longrightarrow 0 \end{aligned}$$

by dominated convergence. Thus the array theorem also recovers the iid central limit theorem.

Remark 6.4 (The largest variance in a row). Let $r_n^2 = \max_{1 \leq k \leq n} \sigma_{n,k}^2$. For each k and each $\varepsilon > 0$,

$$\begin{aligned} \sigma_{n,k}^2 &= \mathbb{E}\left[\widehat{X}_{n,k}^2 \mathbf{1}_{\{|\widehat{X}_{n,k}| \leq \varepsilon\}}\right] + \mathbb{E}\left[\widehat{X}_{n,k}^2 \mathbf{1}_{\{|\widehat{X}_{n,k}| > \varepsilon\}}\right] \\ &\leq \varepsilon^2 + g_n(\varepsilon). \end{aligned}$$

Hence $r_n^2 \le \varepsilon^2 + g_n(\varepsilon)$. Under Lindeberg's condition, first let $n \to \infty$ and then $\varepsilon \downarrow 0$ to obtain $r_n \to 0$. Thus the largest standard deviation in a row tends to zero.

Theorem 6.5 (Feller's converse)**.** *Under the same row independence, centering and total-variance assumption* (6.1)*, if*

$$\widehat{S}_n \xrightarrow{d} N(0,1) \quad \textit{and} \quad r_n \longrightarrow 0,$$

then Lindeberg's condition (6.2) *holds.*

We state this converse without proof. Together with Lindeberg's theorem it is the Lindeberg–Feller characterization under the maximal-variance condition. The proof that follows establishes Lindeberg's sufficient condition.

6.3 Gaussian replacement

The idea is to replace each summand by a Gaussian with the same first two moments, and to control the effect of each replacement on the expectation of a smooth function. On a product extension of the probability space for each row, choose mutually independent variables

$$Y_{n,k} \sim N(0, \sigma_{n,k}^2), \qquad 1 \le k \le n,$$

whose entire row is independent of the original row. Set

$$G_n = \sum_{k=1}^{n} Y_{n,k} \sim N(0, \widehat{V}_n).$$

Here $N(0,0)$ denotes the point mass at zero. Unlike $\widehat{S}_n$, whose summands need not have identical or Gaussian laws, G_n has an explicit distribution.

Lemma 6.6 (Quantitative replacement estimate)**.** *Let $h : \mathbb{R} \to \mathbb{R}$ have continuous, bounded second and third derivatives. For every $n \ge 1$ and every $\varepsilon > 0$,*

$$\begin{aligned}\left|\mathbb{E}h(\widehat{S}_n) - \mathbb{E}h(G_n)\right| \le \left(\frac{\varepsilon}{6} + \frac{r_n}{2}\right) \widehat{V}_n \, \|h'''\|_\infty \\ + g_n(\varepsilon) \, \|h''\|_\infty .\end{aligned} \tag{6.3}$$

No boundedness assumption on h' is needed for this estimate. Indeed, boundedness of h'' gives

$$|h'(x)| \le |h'(0)| + \|h''\|_\infty |x|, \qquad |h(x)| \le |h(0)| + |h'(0)||x| + \tfrac{1}{2} \|h''\|_\infty x^2,$$

so all the expectations in the lemma are finite.

Proof of Theorem 6.2 from the lemma. Fix h with continuous, bounded first three derivatives. Under Lindeberg's condition, $r_n \to 0$; also $\widehat{V}_n \to 1$ by assumption. For each fixed $\varepsilon > 0$, the lemma gives

$$\limsup_{n\to\infty} \left|\mathbb{E}h(\widehat{S}_n) - \mathbb{E}h(G_n)\right| \le \frac{\varepsilon}{6} \, \|h'''\|_\infty .$$

Letting $\varepsilon \downarrow 0$ proves that this difference tends to zero. If $G \sim N(0,1)$, then G_n has the law of $\sqrt{\widehat{V}_n}\, G$. By the mean value estimate,

$$|h(\sqrt{\widehat{V}_n}\, x) - h(x)| \le \|h'\|_\infty \, |\sqrt{\widehat{V}_n} - 1| \, |x|.$$

Since $\mathbb{E}|G| < \infty$, it follows that

$$\mathbb{E}h(\widehat{S}_n) \longrightarrow \mathbb{E}h(G). \tag{6.4}$$

To pass to distribution functions, fix $b \in \mathbb{R}$. Choose a smooth nondecreasing function $\rho : \mathbb{R} \to [0,1]$ with $\rho(x) = 0$ for $x \le 0$ and $\rho(x) = 1$ for $x \ge 1$. Such a function is obtained by integrating and normalizing a nonnegative smooth function supported on $[0,1]$ and positive on $(0,1)$. Define, for $j \ge 1$,

$$h_j^-(x) = \rho(j(b-x)), \qquad h_j^+(x) = 1 - \rho(j(x-b)).$$

Each fixed function has bounded first three derivatives, and

$$0 \le h_j^- \uparrow \mathbf{1}_{(-\infty,b)}, \qquad 1 \ge h_j^+ \downarrow \mathbf{1}_{(-\infty,b]}.$$

The derivative bounds may depend on j. Keeping j fixed while taking the limit in n, (6.4) yields

$$\liminf_{n\to\infty} P_n(\widehat{S}_n < b) \ge \lim_{n\to\infty} \mathbb{E}h_j^-(\widehat{S}_n) = \mathbb{E}h_j^-(G),$$
$$\limsup_{n\to\infty} P_n(\widehat{S}_n \le b) \le \lim_{n\to\infty} \mathbb{E}h_j^+(\widehat{S}_n) = \mathbb{E}h_j^+(G).$$

Now let $j \to \infty$. Monotone convergence for the lower approximation and bounded convergence for the upper one give the bounds $P(G < b)$ and $P(G \le b)$, respectively. The normal distribution has no atom at b, so

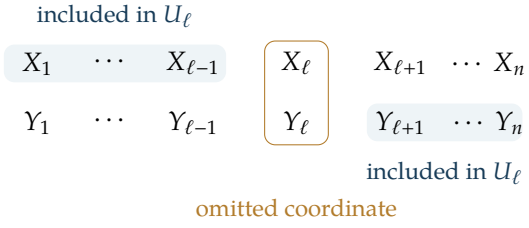
$$P_n(\widehat{S}_n \le b) \longrightarrow P(G \le b).$$

This holds for every $b \in \mathbb{R}$ and proves the theorem. □

Proof of Lemma 6.6. Fix a row n and write $X_k = \widehat{X}_{n,k}$ and $Y_k = Y_{n,k}$. For every $\ell = 1, \dots, n$, define the mixed sum with its ℓth coordinate omitted:

$$U_\ell = \sum_{k=1}^{\ell-1} X_k + \sum_{k=\ell+1}^{n} Y_k. \tag{6.5}$$

Empty sums are zero. The variable U_ℓ is independent of the pair (X_ℓ, Y_ℓ).



The endpoints and consecutive mixed sums satisfy

$$G_n = U_1 + Y_1, \qquad \widehat{S}_n = U_n + X_n, \qquad U_\ell + X_\ell = U_{\ell+1} + Y_{\ell+1} \quad (\ell < n).$$

Inserting the successive mixed sums therefore telescopes all n replacements:

$$\mathbb{E}h(G_n) - \mathbb{E}h(\widehat{S}_n) = \sum_{\ell=1}^{n} \mathbb{E}\big[h(U_\ell + Y_\ell) - h(U_\ell + X_\ell)\big],$$

$$\left|\mathbb{E}h(G_n) - \mathbb{E}h(\widehat{S}_n)\right| \le \sum_{\ell=1}^{n} \Delta_\ell,$$

where

$$\Delta_\ell = \left|\mathbb{E}\big[h(U_\ell + Y_\ell) - h(U_\ell + X_\ell)\big]\right|.$$

For example, the first insertion subtracts $h(U_1 + X_1)$ and then adds the same quantity, now written as $h(U_2 + Y_2)$. Repeating this through the nth coordinate leaves only the two endpoint sums.

Define the Taylor remainder

$$R_\ell(z) = h(U_\ell + z) - h(U_\ell) - zh'(U_\ell) - \frac{z^2}{2}h''(U_\ell).$$

Taylor's formula gives two useful bounds:

$$|R_\ell(z)| \le \frac{\|h'''\|_\infty}{6}|z|^3, \qquad |R_\ell(z)| \le \|h''\|_\infty |z|^2. \tag{6.6}$$

For the second bound, the remainder after the linear term is at most $\frac{1}{2}\|h''\|_\infty |z|^2$. The definition of R_ℓ subtracts one more quadratic term, whose absolute value is also at most $\frac{1}{2}\|h''\|_\infty |z|^2$. Their sum explains why the quadratic bound has no factor $1/2$.

Expand at U_ℓ in both cases:

$$h(U_\ell + X_\ell) = h(U_\ell) + X_\ell h'(U_\ell) + \tfrac{1}{2}X_\ell^2 h''(U_\ell) + R_\ell(X_\ell),$$
$$h(U_\ell + Y_\ell) = h(U_\ell) + Y_\ell h'(U_\ell) + \tfrac{1}{2}Y_\ell^2 h''(U_\ell) + R_\ell(Y_\ell).$$

Here $\mathbb{E}X_\ell = \mathbb{E}Y_\ell = 0$ and $\mathbb{E}X_\ell^2 = \mathbb{E}Y_\ell^2 = \sigma_{n,\ell}^2$. Subtraction gives

$$\Delta_\ell \le |\mathbb{E}[(X_\ell - Y_\ell)h'(U_\ell)]| + \frac{1}{2}\left|\mathbb{E}[(X_\ell^2 - Y_\ell^2)h''(U_\ell)]\right| + |\mathbb{E}[R_\ell(X_\ell) - R_\ell(Y_\ell)]|.$$

Independence from U_ℓ makes the first two terms zero. All products are integrable: h' has at most linear growth, U_ℓ has a finite second moment, and h'' is bounded. It follows that

$$\Delta_\ell \le \mathbb{E}|R_\ell(X_\ell)| + \mathbb{E}|R_\ell(Y_\ell)|.$$

Split the original summand at ε and use the cubic bound on the smaller values and the quadratic bound on the larger ones:

$$\begin{aligned}\mathbb{E}|R_\ell(X_\ell)| &\le \frac{\|h'''\|_\infty}{6}\mathbb{E}\left[|X_\ell|^3\mathbf{1}_{\{|X_\ell|\le\varepsilon\}}\right] + \|h''\|_\infty\,\mathbb{E}\left[X_\ell^2\mathbf{1}_{\{|X_\ell|>\varepsilon\}}\right]\\ &\le \frac{\varepsilon}{6}\,\|h'''\|_\infty\,\sigma_{n,\ell}^2 + \|h''\|_\infty\,\mathbb{E}\left[X_\ell^2\mathbf{1}_{\{|X_\ell|>\varepsilon\}}\right].\end{aligned}$$

Summing gives

$$\sum_{\ell=1}^{n}\mathbb{E}|R_\ell(X_\ell)| \le \frac{\varepsilon}{6}\,\|h'''\|_\infty\,\widehat{V}_n + \|h''\|_\infty\,g_n(\varepsilon). \tag{6.7}$$

For the Gaussian summands, recall the moments of $G\sim N(0,1)$. Writing $\varphi(x)=(2\pi)^{-1/2}\mathrm{e}^{-x^2/2}$, integration by parts using $\varphi'(x)=-x\varphi(x)$ gives

$$\mathbb{E}G^2=\int_{\mathbb{R}}\varphi(x)\,\mathrm{d}x=1,\qquad \mathbb{E}G^4=3\int_{\mathbb{R}}x^2\varphi(x)\,\mathrm{d}x=3.$$

The boundary terms vanish because a polynomial times φ tends to zero at both ends. Cauchy–Schwarz now yields $\mathbb{E}|G|^3\le(\mathbb{E}G^2\,\mathbb{E}G^4)^{1/2}=\sqrt{3}\le 3$. If $\sigma_{n,\ell}>0$, then $Y_\ell/\sigma_{n,\ell}$ is standard normal, and hence

$$\mathbb{E}|Y_\ell|^3\le 3\sigma_{n,\ell}^3\le 3r_n\sigma_{n,\ell}^2.$$

If $\sigma_{n,\ell}=0$, then $Y_\ell=0$ almost surely and the same inequality holds directly. Thus

$$\sum_{\ell=1}^{n}\mathbb{E}|R_\ell(Y_\ell)|\le\frac{\|h'''\|_\infty}{6}\sum_{\ell=1}^{n}\mathbb{E}|Y_\ell|^3\le\frac{r_n}{2}\,\|h'''\|_\infty\,\widehat{V}_n. \tag{6.8}$$

Combining (6.7) and (6.8) with the telescoping estimate proves (6.3). □

6.4 Lyapunov's condition

Corollary 6.7 (Lyapunov)**.** *Let $X_1,X_2,\dots$ be mutually independent random variables with finite variances, and put*

$$S_n=\sum_{k=1}^{n}X_k,\qquad \sigma_n^2=\operatorname{Var}(S_n).$$

Suppose $\sigma_n>0$ for all sufficiently large n and, for some $q>2$,

$$\frac{1}{\sigma_n^q}\sum_{k=1}^{n}\mathbb{E}|X_k-\mathbb{E}X_k|^q\longrightarrow 0. \tag{6.9}$$

Then the normalized array satisfies Lindeberg's condition, and

$$\frac{S_n-\mathbb{E}S_n}{\sigma_n}\xrightarrow{\;d\;}N(0,1).$$

Proof. For the rows with positive total variance, set $\widehat{X}_{n,k} = (X_k - \mathbb{E}X_k)/\sigma_n$. Their total variance is one. On $\{|\widehat{X}_{n,k}| > \varepsilon\}$, the inequality $q > 2$ gives

$$\widehat{X}_{n,k}^2 \le \varepsilon^{2-q} |\widehat{X}_{n,k}|^q.$$

Consequently,

$$g_n(\varepsilon) \le \frac{\varepsilon^{2-q}}{\sigma_n^q} \sum_{k=1}^{n} \mathbb{E}|X_k - \mathbb{E}X_k|^q \longrightarrow 0.$$

Apply Theorem 6.2. Discarding finitely many initial rows does not affect the limit. □

For $q = 2$, the expression on the left of (6.9) equals $\operatorname{Var}(S_n)/\sigma_n^2 = 1$, so it cannot tend to zero.

6.5 An application: the three-series theorem

The central limit theorem also supplies a necessity argument for convergence of independent series. Together with the one- and two-series theorems, it gives a complete criterion in terms of bounded truncations.

Theorem 6.8 (Kolmogorov's three-series theorem)**.** *Let $X_1, X_2, \dots$ be independent real random variables. The series $\sum_{n=1}^{\infty} X_n$ converges almost surely to a finite limit if and only if there exists $c > 0$ for which all three conditions hold:*

(i) $\displaystyle\sum_{n=1}^{\infty} P(|X_n| > c) < \infty;$

(ii) $\displaystyle\sum_{n=1}^{\infty} \mathbb{E}[X_n \mathbf{1}_{\{|X_n| \le c\}}]$ *converges in* $\mathbb{R}$;

(iii) $\displaystyle\sum_{n=1}^{\infty} \operatorname{Var}(X_n \mathbf{1}_{\{|X_n| \le c\}}) < \infty.$

Equivalently, the three conditions hold for every $c > 0$. In particular, checking the single cutoff $c = 1$ suffices.

Proof. Sufficiency. Fix a cutoff c satisfying the three conditions and put $X_n' = X_n \mathbf{1}_{\{|X_n| \le c\}}$. The variables X_n' are independent and bounded by c. Conditions (ii) and (iii), together with the two-series theorem (Theorem 5.13), imply that $\sum_n X_n'$ converges almost surely. Moreover,

$$\sum_{n} P(X_n \ne X_n') = \sum_{n} P(|X_n| > c) < \infty.$$

By the first Borel–Cantelli lemma, $X_n = X_n'$ eventually almost surely. On this probability-one event the two series differ in only finitely many terms, so one converges if and only if the other does. Therefore $\sum_n X_n$ converges almost surely.

Necessity. Suppose $\sum_n X_n$ converges almost surely, and fix an arbitrary $c > 0$. Then $X_n \to 0$ almost surely, so $|X_n| \le c$ eventually almost surely. The events $\{|X_n| > c\}$ are independent. The second Borel–Cantelli lemma therefore forces

$$\sum_n P(|X_n| > c) < \infty.$$

Thus (i) holds. For X'_n as above, the same eventual-equality argument shows that $\sum_n X'_n$ also converges almost surely.

It remains to prove

$$\sum_{n=1}^{\infty} \operatorname{Var}(X'_n) < \infty. \tag{6.10}$$

This will finish the proof: with $\overline{X}'_n = X'_n - \mathbb{E}X'_n$, the one-series theorem (Theorem 5.8) will imply almost-sure convergence of $\sum_n \overline{X}'_n$. Subtracting this convergent series from $\sum_n X'_n$ gives convergence in $\mathbb{R}$ of the deterministic series $\sum_n \mathbb{E}X'_n$, which is (ii).

To prove (6.10), suppose instead that $V_n := \sum_{k=1}^n \operatorname{Var}(X'_k) \longrightarrow \infty$. For all sufficiently large n, define

$$\widehat{X}_{n,k} = \frac{\overline{X}'_k}{\sqrt{V_n}}, \qquad \widehat{S}_n = \frac{1}{\sqrt{V_n}} \sum_{k=1}^n \overline{X}'_k.$$

The row variables are independent and centered, and their total variance is exactly one. Since $|X'_k| \le c$ and $|\mathbb{E}X'_k| \le c$,

$$\max_{1 \le k \le n} |\widehat{X}_{n,k}| \le \frac{2c}{\sqrt{V_n}} \longrightarrow 0.$$

For each $\varepsilon > 0$, every indicator in the Lindeberg sum is therefore zero for all sufficiently large n. Theorem 6.2 gives

$$\widehat{S}_n \xrightarrow{d} G, \qquad G \sim N(0,1).$$

On the other hand, the almost-sure convergence of $\sum_k X'_k$ gives $A_n := V_n^{-1/2} \sum_{k=1}^n X'_k \longrightarrow 0$ almost surely. The identity $\widehat{S}_n = A_n - V_n^{-1/2} \sum_{k=1}^n \mathbb{E}X'_k$ and the converging-together form of Slutsky's lemma (Lemma 4.21) imply

$$d_n := \frac{1}{\sqrt{V_n}} \sum_{k=1}^n \mathbb{E}X'_k = -\widehat{S}_n + A_n \xrightarrow{d} -G.$$

The law of $-G$ is again $N(0,1)$. But each d_n is deterministic, so $P(d_n > 0) \in \{0,1\}$. Convergence in distribution to the normal law, whose distribution function is continuous at zero, would require $P(d_n > 0) \longrightarrow P(G > 0) = \frac{1}{2}$, which is impossible. This contradiction proves (6.10), hence (iii) and then (ii). Because $c > 0$ was arbitrary, necessity holds for every cutoff, completing both formulations of the theorem. □

A related bounded-series criterion is discussed in S. R. S. Varadhan, *Probability Theory*, Courant Lecture Notes in Mathematics 7 (2001).

Problems

In these problems, a result from another problem may be used without proving it again.

Problem 6.1 (A necessary second moment)**.** Suppose that $X_1, X_2, \ldots, X_1', X_2', \ldots$ are iid and

$$n^{-1/2} \sum_{k=1}^{n} X_k \xrightarrow{d} Z$$

for a finite real-valued random variable Z.

(i) Show that $Y_k = X_k - X_k'$ satisfies

$$n^{-1/2} \sum_{k=1}^{n} Y_k \xrightarrow{d} Z - Z',$$

where Z and Z' are iid.

(ii) For $m > 0$, let

$$Y_k^{(m)} = Y_k \mathbf{1}_{\{|Y_k| \le m\}}, \qquad \widetilde{Y}_k^{(m)} = Y_k \mathbf{1}_{\{|Y_k| > m\}}.$$

Show that, for every $a \in \mathbb{R}$ and $n \ge 1$,

$$P\left(\sum_{k=1}^{n} Y_k \ge a\sqrt{n}\right) \ge P\left(\sum_{k=1}^{n} Y_k^{(m)} \ge a\sqrt{n},\ \sum_{k=1}^{n} \widetilde{Y}_k^{(m)} \ge 0\right) \ge \frac{1}{2} P\left(\sum_{k=1}^{n} Y_k^{(m)} \ge a\sqrt{n}\right).$$

(iii) Use the Portmanteau theorem and the CLT for the bounded variables $Y_k^{(m)}$ to conclude that, for every $a \in \mathbb{R}$ and $m > 0$,

$$P(Z - Z' \ge a) \ge \frac{1}{2} P\left(G \ge \frac{a}{s_m}\right), \qquad s_m^2 = \mathbb{E}[(Y_1^{(m)})^2], \quad G \sim N(0,1),$$

when $s_m > 0$. When $s_m = 0$, read the Gaussian probability as $P(s_m G \ge a)$, without division by s_m, and handle this case directly. Further deduce that $\mathbb{E}Y_1^2 < \infty$ by letting $m \to \infty$ first and then $a \to \infty$. The closed-set implication of Portmanteau that may be used here is: if $W_n \xrightarrow{d} W$ and $F \subseteq \mathbb{R}$ is closed, then $\limsup_n P(W_n \in F) \le P(W \in F)$.

(iv) Conclude that $\mathbb{E}X_1^2 < \infty$.

Problem 6.2 (Lyapunov's criterion and its limitation)**.** Let (X_k) be mutually independent random variables, and suppose $S_n = \sum_{k=1}^n X_k$ has variance $\sigma_n^2 = \operatorname{Var}(S_n) < \infty$. Normalized expressions are considered for sufficiently large n with $\sigma_n > 0$; assume such positivity in (i).

(i) Prove that, if for some $q > 2$,

$$\lim_{n\to\infty} \sigma_n^{-q} \sum_{k=1}^n \mathbb{E}|X_k - \mathbb{E}X_k|^q = 0, \tag{6.11}$$

then $(S_n - \mathbb{E}S_n)/\sigma_n \xrightarrow{d} N(0,1)$.

(ii) Show that if $\sigma_n \to \infty$ and there exist $C > 0$ and $q > 2$ such that

$$\mathbb{E}|X_k - \mathbb{E}X_k|^q \le C \operatorname{Var}(X_k) \qquad \text{for every } k,$$

then (6.11) holds.

(iii) Give an example with constants $C > 0$ and $q > 2$ such that

$$\mathbb{E}|X_k - \mathbb{E}X_k|^q \le C \operatorname{Var}(X_k)^{q/2} \qquad \text{for every } k,$$

and $\sigma_n \to \infty$, yet $(S_n - \mathbb{E}S_n)/\sigma_n$ does not converge in distribution.

Problem 6.3 (Rare jumps and central limits)**.** Let (X_k) be mutually independent random variables and let $S_n = \sum_{k=1}^n X_k$.

(i) Suppose that, for fixed $\alpha \in \mathbb{R}$ and $\beta > 1$,

$$\begin{aligned} P(X_k = k^\alpha) = P(X_k = -k^\alpha) &= \tfrac{1}{2} k^{-\beta}, \\ P(X_k = 0) &= 1 - k^{-\beta}. \end{aligned} \tag{6.12}$$

Show that S_n converges almost surely as $n \to \infty$ in two ways: first by a direct application of Borel–Cantelli, and then by Kolmogorov's three-series theorem.

(ii) Suppose (6.12) holds instead with $0 \le \beta < \min\{1, 2\alpha + 1\}$. Show that there exists a nonrandom sequence (b_n) of positive numbers such that $b_n^{-1} S_n \xrightarrow{d} Z$ for a random variable Z satisfying $0 < F_Z(z) < 1$ for some $z \in \mathbb{R}$, and compute its characteristic function $\Phi_Z(t)$.

(iii) Now suppose that

$$P(X_k = 2k) = P(X_k = -2k) = \frac{1}{2k^2}, \qquad P(X_k = 1) = P(X_k = -1) = \frac{1}{2}\left(1 - \frac{1}{k^2}\right).$$

Show that $S_n/\sqrt{n} \xrightarrow{d} N(0,1)$.

Problem 6.4 (Self-normalization and the sphere)**.** Let (X_k) be iid with $\mathbb{E}X_1 = 0$ and $\mathrm{Var}(X_1) = 1$. Put

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \qquad Y_{n,k} = \begin{cases} X_k/\sqrt{Q_n}, & Q_n > 0, \\ 0, & Q_n = 0, \end{cases} \qquad 1 \le k \le n.$$

(i) Show that $Y_{n,1} \xrightarrow{d} X_1$.

(ii) Show that $n^{-1/2}\sum_{k=1}^{n} Y_{n,k} \xrightarrow{d} N(0,1)$.

(iii) Show that if $X_1 \sim N(0,1)$, then the vector $(Y_{n,1}, \dots, Y_{n,n})$ is uniformly distributed on

$$S = \{x \in \mathbb{R}^n : |x| = \sqrt{n}\},$$

where $|x|$ is the Euclidean norm. Here the uniform law means the unique probability law on S invariant under orthogonal transformations. Interpret both (i) and (ii) in this special case.

Problem 6.5 (A characteristic-function proof of Lindeberg's theorem)**.** For each $n \ge 1$, let $X_{n,1}, \dots, X_{n,n}$ be mutually independent random variables with $\mathbb{E}X_{n,k} = 0$ and $\mathbb{E}X_{n,k}^2 = \sigma_{n,k}^2 < \infty$. Let

$$S_n = \sum_{k=1}^{n} X_{n,k}, \qquad v_n = \mathrm{Var}(S_n) = \sum_{k=1}^{n} \sigma_{n,k}^2 \longrightarrow 1,$$

and define

$$g_n(\varepsilon) = \sum_{k=1}^{n} \mathbb{E}[X_{n,k}^2 \mathbf{1}_{\{|X_{n,k}|>\varepsilon\}}], \qquad \varepsilon > 0.$$

(i) For each $t \in \mathbb{R}$, show that

$$\Phi_{S_n}(t) = \prod_{k=1}^{n} (1 + a_{n,k}(t)), \qquad a_{n,k}(t) = \Phi_{X_{n,k}}(t) - 1.$$

Use the Taylor estimates in Problem 6.6 to show further that $|a_{n,k}(t)| \le 2t^2\sigma_{n,k}^2$.

(ii) Set $a_n(t) = \sum_{k=1}^{n} a_{n,k}(t)$. For every $t \in \mathbb{R}$ and $\varepsilon > 0$, prove the complete error bound

$$\begin{aligned} \left|a_n(t) + \tfrac{1}{2}t^2 v_n\right| &\le \sum_{k=1}^{n} \left|a_{n,k}(t) + \tfrac{1}{2}t^2\sigma_{n,k}^2\right| \\ &\le t^2 g_n(\varepsilon) + \frac{|t|^3}{6}\varepsilon v_n. \end{aligned}$$

(iii) Assume that $g_n(\varepsilon) \to 0$ for every $\varepsilon > 0$. Recall why this implies

$$r_n^2 := \max_{1\le k\le n} \sigma_{n,k}^2 \longrightarrow 0.$$

Deduce that, for each $t \in \mathbb{R}$,

$$a_n(t) \longrightarrow -\frac{t^2}{2}, \qquad \eta_n(t) := \sum_{k=1}^{n} |a_{n,k}(t)|^2 \longrightarrow 0.$$

Apply the following product-limit fact and the continuity theorem to conclude that $S_n \xrightarrow{d} N(0,1)$: for complex numbers $z_{n,k}$, if

$$\sum_{k=1}^{n} z_{n,k} \longrightarrow z, \qquad \sum_{k=1}^{n} |z_{n,k}|^2 \longrightarrow 0, \quad \text{then} \quad \prod_{k=1}^{n} (1 + z_{n,k}) \longrightarrow \mathrm{e}^{z}.$$

Problem 6.6 (Taylor remainders and a central limit theorem)**.** For $x \in \mathbb{R}$ and an integer $n \geq 0$, define

$$f_n(x) = \mathrm{e}^{\mathrm{i}x} - \sum_{k=0}^{n} \frac{(\mathrm{i}x)^k}{k!}.$$

Use the convention $x^0 = 1$, including at $x = 0$.

(a) Show that, for $n \geq 1$,

$$f_n(x) = \int_0^x \mathrm{i} f_{n-1}(y)\, \mathrm{d}y.$$

Deduce by induction on n that

$$|f_n(x)| \leq \min\left\{ \frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!} \right\} \qquad (x \in \mathbb{R},\ n \geq 0).$$

(b) Use this to show that, if $\mathbb{E}|X|^n < \infty$, then

$$\left| \Phi_X(t) - \sum_{k=0}^{n} \frac{(\mathrm{i}t)^k \mathbb{E}[X^k]}{k!} \right| \leq |t|^n \mathbb{E}\left[\min\left\{ \frac{2|X|^n}{n!}, \frac{|t|\,|X|^{n+1}}{(n+1)!} \right\} \right].$$

Explain the implication this estimate has for a CLT for iid random variables X_k with $\mathbb{E}|X_k|^n < \infty$, specifying which moment orders permit the classical finite-variance normalization.

Chapter 7

Conditional Expectation and Conditional Distributions

7.1 Conditioning on finitely many atoms

Let X and Z be simple random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, taking values $x_1, \dots, x_m$ and $z_1, \dots, z_n$, respectively. Suppose these values have positive probability. For each fixed j, the numbers

$$\mathbb{P}(X = x_i \mid Z = z_j) = \frac{\mathbb{P}(X = x_i, Z = z_j)}{\mathbb{P}(Z = z_j)}, \qquad 1 \le i \le m,$$

form a probability distribution. Its expectation is

$$\mathbb{E}[X \mid Z = z_j] = \sum_{i=1}^{m} x_i \mathbb{P}(X = x_i \mid Z = z_j).$$

Define the random variable $Y = \mathbb{E}[X \mid Z]$ by assigning this value to every ω for which $Z(\omega) = z_j$.

Thus Y is constant on each atom

$$G_j = \{\omega : Z(\omega) = z_j\}$$

and is measurable with respect to $\mathcal{G} = \sigma(Z)$. Here $\sigma(Z)$ consists of the preimages of the subsets of $\{z_1, \dots, z_n\}$. If $G \in \mathcal{G}$, write $G = \bigcup_{j \in J} G_j$ for some $J \subseteq \{1, \dots, n\}$. Then

$$\begin{aligned} \mathbb{E}[Y \mathbf{1}_G] &= \sum_{j \in J} \mathbb{E}[Y \mathbf{1}_{G_j}] \\ &= \sum_{j \in J} \mathbb{E}[X \mid Z = z_j] \mathbb{P}(Z = z_j) \\ &= \sum_{j \in J} \sum_{i=1}^{m} x_i \mathbb{P}(X = x_i, Z = z_j) = \mathbb{E}[X \mathbf{1}_G]. \end{aligned}$$

These two features, measurability with respect to the information supplied by Z and preservation of integrals over its events, give the general definition. If a listed atom has probability zero, any finite constant on that atom gives the same integral identities.

7.2 Definition, uniqueness, and existence

Theorem 7.1 (Conditional expectation)**.** *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field. There exists $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ such that*

$$\mathbb{E}[Y\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_G] \qquad \text{for every } G \in \mathcal{G}. \tag{7.1}$$

Any two such random variables agree almost surely. A random variable with these properties is called a version of the conditional expectation *of X given $\mathcal{G}$, and is denoted by $\mathbb{E}[X \mid \mathcal{G}]$.*

We write

$$\mathbb{E}[X \mid Z] = \mathbb{E}[X \mid \sigma(Z)], \qquad \mathbb{E}[X \mid Z_1, Z_2, \ldots] = \mathbb{E}[X \mid \sigma(Z_1, Z_2, \ldots)].$$

The notation determines an almost-sure equivalence class, rather than values at every point. Whenever a version is chosen, it is $\mathcal{G}$-measurable. Replacing its values by a fixed finite constant on a $\mathcal{G}$-measurable null set preserves the defining properties.

Proof of uniqueness. Suppose $Y, Y' \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ satisfy (7.1). For $\delta > 0$, set $G_\delta = \{Y - Y' > \delta\} \in \mathcal{G}$. The two defining identities give

$$0 = \mathbb{E}[(Y - Y')\mathbf{1}_{G_\delta}] \geq \delta \mathbb{P}(G_\delta).$$

Hence $\mathbb{P}(G_\delta) = 0$. As $\delta \downarrow 0$, these events increase to $\{Y > Y'\}$, so continuity from below gives $\mathbb{P}(Y > Y') = 0$. Interchanging Y and Y' gives $\mathbb{P}(Y' > Y) = 0$. □

Lemma 7.2 (Testing on a generating class)**.** *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$. Suppose $\mathcal{P}$ is a π-system containing Ω, $\sigma(\mathcal{P}) = \mathcal{G}$, and*

$$\mathbb{E}[X\mathbf{1}_G] = \mathbb{E}[Y\mathbf{1}_G] \qquad (G \in \mathcal{P}).$$

Then Y is a version of $\mathbb{E}[X \mid \mathcal{G}]$.

Proof. Let

$$\mathcal{L} = \{A \in \mathcal{F} : \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A]\}.$$

By hypothesis $\mathcal{P} \subseteq \mathcal{L}$, so $\Omega \in \mathcal{L}$. If $A \subseteq B$ and $A, B \in \mathcal{L}$, subtracting their integral identities shows that $B \setminus A$ is in $\mathcal{L}$. If $A_k \uparrow A$ and every $A_k \in \mathcal{L}$, dominated convergence, with dominators $|X|$ and $|Y|$, gives $A \in \mathcal{L}$. These properties make $\mathcal{L}$ a λ-system. The π-λ theorem (Theorem 1.17) therefore gives $\mathcal{G} = \sigma(\mathcal{P}) \subseteq \mathcal{L}$. □

The threshold argument for uniqueness has a useful companion: an inequality between all eventwise integrals determines the order of the random variables.

Lemma 7.3 (Order from integrals)**.** *Let X, Y be $\mathcal{F}$-measurable real random variables, each having at least one integrable sign part. If*

$$\mathbb{E}[X\mathbf{1}_A] \leq \mathbb{E}[Y\mathbf{1}_A] \qquad \text{for every } A \in \mathcal{F},$$

where the expectations may be infinite, then $X \leq Y$ almost surely. The same assertion holds with $\mathcal{F}$ replaced throughout by a sub-σ-field.

Proof. For $n \geq 1$, take $A_n = \{Y < X,\ |X| \leq n,\ |Y| \leq n\}$ and put $Z_n = (Y - X)\mathbf{1}_{A_n} = -(Y - X)^{-}\mathbf{1}_{A_n}$. Both $X\mathbf{1}_{A_n}$ and $Y\mathbf{1}_{A_n}$ are bounded, so

$$0 \geq \mathbb{E}Z_n = \mathbb{E}[Y\mathbf{1}_{A_n}] - \mathbb{E}[X\mathbf{1}_{A_n}] \geq 0.$$

The nonpositive integrable random variable Z_n therefore vanishes almost surely, giving $\mathbb{P}(A_n) = 0$. Since X, Y are real, $A_n \uparrow \{Y < X\}$, and consequently $\mathbb{P}(Y < X) = 0$. □

7.2.1 The Radon–Nikodym route

Recall that, for measures μ, ν on a measurable space $(S, \mathcal{S})$, *absolute continuity*, written $\nu \ll \mu$, means

$$\mu(A) = 0 \implies \nu(A) = 0 \qquad (A \in \mathcal{S}).$$

Every nonnegative measurable function f induces a measure $f\mu$ by

$$(f\mu)(A) = \int_A f \, \mathrm{d}\mu.$$

Indeed, this set function is nonnegative, vanishes at the empty set, and is finitely additive. If $A_k \uparrow A$, then $f\mathbf{1}_{A_k} \uparrow f\mathbf{1}_A$, so monotone convergence gives $(f\mu)(A_k) \uparrow (f\mu)(A)$. Finite additivity and continuity from below imply countable additivity.

We use the following measure-theoretic existence theorem.

Theorem 7.4 (Radon–Nikodym)**.** *If μ and ν are σ-finite measures on $(S, \mathcal{S})$ and $\nu \ll \mu$, there is a nonnegative measurable function f such that*

$$\nu(A) = \int_A f \, \mathrm{d}\mu \qquad (A \in \mathcal{S}).$$

One can choose f finite everywhere. It is unique μ-almost everywhere and is denoted by $d\nu/d\mu$, the Radon–Nikodym derivative*.*

The finiteness qualification follows from σ-finiteness of ν: if $S = \bigcup_n S_n$ with $\nu(S_n) < \infty$, then $\mu(S_n \cap \{f = \infty\}) = 0$ for each n. Thus f may be set to zero on the measurable null set $\{f = \infty\}$. When $\nu(S) < \infty$, the derivative is also μ-integrable, with integral $\nu(S)$.

Existence in Theorem 7.1. First suppose $X \geq 0$. On $(\Omega, \mathcal{G})$ define

$$\mu = \mathbb{P}|_{\mathcal{G}}, \qquad \nu(A) = \int_A X \, \mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbf{1}_A], \quad A \in \mathcal{G}.$$

The second integral is taken with respect to the original probability measure on $\mathcal{F}$; X need not be $\mathcal{G}$-measurable. The preceding measure construction on $\mathcal{F}$, followed by restriction to $\mathcal{G}$, shows that ν is a measure. It is finite, since $\nu(\Omega) = \mathbb{E}X < \infty$. Moreover, $\mu(A) = 0$ implies $\nu(A) = 0$.

The Radon–Nikodym theorem supplies a nonnegative $\mathcal{G}$-measurable function Y such that $\nu = Y\mu$. In particular,

$$\mathbb{E}[Y\mathbf{1}_G] = \nu(G) = \mathbb{E}[X\mathbf{1}_G] \qquad (G \in \mathcal{G}), \qquad \mathbb{E}Y = \mathbb{E}X < \infty.$$

Thus Y is finite almost surely and is an integrable version of the desired conditional expectation.

For general $X \in L^1$, write $X = X^+ - X^-$ and let

$$Y^+ = \mathbb{E}[X^+ \mid \mathcal{G}], \qquad Y^- = \mathbb{E}[X^- \mid \mathcal{G}].$$

Here the superscripts on Y^+ and Y^- name two auxiliary nonnegative random variables. They are integrable, and $Y = Y^+ - Y^-$ satisfies $\mathbb{E}|Y| \le \mathbb{E}(Y^+ + Y^-) = \mathbb{E}|X|$. For every $G \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}[Y\mathbf{1}_G] &= \mathbb{E}[Y^+\mathbf{1}_G] - \mathbb{E}[Y^-\mathbf{1}_G] \\ &= \mathbb{E}[X^+\mathbf{1}_G] - \mathbb{E}[X^-\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_G].\end{aligned}$$

This proves existence. □

Example 7.5 (Conditional expectation and sign parts)**.** The auxiliary variables in this construction need not be the positive and negative parts of Y. Let

$$\mathbb{P}(X = 2) = \mathbb{P}(X = -2) = \tfrac{1}{2}, \qquad \mathcal{G} = \{\varnothing, \Omega\}.$$

Then $\mathbb{E}[X \mid \mathcal{G}] = 0$, whereas

$$\mathbb{E}[X^+ \mid \mathcal{G}] = \mathbb{E}[X^- \mid \mathcal{G}] = 1.$$

In general, neither $(\mathbb{E}[X \mid \mathcal{G}])^+ = \mathbb{E}[X^+ \mid \mathcal{G}]$ nor $(\mathbb{E}[X \mid \mathcal{G}])^- = \mathbb{E}[X^- \mid \mathcal{G}]$ is valid.

7.3 Properties of conditional expectation

All conditional-expectation equalities and inequalities below are understood almost surely. Write $Y = \mathbb{E}[X \mid \mathcal{G}]$ for $X \in L^1$.

Proposition 7.6 (Expectation, positivity, and linearity)**.** *Conditional expectation has the following properties.*

(i) $\mathbb{E}Y = \mathbb{E}X$.

(ii) If $X \ge 0$, *then* $Y \ge 0$ *and* $\mathbb{P}(X > 0, Y \le 0) = 0$.

(iii) For $X_1, X_2 \in L^1$ *and* $\alpha, \beta \in \mathbb{R}$,

$$\mathbb{E}[\alpha X_1 + \beta X_2 \mid \mathcal{G}] = \alpha\mathbb{E}[X_1 \mid \mathcal{G}] + \beta\mathbb{E}[X_2 \mid \mathcal{G}].$$

Also, $\mathbb{E}[X \mid \mathcal{G}] = X$ *if* X *is* $\mathcal{G}$*-measurable, and* $X_1 \le X_2$ *implies* $\mathbb{E}[X_1 \mid \mathcal{G}] \le \mathbb{E}[X_2 \mid \mathcal{G}]$.

Proof. Taking $G = \Omega$ in the defining identity proves (i). For (ii), set $G = \{Y \le 0\} \in \mathcal{G}$. Then

$$0 \le \mathbb{E}[X\mathbf{1}_G] = \mathbb{E}[Y\mathbf{1}_G] \le 0.$$

Thus $Y\mathbf{1}_G = 0$ almost surely, proving $Y \ge 0$. For every $\delta > 0$ we also have

$$\delta\mathbb{P}(X > \delta, Y \le 0) \le \mathbb{E}[X\mathbf{1}_{\{X>\delta\}}\mathbf{1}_G] \le \mathbb{E}[X\mathbf{1}_G] = 0.$$

Letting $\delta \downarrow 0$ proves the additional assertion. In particular, $X > 0$ almost surely implies $Y > 0$ almost surely.

For (iii), the proposed linear combination is integrable and $\mathcal{G}$-measurable. Linearity of integration verifies its defining identity, and uniqueness finishes the proof. If X is already $\mathcal{G}$-measurable, it itself satisfies that identity. Finally apply (ii) and (iii) to $X_2 - X_1$ to obtain monotonicity. □

Proposition 7.7 (Conditional Jensen inequality)**.** *Let φ be a finite convex function on an open interval $\Theta \subseteq \mathbb{R}$. If $\mathbb{P}(X \in \Theta) = 1$ and $X, \varphi(X) \in L^1$, then $Y = \mathbb{E}[X \mid \mathcal{G}] \in \Theta$ almost surely, $\varphi(Y)$ is integrable, and*

$$\varphi(Y) \le \mathbb{E}[\varphi(X) \mid \mathcal{G}], \qquad \mathbb{E}[\varphi(Y)] \le \mathbb{E}[\varphi(X)].$$

Proof. At each finite endpoint of Θ, apply the strict-positivity conclusion above to the distance of X from that endpoint. This gives $Y \in \Theta$ almost surely. A finite convex function on an open interval is the supremum of countably many supporting affine functions: choosing supports at a countable dense set gives

$$\varphi(x) = \sup_{j \ge 1} (a_j x + b_j), \qquad x \in \Theta.$$

For every j, monotonicity and linearity give

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \ge a_j Y + b_j.$$

Outside a single null set these inequalities hold for every j; taking the supremum proves the conditional inequality. A fixed supporting affine function bounds $\varphi(Y)$ from below by an integrable random variable, while the conditional inequality bounds it from above by an integrable random variable. Thus $\varphi(Y)$ is integrable, and taking expectations proves the last assertion. □

Proposition 7.8 (Tower property)**.** *For $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ and $X \in L^1$,*

$$\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\big] = \mathbb{E}[X \mid \mathcal{H}].$$

Proof. Put $Y = \mathbb{E}[X \mid \mathcal{G}]$ and $Z = \mathbb{E}[Y \mid \mathcal{H}]$. Both are integrable, and Z is $\mathcal{H}$-measurable. For any $H \in \mathcal{H}$, we have $H \in \mathcal{G}$, and hence

$$\mathbb{E}[Z\mathbf{1}_H] = \mathbb{E}[Y\mathbf{1}_H] = \mathbb{E}[X\mathbf{1}_H].$$

Uniqueness identifies Z with $\mathbb{E}[X \mid \mathcal{H}]$. □

Proposition 7.9 (Independent extra information)**.** *If $X \in L^1$ and $\mathcal{H}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$, then*

$$\mathbb{E}[X \mid \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}].$$

In particular, if $\mathcal{H}$ is independent of $\sigma(X)$, then $\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}X$.

Proof. Let $Y = \mathbb{E}[X \mid \mathcal{G}]$. For $G \in \mathcal{G}$ and $H \in \mathcal{H}$, the random variable $\mathbf{1}_H$ is independent of both $X\mathbf{1}_G$ and $Y\mathbf{1}_G$. Therefore

$$\begin{aligned}\mathbb{E}[X\mathbf{1}_{G\cap H}] &= \mathbb{E}[X\mathbf{1}_G]\mathbb{E}[\mathbf{1}_H] \\ &= \mathbb{E}[Y\mathbf{1}_G]\mathbb{E}[\mathbf{1}_H] = \mathbb{E}[Y\mathbf{1}_{G\cap H}].\end{aligned}$$

The collection $\mathcal{P} = \{G \cap H : G \in \mathcal{G}, H \in \mathcal{H}\}$ is a π-system containing Ω and generating $\sigma(\mathcal{G}, \mathcal{H})$. Lemma 7.2 proves the assertion. Taking $\mathcal{G} = \{\varnothing, \Omega\}$ gives the final claim, because its conditional expectation is the constant $\mathbb{E}X$. □

Proposition 7.10 (Taking out a measurable factor)**.** *Let Z be a real $\mathcal{G}$-measurable random variable. If $X, XZ \in L^1$, then $Z\mathbb{E}[X \mid \mathcal{G}] \in L^1$ and*

$$\mathbb{E}[XZ \mid \mathcal{G}] = Z\mathbb{E}[X \mid \mathcal{G}].$$

Proof. For $Z = \mathbf{1}_A$ with $A \in \mathcal{G}$, test on $G \in \mathcal{G}$ and use the defining identity on $A \cap G$. Linearity gives the result for simple Z, and uniform simple approximation followed by dominated convergence gives it for bounded Z.

Write $Y = \mathbb{E}[X \mid \mathcal{G}]$ and $W = \mathbb{E}[|X| \mid \mathcal{G}]$. Monotonicity and linearity applied to $-|X| \le X \le |X|$ give $|Y| \le W$. Apply the bounded case to $|Z| \wedge n$ and $|X|$, take expectations, and use monotone convergence:

$$\mathbb{E}[|Z|W] = \lim_n \mathbb{E}[(|Z| \wedge n)|X|] = \mathbb{E}|ZX| < \infty.$$

It follows that ZY is integrable. Finally take $Z_n = Z\mathbf{1}_{\{|Z|\le n\}}$. The bounded case yields

$$\mathbb{E}[Z_n Y\mathbf{1}_G] = \mathbb{E}[Z_n X\mathbf{1}_G] \qquad (G \in \mathcal{G}).$$

Dominated convergence, using $|ZY|$ and $|ZX|$, passes this identity to the limit. Since ZY is $\mathcal{G}$-measurable, uniqueness applies. □

7.4 Conditional expectation as an orthogonal projection

For real $U, V \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, define

$$\langle U, V\rangle = \mathbb{E}[UV], \qquad \|U\|_2^2 = \langle U, U\rangle .$$

Cauchy–Schwarz makes the inner product finite. We identify random variables that agree almost surely, since $\|U\|_2 = 0$ if and only if $U = 0$ almost surely. Write $U \perp V$ when $\langle U, V\rangle = 0$. Expanding the inner products gives Pythagoras' identity and the parallelogram identity:

$$\|U + V\|_2^2 = \|U\|_2^2 + \|V\|_2^2 \qquad \text{if } U \perp V, \tag{7.2}$$

$$\|U + V\|_2^2 + \|U - V\|_2^2 = 2\,\|U\|_2^2 + 2\,\|V\|_2^2 . \tag{7.3}$$

Theorem 7.11 (Completeness of L^2)**.** *Every Cauchy sequence (X_m) in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ converges in L^2 to a member of that space.*

Proof. Choose integers $k_N \uparrow \infty$ such that

$$\sup_{m,n \geq k_N} \|X_m - X_n\|_2 \leq 2^{-N}.$$

Cauchy–Schwarz gives

$$\mathbb{E}|X_{k_{N+1}} - X_{k_N}| \leq \left\|X_{k_{N+1}} - X_{k_N}\right\|_2 \leq 2^{-N}.$$

By monotone convergence,

$$\mathbb{E} \sum_{N=1}^{\infty} |X_{k_{N+1}} - X_{k_N}| \leq \sum_{N=1}^{\infty} 2^{-N} < \infty.$$

The sum of absolute increments is therefore finite almost surely. The series of increments converges, so X_{k_N} has a finite limit almost surely. Let X be that limit on the measurable set of convergence, and set $X = 0$ on its null complement. Then X is $\mathcal{F}$-measurable.

For $m \geq k_N$ and $t \geq N$,

$$\mathbb{E}|X_m - X_{k_t}|^2 \leq 2^{-2N}.$$

Letting $t \to \infty$ and applying Fatou's lemma gives

$$\mathbb{E}|X_m - X|^2 \leq 2^{-2N}.$$

In particular, $X - X_{k_1} \in L^2$, so $X \in L^2$; the same estimate shows $X_m \to X$ in L^2. □

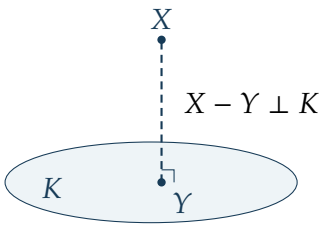
Theorem 7.12 (Orthogonal projection)**.** *Let K be a complete linear subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. For every $X \in L^2$ there is a unique element $Y \in K$ such that*

$$\|X - Y\|_2 = \inf_{W \in K} \|X - W\|_2 . \tag{7.4}$$

For $Y \in K$, condition (7.4) is equivalent to

$$\langle X - Y, Z \rangle = 0 \qquad \text{for every } Z \in K. \tag{7.5}$$

Uniqueness of random-variable representatives is almost sure.



The nearest point Y leaves an error orthogonal to K.

Proof. Let $\Delta = \inf_{W\in K} \|X - W\|_2$ and choose $Y_n \in K$ such that $\|X - Y_n\|_2 \to \Delta$. The parallelogram identity gives

$$\|X - Y_m\|_2^2 + \|X - Y_n\|_2^2 = 2\left\|X - \frac{Y_m + Y_n}{2}\right\|_2^2 + \frac{1}{2}\|Y_m - Y_n\|_2^2 .$$

Since $(Y_m + Y_n)/2 \in K$, the first term on the right is at least $2\Delta^2$. Consequently

$$\frac{1}{2}\|Y_m - Y_n\|_2^2 \le \|X - Y_m\|_2^2 + \|X - Y_n\|_2^2 - 2\Delta^2 \longrightarrow 0.$$

Thus (Y_n) is Cauchy. Completeness gives $Y \in K$ with $Y_n \to Y$ in L^2. The inequalities

$$\Delta \le \|X - Y\|_2 \le \|X - Y_n\|_2 + \|Y_n - Y\|_2$$

show that $\|X - Y\|_2 = \Delta$.

For $Z \in K$ and every $\alpha \in \mathbb{R}$, minimality at Y implies

$$\|X - (Y + \alpha Z)\|_2^2 \ge \|X - Y\|_2^2 ,$$

or, after expansion,

$$\alpha^2 \|Z\|_2^2 - 2\alpha \langle Z, X - Y\rangle \ge 0.$$

This can hold for both signs of every sufficiently small α only if $\langle Z, X - Y\rangle = 0$.

Conversely, if $Y \in K$ satisfies (7.5), then $Y - W \in K$ for every $W \in K$, so Pythagoras gives

$$\|X - W\|_2^2 = \|X - Y\|_2^2 + \|Y - W\|_2^2 .$$

This proves minimality. It also shows that any other minimizer W satisfies $\|Y - W\|_2 = 0$, which proves uniqueness. □

Theorem 7.13 (The projection is conditional expectation)**.** *For $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, the orthogonal projection of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a version of $\mathbb{E}[X \mid \mathcal{G}]$. Equivalently, this version minimizes*

$$\mathbb{E}[(X - W)^2], \qquad W \in L^2(\Omega, \mathcal{G}, \mathbb{P}),$$

and satisfies $\mathbb{E}[(X - Y)Z] = 0$ for every $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Proof. Apply the completeness theorem on $(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$. It shows that $K = L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a complete linear subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Its elements are equivalence classes admitting $\mathcal{G}$-measurable representatives. The projection theorem gives $Y \in K$ with $\mathbb{E}[(X - Y)Z] = 0$ for every $Z \in K$. Taking $Z = \mathbf{1}_G$, $G \in \mathcal{G}$, gives

$$\mathbb{E}[Y\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_G].$$

Also $Y \in L^1$ by Cauchy–Schwarz. Thus the projection satisfies exactly the defining conditions for conditional expectation, whose uniqueness was proved directly. □

7.4.1 A second existence route: from L^2 to L^1

The projection construction gives existence without using the Radon–Nikodym theorem. To extend it, first take $X \geq 0$ with $\mathbb{E}X < \infty$ and choose bounded nonnegative $X_n \uparrow X$, for example $X_n = X \wedge n$. Each X_n belongs to L^2, so its projection gives $Y_n = \mathbb{E}[X_n \mid \mathcal{G}]$.

Positivity and linearity followed directly from the defining integral identities and uniqueness. They therefore apply to these projections without presupposing existence for general L^1 variables. They give $0 \leq Y_n \leq Y_{n+1}$ almost surely. There is one $\mathcal{G}$-measurable null set outside which all these inequalities hold; set every Y_n to zero on that set. Now $Y = \lim_n Y_n$ is a nonnegative $\mathcal{G}$-measurable extended random variable. Monotone convergence gives

$$\mathbb{E}Y = \lim_n \mathbb{E}Y_n = \lim_n \mathbb{E}X_n = \mathbb{E}X < \infty.$$

Thus Y is finite almost surely and integrable. Set it to zero on its $\mathcal{G}$-measurable set of infinite values. For every $G \in \mathcal{G}$, another application of monotone convergence gives

$$\mathbb{E}[Y\mathbf{1}_G] = \lim_n \mathbb{E}[Y_n\mathbf{1}_G] = \lim_n \mathbb{E}[X_n\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_G].$$

Finally, apply this construction to X^+ and X^- and subtract. Their finite expectations justify the subtraction exactly as in the first existence proof.

The same limit argument shows that if $0 \leq U_n \uparrow U$ almost surely and $U \in L^1$, then

$$\mathbb{E}[U_n \mid \mathcal{G}] \uparrow \mathbb{E}[U \mid \mathcal{G}] \quad \text{almost surely.} \tag{7.6}$$

7.5 Conditional densities

Suppose $(X, Z) \in \mathbb{R}^2$ has a joint probability density $f_{X,Z}(x, z)$. Choose a nonnegative finite Borel version of this density. The marginal densities are

$$f_Z(z) = \int_{\mathbb{R}} f_{X,Z}(x, z)\,\mathrm{d}x, \qquad f_X(x) = \int_{\mathbb{R}} f_{X,Z}(x, z)\,\mathrm{d}z.$$

Both are Borel measurable by Tonelli's theorem. Their integrals are one, so their infinite values occur only on Lebesgue null sets. We may set f_X to zero at such points to obtain a finite Borel version. Put $D = \{z : 0 < f_Z(z) < \infty\}$ and define

$$f_{X|Z}(x \mid z) = \begin{cases} f_{X,Z}(x, z)/f_Z(z), & z \in D, \\ f_X(x), & z \notin D. \end{cases} \tag{7.7}$$

The fallback covers both zero and infinite marginal values. This function is jointly Borel measurable and, for every $z \in \mathbb{R}$,

$$\int_{\mathbb{R}} f_{X|Z}(x \mid z)\,\mathrm{d}x = 1.$$

Moreover, $\mathbb{P}(Z \in D) = 1$: the zero-density set has zero Z-probability, and the infinite-density set is Lebesgue null.

Proposition 7.14 (Integration against a conditional density)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be Borel measurable and $\mathbb{E}|g(X)| < \infty$. Define*

$$\widehat{g}(z) = \begin{cases} \displaystyle\int_{\mathbb{R}} g(x) f_{X|Z}(x \mid z)\,\mathrm{d}x, & \displaystyle\int_{\mathbb{R}} |g(x)| f_{X|Z}(x \mid z)\,\mathrm{d}x < \infty, \\ 0, & \text{otherwise}. \end{cases}$$

Then $\widehat{g}$ is Borel measurable, $\widehat{g}(Z)$ is integrable, and

$$\mathbb{E}[g(X) \mid Z] = \widehat{g}(Z) \quad \text{almost surely}.$$

Proof. Joint measurability of the density and Tonelli's theorem show that

$$h(z) = \int_{\mathbb{R}} |g(x)| f_{X|Z}(x \mid z)\,\mathrm{d}x$$

is Borel measurable. Put $A = \{z : h(z) < \infty\}$. Integrating the positive and negative parts separately shows that $\widehat{g}$ is Borel measurable on $\{h < \infty\}$, and hence everywhere with the stated fallback. Since $\mathbb{P}(Z \in D) = 1$, Tonelli gives

$$\begin{aligned} \mathbb{E}[h(Z)] &= \int_D \int_{\mathbb{R}} |g(x)| f_{X,Z}(x, z)\,\mathrm{d}x\,\mathrm{d}z \\ &= \mathbb{E}|g(X)| < \infty. \end{aligned}$$

Thus $h(Z) < \infty$ almost surely. The elementary Jensen inequality for the probability density $f_{X|Z}(\cdot \mid z)$ gives $|\widehat{g}(z)| \le h(z)$ wherever $h(z) < \infty$, and proves integrability of $\widehat{g}(Z)$.

For a Borel set $B \subseteq \mathbb{R}$, Fubini is justified by the absolute integrability just established and yields

$$\begin{aligned} \mathbb{E}[\widehat{g}(Z)\mathbf{1}_{\{Z\in B\}}] &= \int_{B\cap D} \widehat{g}(z) f_Z(z)\,\mathrm{d}z \\ &= \int_{B\cap D\cap A} \int_{\mathbb{R}} g(x) f_{X,Z}(x, z)\,\mathrm{d}x\,\mathrm{d}z \\ &= \mathbb{E}[g(X)\mathbf{1}_{\{Z\in B\}}]. \end{aligned}$$

In the second equality the exceptional set $\{h = \infty\}$ may be omitted, since it has zero Z-probability. Every event of $\sigma(Z)$ has the form $\{Z \in B\}$, so these are the defining identities. □

In particular, taking $g = \mathbf{1}_B$ gives conditional probabilities

$$\widehat{P}_{X|Z}(B, \omega) = \int_B f_{X|Z}(x \mid Z(\omega))\,\mathrm{d}x.$$

For fixed B this is a version of $\mathbb{E}[\mathbf{1}_{\{X\in B\}} \mid Z]$, and for fixed ω it is a probability measure in B.

7.6 Regular conditional distributions

Definition 7.15 (Regular conditional probability distribution)**.** Let $Y : \Omega \to S$ be an $(S, \mathcal{S})$-valued random variable, and let $\mathcal{G} \subseteq \mathcal{F}$. A collection $\widehat{P}_{Y|\mathcal{G}} : \mathcal{S} \times \Omega \to [0,1]$ is a *regular conditional probability distribution* of Y given $\mathcal{G}$ if:

(a) for each $B \in \mathcal{S}$, $\widehat{P}_{Y|\mathcal{G}}(B, \cdot)$ is a version of $\mathbb{E}[\mathbf{1}_{\{Y \in B\}} \mid \mathcal{G}]$;

(b) for each $\omega \in \Omega$, $\widehat{P}_{Y|\mathcal{G}}(\cdot, \omega)$ is a probability measure on $(S, \mathcal{S})$.

In particular, taking $S = \Omega$, $\mathcal{S} = \mathcal{F}$, and $Y(\omega) = \omega$ gives a *regular conditional probability* on $\mathcal{F}$ given $\mathcal{G}$, also denoted by $\widehat{P}(B \mid \mathcal{G})(\omega)$. Such a regular conditional probability need not exist on an arbitrary measurable space.

The issue is simultaneous countable additivity. For each fixed disjoint sequence (B_n), arbitrary conditional-probability versions satisfy

$$\mathbb{P}\Big(Y \in \bigcup_n B_n \;\Big|\; \mathcal{G}\Big) = \sum_n \mathbb{P}(Y \in B_n \mid \mathcal{G})$$

almost surely, but the exceptional null set can depend on the whole sequence. There may be uncountably many sequences to consider, so one cannot simply take the union of all exceptional sets and assume it is null. Real-valued random variables admit a construction using only countably many initial conditions.

Theorem 7.16 (Existence of regular conditional distributions)**.** *For any real-valued random variable X and any sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$, a regular conditional distribution $\widehat{P}_{X|\mathcal{G}}$ exists.*

Proof. For each $q \in \mathbb{Q}$, choose a version

$$H(q, \omega) = \mathbb{E}[\mathbf{1}_{\{X \le q\}} \mid \mathcal{G}](\omega).$$

Monotonicity and (7.6), also applied to complementary indicators, show that outside a single $\mathcal{G}$-measurable null set these values lie in $[0,1]$, are nondecreasing in q, and satisfy

$$H(q + 1/n, \omega) \downarrow H(q, \omega), \qquad H(-n, \omega) \downarrow 0, \qquad H(n, \omega) \uparrow 1.$$

Only countably many rational pairs and sequences are involved. On this null set replace all $H(q, \omega)$ by $\mathbf{1}_{\{q \ge 0\}}$, the distribution function of a point mass at zero. Define, for every real x,

$$F(x, \omega) = \inf_{\substack{q \in \mathbb{Q} \\ q > x}} H(q, \omega).$$

For each ω, this is nondecreasing, right-continuous, and has limits zero and one at the two ends of the real line. Right continuity follows because, if $x_n \downarrow x$, the sets $\{q \in \mathbb{Q} : q > x_n\}$ increase to $\{q \in \mathbb{Q} : q > x\}$. It therefore determines a probability measure $\widehat{P}_{X|\mathcal{G}}(\cdot, \omega)$ with this distribution function. For rational q, the inequalities

$$H(q, \omega) \le F(q, \omega) \le H(q + 1/n, \omega)$$

show that $F(q,\omega) = H(q,\omega)$.

For each x, $F(x,\cdot)$ is $\mathcal{G}$-measurable, being a countable infimum. The class of Borel sets B for which $\widehat{P}_{X|\mathcal{G}}(B,\cdot)$ is $\mathcal{G}$-measurable is a λ-system containing the rational half-lines. The π-λ theorem extends measurability to every Borel set. Finally, fix $G \in \mathcal{G}$. The two finite measures in B,

$$B \longmapsto \int_G \widehat{P}_{X|\mathcal{G}}(B,\omega)\,\mathrm{d}\mathbb{P}(\omega), \qquad B \longmapsto \mathbb{P}(X \in B, G),$$

agree on the rational half-lines and on $\mathbb{R}$, and hence on every Borel set. These identities verify condition (a) of the definition; condition (b) holds by construction for every ω. □

More generally, the same existence conclusion holds when $(S,\mathcal{S})$ is *Borel-isomorphic*: there is a Borel subset $T \subseteq \mathbb{R}$ and a bijection $b : S \to T$ such that both b and b^{-1} are measurable. In particular, this includes spaces measurably isomorphic to $(\mathbb{R},\mathcal{B})$. Indeed, apply the theorem to $b(Y)$. Its conditional distribution assigns mass one to T almost surely, since $b(Y) \in T$. On the measurable null exceptional set, replace it by a point mass at some fixed point of T. Then

$$\widehat{P}_{Y|\mathcal{G}}(B,\omega) = \widehat{P}_{b(Y)|\mathcal{G}}(b(B),\omega), \qquad B \in \mathcal{S},$$

has the required properties. Here $b(B)$ is Borel by measurability of b^{-1} and the fact that T is Borel. The existence of the random variable Y ensures that S, and hence T, is nonempty.

Problems

The result of another problem may be cited without proving it.

Problem 7.1 (Product Radon–Nikodym derivatives)**.** Let (ν_k, μ_k) be pairs of σ-finite measures on $(\Omega_k, \mathcal{F}_k)$, with $\nu_k \ll \mu_k$ for $k = 1, \dots, n$, and let $f_k = d\nu_k/d\mu_k$. On the product measurable space

$$(\Omega, \mathcal{F}) = \left(\prod_{k=1}^{n} \Omega_k, \bigotimes_{k=1}^{n} \mathcal{F}_k \right),$$

write $\nu = \nu_1 \otimes \cdots \otimes \nu_n$ and $\mu = \mu_1 \otimes \cdots \otimes \mu_n$.

(a) Show that $\nu \ll \mu$ and that

$$\frac{d\nu}{d\mu}(\omega) = \prod_{k=1}^{n} f_k(\omega_k) \quad \text{for } \mu\text{-almost every } \omega = (\omega_1, \dots, \omega_n) \in \Omega.$$

(b) Suppose that the product measures μ and ν are probability measures. Show that the random variables $X_k(\omega) = f_k(\omega_k)$, $k = 1, \dots, n$, are mutually independent with respect to μ and also with respect to ν.

Problem 7.2 (Conditional variance)**.** Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}X^2 < \infty$, and define

$$\operatorname{Var}(X \mid \mathcal{G}) = \mathbb{E}\big[(X - \mathbb{E}[X \mid \mathcal{G}])^2 \mid \mathcal{G}\big].$$

(a) Prove that

$$\mathbb{E}[\operatorname{Var}(X \mid \mathcal{G})] \le \mathbb{E}[\operatorname{Var}(X \mid \mathcal{H})]$$

whenever $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ are σ-fields. Thus the expected squared deviation from the conditional mean decreases as the conditioning σ-field grows.

(b) Prove that, for every sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$,

$$\operatorname{Var}(X) = \mathbb{E}[\operatorname{Var}(X \mid \mathcal{G})] + \operatorname{Var}(\mathbb{E}[X \mid \mathcal{G}]).$$

Problem 7.3 (Equality to a conditional expectation)**.** Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $Y = \mathbb{E}[X \mid \mathcal{G}]$ for a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$.

(a) Prove that if $\mathbb{E}Y^2 = \mathbb{E}X^2 < \infty$, then $Y = X$ almost surely.

(b) Prove that if Y and X have the same law, then $Y = X$ almost surely, including the case $\mathbb{E}X^2 = \infty$.

Hint for (b). Show that $\mathbb{E}[(|X| - X)\mathbf{1}_A] = 0$ for $A = \{Y \ge 0\} \in \mathcal{G}$, so that $X \ge 0$ almost surely on A. Apply this to $X - c$, with $c \in \mathbb{R}$ nonrandom, to deduce $\mathbb{P}(X < c \le Y) = 0$ and then $X \ge Y$ almost surely.

Problem 7.4 (Conditional tails, Markov, and Hölder inequalities)**.** Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) Fix $p > 0$ and let X satisfy $\mathbb{E}|X|^p < \infty$. Show that, almost surely,

$$\mathbb{E}[|X|^p \mid \mathcal{G}] = \int_0^\infty p x^{p-1} \mathbb{P}(|X| > x \mid \mathcal{G}) \, \mathrm{d}x.$$

In this integral, use the conditional tails from one regular conditional distribution of $|X|$ given $\mathcal{G}$. Deduce that for every $a > 0$,

$$\mathbb{P}(|X| \geq a \mid \mathcal{G}) \leq a^{-p} \mathbb{E}[|X|^p \mid \mathcal{G}] \quad \text{almost surely.}$$

(b) Let $p, q > 1$ satisfy $1/p + 1/q = 1$, and suppose $\mathbb{E}|X|^p < \infty$ and $\mathbb{E}|Y|^q < \infty$. Prove the conditional Hölder inequality

$$\mathbb{E}[|XY| \mid \mathcal{G}] \leq \big(\mathbb{E}[|X|^p \mid \mathcal{G}]\big)^{1/p} \big(\mathbb{E}[|Y|^q \mid \mathcal{G}]\big)^{1/q} \quad \text{almost surely.}$$

Problem 7.5 (Reduction of conditional independence)**.** On a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, real random variables U, V are called *conditionally independent given* a sub-σ-field $\mathcal{H}$ if, for all bounded Borel functions $h_1, h_2 : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[h_1(U) h_2(V) \mid \mathcal{H}] = \mathbb{E}[h_1(U) \mid \mathcal{H}] \mathbb{E}[h_2(V) \mid \mathcal{H}] \quad \text{almost surely.}$$

Let X, Y, Z be real random variables on this space. Suppose X and Y are conditionally independent given $\sigma(Z)$, and X and Z are conditionally independent given $\mathcal{F}$, where $\mathcal{F} \subseteq \sigma(Z)$. Prove that X and Y are conditionally independent given $\mathcal{F}$.

Chapter 8

Markov Chains and Recurrence

8.1 Filtrations and transition probabilities

A filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ is a nondecreasing sequence of sub-sigma-fields

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_\infty, \qquad \mathcal{F}_\infty = \sigma\left(\bigcup_{n\geq 0} \mathcal{F}_n\right) \subseteq \mathcal{F}.$$

A stochastic process $(X_n)_{n\geq 0}$ is *adapted* to this filtration if $\sigma(X_n) \subseteq \mathcal{F}_n$ for every n. Its natural filtration is

$$\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n).$$

Definition 8.1 (Markov chain)**.** An adapted process with values in a measurable space $(\mathsf{S}, \mathcal{S})$ is a Markov chain relative to $(\mathcal{F}_n)$ if, for every $n \geq 0$ and $A \in \mathcal{S}$,

$$\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A \mid X_n) \qquad \text{almost surely.}$$

Unless a different filtration is specified, the natural filtration is used.

A chain that is Markov relative to $(\mathcal{F}_n)$ is also Markov relative to $(\mathcal{F}_n^X)$. Indeed, adaptedness gives $\mathcal{F}_n^X \subseteq \mathcal{F}_n$, and the tower property (Proposition 7.8) gives

$$\begin{aligned}\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n^X) &= \mathbb{E}\left[\mathbb{E}[\mathbf{1}_{\{X_{n+1}\in A\}} \mid \mathcal{F}_n] \mid \mathcal{F}_n^X\right] \\ &= \mathbb{E}\left[\mathbb{E}[\mathbf{1}_{\{X_{n+1}\in A\}} \mid X_n] \mid \mathcal{F}_n^X\right] \\ &= \mathbb{P}(X_{n+1} \in A \mid X_n).\end{aligned}$$

Definition 8.2 (Transition probabilities)**.** A transition probability, or probability kernel, on $(\mathsf{S}, \mathcal{S})$ is a map $p : \mathsf{S} \times \mathcal{S} \to [0, 1]$ such that

(i) for each $x \in \mathsf{S}$, $A \mapsto p(x, A)$ is a probability measure on $(\mathsf{S}, \mathcal{S})$;

(ii) for each $A \in \mathcal{S}$, $x \mapsto p(x, A)$ is measurable.

A chain has transition probabilities (p_n) if

$$\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = p_n(X_n, A) \quad \text{almost surely,} \qquad n \geq 0, \quad A \in \mathcal{S}.$$

It is *homogeneous* when $p_n = p$ for every n: the transition probabilities do not depend on time.

For a bounded measurable function h on S, write

$$(p_n h)(x) = \int h(y)\, p_n(x,\, \mathrm{d}y).$$

This is a bounded measurable function. The assertion follows first for indicators from the kernel definition, then for simple functions and bounded measurable functions by approximation. The same approximation in the conditional-expectation identity gives

$$\mathbb{E}[h(X_{n+1}) \mid \mathcal{F}_n] = (p_n h)(X_n). \tag{8.1}$$

Proposition 8.3 (Finite joint laws)**.** *Let ν be a sigma-finite measure on $(\mathsf{S}, \mathcal{S})$ and let (p_n) be probability kernels. For every $n \geq 0$ there is a unique sigma-finite measure*

$$\mu_n = \nu \otimes p_0 \otimes \cdots \otimes p_{n-1} \quad \text{on } (\mathsf{S}^{n+1}, \mathcal{S}^{\otimes(n+1)})$$

such that $\mu_0 = \nu$ and

$$\mu_n(A_0 \times \cdots \times A_n) = \int_{A_0} \nu(\mathrm{d}x_0) \int_{A_1} p_0(x_0,\, \mathrm{d}x_1) \cdots \int_{A_n} p_{n-1}(x_{n-1},\, \mathrm{d}x_n). \tag{8.2}$$

If (X_n) is a Markov chain with initial probability law ν and transition probabilities (p_n), then for bounded measurable h_i,

$$\mathbb{E}\left[\prod_{i=0}^{n} h_i(X_i)\right] = \int \nu(\mathrm{d}x_0) h_0(x_0) \int p_0(x_0,\, \mathrm{d}x_1) h_1(x_1) \cdots \int p_{n-1}(x_{n-1},\, \mathrm{d}x_n) h_n(x_n). \tag{8.3}$$

In particular, $(X_0, \ldots, X_n)$ has law μ_n.

Proof. If μ is a measure and q a probability kernel, define

$$(\mu \otimes q)(C) = \int \int \mathbf{1}_C(u, v)\, q(u,\, \mathrm{d}v)\, \mu(\mathrm{d}u).$$

The inner integral is measurable: this holds for rectangles, and the class of sets for which it holds is a lambda-system containing the rectangles. Monotone convergence proves countable additivity. A countable cover of the first space by sets of finite μ-measure proves sigma-finiteness. To prove uniqueness, refine this cover to a disjoint measurable partition and restrict each candidate measure to one partition set times the second space. These restrictions are finite and agree on rectangles, so Corollary 1.18 identifies them. Summing over the partition proves uniqueness. Iterating this construction, with $q((x_0, \ldots, x_j), \cdot) = p_j(x_j, \cdot)$, proves (8.2).

For the expectation formula, condition on $\mathcal{F}_{n-1}$:

$$\mathbb{E}\left[\prod_{i=0}^{n} h_i(X_i)\right] = \mathbb{E}\left[\prod_{i=0}^{n-1} h_i(X_i)(p_{n-1}h_n)(X_{n-1})\right].$$

Induction, starting with $\mathbb{E}h_0(X_0) = \int h_0\, \mathrm{d}\nu$, gives the iterated integral. Taking $h_i = \mathbf{1}_{A_i}$ gives the joint probabilities. □

Regular conditional distributions provide transition kernels for a Markov chain on a Borel-isomorphic state space, by Theorem 7.16. Here Borel-isomorphic means measurably isomorphic to a Borel subset of $\mathbb{R}$; in particular, countable spaces with all subsets measurable and Euclidean Borel spaces are included. We assume this state-space condition in the path-law construction below. The finite-dimensional construction above did not require it.

8.2 The law on path space

Put

$$\mathsf{S}_\infty = \mathsf{S}^{\{0,1,2,\dots\}}, \qquad \mathcal{S}_c = \sigma\{\omega : \omega_0 \in A_0, \dots, \omega_n \in A_n;\ A_i \in \mathcal{S},\ n \ge 0\}.$$

The sigma-field $\mathcal{S}_c$ is the cylinder sigma-field. Write $b\mathcal{S}_c$ for the bounded measurable real functions on this space.

Proposition 8.4 (Canonical path law)**.** *On a Borel-isomorphic state space, an initial probability law ν and transition kernels (p_n) determine a unique probability law $\mathbb{P}_\nu$ on $(\mathsf{S}_\infty, \mathcal{S}_c)$ with finite joint laws (8.2). The coordinate process $X_n(\omega) = \omega_n$ is a Markov chain with these transitions and initial law.*

Proof. The finite joint laws are probabilities, because each kernel has total mass one. Moreover, integrating out the last coordinate gives

$$\mu_{n+1}(C \times \mathsf{S}) = \mu_n(C), \qquad C \in \mathcal{S}^{\otimes(n+1)}.$$

Thus they are consistent. Kolmogorov's extension theorem states that consistent finite-dimensional probability laws on a Borel-isomorphic space extend uniquely to its countable product with the cylinder sigma-field. Its state-space and consistency hypotheses have both been verified here, so it gives $\mathbb{P}_\nu$. Uniqueness also follows directly because cylinders form a generating pi-system.

For $C \in \mathcal{S}^{\otimes(n+1)}$ and $A \in \mathcal{S}$, the defining finite joint laws give

$$\begin{aligned}\mathbb{E}_\nu[\mathbf{1}_{\{(X_0,\dots,X_n)\in C\}}\mathbf{1}_{\{X_{n+1}\in A\}}] &= \int_C p_n(x_n, A)\,\mu_n(\,\mathrm{d}x_0 \cdots \mathrm{d}x_n) \\ &= \mathbb{E}_\nu[\mathbf{1}_{\{(X_0,\dots,X_n)\in C\}} p_n(X_n, A)].\end{aligned}$$

These are exactly the test-event identities for conditioning on $\mathcal{F}_n^X$. They establish the Markov property of the coordinate process. □

For a homogeneous kernel, $\mathbb{P}_x$ denotes this law with initial distribution δ_x, and $\mathbb{E}_x$ its expectation. In particular, $\mathbb{E}_x h$ is now defined for every $h \in b\mathcal{S}_c$, including functions of the entire infinite path. The map $x \mapsto \mathbb{E}_x h$ is measurable. For a cylinder product $h(\omega) = \prod_{i=0}^r h_i(\omega_i)$ this follows from the iterated kernel integral. The class of bounded functions for which it holds is linear and closed under uniformly bounded monotone

limits, so the functional monotone-class theorem (Theorem 2.2) extends it to $b\mathcal{S}_c$. The same argument gives

$$\mathbb{E}_\nu h = \int \mathbb{E}_x h \, \nu(\mathrm{d}x).$$

The shift operator removes the first coordinate:

$$\theta(\omega_0, \omega_1, \ldots) = (\omega_1, \omega_2, \ldots), \qquad \theta^n \omega = (\omega_n, \omega_{n+1}, \ldots).$$

It is measurable, since the inverse image of a cylinder is a cylinder. For a process on another probability space, $h(\theta^n X)$ means $h(X_n, X_{n+1}, \ldots)$.

8.3 Stopping times and stopped sigma-fields

Definition 8.5 (Stopping time)**.** A random time $\tau : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is a stopping time for $(\mathcal{F}_n)$ if

$$\{\tau \le n\} \in \mathcal{F}_n \qquad \text{for every finite } n \ge 0.$$

Equivalently, $\{\tau = n\} \in \mathcal{F}_n$ for all $n \ge 0$. Throughout, the infimum of an empty set of times is ∞.

Example 8.6 (Visits to a set)**.** For an adapted process and a measurable set A, the first hitting time including time zero is

$$H_A = \inf\{n \ge 0 : X_n \in A\}.$$

It is a stopping time, because $\{H_A \le n\} = \bigcup_{j=0}^{n} \{X_j \in A\} \in \mathcal{F}_n$. The second visit is also a stopping time: it has occurred by time n exactly when $\sum_{j=0}^{n} \mathbf{1}_{\{X_j \in A\}} \ge 2$. The last visit is generally not a stopping time, since deciding that a visit is the last one requires future information. For example, for independent fair $\{0, 1\}$-valued X_0, X_1, let the last visit to $\{1\}$ among times $0, 1$ have value ∞ if there is no visit. It equals zero precisely on $\{X_0 = 1, X_1 = 0\}$, an event not determined by X_0.

Definition 8.7 (Stopped sigma-field)**.** The sigma-field of information available when stopping at τ is

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \le n\} \in \mathcal{F}_n \text{ for every finite } n \ge 0\}.$$

Equivalently, $A \in \mathcal{F}_\tau$ if $A \in \mathcal{F}_\infty$ and $A \cap \{\tau = k\} \in \mathcal{F}_k$ for every finite $k \ge 0$.

Proposition 8.8 (Properties of the stopped sigma-field)**.** *Let τ be a stopping time.*

(i) $\mathcal{F}_\tau$ is a sigma-field.

(ii) If $\tau = m$ everywhere, then $\mathcal{F}_\tau = \mathcal{F}_m$, also for $m = \infty$. If $\tau = m$ only almost surely, the equality holds modulo null sets: each event in either sigma-field agrees almost surely with an event in the other.

(iii) If X_n is $\mathcal{F}_n$-measurable for every $n \in \{0, 1, \ldots, \infty\}$, then X_τ is $\mathcal{F}_\tau$-measurable. In particular, $\sigma(\tau) \subseteq \mathcal{F}_\tau$. For real-valued adapted X and each fixed finite k, $X_k \mathbf{1}_{\{\tau = k\}}$ is $\mathcal{F}_\tau$-measurable.

(iv) For any fixed finite k and integrable random variable Y_k,

$$\mathbb{E}[Y_k \mathbf{1}_{\{\tau=k\}} \mid \mathcal{F}_\tau] = \mathbf{1}_{\{\tau=k\}} \mathbb{E}[Y_k \mid \mathcal{F}_k]. \tag{8.4}$$

In particular, the left side may be written $\mathbb{E}[Y_\tau \mathbf{1}_{\{\tau=k\}} \mid \mathcal{F}_\tau]$ for a family (Y_j), interpreting that localized variable as Y_k on $\{\tau = k\}$ and zero elsewhere. The identity also holds for nonnegative variables, allowing infinite conditional expectations.

(v) If $\sigma \le \tau$ are stopping times, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

Proof. For (i), $\Omega \in \mathcal{F}_\tau$ by the stopping-time property. Taking complements within $\{\tau \le n\}$ and countable unions verifies the remaining sigma-field axioms. The equivalent criterion in Definition 8.7 follows by differences in one direction and finite unions in the other.

For a pointwise constant finite time, the condition at $n = m$ is exactly $A \in \mathcal{F}_m$; all other conditions follow. For $\tau = \infty$ it imposes no restriction beyond $A \in \mathcal{F}_\infty$. If $\tau = m$ almost surely and $m < \infty$, then $A \cap \{\tau = m\}$ belongs to $\mathcal{F}_m$ for $A \in \mathcal{F}_\tau$, and $B \cap \{\tau = m\}$ belongs to $\mathcal{F}_\tau$ for $B \in \mathcal{F}_m$. These intersections agree almost surely with A and B, respectively. For $m = \infty$ use $B \cap \{\tau = \infty\}$ instead. No completeness assumption on the filtration is needed.

For (iii), if B is measurable in the state space, then

$$\{X_\tau \in B\} \cap \{\tau \le n\} = \bigcup_{j=0}^{n} \big(\{X_j \in B\} \cap \{\tau = j\}\big) \in \mathcal{F}_n.$$

The event $\{X_\tau \in B\}$ also belongs to $\mathcal{F}_\infty$, by its countable decomposition over finite times together with $\{X_\infty \in B, \tau = \infty\}$. This proves the claim, including the infinite-time case. Apply it to the deterministic process $X_n = n$, $X_\infty = \infty$, to obtain measurability of τ. More generally, any $\mathcal{F}_k$-measurable real variable localized to $\{\tau = k\}$ is $\mathcal{F}_\tau$-measurable, by the same event criterion. When X_∞ has no given meaning, a fixed state can be chosen for its value; this convention defines X_τ on $\{\tau = \infty\}$.

For (iv), the proposed right side is measurable by this localization observation and integrable. For $A \in \mathcal{F}_\tau$, the event $A \cap \{\tau = k\}$ is in $\mathcal{F}_k$, so

$$\mathbb{E}[\mathbf{1}_A \mathbf{1}_{\{\tau=k\}} \mathbb{E}[Y_k \mid \mathcal{F}_k]] = \mathbb{E}[\mathbf{1}_A \mathbf{1}_{\{\tau=k\}} Y_k].$$

This is the defining conditional-expectation identity. Truncation gives the nonnegative version.

Finally, for $A \in \mathcal{F}_\sigma$,

$$A \cap \{\tau \le n\} = \big(A \cap \{\sigma \le n\}\big) \cap \{\tau \le n\} \in \mathcal{F}_n,$$

which proves (v). □

8.4 The ordinary and strong Markov properties

Proposition 8.9 (Markov property for the future path)**.** *Let (X_n) be a homogeneous chain on a Borel-isomorphic state space, with initial law ν. For every $h \in b\mathcal{S}_c$ and $n \ge 0$,*

$$\mathbb{E}_\nu[h(\theta^n X) \mid \mathcal{F}_n^X] = \mathbb{E}_{X_n} h \qquad \textit{almost surely}. \tag{8.5}$$

Proof. First let $h(\omega) = \prod_{j=0}^{r} f_j(\omega_j)$, with bounded measurable f_j. For $B \in \mathcal{S}^{\otimes(n+1)}$, split the iterated integral in (8.3) after x_n:

$$\begin{aligned}
&\mathbb{E}_\nu[\mathbf{1}_{\{(X_0,\dots,X_n)\in B\}} h(\theta^n X)] \\
&= \int_B \mu_n(\mathrm{d}x_0 \cdots \mathrm{d}x_n) f_0(x_n) \int p(x_n, \mathrm{d}x_{n+1}) f_1(x_{n+1}) \cdots \int p(x_{n+r-1}, \mathrm{d}x_{n+r}) f_r(x_{n+r}) \\
&= \mathbb{E}_\nu[\mathbf{1}_{\{(X_0,\dots,X_n)\in B\}} \mathbb{E}_{X_n} h].
\end{aligned}$$

The integration identity for arbitrary B follows from rectangles by measure uniqueness. Since $x \mapsto \mathbb{E}_x h$ is measurable, these test-event identities prove (8.5) for cylinder products. Their linear span is an algebra generating $\mathcal{S}_c$. The class of bounded h satisfying the identity is linear and closed under uniformly bounded monotone limits, using monotone convergence for both ordinary and conditional expectations. The functional monotone-class theorem therefore proves it for all $h \in b\mathcal{S}_c$. □

Thus, given $(X_0, \dots, X_n)$, the expectation of $h(X_n, X_{n+1}, \dots)$ is the expectation of the same function of a chain started afresh at X_n. A stopping time can replace the fixed time n.

Theorem 8.10 (Strong Markov property)**.** *Let τ be a stopping time for the natural filtration of a homogeneous chain as above. Let $h_n \in b\mathcal{S}_c$ for every $n \geq 0$, with $\sup_{n,\omega} |h_n(\omega)| < \infty$, and put $g(n, x) = \mathbb{E}_x h_n$. Define*

$$Z(\omega) = \begin{cases} h_{\tau(\omega)}(\theta^{\tau(\omega)} X(\omega)), & \tau(\omega) < \infty, \\ 0, & \tau(\omega) = \infty. \end{cases}$$

Then

$$\mathbb{E}_\nu[Z \mid \mathcal{F}_\tau^X] = g(\tau, X_\tau)\mathbf{1}_{\{\tau<\infty\}} \qquad \text{almost surely}, \tag{8.6}$$

where the right side is defined to be zero on $\{\tau = \infty\}$. Equivalently, on the event of finite stopping,

$$\mathbb{E}_\nu[h_\tau(\theta^\tau X) \mid \mathcal{F}_\tau^X]\mathbf{1}_{\{\tau<\infty\}} = \mathbb{E}_{X_\tau}[h_\tau]\mathbf{1}_{\{\tau<\infty\}},$$

with the same zero extension on both sides.

Proof. Each $g(k, \cdot)$ is measurable. The right side of (8.6) is the $\mathcal{F}_\tau^X$-measurable sum

$$\sum_{k\geq 0} g(k, X_k)\mathbf{1}_{\{\tau=k\}}.$$

The ordinary Markov property and the localization identity (8.4) give

$$\mathbb{E}_\nu[h_k(\theta^k X)\mathbf{1}_{\{\tau=k\}} \mid \mathcal{F}_\tau^X] = \mathbf{1}_{\{\tau=k\}}\mathbb{E}_\nu[h_k(\theta^k X) \mid \mathcal{F}_k^X] = \mathbf{1}_{\{\tau=k\}} g(k, X_k).$$

Summing over $k \geq 0$ proves the theorem. More explicitly, the partial sums on the left converge to Z in L^1, because they are uniformly bounded and agree with Z as soon as their upper index reaches the finite value of τ. Conditional expectation is an L^1 contraction, so their conditional expectations converge as well. The sum on the right is exactly $g(\tau, X_\tau)\mathbf{1}_{\{\tau<\infty\}}$. □

Nonnegative path functions are also allowed, with extended expectations, by truncation and conditional monotone convergence. This version will be used for visit counts and return times.

8.5 Countable chains and communication

For the remainder of the recurrence discussion, S is countable, $\mathcal{S} = 2^{\mathsf{S}}$, and the chain is homogeneous. Its transition probabilities are the stochastic matrix

$$p(x, y) = p(x, \{y\}) \geq 0, \qquad \sum_{y \in \mathsf{S}} p(x, y) = 1.$$

Write $p^{(n)}(x, y) = \mathbb{P}_x(X_n = y)$, with $p^{(0)}(x, y) = \mathbf{1}_{\{x=y\}}$. Conditioning at a fixed time gives

$$p^{(m+n)}(x, z) = \sum_{y \in \mathsf{S}} p^{(m)}(x, y) p^{(n)}(y, z).$$

Here we use *positive-time* visit times:

$$\tau_y = \inf\{n \geq 1 : X_n = y\}, \qquad \tau_A = \inf\{n \geq 1 : X_n \in A\}, \qquad \rho(x, y) = \mathbb{P}_x(\tau_y < \infty). \tag{8.7}$$

In particular, under $\mathbb{P}_x$, τ_x is the first return to x. This differs from H_x in Example 8.6, which equals zero when the chain starts at x.

Definition 8.11 (Communication)**.** The state y is accessible from x if $\rho(x, y) > 0$. Two states communicate, written $x \leftrightarrow y$, if either $x = y$ or both $\rho(x, y) > 0$ and $\rho(y, x) > 0$. A nonempty subset $C \subseteq \mathsf{S}$ is *irreducible* if every two distinct states in it communicate. It is *closed* if there are no $x \in C$, $y \notin C$ with $\rho(x, y) > 0$.

Accessibility is transitive. Indeed, if $\rho(x, y) > 0$ and $\rho(y, z) > 0$, there are positive integers r, s with $p^{(r)}(x, y) > 0$ and $p^{(s)}(y, z) > 0$. Therefore $p^{(r+s)}(x, z) \geq p^{(r)}(x, y) p^{(s)}(y, z) > 0$. It follows that communication is an equivalence relation; its equivalence classes are exactly the maximal irreducible subsets. They are also the strongly connected components of the directed accessibility graph, whose edges $x \to y$ mean $\rho(x, y) > 0$. Equivalently, use the graph of one-step transitions $p(x, y) > 0$: its directed paths give the same accessibility relation.

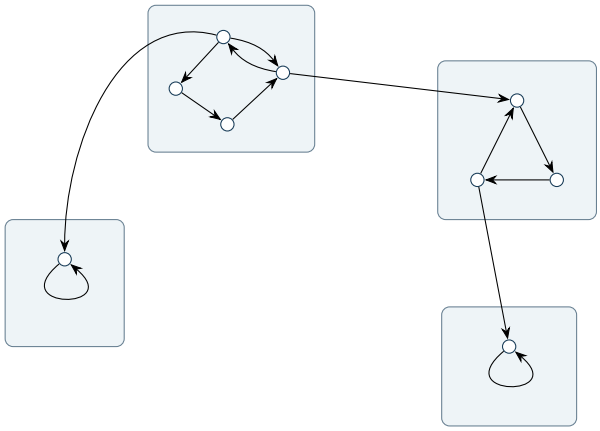


Figure 8.1: Four communicating classes. The arrows shown generate the accessibility relation; arrows implied by transitivity are suppressed. The two singleton classes are absorbing and closed.

8.6 Recurrence, visits, and the Green function

Definition 8.12 (Recurrence and transience)**.** A state x is *recurrent* if $\rho(x, x) = 1$, and *transient* if $\rho(x, x) < 1$. A recurrent state is *positive recurrent* if $\mathbb{E}_x \tau_x < \infty$, and *null recurrent* if $\mathbb{E}_x \tau_x = \infty$.

Example 8.13 (Simple random walk)**.** For simple symmetric random walk on $\mathbb{Z}^d$, each of the $2d$ nearest-neighbor steps has probability $1/(2d)$. The walk is recurrent in dimensions $d = 1, 2$ and transient in dimensions $d \geq 3$. This dimension-dependent classification is stated here without proof.

Lemma 8.14 (Hitting before returning)**.** *If $x \neq y$ and $\rho(x, y) > 0$, then*

$$\mathbb{P}_x(\tau_y < \tau_x) > 0.$$

Proof. Suppose instead that $\mathbb{P}_x(\tau_y < \tau_x) = 0$. For every integer $k \geq 1$, decompose a first visit to y at time k according to whether the chain has already returned to x. Strong Markov at the first return gives the exact identity

$$\begin{aligned} \mathbb{P}_x(\tau_y = k) &= \mathbb{P}_x(\tau_y = k, \tau_y < \tau_x) \\ &\quad + \sum_{j=1}^{k-1} \mathbb{P}_x(\tau_x = j, \tau_y > j) \mathbb{P}_x(\tau_y = k - j). \end{aligned}$$

The event $\{\tau_y > j\}$ in the summand ensures this is a first visit after the first return to x; $X_{\tau_x} = x$ on that event. The first term is zero by assumption. Induction on k, starting with $k = 1$, now makes every term zero. Summing over k yields $\rho(x, y) = 0$, a contradiction. In particular, the argument does not assume that paths avoiding y necessarily return to x forever. □

Define the time of the kth positive-time visit to y by

$$\tau_y^{(0)} = 0, \qquad \tau_y^{(k)} = \inf\{n > \tau_y^{(k-1)} : X_n = y\}, \quad k \geq 1.$$

If the previous time is infinite, the next is infinite as well. These are stopping times: $\{\tau_y^{(k)} \leq n\}$ is the event that $X_1, \dots, X_n$ visit y at least k times.

Proposition 8.15 (Successive visits)**.** *For every $x, y \in \mathsf{S}$ and $k \geq 1$,*

$$\mathbb{P}_x(\tau_y^{(k)} < \infty) = \rho(x, y)\rho(y, y)^{k-1}. \tag{8.8}$$

The exponent-zero factor for $k = 1$ is interpreted as one.

Proof. The case $k = 1$ is the definition of $\rho(x, y)$. For $k \geq 2$, apply strong Markov at $\sigma = \tau_y^{(k-1)}$ to the path event $\{\tau_y < \infty\}$. This event is measurable in $\mathcal{S}_c$, being a countable union of finite-time visit events. On $\{\sigma < \infty\}$ the chain is at y, and a further visit has conditional probability $\rho(y, y)$. Consequently,

$$\begin{aligned}\mathbb{P}_x(\tau_y^{(k)} < \infty) &= \mathbb{E}_x\left[\mathbf{1}_{\{\sigma<\infty\}}\mathbb{E}_x[\mathbf{1}_{\{\tau_y\circ\theta^\sigma<\infty\}} \mid \mathcal{F}_\sigma^X]\right] \\ &= \rho(y, y)\mathbb{P}_x(\tau_y^{(k-1)} < \infty).\end{aligned}$$

All shifted expressions are used only on $\{\sigma < \infty\}$ and extended by zero elsewhere. Induction proves the formula. □

Let

$$N_\infty(y) = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=y\}}, \qquad G(x, y) = \mathbb{E}_x N_\infty(y).$$

Thus G is the Green function counting visits from time 1, without the possible visit at time zero. Since $\{N_\infty(y) \geq k\} = \{\tau_y^{(k)} < \infty\}$, monotone convergence and (8.8) give

$$\begin{aligned}G(x, y) = \sum_{n=1}^{\infty} p^{(n)}(x, y) &= \sum_{k=1}^{\infty} \mathbb{P}_x(N_\infty(y) \geq k) \\ &= \sum_{k=1}^{\infty} \rho(x, y)\rho(y, y)^{k-1} \\ &= \begin{cases} 0, & \rho(x, y) = 0, \\ \dfrac{\rho(x, y)}{1 - \rho(y, y)}, & \rho(x, y) > 0,\ \rho(y, y) < 1, \\ \infty, & \rho(x, y) > 0,\ \rho(y, y) = 1. \end{cases}\end{aligned} \tag{8.9}$$

The first case applies even when y is recurrent but inaccessible from x.

Corollary 8.16 (Equivalent recurrence criteria)**.** *For a state y, the following are equivalent:*

(i) $\rho(y,y) = 1$;

(ii) $\mathbb{P}_y(\tau_y^{(k)} < \infty) = 1$ *for every* $k \geq 1$;

(iii) $\mathbb{P}_y(X_n = y \text{ infinitely often}) = 1$;

(iv) $\mathbb{P}_y(N_\infty(y) = \infty) = 1$;

(v) $G(y,y) = \infty$.

Proof. Formula (8.8) gives (i)⇒(ii). The countable intersection of the probability-one events in (ii) gives infinitely many visits, so (ii)⇒(iii). The events in (iii) and (iv) are identical, and (iv) implies (v). Finally, if $\rho(y,y) < 1$, (8.9) makes $G(y,y)$ finite, proving (v)⇒(i). □

8.7 Types of communicating classes

Theorem 8.17 (Recurrence is a class property)**.** *Communicating states have the same type: they are either all transient, all positive recurrent, or all null recurrent. Every communicating class containing a recurrent state is closed.*

Proof. First suppose $x \leftrightarrow y$, with $x \neq y$. Choose $r, s \geq 1$ such that $a = p^{(r)}(x,y) > 0$ and $b = p^{(s)}(y,x) > 0$. The Markov property gives, for $n \geq 1$,

$$p^{(r+n+s)}(x,x) \geq ab\, p^{(n)}(y,y).$$

Summing gives $G(x,x) \geq abG(y,y)$. Interchanging x, y and using Corollary 8.16 proves that recurrence, and hence transience, is common to a communicating class.

If x is recurrent and $y \neq x$ is accessible from x, put $q = \mathbb{P}_x(\tau_y < \tau_x) > 0$, by Lemma 8.14. On hitting y before returning to x, the probability of never subsequently reaching x is $1 - \rho(y,x)$. Strong Markov at τ_y therefore gives

$$0 = \mathbb{P}_x(\tau_x = \infty) \geq q\big(1 - \rho(y,x)\big).$$

Thus $\rho(y,x) = 1$, and y is in the same communicating class as x. This proves closedness. It also shows that any two states in a recurrent class hit one another with probability one, by reversing their roles once recurrence of y is established.

It remains to distinguish positive and null recurrence. Suppose $m_x = \mathbb{E}_x\tau_x < \infty$ and $x \leftrightarrow y$, $x \neq y$. With the same $q > 0$, the portion of the first x-return cycle after its first visit to y has expected length $q\mathbb{E}_y\tau_x$. Strong Markov, first with truncated times and then by monotone convergence, gives

$$m_x \geq \mathbb{E}_x[(\tau_x - \tau_y)\mathbf{1}_{\{\tau_y < \tau_x\}}] = q\mathbb{E}_y\tau_x. \tag{8.10}$$

The difference in this expression is evaluated only on the indicated event. Hence $\mathbb{E}_y\tau_x < \infty$.

For the other direction, let $T_0 = 0$ and $T_j = \tau_x^{(j)}$ under $\mathbb{P}_x$, and let K be the first return cycle which visits y. The event $\{K \geq j\}$ is determined at T_{j-1}. Conditional on this event and the past at T_{j-1}, the next cycle visits y with probability q and has expected length

m_x, by strong Markov. Induction gives $\mathbb{P}_x(K \geq j) = (1-q)^{j-1}$, so $K < \infty$ almost surely. Moreover, by monotone convergence,

$$\begin{aligned}\mathbb{E}_x \tau_y \leq \mathbb{E}_x T_K &= \sum_{j \geq 1} \mathbb{E}_x[(T_j - T_{j-1})\mathbf{1}_{\{K \geq j\}}] \\ &= m_x \sum_{j \geq 1} (1-q)^{j-1} = \frac{m_x}{q} < \infty.\end{aligned}$$

Starting from y, first reach x and then reach y. The first return to y is no later than this combined journey. Strong Markov at τ_x now gives

$$\mathbb{E}_y \tau_y \leq \mathbb{E}_y \tau_x + \mathbb{E}_x \tau_y < \infty.$$

Thus positive recurrence passes between communicating states. Consequently a recurrent class with no positive recurrent state consists entirely of null recurrent states. □

Corollary 8.18 (Decomposition of the state space)**.** *The countable state space has a unique decomposition, up to the order of its classes,*

$$\mathsf{S} = T \dot{\cup} R_1 \dot{\cup} R_2 \dot{\cup} \cdots,$$

where T is the set of transient states and the R_i are the recurrent communicating classes. Each R_i is irreducible and closed. There may be no recurrent classes, finitely many, or countably many.

Proof. Communication partitions the space into its maximal irreducible subsets. By Theorem 8.17, each such subset consists entirely of transient states or entirely of recurrent states, and the recurrent ones are closed. Their disjoint union with T is the whole space. The equivalence classes, and therefore this decomposition, are unique. □

8.8 Martingales at bounded stopping times

The following terminology and bounded-time fact will also be used in the problems.

Definition 8.19 (Martingale and submartingale)**.** A real adapted process $(M_n, \mathcal{F}_n)$ is a martingale if $\mathbb{E}|M_n| < \infty$ for every n and

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n \qquad \text{almost surely.}$$

It is a submartingale if the equality is replaced by $\geq$, and a supermartingale if it is replaced by $\leq$. It is L^1-bounded if $\sup_n \mathbb{E}|M_n| < \infty$.

Proposition 8.20 (Bounded-time stopping)**.** *Let $(M_n, \mathcal{F}_n)$ be a martingale, and let $\sigma \leq \tau$ be stopping times, possibly infinite. For every finite n,*

$$\mathbb{E}M_{n\wedge\sigma} = \mathbb{E}M_{n\wedge\tau}.$$

For a submartingale the left side is at most the right side; for a supermartingale the inequality is reversed. In particular, if $0 \leq T \leq n$ is a stopping time and M a martingale, then $\mathbb{E}M_T = \mathbb{E}M_0$.

Proof. All stopped variables are integrable, being finite sums of localized M_k. Also

$$M_{n\wedge\tau} - M_{n\wedge\sigma} = \sum_{k=1}^{n} \mathbf{1}_{\{\sigma<k\leq\tau\}}(M_k - M_{k-1}).$$

The indicator is $\mathcal{F}_{k-1}$-measurable, since $\{\sigma < k \leq \tau\} = \{\sigma \leq k-1\} \setminus \{\tau \leq k-1\}$. Conditioning each summand on $\mathcal{F}_{k-1}$ gives mean zero for a martingale, a nonnegative mean for a submartingale, and a nonpositive mean for a supermartingale. Summing proves the assertions. □

Problems

In the first four problems, results from other problems may be cited without proof.

Problem 8.1 (Passage times and renewal overshoots)**.** Let $(X_i)_{i\ge 1}$ be iid random variables with

$$\mathbb{P}(X_1 = k) = q_k, \qquad k = 1, 2, \ldots, \qquad \sum_{k=1}^{\infty} q_k = 1, \qquad \mathbb{E}X_1 < \infty.$$

Set $\mathcal{F}_0 = \{\varnothing, \Omega\}$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, $S_0 = 0$, and $S_n = \sum_{i=1}^{n} X_i$. For nonnegative integer levels t, define

$$\tau_t = \inf\{n \ge 0 : S_n \ge t\}.$$

(a) Show that, for every $t \ge 0$, τ_t is a stopping time for $(\mathcal{F}_n)$: for every $n \ge 0$, $\{\tau_t \le n\}$ is $\mathcal{F}_n$-measurable. Further, show that every $(\mathcal{F}_n)$-stopping time τ has the following property. If there exist a positive integer m and $\varepsilon > 0$ such that

$$\mathbb{P}(\tau \le n + m \mid \mathcal{F}_n) > \varepsilon \quad \text{almost surely for every } n \ge 0,$$

then

$$\mathbb{P}(\tau > km) \le (1 - \varepsilon)^k \qquad (k \ge 1),$$

and hence $\mathbb{E}\tau < \infty$.

(b) Show that, for every nonrandom nonnegative integer t, X_{τ_t+1} is independent of $\mathcal{F}_{\tau_t}$ and $X_{\tau_t+1} \overset{d}{=} X_1$.

(c) Show that $D_t = S_{\tau_t} - t$, $t = 0, 1, 2, \ldots$, is a homogeneous Markov chain on $\{0, 1, 2, \ldots\}$. Find its transition probabilities and show that

$$\mu(i) = \frac{1}{\mathbb{E}X_1} \sum_{k>i} q_k, \qquad i \ge 0,$$

is an invariant measure for it. Here invariance means that $\mu(j) = \sum_{i\ge 0} \mu(i)p(i, j)$ for every $j \ge 0$, where p is the transition matrix of D.

Problem 8.2 (A walk with a random bias)**.** Let $Z, X_1, X_2, \ldots$ be independent Uniform$(0, 1)$ random variables. Put

$$\xi_i = \operatorname{sgn}(Z - X_i), \qquad S_0 = 0, \qquad S_n = \sum_{i=1}^{n} \xi_i.$$

Thus a single uniform bias Z is first chosen, and conditional on Z each step is $+1$ with probability Z and -1 otherwise.

(a) Compute $\mathbb{P}(\xi_{n+1} = 1 \mid \xi_1, \ldots, \xi_n)$.

(b) Show that (S_n) is a Markov chain and find its transition probabilities. Is it homogeneous?

Problem 8.3 (First entrance and occupation sums)**.** Let (X_n) be a homogeneous Markov chain on a countable state space S, with all subsets measurable. For $x \in \mathsf{S}$ and $k \geq 0$, put

$$\tau_{x,k} = \inf\{n \geq k : X_n = x\}.$$

As usual, $\mathbb{P}_{x_0}$ is the path law with initial state x_0.

(a) Prove that, for every $x_0, x \in \mathsf{S}$, every $B \subseteq \mathsf{S}$, and positive integers $k \leq n$,

$$\mathbb{P}_{x_0}(X_n \in B, \tau_{x,k} \leq n) = \sum_{i=0}^{n-k} \mathbb{P}_{x_0}(\tau_{x,k} = n - i)\mathbb{P}_x(X_i \in B).$$

(b) Deduce that, for the same positive integers $k \leq n$,

$$\mathbb{P}_{x_0}(X_n = x) = \sum_{i=k}^{n} \mathbb{P}_{x_0}(\tau_{x,k} = i)\mathbb{P}_x(X_{n-i} = x).$$

(c) Conclude that, for every $x \in \mathsf{S}$ and nonnegative integers k, ℓ,

$$\sum_{i=0}^{\ell} \mathbb{P}_x(X_i = x) \geq \sum_{n=k}^{\ell+k} \mathbb{P}_x(X_n = x).$$

Problem 8.4 (Harmonic boundary values)**.** Let (X_n) be a homogeneous Markov chain on a countable state space $\mathsf{S} = S \cup T$, where $S \cap T = \varnothing$, S is finite, and $T = A \cup B$ for disjoint nonempty sets A, B. For $E \subseteq \mathsf{S}$, use the hitting time including zero,

$$H_E = \inf\{n \geq 0 : X_n \in E\}.$$

Suppose

$$\inf_{s \in S} \mathbb{P}_s(H_T < \infty) > 0.$$

If S is empty, this hypothesis is vacuous. A bounded function $h : \mathsf{S} \to \mathbb{R}$ is harmonic at x for the transition matrix p if $h(x) = \sum_{y \in \mathsf{S}} p(x, y)h(y)$.

(a) Prove that there exist $\varepsilon > 0$ and a finite positive integer m such that

$$\mathbb{P}_x(H_T > km) \leq (1 - \varepsilon)^k \qquad (k \geq 1,\ x \in \mathsf{S}),$$

and conclude that $H_T < \infty$ almost surely for every initial state.

(b) Prove that $h(x) = \mathbb{P}_x(H_A < H_B)$ satisfies

$$h = 1 \text{ on } A, \qquad h = 0 \text{ on } B, \qquad h \text{ is harmonic at every } x \in S. \tag{8.11}$$

(c) Prove that any bounded function satisfying (8.11) equals $x \mapsto \mathbb{P}_x(H_A < H_B)$. To do so, first show that if $\mathcal{F}_n^X$ is the natural filtration and $g : \mathsf{S} \to \mathbb{R}$ is bounded and harmonic at every $x \in S$, then $Y_n = g(X_{n \wedge H_T})$ is a martingale; that is,

$$\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n^X] = Y_n \qquad (n \geq 0).$$

Problem 8.5 (First-order autoregression)**.** The first-order autoregressive process (X_n) is defined by

$$X_n = \alpha X_{n-1} + \xi_n, \qquad n \geq 1,$$

where α is a nonrandom scalar and $(\xi_n)_{n\geq 1}$ are iid $\mathbb{R}^d$-valued random variables independent of X_0.

(a) With $\mathcal{F}_n = \sigma(X_0, \xi_1, \ldots, \xi_n)$, verify that (X_n) is a homogeneous $(\mathcal{F}_n)$-Markov chain on $\mathbb{R}^d$ equipped with its Borel sigma-field, and give its transition probabilities.

(b) Suppose $|\alpha| < 1$ and $X_0 = \beta\xi_0$ for a nonrandom scalar β, where $(\xi_n)_{n\geq 0}$ are iid with the centered multivariate Gaussian law $\mathcal{N}(0, V)$, of covariance matrix V. Find all values of β for which the law of X_n is independent of $n \geq 0$. Here V may be singular: the Gaussian convention means that $t^{\mathsf{T}}\xi_n$ is centered normal with variance $t^{\mathsf{T}}Vt$ for every $t \in \mathbb{R}^d$, including the constant-zero law when this variance vanishes.

Problem 8.6 (Return excursions)**.** Fix a recurrent state y of a homogeneous countable-state Markov chain. Let $R_0 = 0$, $R_k = \tau_y^{(k)}$ for $k \geq 1$, and $r_k = R_k - R_{k-1}$, the number of steps between consecutive returns. On the event $R_1 < \infty$, recurrence makes every R_k finite. Define the duration and path vectors

$$\boldsymbol{Y}_k = (r_k, X_{R_{k-1}}, \ldots, X_{R_k-1}), \qquad k \geq 1.$$

These take values in the disjoint union $\bigcup_{r\geq 1}(\{r\} \times \mathsf{S}^r)$, equipped with all subsets.

(a) Deduce from the strong Markov property that under $\mathbb{P}_y$ the random vectors $(\boldsymbol{Y}_k)_{k\geq 1}$ are independent and identically distributed.

(b) Show that, for any initial probability law ν with $\mathbb{P}_\nu(R_1 < \infty) > 0$, under $\mathbb{P}_\nu$ conditional on $\{R_1 < \infty\}$, the random vectors $\boldsymbol{Y}_k$ are independent of each other. Moreover, for all $k \geq 2$, $\boldsymbol{Y}_k$ has the same conditional law as $\boldsymbol{Y}_2$, and this is the law of $\boldsymbol{Y}_1$ under $\mathbb{P}_y$. If $\mathbb{P}_\nu(R_1 < \infty) = 0$, no conditional-law assertion on that event is required.

Further martingale problems

Recall Definition 8.19, conditional expectation (Theorem 7.1), conditional Jensen (Proposition 7.7), and monotone convergence (Theorem 3.11 and (7.6)). The bounded-time stopping fact is Proposition 8.20.

Problem 8.7 (An L^1-bounded martingale decomposition)**.** Let $(X_n, \mathcal{F}_n)$ be a martingale with $\sup_n \mathbb{E}|X_n| < \infty$. Show that there is a representation $X_n = Y_n - Z_n$ in which $(Y_n, \mathcal{F}_n)$ and $(Z_n, \mathcal{F}_n)$ are nonnegative martingales and

$$\sup_n \mathbb{E}|Y_n| < \infty, \qquad \sup_n \mathbb{E}|Z_n| < \infty.$$

Problem 8.8 (A bound on a small maximum)**.** Let $S_0 = 0$ and $S_n = \sum_{k=1}^{n} \xi_k$, where the ξ_k are independent, $\mathbb{E}\xi_k = 0$, and $|\xi_k| \le K < \infty$ for all k. Put $s_0^2 = 0$ and $s_n^2 = \sum_{k=1}^{n} \mathbb{E}\xi_k^2$, and use the filtration $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, with trivial $\mathcal{F}_0$. Using Proposition 8.20 for the martingale $S_n^2 - s_n^2$ and a suitable stopping time, show that, for $x \ge 0$ and $n \ge 1$ with $s_n^2 > 0$,

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k| \le x\right) \le \frac{(x+K)^2}{s_n^2}.$$

When $s_n^2 = 0$, formulate the same bound without division as $s_n^2 \mathbb{P}(\max_{1\le k\le n} |S_k| \le x) \le (x+K)^2$.

Appendix A

Integration Against Distribution Functions

Let F be a distribution function and let P be the corresponding probability measure on $(\mathbb{R}, \mathcal{B})$, so that

$$P((a, b]) = F(b) - F(a).$$

For $g \in C_b(\mathbb{R})$, the space of bounded continuous real-valued functions, we can define an integral directly from these increments of F.

For each N, choose a finite partition

$$-\infty < a_1^N < a_2^N < \cdots < a_{N+1}^N < \infty$$

and tags $x_j^N \in (a_j^N, a_{j+1}^N]$, $1 \le j \le N$. We take limits along partitions for which

$$a_1^N \longrightarrow -\infty, \qquad a_{N+1}^N \longrightarrow +\infty, \qquad \max_{1 \le j \le N} \operatorname{osc}_{[a_j^N, a_{j+1}^N]} g \longrightarrow 0,$$

where $\operatorname{osc}_I g = \sup_{x,y \in I} |g(x) - g(y)|$. Equivalently, one may index a sequence of partitions by n and allow its number N_n of subintervals to vary. Such partitions exist: on each expanding compact interval, uniform continuity permits a sufficiently fine finite subdivision.

Theorem A.1. *For $g \in C_b(\mathbb{R})$, the sums*

$$\sum_{j=1}^{N} g(x_j^N)\big[F(a_{j+1}^N) - F(a_j^N)\big]$$

converge, independently of the tags and of the partitions satisfying the preceding conditions, to $\int_{\mathbb{R}} g \, \mathrm{d}P$. Their limit is the Riemann–Stieltjes integral *of g against F:*

$$\int_{-\infty}^{\infty} g(x) \, \mathrm{d}F(x) = \int_{\mathbb{R}} g(x) \, \mathrm{d}P(x).$$

Proof. Write $M = \sup_{x \in \mathbb{R}} |g(x)|$ and fix $\varepsilon, \delta > 0$. Choose continuity points $a < b$ of F such that

$$F(a) \le \varepsilon, \qquad 1 - F(b) \le \varepsilon.$$

The tail limits make the inequalities possible, and the continuity points can be chosen because a nondecreasing function has at most countably many discontinuities. Indeed, here every discontinuity is a positive jump, and there are only finitely many jumps of size at least $1/k$, for each positive integer k, because their sum is at most one.

By uniform continuity of g on $[a, b]$, choose a finite partition

$$a = a_1 < a_2 < \cdots < a_{N+1} = b$$

such that the oscillation of g on each $[a_j, a_{j+1}]$ is less than δ. We may choose all the endpoints to be continuity points of F, since the complement of a countable set is dense. Put $I_j = (a_j, a_{j+1}]$, choose arbitrary $x_j \in I_j$, and define the simple function

$$h(x) = \sum_{j=1}^{N} g(x_j) \mathbf{1}_{I_j}(x).$$

It vanishes outside $(a, b]$, and

$$|g(x) - h(x)| < \delta \qquad \text{for } x \in (a, b].$$

Also,

$$\int_{\mathbb{R}} h \, \mathrm{d}P = \sum_{j=1}^{N} g(x_j) P(I_j) = \sum_{j=1}^{N} g(x_j) \big[F(a_{j+1}) - F(a_j) \big].$$

Consequently,

$$\begin{aligned} &\left| \int_{\mathbb{R}} g(x) \, \mathrm{d}P(x) - \sum_{j=1}^{N} g(x_j) \big[F(a_{j+1}) - F(a_j) \big] \right| \\ &\quad = \left| \int_{(a,b]} (g - h) \, \mathrm{d}P + \int_{\mathbb{R} \setminus (a,b]} g \, \mathrm{d}P \right| \\ &\quad \le \delta \, P((a, b]) + M P(\mathbb{R} \setminus (a, b]) \\ &\quad \le \delta + M \big[1 - F(b) + F(a) \big] \\ &\quad \le \delta + 2M\varepsilon. \end{aligned}$$

Since ε and δ are arbitrary, these sums approximate $\int g \, \mathrm{d}P$ to any prescribed accuracy.

For an arbitrary sequence of partitions in the theorem, the same estimate applies with $a = a_1^N$, $b = a_{N+1}^N$, and δ replaced by the maximum oscillation. Continuity of the endpoints was useful in making the preceding choices, but the estimate itself only uses the half-open interval identity for P. The tail term and the maximum oscillation both tend to zero. This proves convergence for every permitted sequence and every choice of tags, and identifies the limit as claimed. □