COMPLETE READING NOTES

Probability Theory and Examples

Rick Durrett, Fifth Edition

Xuanzhe Hou

October 2026

Contents

1 Measure Theory **1**
1.1 Probability Spaces . . . 1
1.2 Distributions . . . 3
1.3 Random Variables . . . 5
1.4 Integration . . . 7
1.5 Properties of the Integral . . . 8
1.6 Expected Value . . . 10
1.6.1 Inequalities . . . 10
1.6.2 Integration to the Limit . . . 11
1.6.3 Computing Expected Values . . . 11
1.7 Product Measures, Fubini's Theorem . . . 12

2 Laws of Large Numbers **14**
2.1 Independence . . . 14
2.1.1 Sufficient Conditions for Independence . . . 14
2.1.2 Independence, Distribution, and Expectation . . . 15
2.1.3 Sums of Independent Random Variables . . . 16
2.1.4 Constructing Independent Random Variables . . . 17
2.2 Weak Laws of Large Numbers . . . 17
2.2.1 L^2 Weak Laws . . . 17
2.2.2 Triangular Arrays . . . 18
2.2.3 Truncation . . . 19
2.3 Borel–Cantelli Lemmas . . . 20
2.4 Strong Law of Large Numbers . . . 22
2.5 Convergence of Random Series* . . . 23
2.5.1 Rates of Convergence . . . 24
2.5.2 Infinite Mean . . . 24
2.6 Renewal Theory* . . . 25
2.7 Large Deviations* . . . 27

3 Central Limit Theorems **29**
3.1 The De Moivre–Laplace Theorem . . . 29
3.2 Weak Convergence . . . 30
3.2.1 Examples . . . 30
3.2.2 Theory . . . 30
3.3 Characteristic Functions . . . 31
3.3.1 Definition, Inversion Formula . . . 31
3.3.2 Weak Convergence . . . 32
3.3.3 Moments and Derivatives . . . 33
3.3.4 Pólya's Criterion* . . . 33
3.3.5 The Moment Problem* . . . 33
3.4 Central Limit Theorems . . . 34
3.4.1 i.i.d. Sequences . . . 34
3.4.2 Triangular Arrays . . . 34
3.4.3 Prime Divisors (Erdős–Kac)* . . . 35
3.4.4 Rates of Convergence (Berry–Esseen)* . . . 35
3.5 Local Limit Theorems* . . . 36
3.6 Poisson Convergence . . . 37
3.6.1 The Basic Limit Theorem . . . 37
3.6.2 Two Examples with Dependence . . . 37
3.7 Poisson Processes . . . 38
3.7.1 Compound Poisson Processes . . . 38

3.7.2 Thinning . . . 39
3.7.3 Conditioning . . . 39
3.8 Stable Laws* . . . 39
3.9 Infinitely Divisible Distributions* . . . 41
3.10 Limit Theorems in $\mathbb{R}^d$ . . . 41

4 Martingales **43**
4.1 Conditional Expectation . . . 43
4.1.1 Examples . . . 43
4.1.2 Properties . . . 44
4.1.3 Regular Conditional Probabilities* . . . 45
4.2 Martingales and Almost Sure Convergence . . . 45
4.3 Examples . . . 47
4.3.1 Bounded Increments . . . 47
4.3.2 Pólya's Urn Scheme . . . 47
4.3.3 Radon–Nikodym Derivatives . . . 48
4.3.4 Branching Processes . . . 48
4.4 Doob's Inequality and Convergence in L^p, $p > 1$ . . . 48
4.5 Square-Integrable Martingales* . . . 49
4.6 Uniform Integrability and Convergence in L^1 . . . 50
4.7 Backwards Martingales . . . 51
4.8 Optional Stopping Theorems . . . 52
4.8.1 Applications to Random Walks . . . 52
4.9 Combinatorics of Simple Random Walk* . . . 53

5 Markov Chains **54**
5.1 Examples . . . 54
5.2 Construction and Markov Properties . . . 55
5.3 Recurrence and Transience . . . 56
5.4 Recurrence of Random Walks* . . . 58
5.5 Stationary Measures . . . 59
5.6 Asymptotic Behavior . . . 61
5.7 Periodicity and the Tail σ-Field* . . . 62
5.8 General State Space* . . . 63
5.8.1 Recurrence and Transience . . . 64
5.8.2 Stationary Measures . . . 64
5.8.3 Convergence Theorem . . . 64
5.8.4 $GI/G/1$ Queue . . . 65

6 Ergodic Theorems **66**
6.1 Definitions and Examples . . . 66
6.2 Birkhoff's Ergodic Theorem . . . 67
6.3 Recurrence . . . 68
6.4 A Subadditive Ergodic Theorem . . . 69
6.5 Applications . . . 71

7 Brownian Motion **72**
7.1 Definition and Construction . . . 72
7.2 Markov Property and Blumenthal's Zero–One Law . . . 74
7.3 Stopping Times and the Strong Markov Property . . . 75
7.4 Path Properties . . . 76
7.4.1 Zeros of Brownian Motion . . . 76
7.4.2 Hitting Times . . . 77
7.5 Brownian Martingales . . . 78
7.6 Itô's Formula* . . . 79

8 Applications to Random Walk **81**
8.1 Donsker's Theorem . . . 81
8.2 Central Limit Theorems for Martingales . . . 85
8.3 Central Limit Theorems for Stationary Sequences . . . 88
8.3.1 Mixing Properties . . . 90
8.4 Empirical Distributions and Brownian Bridge . . . 92
8.5 Laws of the Iterated Logarithm . . . 95

9 Multidimensional Brownian Motion **97**
9.1 Martingales . . . 97
9.2 Heat Equation . . . 98
9.3 Inhomogeneous Heat Equation . . . 99
9.4 Feynman–Kac Formula . . . 100
9.5 Dirichlet Problem . . . 101
9.5.1 Exit Distributions . . . 103
9.6 Green's Functions and Potential Kernels . . . 103
9.7 Poisson's Equation . . . 104
9.7.1 Occupation Times . . . 105
9.8 Schrödinger Equation . . . 106

A Measure Theory Details **108**
A.1 Carathéodory's Extension Theorem . . . 108
A.2 Which Sets Are Measurable? . . . 110
A.3 Kolmogorov's Extension Theorem . . . 112
A.4 Radon–Nikodym Theorem . . . 112
A.5 Differentiating under the Integral . . . 114

1 Measure Theory

1.1 Probability Spaces

Definition 1.1 (Probability space and measure). A *measurable space* is a pair $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-field: $\Omega \in \mathcal{F}$, $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$, and $A_n \in \mathcal{F}$ implies $\bigcup_{n\geq 1} A_n \in \mathcal{F}$. A measure is a map $\mu : \mathcal{F} \to [0, \infty]$ with $\mu(\varnothing) = 0$ and countable additivity on pairwise disjoint sets. A *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ has $\mathbb{P}(\Omega) = 1$; its measure $\mathbb{P}$ assigns probabilities to events.

A σ-field is also closed under countable intersections and set differences. Closure under uncountable unions need not hold.

Theorem 1.2 (Basic consequences of countable additivity; Book 1.1.1). Let μ be a measure and let all displayed sets be measurable.

1. If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
2. If $A \subseteq \bigcup_{m\geq 1} A_m$, then $\mu(A) \leq \sum_{m\geq 1} \mu(A_m)$.
3. If $A_n \uparrow A$, then $\mu(A_n) \uparrow \mu(A)$.
4. If $A_n \downarrow A$ and $\mu(A_1) < \infty$, then $\mu(A_n) \downarrow \mu(A)$.

Proof sketch. For monotonicity, decompose B as A plus $B \setminus A$. For subadditivity, disjointify the covering sets. For continuity from below, write an increasing union as the disjoint union of its successive increments. Continuity from above follows by applying continuity from below to $A_1 \setminus A_n$; finiteness of $\mu(A_1)$ is what makes the subtraction valid.

Remark 1.3 (Continuity from above under counting measure). Continuity from above can fail when $\mu(A_1) = \infty$. Under counting measure on $\mathbb{N}$, the sets $A_n = \{n, n+1, \ldots\}$ decrease to $\varnothing$, but every $\mu(A_n)$ is infinite.

Example 1.4 (Discrete probability space; Book 1.1.2). If Ω is countable and $p(\omega) \geq 0$ with $\sum_{\omega\in\Omega} p(\omega) = 1$, then

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

defines a probability measure on the power set of Ω. Conversely, every probability measure on this discrete measurable space is of this form, with $p(\omega) = \mathbb{P}(\{\omega\})$.

Definition 1.5 (Generated and Borel σ-fields). For a family $\mathcal{A}$ of subsets of Ω, $\sigma(\mathcal{A})$ is the intersection of all σ-fields containing $\mathcal{A}$. The Borel σ-field $\mathcal{B}(\mathbb{R}^d)$ is generated by the open subsets of $\mathbb{R}^d$. Equivalently, it is generated by open rectangles with rational endpoints.

A measure can often be specified on a generating class and then extended to the generated σ-field. Measurability can likewise be checked on a generating class.

Definition 1.6 (Stieltjes measure function). A function $F : \mathbb{R} \to \mathbb{R}$ is a Stieltjes measure function when it is nondecreasing and right-continuous. Its endpoint limits $F(-\infty)$ and $F(+\infty)$ are understood in the extended sense.

Theorem 1.7 (Lebesgue–Stieltjes construction; Book 1.1.4). For each Stieltjes measure function F there is a unique Borel measure μ_F such that

$$\mu_F((a, b]) = F(b) - F(a), \qquad -\infty \le a < b \le \infty,$$

whenever the difference is meaningful. The choice $F(x) = x$ gives Lebesgue measure on $\mathbb{R}$.

Remark 1.8. The choice of half-open intervals agrees with right continuity: if $b_n \downarrow b$, then $\bigcap_n (a, b_n] = (a, b]$ and $F(b_n) - F(a) \to F(b) - F(a)$.

Definition 1.9 (Semialgebra, algebra, and σ-finiteness). A semialgebra $\mathcal{S}$ contains $\varnothing$, is closed under intersections, and the complement of each member is a finite disjoint union of members. An algebra is closed under finite unions and complements. A measure is σ-finite if the whole space is a countable union of measurable sets of finite measure.

The basic semialgebra in $\mathbb{R}^d$ consists of $\varnothing$ and the half-open rectangles

$$(a_1, b_1] \times \cdots \times (a_d, b_d].$$

Finite disjoint unions of members of a semialgebra form its generated algebra Book 1.1.7. For example, finite disjoint unions of half-open intervals form an algebra on $\mathbb{R}$, whereas the finite and cofinite subsets of $\mathbb{Z}$ show that an algebra need not be a σ-field.

Theorem 1.10 (Extension from a semialgebra; Book 1.1.9). Let μ_0 be nonnegative on a semialgebra $\mathcal{S}$, with $\mu_0(\varnothing) = 0$. Suppose it is finitely additive under finite disjoint decompositions inside $\mathcal{S}$ and countably subadditive whenever a member of $\mathcal{S}$ is covered by countably many members. Then μ_0 extends to a measure on the algebra of finite disjoint unions. If this extension is σ-finite, its extension to $\sigma(\mathcal{S})$ is unique.

Proof sketch. Refining disjoint decompositions shows that the value assigned to a finite union is independent of its representation. The lemma on finite covers Book 1.1.10 gives monotonicity and finite subadditivity on the algebra. The outer measure construction extends the measure to the generated σ-field. Uniqueness on sets of finite measure gives uniqueness on the whole space by σ-finiteness.

For a finite rectangle $A = \prod_{i=1}^{d} (a_i, b_i]$, define its F-increment by

$$\Delta_A F = \sum_{v \in \{a_1, b_1\} \times \cdots \times \{a_d, b_d\}} (-1)^{\#\{i : v_i = a_i\}} F(v).$$

Theorem 1.11 (Distribution functions in $\mathbb{R}^d$; Book 1.1.11). Suppose $F : \mathbb{R}^d \to [0,1]$ is nondecreasing in each coordinate and right-continuous under coordinatewise decrease. Assume $F(x) \to 0$ whenever any one coordinate of x tends to $-\infty$, while $F(x) \to 1$ when all coordinates tend to $+\infty$. If F is also d-increasing, in the sense that $\Delta_A F \geq 0$ for every finite rectangle A, then there is a unique Borel probability measure μ satisfying $\mu(A) = \Delta_A F$ for every such rectangle.

Remark 1.12 (Rectangle increments). In dimensions $d \geq 2$, a coordinatewise nondecreasing, right-continuous function can assign a negative alternating increment to a rectangle. The condition $\Delta_A F \geq 0$ requires every rectangle to have nonnegative mass.

Example 1.13 (Product distribution). If $F(x) = \prod_{i=1}^{d} F_i(x_i)$, then

$$\Delta_A F = \prod_{i=1}^{d} \bigl(F_i(b_i) - F_i(a_i)\bigr) \geq 0.$$

This is the distribution function associated with independent coordinates having marginal distribution functions F_i.

1.2 Distributions

Definition 1.14 (Random variable and indicator). A real-valued random variable is a measurable map $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Thus $X^{-1}(B) \in \mathcal{F}$ for every Borel set B. For $A \in \mathcal{F}$, the indicator $\mathbf{1}_A$ is the random variable equal to one on A and zero off A.

Definition 1.15 (Law or pushforward distribution). The law of X is the probability measure

$$\mathcal{L}(X) = \mathbb{P} \circ X^{-1}, \qquad \mathcal{L}(X)(B) = \mathbb{P}(X \in B).$$

Two random variables defined on different probability spaces can have the same law. Equality in distribution is denoted by $X \overset{d}{=} Y$.

Remark 1.16 (Equality in distribution). If X and Y are independent fair signs, then $X \overset{d}{=} Y$ while $\mathbb{P}(X = Y) = 1/2$. Equality in distribution therefore differs from equality almost surely.

Definition 1.17 (Distribution function). The distribution function of a real random variable X is

$$F_X(x) = \mathbb{P}(X \leq x).$$

Theorem 1.18 (Properties of distribution functions; Book 1.2.1). Every F_X is nondecreasing and right-continuous, with $F_X(x) \to 0$ as $x \to -\infty$ and $F_X(x) \to 1$ as $x \to +\infty$. Moreover,

$$F_X(x-) = \mathbb{P}(X < x), \qquad \mathbb{P}(X = x) = F_X(x) - F_X(x-).$$

Proof sketch. Translate numerical limits into monotone limits of events. For example, $\{X \le y\} \downarrow \{X \le x\}$ as $y \downarrow x$, so continuity from above gives right continuity. Approaching from the left produces $\{X < x\}$, and the jump is the mass of the singleton.

Theorem 1.19 (Quantile realization; Book 1.2.2). Conversely, every nondecreasing, right-continuous $F : \mathbb{R} \to [0, 1]$ with endpoint limits 0 and 1 is a distribution function. If U is uniform on $(0, 1)$ and

$$F^{-1}(u) = \inf\{x : F(x) \ge u\},$$

then $F^{-1}(U)$ has distribution function F.

Proof sketch. The key equivalence is $\{F^{-1}(U) \le x\} = \{U \le F(x)\}$. Taking probabilities gives $F(x)$. The generalized inverse gives this equivalence even when F has flat portions or jumps.

Example 1.20 (Simulation and coupling). The same uniform variable can generate many laws: $X = F^{-1}(U)$ and $Y = G^{-1}(U)$. If $F^{-1}(u) \le G^{-1}(u)$ for every u, then this construction couples the variables with $X \le Y$ almost surely. This is the basic quantile coupling.

Definition 1.21 (Density, discrete law, and singular law). A law is absolutely continuous when $F(x) = \int_{-\infty}^{x} f(y)\,dy$ for a nonnegative f with integral one. It is discrete when it is concentrated on a countable set. A law μ is singular with respect to Lebesgue measure λ when $\mu \perp \lambda$, equivalently when it is concentrated on a Lebesgue-null set. A singular law may be discrete; if it is atomless, it is singular continuous, as in the Cantor distribution.

Example 1.22 (Three standard densities). The uniform law on $(0, 1)$ has $f(x) = \mathbf{1}_{(0,1)}(x)$. The exponential law of rate $\lambda > 0$ has $f(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0,\infty)}(x)$. The standard normal law has

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

All three assign probability zero to each singleton even though their densities are positive at many points.

Theorem 1.23 (Normal tail estimate; Book 1.2.6). For $x > 0$,

$$(x^{-1} - x^{-3})e^{-x^2/2} \le \int_x^\infty e^{-y^2/2}\,dy \le x^{-1}e^{-x^2/2}.$$

After multiplication by $(2\pi)^{-1/2}$ these are bounds for the standard normal upper tail.

Proof sketch. Integration by parts gives

$$\int_x^\infty e^{-y^2/2}\,dy = x^{-1}e^{-x^2/2} - \int_x^\infty y^{-2}e^{-y^2/2}\,dy.$$

The remainder is nonnegative and at most $x^{-2}\int_x^\infty e^{-y^2/2}\,dy \le x^{-3}e^{-x^2/2}$. Consequently, $\mathbb{P}(Z > x) \sim (2\pi)^{-1/2}x^{-1}e^{-x^2/2}$ as $x \to \infty$.

Example 1.24 (Continuous but singular). The Cantor distribution has a continuous distribution function and no atoms, yet all of its mass lies on the Cantor set, which has Lebesgue measure zero. Thus "continuous distribution function" does not mean "has a density."

Example 1.25 (Discrete law with dense jumps). Enumerate $\mathbb{Q}$ as (q_i) and choose $a_i > 0$ with $\sum_i a_i = 1$. Then $F(x) = \sum_i a_i \mathbf{1}_{[q_i,\infty)}(x)$ is a valid distribution function whose discontinuities are dense. Monotone functions still have only countably many jumps.

1.3 Random Variables

Definition 1.26 (Measurable map and generated σ-field). For measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$, a map $X : \Omega \to S$ is measurable if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{S}$. The σ-field

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

is the smallest σ-field on Ω that makes X measurable.

Theorem 1.27 (Generator test; Book 1.3.1). If $\mathcal{A}$ generates $\mathcal{S}$ and $X^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{A}$, then X is measurable.

Proof sketch. The class of $B \subseteq S$ whose inverse images are in $\mathcal{F}$ is itself a σ-field. Since it contains the generator, it contains all of $\sigma(\mathcal{A}) = \mathcal{S}$.

For real-valued maps it is enough to test sets $(-\infty, x]$, or even $(-\infty, q)$ with $q \in \mathbb{Q}$. For $\mathbb{R}^d$, open rectangles form a convenient generator. The rational version is especially useful because it keeps all unions countable.

Theorem 1.28 (Composition; Book 1.3.4). If $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ are measurable, then $f \circ X$ is measurable.

Theorem 1.29 (Measurable functions of random vectors; Book 1.3.5). If $X_1, \dots, X_n$ are random variables and $f : \mathbb{R}^n \to \mathbb{R}$ is Borel measurable, then $f(X_1, \dots, X_n)$ is a random variable. In particular, finite sums, products, maxima, and minima are measurable.

Proof sketch. First show $(X_1, \dots, X_n)$ is measurable by checking inverse images of rectangles; then apply the composition theorem. Continuity of addition gives the special sum result Book 1.3.6.

Theorem 1.30 (Countable extrema and limits; Book 1.3.7). For random variables (X_n), each of

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

is measurable with values in the extended real line.

Proof sketch. Use $\{\inf_n X_n < a\} = \bigcup_n \{X_n < a\}$ and the analogous identity for a supremum. Then express liminf as $\sup_n \inf_{m\geq n} X_m$ and limsup as $\inf_n \sup_{m\geq n} X_m$.

Definition 1.31 (Almost sure convergence). For a real random variable X, the event $\{X_n \to X\}$ is measurable because

$$\{X_n \to X\} = \left\{ \limsup_n X_n = \liminf_n X_n = X \right\}.$$

We write $X_n \to X$ a.s. when this event has probability one.

Remark 1.32 (Extended values). Limsup and liminf take values in $[-\infty, \infty]$. Equality of the two gives convergence in the extended real line; convergence to a real limit additionally requires their common value to be finite.

Proposition 1.33 (Doob–Dynkin principle, from the section exercises). A real random variable Y is $\sigma(X)$-measurable if and only if there is a Borel map f such that $Y = f(X)$.

Remark 1.34 (Uncountable suprema). An uncountable pointwise supremum of measurable functions need not be measurable. For example, on $\mathbb{R}$ with its Borel σ-field, let A be non-Borel. Each $\mathbf{1}_{\{a\}}$ is measurable, whereas $\sup_{a\in A} \mathbf{1}_{\{a\}} = \mathbf{1}_A$ is not.

1.4 Integration

Let $(\Omega, \mathcal{F}, \mu)$ be σ-finite. The integral is first defined for simple functions, then extended to nonnegative functions by approximation and to integrable signed functions by their positive and negative parts.

Definition 1.35 (Almost everywhere). The relation $f \geq g$ μ-a.e. means $\mu(\{\omega : f(\omega) < g(\omega)\}) = 0$. Functions equal a.e. represent the same element of an L^p space and have the same integral whenever the integral is defined.

Definition 1.36 (Simple functions). A finite simple function has a representation

$$\phi = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i},$$

where the A_i are disjoint measurable sets of finite measure. Define $\int \phi \, \mathrm{d}\mu = \sum_i a_i \mu(A_i)$.

Refining two representations by intersections proves that the definition is independent of representation. On simple functions the integral is positive, homogeneous, and additive Book 1.4.1; consequently it is monotone, invariant under a.e. equality, and satisfies

$$\left| \int \phi \, \mathrm{d}\mu \right| \leq \int |\phi| \, \mathrm{d}\mu$$

Book 1.4.2.

Definition 1.37 (Bounded functions supported on a set of finite measure). If f is bounded and vanishes outside E with $\mu(E) < \infty$, set

$$\int f \, \mathrm{d}\mu = \sup_{\phi \leq f} \int \phi \, \mathrm{d}\mu = \inf_{\psi \geq f} \int \psi \, \mathrm{d}\mu,$$

where ϕ, ψ are simple and supported on E.

Proof sketch. Partition the range of a bounded f into a mesh of width M/n. The lower and upper simple approximations differ by at most $(M/n)\mathbf{1}_E$, whose integral tends to zero. This proves equality of the lower and upper definitions and carries the six elementary properties to this class Book 1.4.3.

Definition 1.38 (Nonnegative functions). For measurable $f \geq 0$, define

$$\int f \, \mathrm{d}\mu = \sup_{h} \int h \, \mathrm{d}\mu,$$

where the supremum is over bounded measurable h with $0 \leq h \leq f$ that vanish outside a set of finite measure. The value $+\infty$ is allowed.

If $E_n \uparrow \Omega$ with $\mu(E_n) < \infty$, then

$$\int_{E_n} (f \wedge n)\, d\mu \uparrow \int f\, d\mu$$

Book 1.4.4. Positivity, positive homogeneity, additivity, monotonicity, and a.e. invariance continue to hold **Book 1.4.5**.

Definition 1.39 (Integrable signed functions). Write $f^+ = f \vee 0$ and $f^- = (-f) \vee 0$. The function f is integrable if $\int |f|\, d\mu < \infty$, and then

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

Theorem 1.40 (Algebra of the integral; Book 1.4.7). For integrable f, g and $a \in \mathbb{R}$,

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

The integral preserves a.e. order and a.e. equality, and

$$\left| \int f\, d\mu \right| \le \int |f|\, d\mu.$$

Remark 1.41 (Existence versus finiteness). For a signed f, the expression $\int f\, d\mu$ is not defined if both $\int f^+\, d\mu$ and $\int f^-\, d\mu$ are infinite. If exactly one is infinite, the integral has the corresponding sign and infinite value.

Special cases recover familiar notation: Lebesgue integrals on $\mathbb{R}^d$, Stieltjes integrals $\int f\, dF$, and sums under counting measure. Hence the convergence theorems for integrals automatically include convergence theorems for series.

1.5 Properties of the Integral

Theorem 1.42 (Jensen's inequality; Book 1.5.1). Let μ be a probability measure, let ϕ be convex, and suppose f and $\phi(f)$ are integrable. Then

$$\phi\left(\int f\, d\mu \right) \le \int \phi(f)\, d\mu.$$

Proof sketch. At $c = \int f\, d\mu$, choose a supporting affine function $\ell(x) = a(x - c) + \phi(c)$ with $\ell \le \phi$. Monotonicity and linearity yield $\int \phi(f)\, d\mu \ge \int \ell(f)\, d\mu = \phi(c)$.

For $1 \le p < \infty$, write $\|f\|_p = (\int |f|^p\, d\mu)^{1/p}$, and let $\|f\|_\infty$ be the essential supremum.

Theorem 1.43 (Hölder's inequality; Book 1.5.2). If $p, q \in (1, \infty)$ and $p^{-1} + q^{-1} = 1$, then

$$\int |fg| \, \mathrm{d}\mu \le \|f\|_p \, \|g\|_q .$$

The case $p = q = 2$ is Cauchy–Schwarz. Together with the endpoint cases, Hölder also implies Minkowski's inequality $\|f + g\|_p \le \|f\|_p + \|g\|_p$.

Proof sketch. After normalizing both nonzero norms to one, Young's inequality $xy \le x^p/p + y^q/q$ gives the result on integration. Equality holds precisely when $|f|^p$ and $|g|^q$ are proportional almost everywhere.

Definition 1.44 (Convergence in measure). We say $f_n \to f$ in measure if for every $\varepsilon > 0$,

$$\mu(\{|f_n - f| > \varepsilon\}) \to 0.$$

On spaces of finite measure, a.e. convergence implies convergence in measure, but the converse need not hold for the whole sequence.

Theorem 1.45 (Bounded convergence on finite measure; Book 1.5.3). Suppose $\mu(E) < \infty$, every f_n vanishes on E^c, $|f_n| \le M$, and $f_n \to f$ in measure. Then $\int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu$.

Proof sketch. Split E into $\{|f_n - f| < \varepsilon\}$ and its complement. The first contribution is at most $\varepsilon \mu(E)$; the second is at most $2M$ times a measure tending to zero.

Remark 1.46 (Bounded convergence on an infinite measure space). For Lebesgue measure on $\mathbb{R}$, let $f_n = n^{-1} \mathbf{1}_{[0,n]}$. Then $f_n \to 0$ uniformly and is uniformly bounded, but $\int f_n = 1$ for every n.

Theorem 1.47 (Fatou's lemma; Book 1.5.5). For measurable $f_n \ge 0$,

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

Proof sketch. Set $g_n = \inf_{m \ge n} f_m$. Then $g_n \uparrow \liminf f_n$ and $g_n \le f_m$ for every $m \ge n$. Approximate g_n on sets of finite measure by bounded truncations, apply bounded convergence, and pass through the two suprema.

Remark 1.48. Fatou's inequality can be strict. For Lebesgue measure, $f_n = n \mathbf{1}_{(0,1/n]}$ has integral one for every n and pointwise limit zero.

Theorem 1.49 (Monotone convergence; Book 1.5.7). If $0 \le f_n \uparrow f$, then $\int f_n \, d\mu \uparrow \int f \, d\mu$, with $+\infty$ allowed.

Theorem 1.50 (Dominated convergence; Book 1.5.8). If $f_n \to f$ a.e., $|f_n| \le g$ for every n, and g is integrable, then

$$\int f_n \, d\mu \to \int f \, d\mu.$$

Proof sketch. Apply Fatou to $g + f_n$ and to $g - f_n$. The two resulting inequalities bound the liminf and limsup of the integrals by $\int f$. Integrability of g permits the subtraction of $\int g$.

1.6 Expected Value

For a nonnegative random variable, its expectation is $\mathbb{E}X = \int X \, d\mathbb{P}$, possibly $+\infty$. For a signed variable, $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$ exists when at least one term is finite; integrability means $\mathbb{E}\,|X| < \infty$.

Theorem 1.51 (Linearity and monotonicity; Book 1.6.1). If X, Y are both nonnegative or both integrable, then $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$. For integrable X, $\mathbb{E}(aX + b) = a\mathbb{E}X + b$. Whenever expectations are well defined, $X \ge Y$ a.s. implies $\mathbb{E}X \ge \mathbb{E}Y$.

1.6.1 Inequalities

Theorem 1.52 (Jensen for random variables; Book 1.6.2). If ϕ is convex and both X and $\phi(X)$ are integrable, then

$$\phi(\mathbb{E}X) \le \mathbb{E}\phi(X).$$

For strictly convex ϕ, equality holds if and only if $X = \mathbb{E}X$ a.s.

Useful cases are $|\mathbb{E}X| \le \mathbb{E}\,|X|$ and $(\mathbb{E}X)^2 \le \mathbb{E}X^2$. Concave functions reverse the inequality.

Theorem 1.53 (Hölder for random variables; Book 1.6.3). For conjugate $p, q \in [1, \infty]$,

$$\mathbb{E}\,|XY| \le \|X\|_p \, \|Y\|_q .$$

In particular, $|\mathrm{Cov}(X, Y)| \le \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$ when the variances are finite.

Theorem 1.54 (General Markov–Chebyshev bound; Book 1.6.4). Let $\phi \ge 0$, let $A \subseteq \mathbb{R}$ be Borel, and put $i_A = \inf_{y \in A} \phi(y)$. Then

$$i_A \mathbb{P}(X \in A) \le \mathbb{E}(\phi(X); X \in A) \le \mathbb{E}\phi(X).$$

Taking $\phi(x) = |x|^p$ and $A = \{|x| \ge a\}$ gives

$$\mathbb{P}(|X| \ge a) \le \frac{\mathbb{E}\,|X|^p}{a^p}.$$

Applying this to $X - \mathbb{E}X$ with $p = 2$ gives the familiar variance bound $\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \mathrm{Var}(X)/a^2$.

Example 1.55 (A complementary lower bound). For $Y \geq 0$ with $0 < \mathbb{E}Y^2 < \infty$, Cauchy–Schwarz applied to $Y\mathbf{1}_{\{Y>0\}}$ yields

$$\mathbb{P}(Y > 0) \geq \frac{(\mathbb{E}Y)^2}{\mathbb{E}Y^2}.$$

In particular, $\mathbb{E}Y > 0$ implies $\mathbb{P}(Y > 0) > 0$.

1.6.2 Integration to the Limit

Theorem 1.56 (Expectation versions; Book 1.6.5–Book 1.6.7). For nonnegative X_n,

$$\mathbb{E}\liminf_n X_n \leq \liminf_n \mathbb{E}X_n.$$

If $0 \leq X_n \uparrow X$, then $\mathbb{E}X_n \uparrow \mathbb{E}X$. If $X_n \to X$ a.s., $|X_n| \leq Y$, and $\mathbb{E}Y < \infty$, then $\mathbb{E}X_n \to \mathbb{E}X$.

Theorem 1.57 (Convergence of expectations under a moment bound; Book 1.6.8). Suppose $X_n \to X$ a.s. Let continuous functions g, h satisfy $g \geq 0$, $g(x) \to \infty$ as $|x| \to \infty$, and $|h(x)|/g(x) \to 0$. If $\sup_n \mathbb{E}g(X_n) < \infty$, then $\mathbb{E}h(X_n) \to \mathbb{E}h(X)$.

Proof sketch. Truncate X_n to $|X_n| \leq M$. Bounded convergence handles the truncated variables. Outside the truncation region, the ratio $\sup_{|x|\geq M} |h(x)|/g(x)$ controls the error uniformly in n. Fatou transfers the g-moment bound to X, after which $M \to \infty$.

Remark 1.58 (A bound on higher moments). Taking $g(x) = |x|^p$ with $p > 1$ and $h(x) = x$ gives $\mathbb{E}X_n \to \mathbb{E}X$ whenever $X_n \to X$ a.s. and $\sup_n \mathbb{E}|X_n|^p < \infty$.

1.6.3 Computing Expected Values

Theorem 1.59 (Change of variables by the law; Book 1.6.9). Let X take values in $(S, \mathcal{S})$ and have law μ. If f is measurable and either $f \geq 0$ or $\mathbb{E}|f(X)| < \infty$, then

$$\mathbb{E}f(X) = \int_S f(y)\mu(\mathrm{d}y).$$

This result is often called the law of the unconscious statistician.

Proof sketch. Verify the identity for indicators directly from the definition of pushforward measure. Extend to simple functions by linearity, to nonnegative functions by monotone simple approximation, and to integrable functions by positive and negative parts.

If $\mu = \mathbb{E}X$ and $\mathbb{E}X^2 < \infty$, then

$$\mathrm{Var}(X) = \mathbb{E}(X - \mu)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2, \qquad \mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X).$$

Example 1.60 (Exponential and normal moments). If $X \sim \mathrm{Exp}(\lambda)$, then $\mathbb{E}X = \lambda^{-1}$ and $\mathrm{Var}(X) = \lambda^{-2}$. If Z is standard normal, then $\mathbb{E}Z = 0$ and $\mathrm{Var}(Z) = 1$; hence $\sigma Z + m$ has mean m and variance σ^2.

Example 1.61 (Bernoulli, Poisson, and geometric laws). For $B \sim \mathrm{Bernoulli}(p)$, $\mathbb{E}B = p$ and $\mathrm{Var}(B) = p(1-p)$. For $N \sim \mathrm{Poisson}(\lambda)$, factorial moments satisfy

$$\mathbb{E}[N(N-1)\cdots(N-k+1)] = \lambda^k,$$

so $\mathbb{E}N = \mathrm{Var}(N) = \lambda$. If G counts trials through the first success with success probability p, then $\mathbb{E}G = p^{-1}$ and $\mathrm{Var}(G) = (1-p)/p^2$.

Remark 1.62 (Two conventions for the geometric law). Some authors define a geometric variable as the number of failures before the first success, with support $\{0, 1, \ldots\}$. Durrett's convention counts the successful trial and has support $\{1, 2, \ldots\}$. The two means differ by one.

1.7 Product Measures, Fubini's Theorem

Let $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ be σ-finite measure spaces. The product σ-field $\mathcal{A} \otimes \mathcal{B}$ is generated by rectangles $A \times B$.

Theorem 1.63 (Product measure; Book 1.7.1). There is a unique measure $\mu_1 \times \mu_2$ on $\mathcal{A} \otimes \mathcal{B}$ satisfying

$$(\mu_1 \times \mu_2)(A \times B) = \mu_1(A)\mu_2(B).$$

Finite products follow by induction.

Proof sketch. Rectangles form a semialgebra. To check countable additivity for a disjoint rectangle decomposition, fix one coordinate, obtain an indicator identity for the other, and integrate. The semialgebra extension theorem then constructs the product measure; σ-finiteness supplies uniqueness.

For $E \subseteq X \times Y$, define the section $E_x = \{y : (x, y) \in E\}$. For measurable E, the section E_x is measurable Book 1.7.3, and $x \mapsto \mu_2(E_x)$ is measurable with

$$\int_X \mu_2(E_x)\mu_1(\mathrm{d}x) = (\mu_1 \times \mu_2)(E)$$

Book 1.7.4.

Theorem 1.64 (Tonelli–Fubini; Book 1.7.2). If $f : X \times Y \to [-\infty, \infty]$ is measurable and either $f \geq 0$ or $\int |f| \, \mathrm{d}(\mu_1 \times \mu_2) < \infty$, then

$$\int_{X\times Y} f \, \mathrm{d}(\mu_1 \times \mu_2) = \int_X \left(\int_Y f(x,y)\mu_2(\,\mathrm{d}y) \right) \mu_1(\,\mathrm{d}x) = \int_Y \left(\int_X f(x,y)\mu_1(\,\mathrm{d}x) \right) \mu_2(\,\mathrm{d}y).$$

For $f \geq 0$ the common value may be infinite; for signed f, absolute integrability makes all three quantities finite.

Proof sketch. Start with $f = \mathbf{1}_E$ and use measurable sections. Extend by linearity to simple functions, by monotone convergence to nonnegative functions, and by the decomposition $f = f^+ - f^-$ to absolutely integrable signed functions.

Example 1.65 (Failure without absolute integrability). Under counting measure on $\mathbb{N}^2$, put $f(m,m) = 1$, $f(m+1,m) = -1$, and zero elsewhere. Summing one coordinate first gives 1, while reversing the order gives 0. The positive and negative parts both have infinite mass, so Fubini does not apply.

Example 1.66 (Failure without σ-finiteness). On $(0,1)^2$, use Lebesgue measure in one coordinate, counting measure on all subsets in the other, and $f(x,y) = \mathbf{1}_{\{x=y\}}$. A vertical section has counting mass one, whereas a horizontal singleton has Lebesgue mass zero. The iterated integrals differ.

Remark 1.67 (Measurable sections and product measurability). Even when every one-dimensional section is measurable, the subset of the product need not belong to the product σ-field.

Corollary 1.68 (Layer-cake formula). For measurable $g \geq 0$,

$$\int_X g \, \mathrm{d}\mu = \int_0^\infty \mu(\{x : g(x) > t\}) \, \mathrm{d}t.$$

Indeed, integrate the indicator of the subgraph $\{(x,t) : 0 \leq t < g(x)\}$ in the two orders.

2 Laws of Large Numbers

2.1 Independence

Definition 2.1 (Independent events, variables, and σ-fields). Events A and B are independent when $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Random variables X and Y are independent when

$$\mathbb{P}(X \in C, Y \in D) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D), \qquad C, D \in \mathcal{B}(\mathbb{R}).$$

Sigma-fields $\mathcal{F}$ and $\mathcal{G}$ are independent when the event identity holds for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$. An infinite family is independent when each finite subfamily is. For events $A_1, \ldots, A_n$, mutual independence requires the product identity for every subcollection. **Book Definitions preceding Theorem 2.1.1**

Theorem 2.2 (Generated sigma-fields, Theorems 2.1.1–2.1.3). Independence of X and Y is equivalent to independence of $\sigma(X)$ and $\sigma(Y)$. Complements of independent events remain independent, and independent events have independent indicators. More generally, replacing any members of an independent finite family by their complements preserves independence.

Proof sketch. Measurability turns inverse images of Borel sets into events in the relevant sigma-fields. For complements, subtract $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ from $\mathbb{P}(B)$; iterate this identity and then enumerate the possible inverse images of indicator variables.

Example 2.3 (Pairwise does not mean mutual, Example 2.1.4). Let X_1, X_2, X_3 be independent fair bits and set $A_1 = \{X_2 = X_3\}$, $A_2 = \{X_3 = X_1\}$, and $A_3 = \{X_1 = X_2\}$. Every pair has intersection probability $1/4$, but

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = \prod_{i=1}^{3} \mathbb{P}(A_i).$$

The intersection of any two of these events is already their common intersection.

2.1.1 Sufficient Conditions for Independence

Definition 2.4 (π-systems and λ-systems). A collection $\mathcal{P}$ is a π-system if it is closed under finite intersections. A collection $\mathcal{L}$ is a λ-system if it contains the whole space, is closed under proper differences, and is closed under increasing countable unions. Collections $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent when choosing one event from each selected collection gives the appropriate product probability.

Lemma 2.5 (Theorem 2.1.6, the π–λ theorem). If a λ-system $\mathcal{L}$ contains a π-system $\mathcal{P}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.

Lemma 2.6 (Lemma 2.1.5). One may adjoin the whole sample space to each independent generating collection. After doing so, it is enough to verify the product formula using one event from every collection; unused coordinates are represented by the whole space.

Theorem 2.7 (Independence from generators, Theorem 2.1.7). If $\mathcal{A}_1, \dots, \mathcal{A}_n$ are independent π-systems, then $\sigma(\mathcal{A}_1), \dots, \sigma(\mathcal{A}_n)$ are independent.

Proof sketch. Fix events from all collections except the first and put their intersection equal to F. The class

$$\mathcal{L} = \{A : \mathbb{P}(A \cap F) = \mathbb{P}(A)\mathbb{P}(F)\}$$

is a λ-system containing $\mathcal{A}_1$, so it contains $\sigma(\mathcal{A}_1)$. Repeating the argument for the remaining collections proves the result.

Corollary 2.8 (Criterion using distribution functions, Theorem 2.1.8). Random variables $X_1, \dots, X_n$ are independent if

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \le x_i)$$

for every $x_i \in (-\infty, \infty]$. Lower half-lines form a π-system that generates the Borel sigma-field.

Theorem 2.9 (Independent blocks, Theorems 2.1.9–2.1.10). Group an independent family of sigma-fields into disjoint blocks and generate one sigma-field from each block; the block sigma-fields are independent. Consequently, measurable functions of disjoint blocks of independent random variables are independent.

Remark 2.10 (Partial sums and later increments). If $S_k = X_1 + \cdots + X_k$ and $m < n$, then every event determined by $X_1, \dots, X_m$ is independent of the future increment $S_n - S_m$. In particular, S_m and $S_n - S_m$ are independent.

2.1.2 Independence, Distribution, and Expectation

Theorem 2.11 (Product law, Theorem 2.1.11). If independent $X_1, \dots, X_n$ have laws $\mu_1, \dots, \mu_n$, then

$$\mathcal{L}(X_1, \dots, X_n) = \mu_1 \otimes \cdots \otimes \mu_n.$$

It is enough to compare both measures on measurable rectangles and then use the π–λ theorem.

Theorem 2.12 (Integration and factorization, Theorems 2.1.12–2.1.13). For independent $X \sim \mu$ and $Y \sim \nu$,

$$\mathbb{E}h(X,Y) = \iint h(x,y)\,\mu(\mathrm{d}x)\nu(\mathrm{d}y)$$

whenever $h \geq 0$ or $\mathbb{E}\,|h(X,Y)| < \infty$. Hence, for nonnegative or integrable factors, $\mathbb{E}[f(X)g(Y)] = \mathbb{E}f(X)\,\mathbb{E}g(Y)$. By induction, the expectation of a product of independent integrable variables is the product of their expectations.

Proof sketch. Use the product law to replace expectation by integration against $\mu \otimes \nu$, then apply Tonelli for nonnegative functions and Fubini for integrable functions. Tonelli also gives $\mathbb{E}|f(X)g(Y)| = \mathbb{E}|f(X)|\,\mathbb{E}|g(Y)| < \infty$ when both factors are integrable.

Example 2.13 (Uncorrelated but dependent, Example 2.1.14). Assign probability $1/4$ to each of $(1,0), (-1,0), (0,1), (0,-1)$. Then $XY = 0$ almost surely and $\mathbb{E}X = \mathbb{E}Y = 0$, so $\mathrm{Cov}(X,Y) = 0$. However, $\mathbb{P}(X = 1, Y = 1) = 0$ while $\mathbb{P}(X = 1)\mathbb{P}(Y = 1) = 1/16$.

Definition 2.14 (Uncorrelated variables). Square-integrable X and Y are uncorrelated when $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = 0$. Independence plus square integrability implies uncorrelatedness, but the converse fails.

2.1.3 Sums of Independent Random Variables

Theorem 2.15 (Convolution formulas, Theorems 2.1.15–2.1.16). If independent X, Y have distribution functions F, G, then

$$\mathbb{P}(X + Y \leq z) = \int F(z - y)\,\mathrm{d}G(y) = (F * G)(z).$$

If X has density f, then $X + Y$ has density $h(x) = \int f(x - y)\,\mathrm{d}G(y)$; if Y has density g, this becomes $h(x) = \int f(x - y)g(y)\,\mathrm{d}y$.

Proof sketch. Integrate the indicator $\mathbf{1}_{\{x+y\leq z\}}$ under the product law. For densities, write $F(z - y)$ as an integral of f and exchange the two nonnegative integrals.

Example 2.16 (Gamma closure, Example 2.1.17 and Theorem 2.1.18). If $X \sim \mathrm{Gamma}(\alpha, \lambda)$ and $Y \sim \mathrm{Gamma}(\beta, \lambda)$ are independent, then $X + Y \sim \mathrm{Gamma}(\alpha + \beta, \lambda)$. The convolution reduces, after $y = xu$, to the beta integral. In particular, a sum of n exponential variables of rate λ is $\mathrm{Gamma}(n, \lambda)$.

Example 2.17 (Normal closure, Example 2.1.19 and Theorem 2.1.20). For independent $X \sim N(\mu, a)$ and $Y \sim N(\nu, b)$, $X+Y \sim N(\mu+\nu, a+b)$. Completing the square in the convolution separates a normalizing Gaussian integral from the factor depending on z.

2.1.4 Constructing Independent Random Variables

For finitely many prescribed laws $\mu_1, \ldots, \mu_n$, use the coordinate maps on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_1 \otimes \cdots \otimes \mu_n)$. Infinite sequences require a consistent probability measure on an infinite product.

Theorem 2.18 (Kolmogorov extension, Theorem 2.1.21). If probability measures μ_n on $\mathbb{R}^n$ are consistent under coordinate projection, then there is a unique probability measure on $\mathbb{R}^{\mathbb{N}}$ having these finite-dimensional marginals. Coordinate projections under that measure realize the desired process.

Definition 2.19 (Nice or standard Borel space). A measurable space $(S, \mathcal{S})$ is nice when it is measurably isomorphic to a measurable subset of $\mathbb{R}$: there is a one-to-one $\phi : S \to \mathbb{R}$ for which both ϕ and its inverse on $\phi(S)$ are measurable.

Theorem 2.20 (Theorem 2.1.22). Every Borel subset of a complete separable metric space, equipped with its Borel sigma-field, is nice.

Proof sketch. First encode a point of $[0,1)^{\mathbb{N}}$ by interlacing binary digits. For a general separable metric space with dense set $\{x_j\}$, map x to $(d(x, x_j)/(1 + d(x, x_j)))_{j \geq 1}$. This is an injective measurable map into $[0,1)^{\mathbb{N}}$, to which the encoding applies.

2.2 Weak Laws of Large Numbers

Definition 2.21 (Convergence in probability). $Y_n \to Y$ in probability if, for every $\varepsilon > 0$, $\mathbb{P}(|Y_n - Y| > \varepsilon) \to 0$.

2.2.1 L^2 Weak Laws

Theorem 2.22 (Variance additivity, Theorem 2.2.1). If $X_1, \ldots, X_n$ are square-integrable and pairwise uncorrelated, then

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

The cross terms vanish after centering.

Lemma 2.23 (L^p implies probability convergence, Lemma 2.2.2). For $p > 0$, if $\mathbb{E}\,|Z_n|^p \to 0$, then $Z_n \to 0$ in probability, since $\mathbb{P}(|Z_n| \geq \varepsilon) \leq \varepsilon^{-p}\mathbb{E}\,|Z_n|^p$.

Theorem 2.24 (L^2 weak law, Theorem 2.2.3). Let X_i be uncorrelated with $\mathbb{E}X_i = \mu$ and $\mathrm{Var}(X_i) \leq C$. For $S_n = \sum_{i=1}^n X_i$,

$$\mathbb{E}\,|S_n/n - \mu|^2 = \mathrm{Var}(S_n/n) \leq C/n \to 0.$$

Thus $S_n/n \to \mu$ in L^2 and in probability. In particular, the conclusion holds for i.i.d. variables with finite variance.

Example 2.25 (Bernstein approximation, Example 2.2.4). For continuous f on $[0,1]$, the Bernstein polynomial satisfies $f_n(p) = \mathbb{E}f(S_n/n)$ with $S_n \sim \mathrm{Binomial}(n,p)$. The variance bound $p(1-p)/n \leq 1/(4n)$ and uniform continuity give the result by splitting the expectation over $\{|S_n/n - p| \leq \delta\}$ and its complement. Thus $\sup_{p\in[0,1]} |f_n(p) - f(p)| \to 0$.

Example 2.26 (High-dimensional cube, Example 2.2.5). For a point uniformly distributed on $(-1,1)^n$, $n^{-1}\sum X_i^2 \to 1/3$ in probability. Hence most cube volume lies in a thin spherical shell of radius approximately $\sqrt{n/3}$.

2.2.2 Triangular Arrays

Theorem 2.27 (Variance criterion, Theorem 2.2.6). For arbitrary S_n with mean μ_n and variance σ_n^2, if $\sigma_n^2/b_n^2 \to 0$, then $(S_n - \mu_n)/b_n \to 0$ in probability. This follows directly from Chebyshev's inequality.

Example 2.28 (Coupon collector, Example 2.2.7). Write the collection time as a sum of independent geometric waiting times. Then

$$\mathbb{E}T_n = n\sum_{m=1}^{n}\frac{1}{m} \sim n\log n, \qquad \mathrm{Var}(T_n) = O(n^2),$$

so $T_n/(n\log n) \to 1$ in probability.

Lemma 2.29 (Cycles in a random permutation, Lemma 2.2.9). The number C_n of cycles in a uniform random permutation has the same law as a sum of independent Bernoulli variables with success probabilities $1/n, 1/(n-1), \ldots, 1$. Sequentially exposing the permutation makes the choice that closes a cycle uniform among the unused images.

Example 2.30 (Random permutations, Example 2.2.8). Since $\mathbb{E}C_n = \sum_{m=1}^{n} 1/m \sim \log n$ and $\mathrm{Var}(C_n) \le \mathbb{E}C_n$, one gets $C_n/\log n \to 1$ in probability, with the sharper centering scale $(\log n)^{1/2+\varepsilon}$.

Example 2.31 (Occupancy, Example 2.2.10). If r balls are placed independently into n boxes and $r/n \to c$, the fraction N_n/n of empty boxes converges in probability to e^{-c}. Indicators of emptiness are dependent, but a direct calculation of the second moment gives $\mathrm{Var}(N_n/n) \to 0$.

2.2.3 Truncation

Definition 2.32 (Truncation). At level b, the truncated variable is $\bar{X} = X\mathbf{1}_{\{|X|\le b\}}$. It has finite variance, and $\mathbb{P}(X \ne \bar{X}) = \mathbb{P}(|X| > b)$.

Theorem 2.33 (Weak law for triangular arrays). Book Theorem 2.2.11. For each n, let $X_{n,1}, \dots, X_{n,n}$ be independent, $b_n \to \infty$, and $\bar{X}_{n,k} = X_{n,k}\mathbf{1}_{\{|X_{n,k}|\le b_n\}}$. If

$$\sum_{k=1}^{n} \mathbb{P}(|X_{n,k}| > b_n) \to 0, \qquad b_n^{-2} \sum_{k=1}^{n} \mathbb{E}\bar{X}_{n,k}^2 \to 0,$$

then, with $a_n = \sum_k \mathbb{E}\bar{X}_{n,k}$, $(\sum_k X_{n,k} - a_n)/b_n \to 0$ in probability.

Proof sketch. The original and truncated row sums disagree with probability at most the first displayed sum. Center the truncated sum, add its independent variances, and apply Chebyshev using the second displayed condition. Here a_n is the sum of the truncated means, which are finite.

Theorem 2.34 (Weak law under a tail condition, Theorem 2.2.12). If X_i are i.i.d. and $x\mathbb{P}(|X_1| > x) \to 0$, then with $\mu_n = \mathbb{E}[X_1\mathbf{1}_{\{|X_1|\le n\}}]$, $S_n/n - \mu_n \to 0$ in probability.

Lemma 2.35 (Tail integral identity, Lemma 2.2.13). For $Y \ge 0$ and $p > 0$,

$$\mathbb{E}Y^p = \int_0^\infty p y^{p-1}\mathbb{P}(Y > y)\, \mathrm{d}y.$$

It follows by writing $Y^p = \int_0^Y p y^{p-1}\, \mathrm{d}y$ and applying Tonelli.

Theorem 2.36 (Classical weak law, Theorem 2.2.14). If X_i are i.i.d. with $\mathbb{E}|X_1| < \infty$ and $\mu = \mathbb{E}X_1$, then $S_n/n \to \mu$ in probability. Integrability gives both $n\mathbb{P}(|X_1| > n) \to 0$ and convergence of the truncated means to μ.

Example 2.37 (Cauchy averages, Example 2.2.15). For i.i.d. Cauchy variables, $x\mathbb{P}(|X_1| > x)$ approaches a positive constant. No deterministic centering makes S_n/n converge in probability; in fact the Cauchy law is stable under averaging.

Example 2.38 (St. Petersburg, Example 2.2.16). If $\mathbb{P}(X = 2^j) = 2^{-j}$, then $\mathbb{E}X = \infty$, yet a carefully chosen truncation yields $S_n/(n \log_2 n) \to 1$ in probability. The normalization is determined by the growth of the truncated mean.

2.3 Borel–Cantelli Lemmas

Definition 2.39 (Event limsup and liminf).

$$\limsup_n A_n = \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n = \{A_n \text{ i.o.}\}, \qquad \liminf_n A_n = \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n.$$

The first event means infinitely many occurrences; the second means all but finitely many. For example, $X_n \to 0$ almost surely iff $\mathbb{P}(|X_n| > \varepsilon \text{ i.o.}) = 0$ for every $\varepsilon > 0$.

Theorem 2.40 (First Borel–Cantelli lemma, Theorem 2.3.1). If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$. Indeed $N = \sum_n \mathbf{1}_{A_n}$ has finite expectation and hence is finite almost surely.

Theorem 2.41 (Subsequence principle, Theorems 2.3.2–2.3.3). $X_n \to X$ in probability iff every subsequence has a further subsequence converging to X almost surely. Deterministically, if every subsequence of y_n has a further subsequence converging to y, then $y_n \to y$.

Proof sketch. From probability convergence, choose a further subsequence with error probabilities at most 2^{-k} and error thresholds decreasing to zero; Borel–Cantelli gives almost sure convergence. Conversely, a subsequence staying outside some neighborhood cannot contain a further subsequence converging to the proposed limit.

Theorem 2.42 (Continuous mapping and bounded expectations, Theorem 2.3.4). If $X_n \to X$ in probability and f is continuous, then $f(X_n) \to f(X)$ in probability. If f is also bounded, then $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$. Apply the subsequence principle and bounded convergence.

Theorem 2.43 (Strong law under a fourth moment assumption, Theorem 2.3.5). For i.i.d. X_i with mean μ and finite fourth moment, $S_n/n \to \mu$ almost surely. After centering, independence gives $\mathbb{E}S_n^4 = O(n^2)$, so $\mathbb{P}(|S_n| > n\varepsilon) = O(n^{-2})$ and Borel–Cantelli applies.

Example 2.44 (Nested events, Example 2.3.6). Nested events $A_n = (0, a_n)$ with $a_n \downarrow 0$ satisfy A_n i.o. $= \varnothing$, even when $\sum_n a_n = \infty$. Thus divergence of $\sum_n \mathbb{P}(A_n)$ can occur with $\mathbb{P}(A_n \text{ i.o.}) = 0$.

Theorem 2.45 (Second Borel–Cantelli lemma, Theorem 2.3.7). For independent events, $\sum_n \mathbb{P}(A_n) = \infty$ implies $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof sketch. Independence and $1 - x \le e^{-x}$ give

$$\mathbb{P}\left(\bigcap_{n=M}^{N} A_n^c\right) = \prod_{n=M}^{N} (1 - \mathbb{P}(A_n)) \le \exp\left(-\sum_{n=M}^{N} \mathbb{P}(A_n)\right) \to 0.$$

Then send $N \to \infty$ and intersect over M.

Theorem 2.46 (Divergence with infinite absolute mean, Theorem 2.3.8). If the X_n are i.i.d. with $\mathbb{E}\,|X_1| = \infty$, then $|X_n| \ge n$ infinitely often almost surely. Consequently S_n/n cannot converge to a finite real number almost surely. Indeed, convergence of S_n/n to a finite limit would imply $X_n/n = S_n/n - [(n-1)/n]S_{n-1}/(n-1) \to 0$.

Theorem 2.47 (Counts of pairwise independent events, Theorem 2.3.9). If A_n are pairwise independent and $\sum_n \mathbb{P}(A_n) = \infty$, then

$$\frac{\sum_{m=1}^{n} \mathbf{1}_{A_m}}{\sum_{m=1}^{n} \mathbb{P}(A_m)} \to 1 \qquad \text{almost surely.}$$

Proof sketch. The indicators are uncorrelated, so the numerator has variance no larger than its mean. Chebyshev gives convergence in probability. At indices where the expected count first exceeds k^2, the error probabilities are summable. Monotonicity bounds the counts between these indices, and the ratios of successive expected counts tend to one.

Example 2.48 (Records, Example 2.3.10 and Theorem 2.3.11). For i.i.d. observations with a continuous law, the event that observation k is a new record has probability $1/k$, and these record events are independent. Therefore the number R_n of records satisfies $R_n/\log n \to 1$ almost surely. Continuity of the law ensures that ties have probability zero.

Example 2.49 (Longest head run, Example 2.3.12). For fair independent signs, the longest run L_n of heads among the first n tosses obeys $L_n/\log_2 n \to 1$ almost surely. An upper bound uses Borel–Cantelli on excessively long runs; a lower bound divides the trials into independent blocks and shows that a slightly shorter block consisting entirely of heads eventually appears.

2.4 Strong Law of Large Numbers

Theorem 2.50 (Etemadi's strong law, Theorem 2.4.1). If $X_1, X_2, \dots$ are pairwise independent, identically distributed, and $\mathbb{E}\,|X_1| < \infty$, then $S_n/n \to \mu = \mathbb{E}X_1$ almost surely.

Lemma 2.51 (Truncation steps, Lemmas 2.4.2–2.4.4). Set $Y_k = X_k \mathbf{1}_{\{|X_k| \le k\}}$ and $T_n = \sum_{k \le n} Y_k$. Integrability and Borel–Cantelli show $X_k = Y_k$ eventually almost surely, so it suffices to study T_n/n. Moreover,

$$\sum_{k=1}^{\infty} \frac{\operatorname{Var}(Y_k)}{k^2} < \infty, \qquad 2y \sum_{k > y} k^{-2} \le 4.$$

The second bound, combined with the tail integral formula and Tonelli, proves the first.

Proof sketch. Reduce to nonnegative variables by positive and negative parts. For $k(n) = \lfloor \alpha^n \rfloor$, Chebyshev plus the variance series makes the deviations of $T_{k(n)}$ summable; hence $T_{k(n)}/k(n) \to \mu$ almost surely. Monotonicity bounds T_m/m using adjacent geometric indices, giving bounds μ/α and $\alpha\mu$. Finally let $\alpha \downarrow 1$.

Theorem 2.52 (An infinite positive mean, Theorem 2.4.5). If the i.i.d. variables satisfy $\mathbb{E}X_1^+ = \infty$ and $\mathbb{E}X_1^- < \infty$, then $S_n/n \to \infty$ almost surely. Apply the strong law to $X_i \wedge M$ and then send $M \to \infty$ using monotone convergence.

Example 2.53 (Inverse renewal count, Example 2.4.6 and Theorem 2.4.7). For positive i.i.d. lifetimes, $T_n = \sum_{k=1}^{n} X_k$ and $N_t = \sup\{n : T_n \le t\}$ satisfy $N_t/t \to 1/\mu$ almost surely, where $\mu = \mathbb{E}X_1 \in (0, \infty]$. The inequalities $T_{N_t} \le t < T_{N_t+1}$ invert the strong law.

Example 2.54 (Empirical distribution, Example 2.4.8 and Theorem 2.4.9). For i.i.d. data with cdf F, the empirical cdf $F_n(x) = n^{-1} \sum_{m=1}^{n} \mathbf{1}_{\{X_m \le x\}}$ obeys $\sup_x |F_n(x) - F(x)| \to 0$ almost surely (Glivenko–Cantelli). Pointwise strong laws at finitely many quantile cutpoints, together with monotonicity, produce uniform control; jumps of F require using both left and right limits.

Example 2.55 (Shannon's asymptotic equipartition property, Example 2.4.10). For i.i.d. letters with probabilities $p(k) > 0$, the probability of the observed word is $\pi_n = \prod_{i=1}^{n} p(X_i)$. The strong law applied to $-\log p(X_i)$ gives

$$-n^{-1} \log \pi_n \to H = -\sum_k p(k) \log p(k) \quad \text{almost surely.}$$

Thus typical words of length n have probabilities on the exponential scale e^{-nH}.

2.5 Convergence of Random Series*

Definition 2.56 (Tail and exchangeable sigma-fields). For independent X_n, let $\mathcal{T} = \bigcap_n \sigma(X_n, X_{n+1}, \ldots)$. Tail events are unchanged by altering finitely many coordinates. For an i.i.d. sequence, the exchangeable sigma-field consists of events invariant under finite permutations; it contains $\mathcal{T}$ and also events involving symmetric functions of partial sums.

Example 2.57 (Examples 2.5.1–2.5.2). Events such as $\{X_n \in B_n \text{ i.o.}\}$, convergence of $\sum_n X_n$, and $\{\limsup S_n/c_n > x\}$ for $c_n \to \infty$ are tail events. By contrast, $\{\limsup S_n > 0\}$ can change when one coordinate changes and need not be tail measurable.

Theorem 2.58 (Zero–one laws, Theorems 2.5.3–2.5.4). For independent variables every $A \in \mathcal{T}$ has probability 0 or 1 (Kolmogorov). For i.i.d. variables every event invariant under finite coordinate permutations has probability 0 or 1 (Hewitt–Savage).

Proof sketch. For Kolmogorov, the sigma-field generated by the first n variables is independent of $\sigma(X_{n+1}, X_{n+2}, \ldots)$; pass to the generated sigma-field to make a tail event independent of itself. For Hewitt–Savage, approximate A by an event in the first n coordinates, swap that block with the next n coordinates, and use invariance plus independence of the two approximants.

Theorem 2.59 (Kolmogorov maximal inequality, Theorem 2.5.5). For independent centered variables $X_1, \ldots, X_n$ with finite variance and $S_k = \sum_{i \le k} X_i$,

$$\mathbb{P}\Big(\max_{1 \le k \le n} |S_k| \ge x\Big) \le x^{-2} \operatorname{Var}(S_n).$$

Partition according to the first crossing time. The future increment has conditional mean zero and is independent of the crossing event, so all mixed terms vanish.

Theorem 2.60 (Summable variances, Theorem 2.5.6). If independent centered X_n satisfy $\sum_n \operatorname{Var}(X_n) < \infty$, then $\sum_n X_n$ converges almost surely. Apply the maximal inequality to tails of partial sums; their oscillations tend to zero in probability and, being decreasing in the starting index, vanish almost surely.

Example 2.61 (Conditional but not absolute convergence, Example 2.5.7). For independent random signs $X_n = \pm n^{-\alpha}$, the series converges almost surely when $\alpha > 1/2$ by the variance test, whereas absolute convergence requires $\alpha > 1$.

Theorem 2.62 (Kolmogorov three-series theorem, Theorem 2.5.8). For independent X_n and fixed $A > 0$, put $Y_n = X_n \mathbf{1}_{\{|X_n| \le A\}}$. Then $\sum_n X_n$ converges almost surely iff

$$\sum_n \mathbb{P}(|X_n| > A) < \infty, \qquad \sum_n \mathbb{E}Y_n \text{ converges}, \qquad \sum_n \operatorname{Var}(Y_n) < \infty.$$

Sufficiency follows from Borel–Cantelli and almost sure convergence of the centered series with summable variances. Necessity follows by symmetrization and the central limit theorem for bounded independent summands.

Lemma 2.63 (Kronecker's lemma, Theorem 2.5.9). If $a_n \uparrow \infty$ and $\sum_n x_n/a_n$ converges, then $a_n^{-1} \sum_{m=1}^n x_m \to 0$. Summation by parts writes the normalized partial sum as a convergent sequence minus a weighted average of its earlier values.

Theorem 2.64 (Series proof of the strong law, Theorem 2.5.10). Truncate $Y_k = X_k \mathbf{1}_{\{|X_k| \le k\}}$ and center $Z_k = Y_k - \mathbb{E}Y_k$. The variance estimate from Section 2.4 makes $\sum_k Z_k/k$ converge almost surely; Kronecker's lemma gives $n^{-1} \sum_{k \le n} Z_k \to 0$, and Cesaro averaging of $\mathbb{E}Y_k \to \mu$ finishes the proof.

2.5.1 Rates of Convergence

Theorem 2.65 (Rate under a finite variance assumption, Theorem 2.5.11). For centered i.i.d. variables with variance $\sigma^2 < \infty$ and every $\varepsilon > 0$, $S_n/[n^{1/2}(\log n)^{1/2+\varepsilon}] \to 0$ almost surely. Divide X_n by this denominator, note that the resulting variances are summable, and use Kronecker's lemma.

Theorem 2.66 (Marcinkiewicz–Zygmund rate, Theorem 2.5.12). If X_i are i.i.d., $1 < p < 2$, $\mathbb{E}X_1 = 0$, and $\mathbb{E}|X_1|^p < \infty$, then $S_n/n^{1/p} \to 0$ almost surely. Truncate at $n^{1/p}$, use Borel–Cantelli for large jumps, sum the truncated variances after scaling, and control the accumulated truncated means.

2.5.2 Infinite Mean

Theorem 2.67 (Feller dichotomy, Theorem 2.5.13). Let X_i be i.i.d. with $\mathbb{E}|X_1| = \infty$, let $a_n > 0$, and suppose a_n/n is increasing. Then

$$\limsup_{n \to \infty} \frac{|S_n|}{a_n} = 0 \text{ or } \infty$$

according as $\sum_n \mathbb{P}(|X_1| \ge a_n)$ converges or diverges. Thus no such normalization can produce a finite nonzero almost sure limit. The divergent case follows from infinitely many large jumps; the convergent case uses truncation, a variance series, and Kronecker's lemma.

2.6 Renewal Theory*

Let $\xi_1, \xi_2, \ldots$ be i.i.d. positive interarrival times, $T_0 = 0$, and $T_n = \sum_{k=1}^{n} \xi_k$. Define $N_t = \inf\{k : T_k > t\}$, which includes the renewal at time zero and is a stopping time. Write $\mu = \mathbb{E}\xi_1 \in (0, \infty]$.

Theorem 2.68 (Renewal frequency, Theorem 2.6.1). $N_t/t \to 1/\mu$ almost surely. The inequalities $T_{N_t-1} \le t < T_{N_t}$ give this limit from the strong law.

Theorem 2.69 (Wald's equation, Theorem 2.6.2). If X_i are i.i.d., $\mathbb{E}|X_1| < \infty$, and N is a stopping time with $\mathbb{E}N < \infty$, then $\mathbb{E}S_N = \mathbb{E}X_1\,\mathbb{E}N$. Expand $S_N = \sum_{m\ge1} X_m \mathbf{1}_{\{N\ge m\}}$; the event $\{N \ge m\}$ is determined before X_m and hence is independent of it. Absolute integrability justifies rearranging the sum.

Theorem 2.70 (Elementary renewal theorem, Theorem 2.6.3). For $U(t) = \mathbb{E}N_t$, $U(t)/t \to 1/\mu$. Wald gives the lower bound. For the upper bound, truncate each interarrival at c, compare renewal counts, then let $t \to \infty$ followed by $c \to \infty$. To see $\mathbb{E}N_t < \infty$, choose $\delta > 0$ with $\mathbb{P}(\xi_1 \ge \delta) > 0$. The time needed for $\lfloor t/\delta \rfloor + 1$ such increments has finite expectation and bounds N_t from above.

Definition 2.71 (Renewal measure and arithmeticity). The renewal measure is $U(A) = \sum_{n\ge0} \mathbb{P}(T_n \in A)$, so $U([0,t]) = \mathbb{E}N_t$. A law is arithmetic if it is supported on $\{\delta, 2\delta, \ldots\}$ for some $\delta > 0$; otherwise it is nonarithmetic.

Theorem 2.72 (Blackwell renewal theorem, Theorem 2.6.4). If the interarrival law is nonarithmetic, then for every $h > 0$, $U([t, t+h]) \to h/\mu$. For finite mean, couple an ordinary renewal process to a stationary delayed process until renewal epochs are arbitrarily close; the stationary process has mean renewal intensity $1/\mu$. The proof for infinite mean instead uses convolution bounds and nonarithmetic support to force the local limit to zero.

Definition 2.73 (Delayed and stationary renewal processes). If T_0 has distribution G and is independent of later increments, then $V = U * G$ and $V = G + V * F$. The equilibrium delay

$$G(t) = \frac{1}{\mu} \int_0^t (1 - F(y))\, \mathrm{d}y$$

makes $V(t) = t/\mu$ when $\mu < \infty$. For exponential interarrivals, this delay equals the original law, reflecting memorylessness.

Example 2.74 (Renewal equations, Examples 2.6.5–2.6.8). The ordinary renewal function obeys $U = \mathbf{1}_{[0,\infty)} + U * F$; an arbitrary delay obeys $V = G + V * F$; the correction $U(t) - t/\mu$ is another solution with a term involving an integral of the tail of F. The tail of the residual life $H(t) = \mathbb{P}(T_{N_t} - t > x)$ satisfies $H(t) = 1 - F(t + x) + (H * F)(t)$.

Theorem 2.75 (Solving the renewal equation, Theorem 2.6.9). For bounded h, the unique solution of $H = h + H * F$ that is bounded on bounded intervals is

$$H(t) = \int_0^t h(t - s)\, \mathrm{d}U(s).$$

Iterate the equation using $U_n = \sum_{m=0}^{n} F^{m*}$, pass uniformly on bounded intervals, and prove uniqueness from $K = K * F^{n*} \to 0$.

Example 2.76 (Pedestrian delay and ruin, Examples 2.6.10–2.6.11). The waiting time for a unit gap in Poisson traffic is represented by a terminating renewal process, allowing its distribution and mean to be read from a defective renewal measure. For the Cramér–Lundberg insurance model, algebra converts the survival probability into $R = R(0) + R * F$. If premium rate c exceeds mean claim rate μ, then $R(0) = 1 - \mu/c$; if $c < \mu$, ruin is almost sure.

Definition 2.77 (Direct Riemann integrability). For h on $[0, \infty)$, form upper and lower sums over every interval $[k\delta, (k+1)\delta)$. The function is directly Riemann integrable if both infinite sums converge to the same finite limit as $\delta \downarrow 0$.

Theorem 2.78 (Key renewal theorem, Theorem 2.6.12). If F is nonarithmetic and h is directly Riemann integrable, then

$$\int_0^t h(t - s)\, \mathrm{d}U(s) \longrightarrow \frac{1}{\mu} \int_0^\infty h(s)\, \mathrm{d}s.$$

Approximate h above and below by summable step functions and apply Blackwell's theorem to each step. Direct integrability supplies uniform tail control.

Lemma 2.79 (A useful criterion, Lemma 2.6.13). A nonnegative decreasing h with $h(0) < \infty$ and finite integral is directly Riemann integrable; its upper and lower sums differ by at most $h(0)\delta$.

Example 2.80 (Renewal correction and residual life, Examples 2.6.14–2.6.15). Assume the interarrival law is nonarithmetic. If $\mathbb{E}\xi_1^2 = \nu < \infty$, then $U(t) - t/\mu \to \nu/(2\mu^2)$. For $x \geq 0$, the residual life satisfies

$$\mathbb{P}(T_{N_t} - t > x) \to \frac{1}{\mu}\int_x^\infty (1 - F(s))\, \mathrm{d}s,$$

the equilibrium distribution of the residual life. At a fixed observation time, longer renewal intervals are more likely to contain that time.

2.7 Large Deviations*

Let X_i be i.i.d., $S_n = \sum_{i=1}^n X_i$, and fix $a > \mu = \mathbb{E}X_1$. Large deviation theory studies the exponential rate of $\mathbb{P}(S_n \geq na)$.

Lemma 2.81 (Superadditive limit, Lemma 2.7.1). If $\gamma_{m+n} \geq \gamma_m + \gamma_n$, then $\gamma_n/n \to \sup_m \gamma_m/m$. Write $n = km + r$ and iterate the inequality. Since independent blocks give $\mathbb{P}(S_{m+n} \geq (m+n)a) \geq \mathbb{P}(S_m \geq ma)\mathbb{P}(S_n \geq na)$, the limit $\gamma(a) = \lim_n n^{-1} \log \mathbb{P}(S_n \geq na)$ always exists in $[-\infty, 0]$.

Definition 2.82 (Log moment-generating function and rate function). Let $\phi(\theta) = \mathbb{E}e^{\theta X_1}$ and $\kappa(\theta) = \log \phi(\theta)$. Markov's inequality gives the Chernoff bound

$$\mathbb{P}(S_n \geq na) \leq \exp\{-n(\theta a - \kappa(\theta))\}, \qquad \theta > 0.$$

The rate function is $I(a) = \sup_{\theta \geq 0}\{\theta a - \kappa(\theta)\}$.

Lemma 2.83 (Strict exponential decay, Lemma 2.7.2). If $\phi(\theta) < \infty$ for some $\theta > 0$ and $a > \mu$, then $a\theta - \kappa(\theta) > 0$ for all sufficiently small positive θ. The key facts are $\kappa(0) = 0$ and $\kappa'(0+) = \mu$, justified by dominated convergence.

Definition 2.84 (Exponential tilting). For θ in the interior of the mgf domain, define

$$\mathrm{d}F_\theta(x) = e^{\theta x - \kappa(\theta)}\, \mathrm{d}F(x).$$

Then $\mathbb{E}_\theta X = \kappa'(\theta)$ and $\kappa''(\theta) = \mathrm{Var}_\theta(X) \geq 0$. Unless X is constant, κ' is strictly increasing and $\theta a - \kappa(\theta)$ is strictly concave in θ.

Example 2.85 (Examples 2.7.3–2.7.6). For $N(0,1)$, $\kappa(\theta) = \theta^2/2$ and tilting shifts the mean to θ. For exponential variables of rate λ, $\phi(\theta) = \lambda/(\lambda - \theta)$ and tilting changes the rate to $\lambda - \theta$. For fair signs, tilting gives probabilities $e^\theta/(e^\theta + e^{-\theta})$ and $e^{-\theta}/(e^\theta + e^{-\theta})$ to 1 and -1. A density proportional to $x^{-3}e^{-x}$ on $[1, \infty)$ shows that the mgf boundary can be finite while the attainable tilted means remain bounded.

Theorem 2.86 (Cramér exponent in the regular case, Theorem 2.7.7). Assume the mgf is finite to the right of zero, the law is nondegenerate, and there is θ_a in the interior with $\kappa'(\theta_a) = a$. Then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -\theta_a a + \kappa(\theta_a) = -I(a).$$

The Chernoff bound gives the upper estimate. For the lower estimate, tilt to a mean slightly larger than a, apply the weak law to a small interval around that mean, and then let the interval shrink toward a.

Lemma 2.87 (Change of measure for sums, Lemma 2.7.8). If F^n and F^n_λ are the laws of the original and tilted sums, then

$$\frac{\mathrm{d}F^n}{\mathrm{d}F^n_\lambda}(x) = e^{-\lambda x}\phi(\lambda)^n.$$

Integrating this Radon–Nikodym derivative over an interval expresses its original probability in terms of the tilted law.

Example 2.88 (Explicit rate functions). For standard normals, $I(a) = a^2/2$. For exponential variables of rate one and $a > 1$, $I(a) = a - 1 - \log a$. For fair signs and $0 < a < 1$,

$$I(a) = \tfrac{1}{2}\big[(1+a)\log(1+a) + (1-a)\log(1-a)\big].$$

These are Legendre transforms of their log moment generating functions.

Theorem 2.89 (Boundary cases, Theorems 2.7.9–2.7.10). If the law has finite upper endpoint x_0 and is not concentrated there, then $\kappa'(\theta) \uparrow x_0$ as $\theta \to \infty$; at $a = x_0$ the rate is simply $\log \mathbb{P}(X_1 = x_0)$. If the mgf domain ends at finite θ_+ and $\kappa'(\theta) \uparrow a_0 < \infty$, then for every $a \geq a_0$,

$$\gamma(a) = -a\theta_+ + \kappa(\theta_+),$$

so the logarithmic probability rate becomes linear beyond the largest attainable tilted mean.

Remark 2.90 (Subexponential tails). For some heavy-tailed laws, $\{S_n \geq na\}$ occurs primarily through one large summand, and $n^{-1}\log \mathbb{P}(S_n \geq na) \to 0$. Such laws can have $\mathbb{E}e^{\theta X_1} = \infty$ for every $\theta > 0$.

3 Central Limit Theorems

3.1 The De Moivre–Laplace Theorem

Let X_i be independent fair signs, $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$, and $S_n = \sum_{j=1}^{n} X_j$. Exact binomial probabilities and Stirling's formula give the Gaussian approximation to the lattice probabilities.

Lemma 3.1 (Exponential limit; Book 3.1.1). If $c_j \to 0$, $a_j \to \infty$, and $a_j c_j \to \lambda$, then $(1 + c_j)^{a_j} \to e^{\lambda}$.

Theorem 3.2 (Local approximation for fair coin tosses; Book 3.1.2). If $2k/\sqrt{2n} \to x$, then

$$\mathbb{P}(S_{2n} = 2k) \sim (\pi n)^{-1/2} e^{-x^2/2}.$$

Thus one lattice mass equals the limiting density times the lattice mesh, to first order.

Proof sketch. Write the mass as $\binom{2n}{n+k} 2^{-2n}$, apply Stirling to the three factorials, and regroup the powers into terms of the form $(1 + c_n)^{a_n}$. The scaling $k = O(\sqrt{n})$ makes the quadratic term survive while terms of higher order vanish.

Theorem 3.3 (De Moivre–Laplace; Book 3.1.3). For $a < b$,

$$\mathbb{P}\left(a \le \frac{S_m}{\sqrt{m}} \le b \right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \, \mathrm{d}x.$$

Proof sketch. Sum the local approximation over lattice points in $[a, b]$. After multiplying and dividing by the mesh size, the sum is a Riemann sum for the standard normal integral. Parity is handled by comparing odd and even times.

Remark 3.4. A central limit theorem controls probabilities over intervals on the $\sqrt{n}$ scale; a local limit theorem controls individual lattice masses. The latter is strictly stronger and requires attention to lattice span.

3.2 Weak Convergence

Definition 3.5 (Weak convergence on $\mathbb{R}$). For distribution functions F_n, F, write $F_n \Rightarrow F$, or $X_n \Rightarrow X$, when $F_n(x) \to F(x)$ at every continuity point of F.

For example, if $X_n = X + 1/n$, then $F_n(x) = F(x - 1/n) \to F(x-)$, which can differ from $F(x)$ at an atom.

3.2.1 Examples

Example 3.6 (Geometric waiting times; Book 3.2.4). If G_p is geometric with success probability p, then for $x \geq 0$,

$$\mathbb{P}(pG_p > x) \to e^{-x}.$$

Hence pG_p converges weakly to an exponential variable of rate one as $p \downarrow 0$.

Example 3.7 (Birthday collision; Book 3.2.5). For independent uniform draws from $\{1, \ldots, N\}$, let T_N be the first repeat. Multiplying the probabilities of avoiding a repeat at each draw gives

$$\mathbb{P}(T_N/\sqrt{N} > x) \to e^{-x^2/2}, \qquad x \geq 0.$$

After k distinct draws, the conditional probability of a repeat at the next draw is k/N.

Example 3.8 (Central order statistic; Book 3.2.6–Book 3.2.7). The median of $2n+1$ independent uniforms has density proportional to $x^n(1-x)^n$. After centering at $1/2$ and scaling by $\sqrt{n}$, its density has a Gaussian limit, as follows by expanding $\log[x(1-x)]$ around $x = 1/2$.

3.2.2 Theory

Theorem 3.9 (Quantile coupling; Book 3.2.8). If $F_n \Rightarrow F$, one can construct $Y_n \sim F_n$ and $Y \sim F$ on one probability space so that $Y_n \to Y$ almost surely.

Proof sketch. Use one uniform U and generalized inverses: $Y_n = F_n^{-1}(U)$. At quantile levels where the limiting inverse is continuous, cdf convergence traps Y_n between nearby values. A monotone function has only countably many discontinuities, so the exceptional set has Lebesgue measure zero.

Theorem 3.10 (Bounded continuous test functions; Book 3.2.9). $X_n \Rightarrow X$ if and only if $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ for every bounded continuous $g : \mathbb{R} \to \mathbb{R}$.

Theorem 3.11 (Continuous mapping; Book 3.2.10). Let D_g be the discontinuity set of a measurable g. If $X_n \Rightarrow X$ and $\mathbb{P}(X \in D_g) = 0$, then $g(X_n) \Rightarrow g(X)$. If g is also bounded, then $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$.

Theorem 3.12 (Portmanteau theorem; Book 3.2.11). The following are equivalent to $X_n \Rightarrow X$: lower bounds on probabilities of open sets, upper bounds on probabilities of closed sets, and convergence of probabilities for every Borel set A with $\mathbb{P}(X \in \partial A) = 0$.

For open G and closed K, these inequalities are

$$\liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G) \quad (G \text{ open}), \qquad \limsup_n \mathbb{P}(X_n \in K) \leq \mathbb{P}(X \in K) \quad (K \text{ closed}).$$

Definition 3.13 (Tightness). A sequence of laws (μ_n) on $\mathbb{R}$ is tight if for every $\varepsilon > 0$ there is $M < \infty$ such that $\mu_n([-M, M]) \geq 1 - \varepsilon$ for every n.

Theorem 3.14 (Helly selection and tightness; Book 3.2.12–Book 3.2.13). Every sequence of distribution functions has a subsequence converging at the continuity points of a nondecreasing right-continuous function. The sequence is tight if and only if every subsequence has a further subsequence whose limit is a probability distribution function.

Proof sketch. Diagonal selection gives convergence at rational points, and right continuity defines the limit elsewhere. Tightness ensures that its limits at $-\infty$ and $+\infty$ are zero and one. Conversely, failure of tightness yields a subsequence with a fixed positive amount of mass outside every bounded interval, so it has no weakly convergent further subsequence.

Theorem 3.15 (Moment criterion for tightness; Book 3.2.14). If $\phi \geq 0$, $\phi(x) \to \infty$ as $|x| \to \infty$, and $\sup_n \mathbb{E}\phi(X_n) < \infty$, then the laws of X_n are tight.

Proposition 3.16 (Subsequence principle; Book 3.2.15). If every subsequence of (X_n) has a further subsequence converging weakly to the same law X, then $X_n \Rightarrow X$.

3.3 Characteristic Functions

3.3.1 Definition, Inversion Formula

Definition 3.17 (Characteristic function). The characteristic function of X is

$$\varphi_X(t) = \mathbb{E}e^{itX} = \mathbb{E}\cos(tX) + i\mathbb{E}\sin(tX), \qquad t \in \mathbb{R}.$$

It always exists because $\left|e^{itX}\right| = 1$.

Theorem 3.18 (Elementary properties; Book 3.3.1). Every characteristic function satisfies $\varphi(0) = 1$, $\varphi(-t) = \overline{\varphi(t)}$, $|\varphi(t)| \leq 1$, and uniform continuity. For constants a, b,

$$\varphi_{aX+b}(t) = e^{itb}\varphi_X(at).$$

Theorem 3.19 (Characteristic function of an independent sum; Book 3.3.2). If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

For several standard laws, the characteristic functions are

law	$\varphi(t)$
fair sign	$\cos t$
Poisson(λ)	$\exp\{\lambda(e^{it}-1)\}$
$N(m,\sigma^2)$	$\exp\{imt-\sigma^2t^2/2\}$
Exp(1)	$(1-it)^{-1}$
Cauchy	$e^{-\lvert t\rvert}$

Theorem 3.20 (Fourier inversion; Book 3.3.11). If $\varphi(t) = \int e^{itx}\mu(\,\mathrm{d}x)$ and $a < b$, then

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

Consequently, the characteristic function determines the law uniquely.

Remark 3.21. The inversion integral need not converge absolutely. Endpoint atoms contribute half their mass, the Fourier analogue of convergence to the midpoint at a jump.

Theorem 3.22 (Integrable transform gives a density; Book 3.3.14). If $\int_{\mathbb{R}} |\varphi(t)|\,\mathrm{d}t < \infty$, then the law has bounded continuous density

$$f(y) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ity}\varphi(t)\,\mathrm{d}t.$$

Remark 3.23. Having a density alone does not imply absolute integrability of the characteristic function. For a uniform variable on $[-1,1]$, $\varphi(t) = \sin(t)/t$, which is not absolutely integrable.

3.3.2 Weak Convergence

Theorem 3.24 (Lévy continuity theorem; Book 3.3.17). If $X_n \Rightarrow X$, then $\varphi_{X_n}(t) \to \varphi_X(t)$ for every t. Conversely, if $\varphi_{X_n}(t)$ converges pointwise to a function φ that is continuous at zero, then φ is a characteristic function and X_n converges weakly to its law.

Proof sketch. The forward direction uses bounded continuous test functions with e^{itx}. For the converse, an averaged bound on $1-\Re\varphi_n$ gives tightness. Every subsequential limit has transform φ; inversion gives uniqueness, and the subsequence principle finishes.

Remark 3.25 (A discontinuous limit of characteristic functions). For $X_n \sim N(0,n)$, the characteristic functions converge to zero away from zero and to one at zero. The discontinuous limit is not a characteristic function: the laws are not tight.

3.3.3 Moments and Derivatives

Theorem 3.26 (Differentiating a characteristic function; Book 3.3.18). If $\mathbb{E}\,|X|^k < \infty$, then φ has k continuous derivatives and

$$\varphi^{(k)}(t) = \mathbb{E}\big[(iX)^k e^{itX}\big].$$

In particular, $\varphi^{(k)}(0) = i^k \mathbb{E}X^k$.

Theorem 3.27 (Quadratic expansion; Book 3.3.20). If $\mathbb{E}X^2 < \infty$, then, as $t \to 0$,

$$\varphi(t) = 1 + it\mathbb{E}X - \frac{t^2}{2}\mathbb{E}X^2 + o(t^2).$$

Proof sketch. Use a Taylor remainder bounded by $t^2 \min\{|t|\,|X|^3, CX^2\}$. After division by t^2, the integrand tends to zero and is dominated by an integrable multiple of X^2.

Theorem 3.28 (Converse criterion for a finite second moment; Book 3.3.21). If the symmetric second difference $[\varphi(h) - 2\varphi(0) + \varphi(-h)]/h^2$ is bounded below along $h \downarrow 0$, then $\mathbb{E}X^2 < \infty$.

3.3.4 Pólya's Criterion*

Theorem 3.29 (Pólya's criterion; Book 3.3.22). Suppose φ is even, nonnegative, continuous at zero with $\varphi(0) = 1$, decreases convexly to zero on $(0, \infty)$. Then φ is a characteristic function; indeed it is a mixture of triangular transforms $(1 - |t/s|)_+$.

Proof sketch. Convexity identifies a mixing measure from the slope changes of φ. Finite atomic mixtures are characteristic functions, and weak approximation plus Lévy's theorem gives the general case. In particular, $e^{-|t|^\alpha}$ satisfies the criterion for $0 < \alpha \le 1$.

3.3.5 The Moment Problem*

Distinct laws can have the same moments. For example, certain perturbations of a lognormal density leave every integer moment unchanged.

Theorem 3.30 (Moment determinacy; Book 3.3.25). If the even moments satisfy

$$\limsup_{k\to\infty} \frac{\mu_{2k}^{1/(2k)}}{2k} < \infty,$$

then at most one probability law has the moment sequence (μ_k).

Theorem 3.31 (Method of moments; Book 3.3.26). If every moment of F_n converges to μ_k and the preceding growth condition holds, then F_n converges weakly to the unique law with those moments.

Proof sketch. The moment bounds yield tightness. Uniform control of higher moments passes each fixed moment to subsequential limits. Determinacy forces all subsequential limits to agree.

3.4 Central Limit Theorems

3.4.1 i.i.d. Sequences

Theorem 3.32 (Classical central limit theorem; Book 3.4.1). If X_i are i.i.d., $\mathbb{E}X_i = \mu$, and $\mathrm{Var}(X_i) = \sigma^2 \in (0, \infty)$, then

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

Proof sketch. For centered X, the quadratic expansion gives $\varphi_X(t/(\sigma\sqrt{n})) = 1 - t^2/(2n) + o(n^{-1})$. Independence raises this to the nth power, which tends to $e^{-t^2/2}$. Lévy's continuity theorem identifies the standard normal limit.

Example 3.33 (Continuity correction). For a lattice sum, approximate $\mathbb{P}(S_n = k)$ by the normal probability of the cell $[k - h/2, k + h/2]$, where h is the lattice spacing. This continuity correction accounts for the width of the lattice cell.

3.4.2 Triangular Arrays

Definition 3.34 (Triangular array). A triangular array is a family $X_{n,m}$, $1 \le m \le k_n$, whose nth row is summed to form S_n. The distribution may change with both indices.

Theorem 3.35 (Lindeberg–Feller; Book 3.4.10). Suppose each row consists of independent centered variables,

$$\sum_m \mathbb{E}X_{n,m}^2 \to \sigma^2 > 0, \qquad \sum_m \mathbb{E}(X_{n,m}^2; |X_{n,m}| > \varepsilon) \to 0$$

for every $\varepsilon > 0$. Then $\sum_m X_{n,m} \Rightarrow N(0, \sigma^2)$.

Proof sketch. Lindeberg's condition makes large terms negligible and implies the largest row variance vanishes. Expand each row characteristic function to second order uniformly, compare the product with the exponential of the summed quadratic terms, and invoke Lévy.

Remark 3.36 (Lyapunov's condition). Writing $s_n^2 = \sum_m \mathbb{E}X_{n,m}^2$, a sufficient condition for Lindeberg's condition is

$$s_n^{-(2+\delta)} \sum_m \mathbb{E}|X_{n,m}|^{2+\delta} \to 0$$

for some $\delta > 0$. This follows from $X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}|>\varepsilon s_n\}} \le |X_{n,m}|^{2+\delta}/(\varepsilon s_n)^\delta$.

Example 3.37 (Records and permutation cycles; Book 3.4.11). Independent Bernoulli variables with success probabilities $1/m$ count records and have the same sum law as the number of cycles in a uniform permutation. Their mean and variance are asymptotic to $\log n$. After subtracting the mean and dividing by $\sqrt{\log n}$, Lindeberg–Feller gives convergence to $N(0,1)$.

Theorem 3.38 (Normal domain of attraction; Book 3.4.14). Suppose the X_i are i.i.d. with a nondegenerate common law. A centered and scaled sum can converge to a nondegenerate normal law if and only if the tail is negligible relative to the truncated second moment:

$$\frac{y^2 \mathbb{P}(|X_1| > y)}{\mathbb{E}(|X_1|^2; |X_1| \le y)} \to 0.$$

The criterion also includes some laws with infinite variance.

3.4.3 Prime Divisors (Erdős–Kac)*

Let $\omega(m)$ be the number of distinct prime divisors of m. Divisibility by a small prime behaves approximately like an independent Bernoulli event of probability $1/p$, although these divisibility events are dependent.

Theorem 3.39 (Erdős–Kac; Book 3.4.16). For a uniform integer $M_n \in \{1, \dots, n\}$,

$$\frac{\omega(M_n) - \log\log n}{\sqrt{\log\log n}} \Rightarrow N(0,1).$$

Proof sketch. Restrict first to primes below a suitable cutoff. The moments of the corresponding sum of divisibility indicators can be compared with those of independent Bernoulli variables of means $1/p$. The central limit theorem applies to the independent sum; the moment comparison and a bound for the omitted primes give the same limit for $\omega(M_n)$.

3.4.4 Rates of Convergence (Berry–Esseen)*

Theorem 3.40 (Berry–Esseen bound; Book 3.4.17). If X_i are i.i.d., centered, $\mathbb{E}X_i^2 = \sigma^2 > 0$, and $\rho = \mathbb{E}\,|X_i|^3 < \infty$, then, for the cdf F_n of $S_n/(\sigma\sqrt{n})$ and the standard normal cdf N,

$$\sup_x |F_n(x) - N(x)| \le \frac{3\rho}{\sigma^3\sqrt{n}}.$$

Proof sketch. Compare the characteristic function of the normalized sum with $e^{-t^2/2}$ using a remainder of order three. Smooth both distributions by convolution with Pólya's density, use Fourier inversion on compact support, and then remove the smoothing. The two errors are balanced at a cutoff of order $\sqrt{n}$.

Remark 3.41. For lattice laws, the jumps of F_n can have order $n^{-1/2}$, so a uniform approximation by a continuous distribution function generally has error of at least that order.

3.5 Local Limit Theorems*

Definition 3.42 (Lattice span). X is lattice if $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$ for some $h > 0$; the largest such h is the span. It is arithmetic when one may take $b = 0$.

Theorem 3.43 (Trichotomy for characteristic functions; Book 3.5.2). Exactly one occurs: $|\varphi|(t) < 1$ for all $t \neq 0$ (nonlattice); there is a least $\lambda > 0$ with $|\varphi|(\lambda) = 1$, giving lattice span $2\pi/\lambda$; or $|\varphi|(t) = 1$ for all t, in which case X is degenerate.

Theorem 3.44 (Lattice local CLT; Book 3.5.3). For centered i.i.d. lattice variables of variance σ^2 and span h, if $p_n(x) = \mathbb{P}(S_n/\sqrt{n} = x)$ on its lattice L_n and $g_\sigma(x) = (2\pi\sigma^2)^{-1/2} e^{-x^2/(2\sigma^2)}$, then

$$\sup_{x \in L_n} \left| \frac{\sqrt{n}}{h} p_n(x) - g_\sigma(x) \right| \to 0.$$

Theorem 3.45 (Nonlattice local CLT; Book 3.5.4). Under the analogous nonlattice hypotheses, if $x_n/\sqrt{n} \to x$, then for $a < b$,

$$\sqrt{n}\, \mathbb{P}(S_n \in (x_n + a, x_n + b)) \to (b - a) g_\sigma(x).$$

Proof sketch. Fourier inversion separates low frequencies, where the quadratic expansion gives a Gaussian, from frequencies bounded away from zero, where the span condition bounds the modulus of the characteristic function below one. In the nonlattice case, convolution with a function whose Fourier transform has compact support allows the inversion argument to be restricted to bounded frequencies.

3.6 Poisson Convergence

3.6.1 The Basic Limit Theorem

Theorem 3.46 (Law of rare events; Book 3.6.1). Let $X_{n,m}$ be independent Bernoulli variables with probabilities $p_{n,m}$. If

$$\sum_m p_{n,m} \to \lambda \in (0,\infty), \qquad \max_m p_{n,m} \to 0,$$

then $\sum_m X_{n,m} \Rightarrow \text{Poisson}(\lambda)$.

Proof sketch. The row characteristic function is $\prod_m [1 + p_{n,m}(e^{it} - 1)]$. Infinitesimality and convergence of the total mean turn its logarithm into $\lambda(e^{it} - 1)$. Alternatively, approximate each Bernoulli law by a Poisson law of the same mean and use total variation under products and sums.

Definition 3.47 (Total variation on a countable space). For probability masses μ, ν,

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_x |\mu(x) - \nu(x)| .$$

It contracts under measurable maps and is subadditive under independent products. For Bernoulli(p) versus Poisson(p), the distance is at most p^2.

3.6.2 Two Examples with Dependence

Example 3.48 (Fixed points of a random permutation; Book 3.6.8). The indicators of fixed points are dependent, but inclusion–exclusion shows that their sum converges to Poisson(1). In particular, the derangement probability tends to e^{-1}.

Theorem 3.49 (Empty boxes; Book 3.6.10). Place r balls uniformly into n boxes. If $ne^{-r/n} \to \lambda$, then the number of empty boxes converges to Poisson(λ).

Proof sketch. For any fixed k, the probability that specified k boxes are empty is $(1 - k/n)^r$. Inclusion–exclusion, or equivalently convergence of factorial moments, produces the Poisson probabilities despite dependence between emptiness indicators.

As a corollary, the coupon collector time satisfies

$$\mathbb{P}\left(\frac{T_n - n \log n}{n} \le x \right) \to e^{-e^{-x}}.$$

The nonnormal limit occurs because the last few waiting times are not negligible, so Lindeberg's condition fails.

3.7 Poisson Processes

Theorem 3.50 (Poisson limit for integer-valued arrays; Book 3.7.1). Let independent nonnegative integer-valued $X_{n,m}$ satisfy $\sum_m \mathbb{P}(X_{n,m} = 1) \to \lambda$, $\max_m \mathbb{P}(X_{n,m} = 1) \to 0$, and $\sum_m \mathbb{P}(X_{n,m} \geq 2) \to 0$. Then $\sum_m X_{n,m}$ converges to Poisson(λ).

Theorem 3.51 (Characterization by counts on short intervals; Book 3.7.2). Suppose counts have stationary independent increments and, as $h \downarrow 0$,

$$\mathbb{P}(N(0,h) = 1) = \lambda h + o(h), \qquad \mathbb{P}(N(0,h) \geq 2) = o(h).$$

Then $N(0,t)$ is Poisson with mean λt.

Definition 3.52 (Poisson process). A Poisson process of rate λ $(N(t))_{t \geq 0}$ starts at zero, has independent increments, and $N(t) - N(s) \sim \text{Poisson}(\lambda(t-s))$ for $s < t$. Equivalently, its interarrival times are i.i.d. exponential(λ).

Proof sketch. Partition $(0,t]$ into n short intervals and apply the preceding limit theorem for integer-valued arrays. For the interarrival construction, the memoryless property gives independent future gaps, and gamma densities yield the Poisson count probabilities.

3.7.1 Compound Poisson Processes

Let Y_i be i.i.d. marks independent of N, and set $S(t) = \sum_{j=1}^{N(t)} Y_j$.

Theorem 3.53 (Moments of a random sum; Book 3.7.3). Let $Y_1, Y_2, \ldots$ be i.i.d. and independent of the nonnegative integer-valued N, and put $S = \sum_{j=1}^{N} Y_j$. If $\mathbb{E}\,|Y_1| < \infty$ and $\mathbb{E}N < \infty$, then

$$\mathbb{E}S = \mathbb{E}N\,\mathbb{E}Y_1.$$

If in addition $\mathbb{E}Y_1^2 < \infty$ and $\mathbb{E}N^2 < \infty$, then

$$\operatorname{Var}(S) = \mathbb{E}N \operatorname{Var}(Y_1) + \operatorname{Var}(N)(\mathbb{E}Y_1)^2.$$

When $N \sim \text{Poisson}(\lambda)$ and $\mathbb{E}Y_1^2 < \infty$, $\operatorname{Var}(S) = \lambda \mathbb{E}Y_1^2$.

Proof sketch. Condition on N. The conditional mean is $N\mathbb{E}Y_1$ and the conditional variance is $N \operatorname{Var}(Y_1)$. The formulas follow from $\mathbb{E}S = \mathbb{E}[\mathbb{E}(S \mid N)]$ and $\operatorname{Var}(S) = \mathbb{E}[\operatorname{Var}(S \mid N)] + \operatorname{Var}[\mathbb{E}(S \mid N)]$.

3.7.2 Thinning

Theorem 3.54 (Independent thinning; Book 3.7.4). Assign each point independently a type j with probability q_j. The type counts are independent Poisson processes with rates λq_j.

Theorem 3.55 (Thinning with a probability depending on location; Book 3.7.5). Given the point locations, retain the points independently, keeping a point at s with probability $p(s)$, where p is measurable and takes values in $[0,1]$. The retained points form a nonhomogeneous Poisson process with intensity $\lambda p(s)$.

Example 3.56 (Poissonization). If the number of balls is Poisson(r) before allocation among n boxes, then box occupancies become independent Poisson(r/n). Concentration of the total count can then be used to compare this model with one having a fixed number of balls.

3.7.3 Conditioning

Theorem 3.57 (Ordered uniform arrival times; Book 3.7.9). Conditional on $N(t) = n$, the arrival vector $(T_1, \dots, T_n)$ has the same law as the order statistics of n independent uniform variables on $(0, t)$.

Corollary 3.58 (Binomial bridge; Book 3.7.10). For $0 < s < t$,

$$N(s) \mid N(t) = n \sim \text{Binomial}(n, s/t).$$

Theorem 3.59 (Ratios of arrival times; Book 3.7.11). The vector $(T_1/T_{n+1}, \dots, T_n/T_{n+1})$ has the law of the ordered values of n independent uniforms on $(0, 1)$.

Remark 3.60. Unconditionally, increments over disjoint intervals are independent. Conditioning on the total count destroys that independence and replaces it with a multinomial allocation.

3.8 Stable Laws*

Definition 3.61 (Regular variation and stable law). L is slowly varying if $L(tx)/L(x) \to 1$ for every $t > 0$. A law is stable if every sum of k independent copies, after affine normalization, has the original law.

Theorem 3.62 (Stable limit under regular variation; Book 3.8.2). Suppose X_i are i.i.d.,

$$\mathbb{P}(|X_1| > x) = x^{-\alpha}L(x), \quad 0 < \alpha < 2, \qquad \frac{\mathbb{P}(X_1 > x)}{\mathbb{P}(|X_1| > x)} \to \theta.$$

Choose a_n so that $n\mathbb{P}(|X_1| > a_n) \to 1$, and let $b_n = n\mathbb{E}[X_1 \mathbf{1}_{\{|X_1| \le a_n\}}]$. Then $(S_n - b_n)/a_n$ converges to a nondegenerate α-stable law.

Proof sketch. Large normalized observations converge to a Poisson point process with positive and negative tail intensities determined by θ. Sum points outside a small neighborhood of zero, control the truncated remainder, and then send the cutoff to zero diagonally. The resulting Lévy exponent yields the stable characteristic function.

The limiting stable characteristic function can be written as

$$\exp\left\{ict - b\,|t|^{\alpha}\left(1 + i\kappa\,\operatorname{sgn}(t) w_\alpha(t)\right)\right\}, \qquad w_\alpha(t) = \begin{cases} -\tan(\pi\alpha/2), & \alpha \neq 1, \\ (2/\pi)\log|t|\,, & \alpha = 1, \end{cases}$$

where $b > 0$, $-1 \le \kappa \le 1$, and tail balance gives $\kappa = 2\theta - 1$. The symmetric case has $\kappa = 0$.

Remark 3.63 (Scaling and centering). The sequence a_n is regularly varying with index $1/\alpha$. For $\alpha < 1$, the uncentered sum S_n/a_n also has a stable limit. For $\alpha > 1$, one may instead center by $n\mathbb{E}X_1$; changing the centering changes the location parameter of the limit. The stable law of index 2 is normal.

Theorem 3.64 (Stability and convergence of types; Book 3.8.8–Book 3.8.9). A nondegenerate law can arise as the limit of centered and scaled i.i.d. sums exactly when it is stable. Moreover, if $W_n \Rightarrow W$ and $\alpha_n W_n + \beta_n \Rightarrow W'$ with $\alpha_n > 0$ and both limits nondegenerate, then $\alpha_n \to \alpha > 0$, $\beta_n \to \beta \in \mathbb{R}$, and $W' \overset{d}{=} \alpha W + \beta$.

Example 3.65 (A Cauchy limit for reciprocals). If X has a symmetric density f that is continuous at zero with $f(0) > 0$, then $\mathbb{P}(1/X > x) \sim f(0)/x$ and $\mathbb{P}(1/X < -x) \sim f(0)/x$. For independent copies X_j, the averages $n^{-1}\sum_{j=1}^{n} 1/X_j$ therefore converge to a symmetric Cauchy law.

3.9 Infinitely Divisible Distributions*

Definition 3.66 (Infinite divisibility). A law is infinitely divisible if, for every n, it is the law of a sum of n i.i.d. random variables whose distribution may depend on n.

Theorem 3.67 (Characterization by triangular arrays; Book 3.9.1). A law is a weak limit of sums of n i.i.d. variables whose law may depend on n if and only if it is infinitely divisible. Normal, stable, Poisson, and compound Poisson laws are examples.

For a compound Poisson sum with rate λ and jump transform ψ,

$$\varphi(t) = \exp\{\lambda(\psi(t) - 1)\}.$$

Theorem 3.68 (Lévy–Khintchine representation; Book 3.9.6). A law is infinitely divisible exactly when

$$\log \varphi(t) = ict - \frac{\sigma^2 t^2}{2} + \int_{\mathbb{R}} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \nu(\,\mathrm{d}x),$$

where $\nu(\{0\}) = 0$ and $\int x^2/(1+x^2)\nu(\,\mathrm{d}x) < \infty$.

Remark 3.69. The exponent separates deterministic drift, Gaussian fluctuation, and jumps. The compensation term near zero makes the jump integral converge; changing the truncation convention changes the drift parameter but not the law.

Theorem 3.70 (Representation with finite variance; Book 3.9.7). A centered infinitely divisible law with finite variance has

$$\log \varphi(t) = \int_{\mathbb{R}} (e^{itx} - 1 - itx) x^{-2} \rho(\,\mathrm{d}x),$$

with the integrand interpreted as $-t^2/2$ at zero and $\mathrm{Var}(Z) = \rho(\mathbb{R})$.

3.10 Limit Theorems in $\mathbb{R}^d$

For random vectors, weak convergence is again characterized by bounded continuous tests, the Portmanteau inequalities for open and closed sets, and convergence on sets whose boundaries have zero probability Book 3.10.1. This definition on metric spaces agrees with multivariate cdf convergence at continuity points Book 3.10.2.

Theorem 3.71 (Multivariate tightness; Book 3.10.3). If for every $\varepsilon > 0$ the laws eventually put mass at least $1 - \varepsilon$ on some cube $[-M, M]^d$, then they have a weakly convergent subsequence.

Definition 3.72 (Vector characteristic function). For $X \in \mathbb{R}^d$,

$$\varphi_X(t) = \mathbb{E} e^{i\langle t, X \rangle}, \qquad t \in \mathbb{R}^d.$$

Theorem 3.73 (Multivariate inversion and continuity; Book 3.10.4–Book 3.10.5). Fourier inversion recovers probabilities of rectangles whose boundaries have zero mass. Moreover, $X_n \Rightarrow X$ if and only if $\varphi_{X_n}(t) \to \varphi_X(t)$ for every $t \in \mathbb{R}^d$; a pointwise transform limit continuous at zero is itself the transform of a weak limit.

Theorem 3.74 (Cramér–Wold device; Book 3.10.6). If $\langle \theta, X_n \rangle \Rightarrow \langle \theta, X \rangle$ for every $\theta \in \mathbb{R}^d$, then $X_n \Rightarrow X$.

Theorem 3.75 (Multivariate CLT; Book 3.10.7). Let X_i be i.i.d. vectors with mean μ and finite covariance matrix Γ. Then

$$\frac{S_n - n\mu}{\sqrt{n}} \Rightarrow Z, \qquad \varphi_Z(\theta) = \exp\left(-\frac{1}{2}\theta^T \Gamma \theta\right).$$

The limit may be degenerate when Γ is singular.

Proof sketch. For each θ, the scalar variables $\langle \theta, X_i \rangle$ have variance $\theta^T \Gamma \theta$. Apply the one-dimensional CLT to every projection and then Cramér–Wold. Positive semidefiniteness of Γ follows from $\theta^T \Gamma \theta = \operatorname{Var}(\langle \theta, X_1 \rangle) \geq 0$.

Example 3.76 (Simple random walk). For nearest-neighbor simple random walk on $\mathbb{Z}^d$, a step is $\pm e_j$ with probability $1/(2d)$. Its covariance is I/d, so $S_n/\sqrt{n}$ converges to a centered Gaussian with covariance I/d.

4 Martingales

4.1 Conditional Expectation

Definition 4.1 (Conditional expectation). Let $X \in L^1$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field. A version of $\mathbb{E}(X \mid \mathcal{G})$ is a $\mathcal{G}$-measurable random variable Y such that

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A], \qquad A \in \mathcal{G}.$$

Versions are identified almost surely. The associated conditional probability is $\mathbb{P}(B \mid \mathcal{G}) = \mathbb{E}(\mathbf{1}_B \mid \mathcal{G})$.

Theorem 4.2 (Existence, uniqueness, and localization; Book 4.1.1–4.1.2). For every $X \in L^1$, $\mathbb{E}(X \mid \mathcal{G})$ exists and is unique a.s., and $\mathbb{E}\,|\mathbb{E}(X \mid \mathcal{G})| \le \mathbb{E}\,|X|$. If $X = Y$ on $B \in \mathcal{G}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(Y \mid \mathcal{G})$ a.s. on B.

Proof sketch. For $X \ge 0$, define $\nu(A) = \int_A X \, d\mathbb{P}$ on $\mathcal{G}$ and take the Radon–Nikodym derivative $d\nu / d(\mathbb{P}|_{\mathcal{G}})$. Apply this separately to X^+ and X^-. For uniqueness, test two proposed versions on $\{Y - Y' > \varepsilon\} \in \mathcal{G}$. The same test intersected with B proves localization.

4.1.1 Examples

Example 4.3 (Measurability and independence). If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X \mid \mathcal{G}) = X$. If X is independent of $\mathcal{G}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}X$. In the first case, conditioning leaves X unchanged; in the second, it gives a constant.

Example 4.4 (A partition). If $\mathcal{G} = \sigma(A_1, A_2, \ldots)$ for a countable partition with $\mathbb{P}(A_i) > 0$, then

$$\mathbb{E}(X \mid \mathcal{G}) = \sum_i \frac{\mathbb{E}[X\mathbf{1}_{A_i}]}{\mathbb{P}(A_i)} \mathbf{1}_{A_i}.$$

On each A_i, the conditional expectation is the average of X over that event.

Example 4.5 (Conditioning through a density). If (X, Y) has joint density f, $\mathbb{E}\,|g(X)| < \infty$, and $f_Y(y) = \int f(x, y)\, dx > 0$, then

$$\mathbb{E}[g(X) \mid Y] = h(Y), \qquad h(y) = \frac{\int g(x) f(x, y)\, dx}{f_Y(y)}.$$

Values of h on a set of Y-probability zero may be chosen arbitrarily. If X and Y are independent and $\mathbb{E}\,|\phi(X, Y)| < \infty$, then $\mathbb{E}(\phi(X, Y) \mid X) = g(X)$ with $g(x) = \mathbb{E}\phi(x, Y)$.

Remark 4.6 (Borel's paradox). For a geometric set of probability zero, the conditional law depends on the random variable used to describe the conditioning. Different parametrizations can therefore give different conditional laws on the same set. When $\mathbb{P}(Y = y) = 0$, the expression $\mathbb{P}(A \mid Y = y)$ is defined through a conditional distribution rather than a ratio of probabilities.

4.1.2 Properties

Theorem 4.7 (Properties of conditional expectation; Book 4.1.9–4.1.14). Whenever the displayed quantities are integrable:

1. $\mathbb{E}(aX + bY \mid \mathcal{G}) = a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(Y \mid \mathcal{G})$;
2. $X \le Y$ implies $\mathbb{E}(X \mid \mathcal{G}) \le \mathbb{E}(Y \mid \mathcal{G})$;
3. $0 \le X_n \uparrow X$ implies $\mathbb{E}(X_n \mid \mathcal{G}) \uparrow \mathbb{E}(X \mid \mathcal{G})$;
4. if $\mathcal{H} \subseteq \mathcal{G}$, then
$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}), \qquad \mathbb{E}(\mathbb{E}(X \mid \mathcal{H}) \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{H});$$
5. if Z is $\mathcal{G}$-measurable, then $\mathbb{E}(ZX \mid \mathcal{G}) = Z\mathbb{E}(X \mid \mathcal{G})$.

In both iterated expectations, the result is the conditional expectation given the smaller σ-field.

Theorem 4.8 (Conditional Jensen and contraction; Book 4.1.10–4.1.11). If ϕ is convex and the terms are integrable, then
$$\phi(\mathbb{E}(X \mid \mathcal{G})) \le \mathbb{E}(\phi(X) \mid \mathcal{G}).$$
Consequently, for $1 \le p \le \infty$,
$$\|\mathbb{E}(X \mid \mathcal{G})\|_p \le \|X\|_p .$$

Proof sketch. Write a real convex function as the supremum of a countable family of affine minorants. Apply monotonicity and linearity to each minorant, then take the supremum. Use $\phi(x) = |x|^p$ for the contraction.

Theorem 4.9 (L^2 projection; Book 4.1.15). If $X \in L^2$, then $\mathbb{E}(X \mid \mathcal{G})$ is the unique element of $L^2(\mathcal{G})$ minimizing $\mathbb{E}(X - Y)^2$. Equivalently,
$$\mathbb{E}\big[Z\{X - \mathbb{E}(X \mid \mathcal{G})\}\big] = 0, \qquad Z \in L^2(\mathcal{G}).$$
For $\mathcal{H} \subseteq \mathcal{G}$, the Pythagorean identity is
$$\mathbb{E}\,|X - \mathbb{E}(X \mid \mathcal{H})|^2 = \mathbb{E}\,|X - \mathbb{E}(X \mid \mathcal{G})|^2 + \mathbb{E}\,|\mathbb{E}(X \mid \mathcal{G}) - \mathbb{E}(X \mid \mathcal{H})|^2 .$$

4.1.3 Regular Conditional Probabilities*

Definition 4.10 (Regular conditional distribution). For a measurable $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$, a kernel $K(\omega, A)$ is a regular conditional distribution of X given $\mathcal{G}$ when $K(\cdot, A)$ is a version of $\mathbb{P}(X \in A \mid \mathcal{G})$ for every A, and $K(\omega, \cdot)$ is a probability measure for a.e. ω.

Theorem 4.11 (Integration and existence; Book 4.1.16–4.1.18). If K is a regular conditional distribution, then

$$\mathbb{E}[f(X) \mid \mathcal{G}](\omega) = \int_S f(x) K(\omega, \mathrm{d}x)$$

for every integrable $f(X)$. Regular conditional distributions exist when the state space is standard Borel. When $\mathcal{G} = \sigma(Y)$ and X, Y take values in such spaces, a version can be chosen as a measurable kernel $K(Y(\omega), A)$.

Remark 4.12 (Simultaneous versions). The exceptional null set for a conditional expectation may depend on the test function. A regular conditional distribution chooses these versions as a single probability kernel. The standard Borel assumption ensures that such a choice is possible.

4.2 Martingales and Almost Sure Convergence

Definition 4.13 (Filtration and martingale). A filtration is an increasing family $(\mathcal{F}_n)_{n \geq 0}$. An integrable adapted process (X_n) is a martingale, submartingale, or supermartingale when

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n, \qquad \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \geq X_n, \qquad \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \leq X_n,$$

respectively. By the tower property the same relation holds between any two times $m < n$.

Example 4.14 (Martingales associated with a random walk). Let $S_n = S_0 + \sum_{k=1}^{n} \xi_k$ with constant S_0, i.i.d. increments, and natural filtration. If $\mathbb{E}\xi_1 = \mu$, then $S_n - n\mu$ is a martingale. If $\mu = 0$ and $\operatorname{Var}(\xi_1) = \sigma^2$, then $S_n^2 - n\sigma^2$ is a martingale. If $\varphi(\theta) = \mathbb{E}e^{\theta \xi_1} < \infty$, then

$$M_n(\theta) = \frac{e^{\theta S_n}}{\varphi(\theta)^n}$$

is a nonnegative martingale.

Theorem 4.15 (Convex transforms; Book 4.2.6–4.2.7). If X_n is a martingale and ϕ is convex, then $\phi(X_n)$ is a submartingale, provided it is integrable. If X_n is a submartingale, the same holds for increasing convex ϕ. In particular, if X_n is a martingale, $p \geq 1$, and $\mathbb{E}\,|X_n|^p < \infty$ for every n, then $|X_n|^p$ is a submartingale. For a submartingale X_n, $(X_n - a)^+$ is a submartingale.

Definition 4.16 (Predictable transform and stopping time). A process H_n is predictable if H_n is $\mathcal{F}_{n-1}$-measurable. Its transform of X is

$$(H \cdot X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1}).$$

A random time T is a stopping time if $\{T \le n\} \in \mathcal{F}_n$ for every n. Then $H_k = \mathbf{1}_{\{T \ge k\}}$ is predictable and $(H \cdot X)_n = X_{T \wedge n} - X_0$.

Theorem 4.17 (Predictable transforms; Book 4.2.8–4.2.9). If X is a supermartingale and H is nonnegative, bounded, and predictable, then $H \cdot X$ is a supermartingale. For a martingale, the sign restriction on H is unnecessary. Consequently, $X_{T \wedge n}$ has the same martingale type as X_n.

Remark 4.18 (Conditional increments of a transform). Predictability allows H_n to be taken outside the conditional expectation:

$$\mathbb{E}[H_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}] = H_n \mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}].$$

Boundedness of H_n ensures integrability. The same calculation applies to an unbounded predictable process whenever the required products are integrable.

Theorem 4.19 (Upcrossing inequality; Book 4.2.10). For a submartingale and $a < b$, let $U_n[a, b]$ count completed upcrossings of $[a, b]$ by time n. Then

$$(b - a)\mathbb{E}U_n[a, b] \le \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+.$$

Theorem 4.20 (Martingale convergence; Book 4.2.11–4.2.12). If X_n is a submartingale and $\sup_n \mathbb{E}X_n^+ < \infty$, then $X_n \to X_\infty$ a.s. for an integrable finite limit. In particular, every nonnegative supermartingale converges a.s. to a finite limit and $\mathbb{E}X_\infty \le \mathbb{E}X_0$.

Proof sketch. The upcrossing bound makes the total number of upcrossings of every rational interval finite a.s. If $\liminf X_n < \limsup X_n$, some rational interval is crossed infinitely often, a contradiction. Fatou's lemma and the submartingale mean inequality rule out infinite limits and give integrability.

Example 4.21 (Almost sure convergence without convergence in L^1). Stop simple symmetric random walk started at 1 when it first hits 0. The stopped process is a nonnegative martingale, converges a.s. to 0, but has expectation 1 at every finite time. Its expectations therefore do not converge to the expectation of its limit, so convergence in L^1 fails.

4.3 Examples

4.3.1 Bounded Increments

Theorem 4.22 (Dichotomy for bounded increments; Book 4.3.1). If a martingale satisfies $|X_{n+1} - X_n| \le C$, then a.s. either X_n has a finite limit or

$$\limsup_n X_n = +\infty, \qquad \liminf_n X_n = -\infty.$$

Theorem 4.23 (Doob decomposition; Book 4.3.2). Every submartingale has the unique representation $X_n = M_n + A_n$, where M is a martingale, $A_0 = 0$, and A is predictable and increasing. Explicitly,

$$A_n = \sum_{k=1}^{n} \{\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) - X_{k-1}\}.$$

Theorem 4.24 (Conditional Borel–Cantelli; Book 4.3.4). For $B_n \in \mathcal{F}_n$,

$$\{B_n \text{ i.o.}\} = \left\{ \sum_{n=1}^{\infty} \mathbb{P}(B_n \mid \mathcal{F}_{n-1}) = \infty \right\} \quad \text{a.s.}$$

The statement does not require independence.

Proof sketch. Decompose the counting process $\sum_{k \le n} \mathbf{1}_{B_k}$ into its compensator $\sum_{k \le n} \mathbb{P}(B_k \mid \mathcal{F}_{k-1})$ and a martingale with increments bounded by 1. Apply the dichotomy for martingales with bounded increments.

4.3.2 Pólya's Urn Scheme

Example 4.25 (The limiting proportion in Pólya's urn). In a Pólya urn with two colors, the fraction X_n of green balls is a bounded martingale, hence converges a.s. With one ball of each color and one added copy per draw, the number of green draws among the first n is uniform on $\{0, \dots, n\}$; the limit is uniform on $[0, 1]$. In general the limit has a beta law with parameters determined by the initial color masses. The draws are exchangeable. Conditionally on the limiting proportion, they are independent Bernoulli trials with that parameter.

4.3.3 Radon–Nikodym Derivatives

Theorem 4.26 (Martingale of likelihood ratios; Book 4.3.5–4.3.6). Let μ, ν be probability measures on $(\Omega, \mathcal{F})$, and let $(\mathcal{F}_n)$ increase with $\sigma(\bigcup_n \mathcal{F}_n) = \mathcal{F}$. Write μ_n, ν_n for the restrictions and assume $\mu_n \ll \nu_n$ for every n. Set $X_n = \mathrm{d}\mu_n / \mathrm{d}\nu_n$. Then (X_n) is a nonnegative ν-martingale. Its limit identifies the absolutely continuous component of μ relative to ν; mass carried by $\{X_n \to \infty\}$ accounts for a singular component.

If, in addition, $\mu = \bigotimes_k \mu_k$ and $\nu = \bigotimes_k \nu_k$ are product laws with independent coordinate fields and densities $f_k = \mathrm{d}\mu_k / \mathrm{d}\nu_k$ satisfying $0 < f_k < \infty$ ν_k-a.e., Kakutani's theorem gives the dichotomy

$$\mu \ll \nu \quad \text{or} \quad \mu \perp \nu,$$

with equivalence in the first case. The alternative is decided by $\prod_k \rho_k$, where $\rho_k = \int \sqrt{f_k}\, \mathrm{d}\nu_k$: a positive product gives equivalence, and a zero product gives singularity.

4.3.4 Branching Processes

Theorem 4.27 (Galton–Watson; Book 4.3.9–4.3.13). For a Galton–Watson process Z_n started from $Z_0 = 1$, with offspring distribution (p_k) and mean $\mu \in (0, \infty)$, $W_n = Z_n / \mu^n$ is a nonnegative martingale. If $\mu < 1$, extinction occurs a.s.; it also occurs for $\mu = 1$ unless offspring is identically one. For $\mu > 1$, the extinction probability is the smallest fixed point in $[0, 1]$ of the offspring generating function. In this supercritical case, the limit W is nontrivial on survival exactly under the $L \log L$ condition $\sum_k p_k k \log k < \infty$.

Remark 4.28 (Expectation of the branching limit). Uniform integrability gives $\mathbb{E}W = \mathbb{E}W_0$. When the $L \log L$ condition fails in the supercritical case, $W = 0$ a.s. despite the constant expectations of W_n.

4.4 Doob's Inequality and Convergence in L^p, $p > 1$

Theorem 4.29 (Bounded optional sampling; Book 4.4.1). If X is a submartingale and $T \le n$ a.s., then

$$\mathbb{E}X_0 \le \mathbb{E}X_T \le \mathbb{E}X_n.$$

Equalities hold for martingales.

Theorem 4.30 (Doob maximal inequalities; Book 4.4.2–4.4.4). For a submartingale, $X_n^* = \max_{0 \le k \le n} X_k^+$, and $\lambda > 0$,

$$\lambda \mathbb{P}(X_n^* \ge \lambda) \le \mathbb{E}[X_n \mathbf{1}_{\{X_n^* \ge \lambda\}}] \le \mathbb{E}X_n^+.$$

If $1 < p < \infty$, then

$$\mathbb{E}(X_n^*)^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_n^+)^p.$$

For a martingale M, the analogous bound holds for $\max_{k \le n} |M_k|$ with terminal value $|M_n|$.

Proof sketch. Stop at the first crossing of λ and apply bounded optional sampling. For the L^p inequality, integrate the tail bound using $\mathbb{E}Z^p = \int_0^\infty p\lambda^{p-1}\mathbb{P}(Z > \lambda)\,\mathrm{d}\lambda$, then apply Hölder's inequality and solve for $\|X_n^*\|_p$.

Theorem 4.31 (L^p convergence; Book 4.4.6). If M_n is a martingale and $\sup_n \mathbb{E}\,|M_n|^p < \infty$ for some $p > 1$, then $M_n \to M_\infty$ a.s. and in L^p. Moreover $M_n = \mathbb{E}(M_\infty \mid \mathcal{F}_n)$.

Theorem 4.32 (Orthogonality and conditional variance; Book 4.4.7–4.4.8). For a square-integrable martingale and $m \le n$,

$$\mathbb{E}[(M_n - M_m)Y] = 0 \quad (Y \in L^2(\mathcal{F}_m)),$$

so disjoint martingale increments are orthogonal, and

$$\mathbb{E}((M_n - M_m)^2 \mid \mathcal{F}_m) = \mathbb{E}(M_n^2 \mid \mathcal{F}_m) - M_m^2.$$

Remark 4.33 (The case $p = 1$). The constant $p/(p-1)$ tends to infinity as $p \downarrow 1$, and the corresponding maximal inequality does not hold in L^1. At this endpoint, uniform integrability characterizes convergence in L^1; an $L \log L$ moment condition can also give maximal estimates.

4.5 Square-Integrable Martingales*

Definition 4.34 (Bracket). For a square-integrable martingale $M_0 = 0$ with increments $\Delta M_k = M_k - M_{k-1}$, define

$$A_n = \sum_{k=1}^{n} \mathbb{E}((\Delta M_k)^2 \mid \mathcal{F}_{k-1}), \qquad A_\infty = \lim_n A_n.$$

Then $M_n^2 - A_n$ is a martingale.

Theorem 4.35 (Bounds from predictable quadratic variation; Book 4.5.1–4.5.3). One has

$$\mathbb{E}\sup_n |M_n|^2 \le 4\mathbb{E}A_\infty.$$

The limit $\lim_n M_n$ exists and is finite on $\{A_\infty < \infty\}$. If $f \ge 1$ is increasing and $\int_0^\infty f(t)^{-2}\,\mathrm{d}t < \infty$, then

$$\frac{M_n}{f(A_n)} \longrightarrow 0 \quad \text{a.s. on } \{A_\infty = \infty\}.$$

Theorem 4.36 (Ratio conditional Borel–Cantelli; Book 4.5.5). For $B_n \in \mathcal{F}_n$ and $p_n = \mathbb{P}(B_n \mid \mathcal{F}_{n-1})$,

$$\frac{\sum_{k=1}^{n} \mathbf{1}_{B_k}}{\sum_{k=1}^{n} p_k} \longrightarrow 1 \quad \text{a.s. on } \left\{ \sum_{k=1}^{\infty} p_k = \infty \right\}.$$

Thus, when the sum of the conditional probabilities diverges, the number of occurrences is asymptotic to that sum.

Theorem 4.37 (An L^1 bound from quadratic variation; Book 4.5.7). For the same square-integrable martingale and predictable quadratic variation,

$$\mathbb{E} \sup_n |M_n| \le 3 \mathbb{E} A_\infty^{1/2}.$$

The estimate follows by stopping when A_n first exceeds a level, applying the weak maximal inequality, and integrating over levels. It is often sharper than applying Cauchy–Schwarz to the L^2 bound.

Remark 4.38 (Two quadratic variations). The predictable quadratic variation is $A_n = \sum_{k \le n} \mathbb{E}[(\Delta M_k)^2 \mid \mathcal{F}_{k-1}]$. The sum of squared increments, $[M]_n = \sum_{k \le n} (\Delta M_k)^2$, is the quadratic variation. Their difference $[M]_n - A_n$ is a martingale.

4.6 Uniform Integrability and Convergence in L^1

Definition 4.39 (Uniform integrability). A family $(X_i)_{i \in I}$ is uniformly integrable if

$$\lim_{K \to \infty} \sup_{i \in I} \mathbb{E}[|X_i| \, \mathbf{1}_{\{|X_i| > K\}}] = 0.$$

It is stronger than $\sup_i \mathbb{E}\,|X_i| < \infty$.

Theorem 4.40 (Criteria for uniform integrability; Book 4.6.1–4.6.3). For $X \in L^1$, the family of all $\mathbb{E}(X \mid \mathcal{G})$ is uniformly integrable as $\mathcal{G}$ varies. If $\Phi : [0, \infty) \to [0, \infty)$ is nondecreasing and convex, $\Phi(t)/t \to \infty$, and $\sup_i \mathbb{E}\Phi(|X_i|) < \infty$, then (X_i) is uniformly integrable. If $X_n \to X$ in probability, then the following are equivalent:

1. (X_n) is uniformly integrable;
2. $X_n \to X$ in L^1;
3. $\mathbb{E}\,|X_n| \to \mathbb{E}\,|X| < \infty$.

Theorem 4.41 (Closed martingales; Book 4.6.4–4.6.7). For a martingale M_n, the following are equivalent:

1. (M_n) is uniformly integrable;
2. M_n converges a.s. and in L^1;
3. M_n converges in L^1;
4. there is $M_\infty \in L^1$ such that $M_n = \mathbb{E}(M_\infty \mid \mathcal{F}_n)$.

For submartingales, the first three conditions remain equivalent.

Theorem 4.42 (Increasing filtrations; Book 4.6.8–4.6.10). If $\mathcal{F}_n \uparrow \mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$, then for $X \in L^1$,

$$\mathbb{E}(X \mid \mathcal{F}_n) \to \mathbb{E}(X \mid \mathcal{F}_\infty) \quad \text{a.s. and in } L^1.$$

Hence, for $A \in \mathcal{F}_\infty$, $\mathbb{P}(A \mid \mathcal{F}_n) \to \mathbf{1}_A$ a.s. (Lévy's upward theorem). A dominated triangular version also holds for $\mathbb{E}(Y_n \mid \mathcal{F}_n)$ when $Y_n \to Y$ a.s. and $|Y_n| \le Z \in L^1$.

Remark 4.43 (Conditional dominated convergence). If $X_n \to X$ a.s. and $|X_n| \le Z$ for an integrable Z, then $\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G})$ a.s. and in L^1. Without domination, uniform integrability still gives convergence in L^1 by the contraction property of conditional expectation.

4.7 Backwards Martingales

Definition 4.44 (Backwards martingale). Let $(\mathcal{F}_n)_{n \le 0}$ increase with n. An adapted integrable process is a backwards martingale when $\mathbb{E}(X_n \mid \mathcal{F}_m) = X_m$ for $n > m$. Thus the conditioning σ-fields decrease as the index tends to $-\infty$.

Theorem 4.45 (Backwards convergence; Book 4.7.1–4.7.3). Every backwards martingale converges a.s. and in L^1 as $n \to -\infty$. If $\mathcal{F}_{-\infty} = \cap_{n \le 0} \mathcal{F}_n$, then

$$X_{-\infty} = \mathbb{E}(X_0 \mid \mathcal{F}_{-\infty}).$$

More generally, if $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$, then $\mathbb{E}(Y \mid \mathcal{F}_n) \to \mathbb{E}(Y \mid \mathcal{F}_{-\infty})$ a.s. and in L^1.

Example 4.46 (Strong law from symmetry). For i.i.d. integrable X_i, condition the sample mean on the σ-field invariant under permutations of the first n coordinates. The averages form a backwards martingale. The limiting exchangeable σ-field is trivial by the Hewitt–Savage zero–one law, giving $n^{-1} \sum_{i=1}^n X_i \to \mathbb{E}X_1$ a.s.

Theorem 4.47 (de Finetti; Book 4.7.9–4.7.10). An exchangeable sequence is conditionally i.i.d. given its exchangeable σ-field. For exchangeable Bernoulli variables, there is a random parameter $\Theta \in [0,1]$ such that, conditionally on Θ, the variables are i.i.d. Bernoulli(Θ); equivalently,

$$\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \int_0^1 \theta^k (1-\theta)^{n-k} \, \mathrm{d}F(\theta).$$

4.8 Optional Stopping Theorems

Theorem 4.48 (Optional stopping; Book 4.8.1–4.8.5). For a martingale, $\mathbb{E}M_T = \mathbb{E}M_0$ holds under any of the following conditions:

1. (M_n) is uniformly integrable and $T \le \infty$;
2. $\mathbb{E}\,|M_T| < \infty$ and $(M_n \mathbf{1}_{\{T>n\}})$ is uniformly integrable;
3. the conditional expected increment sizes are uniformly bounded and $\mathbb{E}T < \infty$.

For submartingales or supermartingales, replace equality by the appropriate inequality. For a nonnegative supermartingale, Fatou's lemma gives $\mathbb{E}M_T \le \mathbb{E}M_0$.

Remark 4.49 (Passage to an unbounded stopping time). The identity holds first at $T \wedge n$. Each condition above permits expectations to pass to the limit as $n \to \infty$. For simple symmetric random walk started at 1 and stopped on reaching 0, the stopping time is finite a.s., but the stopped expectations remain 1 while the limiting value is 0.

4.8.1 Applications to Random Walks

Theorem 4.50 (Wald's equation; Book 4.8.6). Let $S_n = \sum_{k=1}^n \xi_k$, where the increments are i.i.d. and integrable with mean μ. For a stopping time T relative to the natural filtration and satisfying $\mathbb{E}T < \infty$,

$$\mathbb{E}S_T = \mu \mathbb{E}T.$$

Stopping the martingale $S_n - n\mu$ at $T \wedge n$ gives the bounded stopping identity. The uniform bound on the expected absolute increments then justifies passage to T.

Example 4.51 (Gambler's ruin). For simple symmetric random walk started at an integer x between integers $a < b$ and $T = \inf\{n : S_n \notin (a,b)\}$,

$$\mathbb{P}_x(S_T = b) = \frac{x-a}{b-a}, \qquad \mathbb{E}_x T = (b-x)(x-a).$$

Stopping S_n gives the hitting probability, while stopping $S_n^2 - n$ gives the mean exit time. For $p \neq q$, the harmonic martingale $(q/p)^{S_n}$ gives the corresponding geometric formula.

Theorem 4.52 (First passage to 1; Book 4.8.8). Let S_n be simple symmetric random walk with $S_0 = 0$, and let $T_1 = \inf\{n \geq 0 : S_n = 1\}$. Then $T_1 < \infty$ a.s. and

$$\mathbb{E}s^{T_1} = \frac{1 - \sqrt{1 - s^2}}{s}, \qquad 0 < s \leq 1.$$

Its distribution is

$$\mathbb{P}(T_1 = 2n - 1) = \frac{1}{2n - 1}\binom{2n}{n} 2^{-2n}, \qquad n \geq 1,$$

and $\mathbb{P}(T_1 = 2n) = 0$.

Proof sketch. Letting the lower boundary tend to $-\infty$ in gambler's ruin shows that T_1 is finite a.s. For $\theta > 0$, stop $e^{\theta S_n}/(\cosh\theta)^n$ at $T_1 \wedge n$. The stopped martingale is bounded by e^{θ}, so bounded convergence gives $\mathbb{E}(\cosh\theta)^{-T_1} = e^{-\theta}$. Set $\cosh\theta = 1/s$ and solve the resulting quadratic for $e^{-\theta}$. The binomial expansion of $\sqrt{1 - s^2}$ gives the displayed probabilities.

Example 4.53 (Exponential bounds). If $\mathbb{E}e^{\theta\xi_1} = 1$ for some $\theta \neq 0$, then $e^{\theta S_n}$ is a nonnegative martingale. For $\theta > 0$ and $b > S_0$, stopping at the first crossing of b gives $\mathbb{P}(\sup_n S_n \geq b) \leq e^{-\theta(b - S_0)}$. For $\theta < 0$ and $a < S_0$, the corresponding bound is $\mathbb{P}(\inf_n S_n \leq a) \leq e^{-\theta(a - S_0)}$. In simple random walk, the absence of overshoot also permits exact hitting formulas.

4.9 Combinatorics of Simple Random Walk*

Theorem 4.54 (Reflection principle and ballot theorem; Book 4.9.1–4.9.2). Reflecting a nearest-neighbor path up to its first hit of 0 bijects paths from $(0, x)$ to (n, y) that hit 0 with paths from $(0, -x)$ to (n, y), for $x, y > 0$. Consequently, if candidate A receives α votes and B receives $\beta < \alpha$, then the probability that A stays strictly ahead throughout the count is

$$\frac{\alpha - \beta}{\alpha + \beta}.$$

Theorem 4.55 (Arcsine laws; Book 4.9.3–4.9.6). Let S_n be simple symmetric random walk started at 0, and let L_{2n} be its last zero by time $2n$. Let A_{2n} count the line segments from $(j-1, S_{j-1})$ to (j, S_j), $1 \leq j \leq 2n$, that lie in the closed upper half-plane. With $u_{2m} = \mathbb{P}(S_{2m} = 0)$, the exact laws are

$$\mathbb{P}(L_{2n} = 2k) = \mathbb{P}(A_{2n} = 2k) = u_{2k}u_{2n-2k}, \qquad 0 \leq k \leq n.$$

For $0 < a < b < 1$,

$$\mathbb{P}\left(a \leq \frac{L_{2n}}{2n} \leq b\right) \longrightarrow \int_a^b \frac{\mathrm{d}x}{\pi\sqrt{x(1-x)}}.$$

The same limit holds for $A_{2n}/(2n)$, the occupation proportion of the linearly interpolated path above zero.

Remark 4.56 (The arcsine density). The arcsine density is largest near 0 and 1. A long path of simple symmetric random walk is therefore more likely to spend nearly all its time on one side of zero than to spend similar amounts of time on each side.

5 Markov Chains

5.1 Examples

Definition 5.1 (Time-homogeneous Markov chain). On a measurable state space $(S, \mathcal{S})$, a transition kernel p assigns a probability $p(x, \cdot)$ to each x and makes $x \mapsto p(x, A)$ measurable for every A. An adapted process (X_n) is Markov with kernel p when

$$\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = p(X_n, A) \quad \text{a.s.}$$

For countable S, write $p(x, y)$. The transition matrix $P = (p(x, y))$ has nonnegative entries and each row sums to one.

Example 5.2 (Random walk). For i.i.d. increments $\xi_n \in \mathbb{Z}^d$, $X_n = x + \sum_{k=1}^{n} \xi_k$ is a Markov chain with

$$p(x, y) = \mathbb{P}(\xi_1 = y - x).$$

Here the transition probabilities depend only on the increment $y - x$, so the kernel is translation invariant.

Example 5.3 (Galton–Watson branching process). If $X_n = i$, the next population is the sum of i independent offspring counts. The state 0 is absorbing. The conditional distribution of the next population depends on the current population size alone.

Example 5.4 (Ehrenfest urn). There are r particles in two boxes; at each step one uniformly selected particle changes box. If X_n is the number in the left box, then

$$p(i, i+1) = \frac{r - i}{r}, \qquad p(i, i-1) = \frac{i}{r}.$$

This finite birth–death chain has period 2, since every transition changes the parity of X_n.

Example 5.5 (Wright–Fisher model). In a population of size N, if $X_n = i$ copies of an allele are present, then

$$X_{n+1} \mid X_n = i \sim \text{Binomial}(N, i/N).$$

The boundary states 0 and N are absorbing; genetic drift eventually produces loss or fixation.

Example 5.6 ($M/G/1$ embedded queue). Observe a single-server queue immediately after service completions. If A_{n+1} is the number of arrivals during the next service, then

$$X_{n+1} = (X_n - 1)^+ + A_{n+1}.$$

At these observation times, the queue length is Markov. To describe the queue between service completions, the state also includes the remaining service time.

5.2 Construction and Markov Properties

Theorem 5.7 (Canonical construction; Book 5.2.1). Given an initial law μ and a kernel p, there is a probability law $\mathbb{P}_\mu$ on the path space $S^{\{0,1,\dots\}}$ whose finite-dimensional distributions satisfy

$$\mathbb{P}_\mu(X_0 \in A_0, \dots, X_n \in A_n) = \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} p(x_0, \mathrm{d}x_1) \cdots \int_{A_n} p(x_{n-1}, \mathrm{d}x_n).$$

Under this law, X_n is Markov with transition kernel p.

Proof sketch. Define consistent finite-dimensional laws by iterated integration and invoke the extension theorem. Verify the conditional expectation identity for the next state first on cylinder events. The monotone class theorem then extends it to the σ-field generated by the past.

Theorem 5.8 (Markov property in path form; Book 5.2.2–5.2.3). Let θ_m delete the first m coordinates of a path. For bounded measurable Y on path space,

$$\mathbb{E}_\mu(Y \circ \theta_m \mid \mathcal{F}_m) = \mathbb{E}_{X_m} Y.$$

Conditionally on X_m, the future has law $\mathbb{P}_{X_m}$ and is independent of the past.

Theorem 5.9 (Chapman–Kolmogorov; Book 5.2.4). For kernels, $p^{m+n} = p^m p^n$; on a countable space,

$$p^{m+n}(x, z) = \sum_{y \in S} p^m(x, y) p^n(y, z).$$

The sum is over all possible states at the intermediate time m.

Definition 5.10 (Hitting and return times). For $A \subseteq S$, set

$$T_A = \inf\{n \geq 0 : X_n \in A\}, \qquad T_A^+ = \inf\{n \geq 1 : X_n \in A\}.$$

For a state y, write T_y, T_y^+ and define successive positive visit times $T_y^{(k)}$ recursively. These are stopping times for the natural filtration.

Theorem 5.11 (Strong Markov property; Book 5.2.5). If T is a stopping time, then on $\{T < \infty\}$ and for a uniformly bounded family of measurable path functionals Y_n,

$$\mathbb{E}_\mu(Y_T \circ \theta_T \mid \mathcal{F}_T) = \mathbb{E}_{X_T} Y_T.$$

In particular, conditionally on $\mathcal{F}_T$, the process after T has the law of a chain started at X_T.

Proof sketch. Decompose every expectation over $\{T = n\}$. On that event apply the Markov property at time n, then sum over n. Integrable functionals can be treated by truncation.

Proposition 5.12 (Repeated visits; Book 5.2.6). Let $\rho_{xy} = \mathbb{P}_x(T_y^+ < \infty)$, where the visit is required to occur at a positive time, and let $T_y^{(k)}$ be the time of the kth such visit. Then

$$\mathbb{P}_x(T_y^{(k)} < \infty) = \rho_{xy}\rho_{yy}^{k-1}.$$

By the strong Markov property, after each visit to y the probability of another visit is ρ_{yy}.

Theorem 5.13 (Reflection bound for symmetric increments; Book 5.2.7). If $S_n = \sum_{k=1}^n \xi_k$ with i.i.d. increments symmetric about 0, then for $a > 0$,

$$\mathbb{P}\left(\max_{k \le n} S_k \ge a\right) \le 2\mathbb{P}(S_n \ge a).$$

Condition on the first crossing. By symmetry, the remaining increment has probability at least $1/2$ of being nonnegative. For nearest-neighbor simple random walk, path reflection yields sharper exact identities.

5.3 Recurrence and Transience

Definition 5.14 (Recurrent and transient states). A state y is recurrent if $\rho_{yy} = 1$ and transient if $\rho_{yy} < 1$. Let

$$N(y) = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = y\}}$$

be the number of visits after time zero. Then y is recurrent iff $\mathbb{E}_y N(y) = \infty$; in the transient case $N(y)$ has a geometric tail.

Theorem 5.15 (Recurrence within a communicating class; Book 5.3.1–5.3.2). If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = 1$. Consequently recurrence or transience is a class property on an irreducible communicating class.

Theorem 5.16 (Finite closed classes and decomposition; Book 5.3.3–5.3.5). Every finite closed set contains a recurrent state; if it is irreducible, every state in it is recurrent. The full recurrent set is a disjoint union of closed irreducible classes. All remaining states are transient and may lead to a recurrent class.

Example 5.17 (Decomposition of a finite chain). A finite chain may contain several closed recurrent classes plus transient states that eventually enter one of them. No transition can leave a recurrent class. In an infinite chain there may be no recurrent class at all.

Theorem 5.18 (Foster–Lyapunov recurrence criterion; Book 5.3.8). Suppose the countable chain is irreducible. If a nonnegative function V has finite sublevel sets, tends to infinity off finite sets, and

$$\mathbb{E}_x V(X_1) \le V(x) \qquad (x \notin F)$$

for some finite F, then the chain is recurrent.

Proof sketch. Stop $V(X_n)$ on first entering F or exceeding a high level M. It is a nonnegative supermartingale, so optional stopping bounds the probability of escaping to level M before hitting F by $V(x)/M$. Let $M \to \infty$. Irreducibility and finiteness of F then imply recurrence.

Remark 5.19 (Superharmonic functions). The recurrence criterion uses a superharmonic function V that tends to infinity outside finite sets. Transience criteria use a different boundary condition, such as a positive superharmonic function that tends to zero at infinity.

Example 5.20 (Birth–death chain). Consider a chain on $\{0, 1, 2, \dots\}$ with $p_0 = p(0,1) > 0$, $p(0,0) = 1 - p_0$, and $q_0 = 0$. For $i \ge 1$, let $p_i = p(i, i+1) > 0$ and $q_i = p(i, i-1) > 0$; the remaining probability is assigned to holding at i. Define the scale function

$$\phi(0) = 0, \qquad \phi(n) = \sum_{m=0}^{n-1} \prod_{j=1}^{m} \frac{q_j}{p_j}.$$

For integers $0 \le a < x < b$ and $T = T_a \wedge T_b$, $\phi(X_{n \wedge T})$ is a martingale, and

$$\mathbb{P}_x(T_a < T_b) = \frac{\phi(b) - \phi(x)}{\phi(b) - \phi(a)}, \qquad \mathbb{P}_x(T_b < T_a) = \frac{\phi(x) - \phi(a)}{\phi(b) - \phi(a)}.$$

State 0 is recurrent exactly when $\phi(n) \to \infty$.

5.4 Recurrence of Random Walks*

Definition 5.21 (Recurrent value). For a random walk S_n in $\mathbb{R}^d$, a point x is recurrent if for every $\varepsilon > 0$,

$$\mathbb{P}(\|S_n - x\|_\infty < \varepsilon \text{ i.o.}) = 1.$$

A point is possible if every neighborhood is hit with positive probability at some time. The set of recurrent values is either empty or equals the closed subgroup of possible values.

Theorem 5.22 (Return criteria; Book 5.4.3, 5.4.5–5.4.7). For a lattice random walk, recurrence is equivalent to each of

$$\mathbb{P}_0(T_0^+ < \infty) = 1, \qquad \mathbb{P}(S_n = 0 \text{ i.o.}) = 1, \qquad \sum_{n=0}^{\infty} \mathbb{P}(S_n = 0) = \infty.$$

For general walks, divergence of $\sum_n \mathbb{P}(\|S_n\|_\infty < \varepsilon)$ for one radius implies recurrence, while convergence implies transience; changing the radius does not change the classification.

Theorem 5.23 (Pólya classification; Book 5.4.4). Nearest-neighbor simple random walk on $\mathbb{Z}^d$ is recurrent for $d = 1, 2$ and transient for $d \geq 3$. The series criterion for recurrence and estimates from the local central limit theorem give

$$\mathbb{P}(S_{2n} = 0) \asymp n^{-d/2}.$$

Proof sketch. In dimension one, use $\binom{2n}{n} 2^{-2n} \asymp n^{-1/2}$. In dimension two, counting paths by their numbers of horizontal and vertical steps gives

$$\mathbb{P}(S_{2n} = 0) = 4^{-2n} \binom{2n}{n}^2 \sim \frac{1}{\pi n}.$$

For $d \geq 3$, compare the return probability with $n^{-d/2}$ and sum. The borderline between divergent and convergent power series occurs at $d = 2$.

Theorem 5.24 (Criteria in dimensions one and two; Book 5.4.8–5.4.9). In one dimension, if $S_n/n \to 0$ in probability, then the walk is recurrent (Chung–Fuchs). In two dimensions, if $S_n/\sqrt{n}$ converges in law to a nondegenerate normal distribution, then the walk is recurrent. Nonzero linear drift in one dimension instead forces transience by the law of large numbers.

Theorem 5.25 (Fourier criterion; Book 5.4.10–5.4.14). Let $\varphi(t) = \mathbb{E}e^{i\langle t, X_1\rangle}$. Recurrence is characterized by divergence near the origin of

$$\int_{(-\delta,\delta)^d} \operatorname{Re} \frac{1}{1-\varphi(t)}\, \mathrm{d}t,$$

equivalently by the corresponding supremum with $1 - r\varphi(t)$ as $r \uparrow 1$. This criterion implies transience when the increments span at least three dimensions.

Remark 5.26 (Dimension of the increment support). A walk in $\mathbb{R}^3$ whose increments lie in a plane can be recurrent as a walk on that plane. The dimension in the transience criterion is the dimension of the linear span of the increment support.

5.5 Stationary Measures

Definition 5.27 (Invariant, stationary, and reversible measures). A nonzero measure μ on a countable state space is stationary when

$$\mu(y) = \sum_x \mu(x)p(x,y), \qquad \text{or } \mu P = \mu.$$

If $\sum_x \mu(x) = 1$, it is a stationary distribution. It is reversible if detailed balance holds:

$$\mu(x)p(x,y) = \mu(y)p(y,x).$$

Reversibility implies stationarity, but not conversely.

Example 5.28 (Three invariant measures). For a random walk on $\mathbb{Z}^d$ with a translation invariant kernel, counting measure is stationary, even when no stationary probability exists. For the Ehrenfest chain, $\pi(i) = 2^{-r}\binom{r}{i}$ is reversible. For a birth–death chain, detailed balance gives recursively

$$\mu(n) = \mu(0) \prod_{j=0}^{n-1} \frac{p_j}{q_{j+1}};$$

this stationary measure can be normalized to a stationary distribution exactly when the sequence is summable. In that case, $\pi(n) = \mu(n)/\sum_j \mu(j)$.

Theorem 5.29 (Time reversal and cycle condition; Book 5.5.5–5.5.6). Under a stationary measure μ, restrict to its support $S_\mu = \{x : \mu(x) > 0\}$. A finite path read backwards is Markov there with dual kernel

$$q(x,y) = \frac{\mu(y)p(y,x)}{\mu(x)}.$$

For an irreducible chain, a reversible measure exists exactly when transitions occur in both directions and the product of the ratios $p(x,y)/p(y,x)$ around every finite loop equals 1.

Theorem 5.30 (Cycle occupation measure; Book 5.5.7). If x is recurrent and T_x^+ is its first positive return, then

$$\mu_x(y) = \mathbb{E}_x \sum_{n=0}^{T_x^+ - 1} \mathbf{1}_{\{X_n = y\}}$$

is stationary and satisfies $\mu_x(x) = 1$. Its total mass is $\mathbb{E}_x T_x^+$.

Proof sketch. Move one step forward inside a return cycle. The initial visit to x that disappears is exactly replaced by the terminal return to x. Taking expectations and using the Markov property proves $\mu_x P = \mu_x$.

Theorem 5.31 (Recurrence under a stationary distribution; Book 5.5.10). If a countable chain has stationary distribution π, every state y with $\pi(y) > 0$ is recurrent.

Proof sketch. If y were transient, the repeated visit formula would give

$$\mathbb{E}_x N(y) = \frac{\rho_{xy}}{1 - \rho_{yy}} \le \frac{1}{1 - \rho_{yy}}.$$

But stationarity gives $\mathbb{E}_\pi N(y) = \sum_{n \ge 1} \pi(y) = \infty$, contradicting this uniform bound.

Theorem 5.32 (Uniqueness and positive recurrence; Book 5.5.9–5.5.15). For an irreducible recurrent chain, stationary measures are unique up to multiplication by a positive constant. For an irreducible chain, the following are equivalent:

1. some state is positive recurrent, $\mathbb{E}_x T_x^+ < \infty$;
2. a stationary distribution exists;
3. every state is positive recurrent.

In that case the stationary distribution is unique and

$$\pi(x) = \frac{1}{\mathbb{E}_x T_x^+}.$$

Remark 5.33 (Null recurrence). An irreducible null recurrent chain has a unique stationary measure up to scale and no stationary probability. Simple symmetric random walk on $\mathbb{Z}$ is the standard example: counting measure is invariant and has infinite mass.

5.6 Asymptotic Behavior

Theorem 5.34 (Limiting occupation frequencies; Book 5.6.1). If y is recurrent, then for every start x,

$$\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{X_k=y\}} \longrightarrow \frac{\mathbf{1}_{\{T_y<\infty\}}}{\mathbb{E}_y T_y^+} \quad \mathbb{P}_x\text{-a.s.},$$

with $1/\infty = 0$. Thus positive recurrent states have positive limiting frequency and null recurrent states have frequency zero.

Proof sketch. Starting from y, the lengths of successive return cycles are i.i.d. by the strong Markov property. Apply the strong law to their partial sums, then express the number of visits by time n in terms of these return times. For a different start, discard the finite segment before the first hit of y.

Definition 5.35 (Period). For recurrent x, define

$$d(x) = \gcd\{n \geq 1 : p^n(x,x) > 0\}.$$

Communicating recurrent states have the same period. An irreducible chain with common period 1 is aperiodic. If $d(x) = 1$, then $p^n(x,x) > 0$ for every sufficiently large n.

Theorem 5.36 (Convergence to equilibrium; Book 5.6.6). If a chain on a countable state space is irreducible, aperiodic, and has stationary distribution π, then

$$p^n(x,y) \longrightarrow \pi(y), \qquad x, y \in S.$$

In fact the coupling proof yields convergence in total variation.

Proof sketch. Run independent copies from x and from stationarity. Aperiodicity makes the product chain irreducible. Its stationary law is $\pi \otimes \pi$, so positive recurrence implies that the two copies meet a.s. After the first meeting, use the same transitions for both copies. The coupling inequality bounds total variation by the probability that they have not yet met.

5.7 Periodicity and the Tail σ-Field*

Theorem 5.37 (Cyclic decomposition; Book 5.7.1–5.7.2). If an irreducible recurrent chain has period d, then

$$S = S_0 \sqcup \cdots \sqcup S_{d-1},$$

and each transition moves S_r to $S_{r+1 \bmod d}$. The kernel P^d restricted to each S_r is irreducible and aperiodic. If a stationary distribution exists, $\pi(S_r) = 1/d$ and, for $x \in S_0$, $y \in S_r$,

$$p^{md+r}(x, y) \longrightarrow d\pi(y).$$

Probabilities at incompatible congruence classes are zero.

Definition 5.38 (Tail field and space–time harmonicity). The path tail field is

$$\mathcal{T} = \bigcap_{n \geq 0} \sigma(X_n, X_{n+1}, \ldots).$$

A bounded function $h : S \times \mathbb{N} \to \mathbb{R}$ is space–time harmonic if

$$h(x, n) = \sum_y p(x, y) h(y, n + 1),$$

equivalently $h(X_n, n)$ is a bounded martingale.

Theorem 5.39 (Tail variables; Book 5.7.3–5.7.4). For an irreducible recurrent chain of period d, the completed tail field records only the initial cyclic phase. For any countable Markov chain, bounded tail variables Z correspond to bounded space–time harmonic functions through

$$h(X_n, n) = \mathbb{E}(Z \mid \mathcal{F}_n), \qquad Z = \lim_{n \to \infty} h(X_n, n).$$

Theorem 5.40 (Tail field of simple random walk; Book 5.7.6). For nearest-neighbor simple random walk on $\mathbb{Z}^d$, $d \geq 1$, set

$$L_0 = \{z \in \mathbb{Z}^d : z_1 + \cdots + z_d \text{ is even}\}, \qquad L_1 = \mathbb{Z}^d \setminus L_0.$$

For every initial distribution μ,

$$\mathcal{T} = \sigma(\{X_0 \in L_0\}, \{X_0 \in L_1\}) \mod \mathbb{P}_\mu.$$

Thus the tail field is trivial when the initial law is concentrated on one parity class, even in the transient dimensions $d \geq 3$.

Proof sketch. Couple two walks started in the same parity class. First run them independently until every coordinate difference is even; the finite chain of parity differences reaches this configuration a.s. Thereafter choose the same coordinate to update in both walks. Use independent signs when that coordinate differs and the same sign once it agrees. Each nonzero coordinate difference is then a recurrent lazy symmetric walk on the even integers, so all coordinates eventually agree and remain equal. The coupling implies that every bounded space–time harmonic function is constant on each parity class at each time. Its two values exchange at each step, so the tail variable correspondence gives the claimed equality. Conversely, initial parity is determined by the parity of X_n and n for every n, so it is tail measurable.

Remark 5.41 (Tail events and invariant events). The cyclic phase is tail measurable, but it changes under a shift by one time step. For the deterministic chain that alternates between two states, the tail field identifies the initial state, whereas the invariant field is trivial under the stationary law.

Example 5.42 (Ornstein coupling and trees). For an irreducible aperiodic random walk on $\mathbb{Z}$, choose a bounded truncation of its step distribution that is still irreducible and aperiodic. Couple large jumps in parallel and use independent bounded jumps until the two walks meet; afterward use identical increments. Their difference is a recurrent walk with mean zero, so they meet a.s. and the tail field is trivial. On a regular tree, a transient walk converges to a boundary direction. The limiting direction gives nonconstant bounded harmonic functions and nontrivial tail events.

5.8 General State Space*

Definition 5.43 (Harris chain). A chain is Harris if there are measurable sets A, B, a probability ρ concentrated on B, and a measurable density $q(x, y) \geq \varepsilon > 0$ on $A \times B$ such that:

1. every starting state has positive probability of reaching A;
2. for $x \in A$ and $C \subseteq B$, $p(x, C) \geq \int_C q(x, y)\rho(\,\mathrm{d}y)$.

The second condition is a local Doeblin minorization.

Example 5.44 (Kernels with densities and queueing chains). If $p(x,\,\mathrm{d}y) = k(x, y)\,\mathrm{d}y$ with k continuous and positive at some (x_0, y_0), small neighborhoods of x_0, y_0 provide the minorization, provided the first is accessible. The discrete Ornstein–Uhlenbeck recursion $X_{n+1} = \theta X_n + \xi_{n+1}$ is a standard example. For Lindley's recursion

$$W_{n+1} = (W_n + \xi_{n+1})^+,$$

the atom at 0 gives regeneration times for the $GI/G/1$ waiting time process when negative increments are possible.

5.8.1 Recurrence and Transience

Definition 5.45 (Splitting). The minorization allows a transition from A to be split: with probability ε jump to a new atom α, then redistribute according to ρ; otherwise use the residual kernel. The split chain $\bar{X}_n$ has a distinguished state α, whose return times permit the same cycle arguments used for chains on countable spaces. The original and split kernels are linked by interlacing a redistribution kernel.

Theorem 5.46 (Harris recurrence dichotomy; Book 5.8.8). Let R be the first return of the split chain to α. The chain is recurrent when $\mathbb{P}_\alpha(R < \infty) = 1$ and transient otherwise. For the irreducibility measure

$$\lambda(C) = \sum_{n=1}^{\infty} 2^{-n} \bar{p}^n(\alpha, C),$$

every set of positive λ-measure is visited infinitely often from α in the recurrent case, and $\mathbb{P}_x(R < \infty) = 1$ for λ-a.e. x.

Remark 5.47 (Starting states outside the recurrent class). The recurrence conclusion holds for λ-almost every starting state. Changing the kernel on a λ-null set can introduce starting states from which the split atom is never reached.

5.8.2 Stationary Measures

Theorem 5.48 (Existence and uniqueness; Book 5.8.9–5.8.11). In the recurrent case, occupation before the return to α defines a σ-finite stationary measure for the split chain, absolutely continuous with respect to λ. Redistributing the mass at α produces a stationary measure for the original kernel. Every σ-finite stationary measure is a constant multiple of this cycle measure.

Proof sketch. Set

$$\bar{\mu}(C) = \mathbb{E}_\alpha \sum_{n=0}^{R-1} \mathbf{1}_{\{\bar{X}_n \in C\}}.$$

Shifting the sum by one time step proves invariance, as in the case of a countable state space. The minorization supplies a countable cover by sets of finite measure, giving σ-finiteness. Represent any other stationary measure on the split chain and compare its mass over return cycles to obtain uniqueness.

5.8.3 Convergence Theorem

Theorem 5.49 (Convergence of Harris chains; Book 5.8.12). Let the recurrent Harris chain be aperiodic and possess stationary distribution π. If the split atom is hit a.s. from x, then

$$\|p^n(x, \cdot) - \pi\|_{d_{\mathrm{TV}}} \longrightarrow 0.$$

The hitting condition holds for π-a.e. x.

Proof sketch. Couple a chain started from x with one started from π at a common regeneration time at α. The difference of their successive regeneration times is a random walk with mean zero. Recurrence and aperiodicity imply that a common regeneration time occurs a.s. Use identical transitions after that time and apply the coupling inequality.

5.8.4 $GI/G/1$ Queue

Example 5.50 (Stationary waiting time). Let $(\eta_n)_{n\geq 0}$ be i.i.d. nonnegative service times and $(\zeta_n)_{n\geq 1}$ i.i.d. nonnegative interarrival times, with the two sequences independent and both means finite. Set $\xi_n = \eta_{n-1} - \zeta_n$ and assume $\mathbb{E}\eta_0 < \mathbb{E}\zeta_1$. The waiting times at arrival epochs satisfy $W_{n+1} = (W_n + \xi_{n+1})^+$ and have a stationary distribution. With $S_0 = 0$ and $S_n = \sum_{k=1}^{n} \xi_k$, Loynes' representation gives

$$W_\infty \stackrel{d}{=} \sup_{n\geq 0} S_n.$$

If service times are exponential with rate β, there is a positive solution $\theta \in (0, \beta)$ of $\mathbb{E}e^{\theta\xi_1} = 1$. Conditional on their occurrence, ascending ladder heights are exponential with rate β; the exponential martingale determines their probability of occurrence. This gives

$$\mathbb{P}(W_\infty > x) = \left(1 - \frac{\theta}{\beta}\right) e^{-\theta x}, \qquad x \geq 0,$$

with an atom of mass θ/β at 0.

If instead interarrival times are exponential with rate α, let $B(s) = \mathbb{E}e^{-s\eta_0}$ and $\rho = \alpha\mathbb{E}\eta_0 < 1$. The stationary law is described by the Pollaczek–Khintchine transform

$$\mathbb{E}e^{-sW_\infty} = \frac{(1-\rho)s}{s - \alpha + \alpha B(s)}, \qquad s > 0.$$

Remark 5.51 (Drift and stationary waiting times). Here ξ_1 is a service time minus the following interarrival time, so $\mathbb{E}\xi_1 < 0$ means that the mean service time is smaller than the mean interarrival time. At positive drift, the waiting time tends to infinity and there is no stationary distribution. The case of zero drift requires separate recurrence analysis.

6 Ergodic Theorems

6.1 Definitions and Examples

Definition 6.1 (Stationary sequence). A sequence $(X_n)_{n\geq 0}$ is stationary if, for every $k, m \geq 0$,

$$(X_0, \ldots, X_m) \stackrel{d}{=} (X_k, \ldots, X_{k+m}).$$

Thus every finite collection of coordinates has a distribution invariant under a common shift of its indices.

Example 6.2 (Basic stationary processes). An i.i.d. sequence is stationary. A Markov chain started from a stationary distribution is stationary. The deterministic chain that alternates between two states is stationary when its initial state is uniform. Its tail field identifies the initial state.

Definition 6.3 (Measure-preserving dynamical system). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A measurable map $T : \Omega \to \Omega$ is measure preserving when

$$\mathbb{P}(T^{-1}A) = \mathbb{P}(A), \qquad A \in \mathcal{F}.$$

For $X \in L^0$, the process $X_n = X \circ T^n$ is stationary. Conversely, every stationary sequence on a standard Borel state space can be represented on sequence space with T the left shift and X the zeroth coordinate.

Definition 6.4 (Invariant field and ergodicity). The invariant σ-field is

$$\mathcal{I} = \{A \in \mathcal{F} : T^{-1}A = A \bmod \mathbb{P}\}.$$

Equivalently, an $\mathcal{I}$-measurable random variable is invariant under T a.s. The system is ergodic if every $A \in \mathcal{I}$ has probability 0 or 1.

Remark 6.5 (Invariant components). If $A \in \mathcal{I}$ has probability strictly between 0 and 1, then A and A^c are two invariant components. A stationary law is then a mixture of component laws. An ergodic law has no such decomposition into two invariant components of positive probability.

Example 6.6 (The shift on an i.i.d. sequence). Every shift-invariant event belongs to the tail field

$$\mathcal{T} = \bigcap_{n\geq 0} \sigma(X_n, X_{n+1}, \ldots).$$

Kolmogorov's zero–one law makes this tail field trivial, so the i.i.d. shift is ergodic. The inclusion $\mathcal{I} \subseteq \mathcal{T}$ may be strict for dependent or periodic processes.

Example 6.7 (Stationary Markov chain). For a chain on a countable state space with stationary law positive on every state, the path shift is ergodic exactly when the recurrent support is one irreducible class. This includes periodic chains. For example, the stationary chain that alternates between two states is ergodic, although its tail field records the initial phase.

Example 6.8 (Rotation of the circle). On $[0,1)$ with Lebesgue measure, let $T(x) = x + \theta \pmod 1$. If $\theta = p/q$ is rational, the set $\{x : \{qx\} < 1/2\}$ is invariant and has measure $1/2$, so the rotation is not ergodic. If θ is irrational, Fourier uniqueness shows every invariant L^2 function is constant, so the rotation is ergodic. It is nevertheless not mixing.

Proof sketch. If $f \circ T = f$ and $f(x) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi i k x}$ in L^2, then $c_k(e^{2\pi i k\theta} - 1) = 0$. For irrational θ, the factor in parentheses is nonzero for $k \neq 0$. Apply the conclusion to $f = \mathbf{1}_A$ for an invariant set A.

6.2 Birkhoff's Ergodic Theorem

Theorem 6.9 (Birkhoff; Book 6.2.1). If T is measure preserving and $X \in L^1$, then

$$A_n X := \frac{1}{n}\sum_{k=0}^{n-1} X \circ T^k \longrightarrow \mathbb{E}(X \mid \mathcal{I}) \quad \text{a.s. and in } L^1.$$

If the system is ergodic, the limit is the constant $\mathbb{E}X$.

Lemma 6.10 (Maximal ergodic lemma; Book 6.2.2). Let $S_k = \sum_{j=0}^{k-1} X \circ T^j$ and $M_n = \max(0, S_1, \ldots, S_n)$. Then

$$\mathbb{E}[X\mathbf{1}_{\{M_n > 0\}}] \geq 0.$$

The result is a deterministic telescoping inequality integrated using measure preservation.

Proof sketch. First subtract $\mathbb{E}(X \mid \mathcal{I})$, reducing to invariant conditional mean zero. For $\varepsilon > 0$, apply the maximal lemma to $X - \varepsilon$ on the invariant event $\{\limsup_n A_n X > \varepsilon\}$. Conditional mean zero makes the resulting nonnegative integral equal to $-\varepsilon$ times that event's probability, so the probability is zero. Repeat for $-X$ to obtain a.s. convergence. For L^1 convergence, truncate X, control the bounded part, and use measure preservation to make the averaged tail uniformly small.

Example 6.11 (A nonconstant invariant limit). Let Z take the values -1 and 1 with equal probability, and set $X_n = Z$ for every n. The sequence is stationary, and its averages equal Z, whereas $\mathbb{E}X_n = 0$. Since Z is invariant under the shift, Birkhoff's limit is $\mathbb{E}(Z \mid \mathcal{I}) = Z$.

Example 6.12 (Strong law and Markov occupation times). For the i.i.d. shift, $\mathcal{I}$ is trivial, so Birkhoff gives the strong law for every integrable coordinate variable. For an irreducible stationary Markov chain and integrable f,

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \sum_x \pi(x) f(x) \quad \text{a.s.}$$

Taking $f = \mathbf{1}_{\{y\}}$ recovers the limiting occupation frequency $\pi(y)$ without requiring aperiodicity.

Example 6.13 (Equidistribution of irrational rotations). For irrational θ and Riemann integrable f,

$$\frac{1}{n}\sum_{k=0}^{n-1} f(x + k\theta) \to \int_0^1 f(u)\, \mathrm{d}u.$$

Birkhoff first gives the conclusion for a.e. starting x. For indicators of intervals, approximation from inside and density of a set of full measure remove the exceptional set, giving convergence for every x. After a logarithmic change of coordinates, this also gives Benford's law for leading digits in suitable sequences.

Remark 6.14 (Simultaneous convergence). Applying Birkhoff to a countable determining class and intersecting the resulting sets of full measure gives simultaneous convergence for that class. Approximation can then extend convergence to other functions.

6.3 Recurrence

Theorem 6.15 (Range growth; Book 6.3.1). Let (X_n) be a stationary $\mathbb{R}^d$-valued sequence, $S_k = \sum_{j=1}^{k} X_j$, and

$$R_n = |\{S_1, \ldots, S_n\}|, \qquad A = \{S_k \neq 0 \text{ for every } k \geq 1\}.$$

Then

$$\frac{R_n}{n} \longrightarrow \mathbb{E}(\mathbf{1}_A \mid \mathcal{I}) \quad \text{a.s.}$$

In the ergodic case the limit is the nonreturn probability $\mathbb{P}(A)$.

Proof sketch. Count times whose current location is never revisited; shifted copies of $\mathbf{1}_A$ give a lower bound for R_n. For an upper bound, count locations not revisited during the next fixed k steps. The discrepancy at the endpoints is $O(k)$. Apply Birkhoff to both counts, then let $k \to \infty$ and use conditional monotone convergence.

Theorem 6.16 (Recurrence for stationary integer increments; Book 6.3.2). Let $X_n \in \mathbb{Z}$ be stationary with $\mathbb{E}\,|X_1| < \infty$ and $S_n = \sum_{k=1}^{n} X_k$. If

$$\mathbb{E}(X_1 \mid \mathcal{I}) = 0,$$

then $\mathbb{P}(S_k \neq 0 \text{ for all } k \geq 1) = 0$ and, in fact, $\mathbb{P}(S_n = 0 \text{ i.o.}) = 1$.

Proof sketch. Birkhoff gives $S_n/n \to 0$ and hence $\max_{k \leq n} |S_k|/n \to 0$. Since an integer path with maximum displacement M visits at most $2M+1$ distinct sites, $R_n/n \to 0$. The range theorem forces the nonreturn probability to vanish. Applying the return argument to shifted paths gives infinitely many returns.

Example 6.17 (Zero mean without recurrence). Choose $X_n \equiv 1$ or $X_n \equiv -1$ with equal probability. This stationary sequence has mean zero, but its partial sums never return to zero. Its conditional mean given $\mathcal{I}$ is $+1$ or -1, so it does not satisfy the recurrence theorem's centering condition.

Theorem 6.18 (Kac recurrence formula; Book 6.3.3). Let (X_n) be stationary, let $\mathbb{P}(X_0 \in A) > 0$, and suppose A is hit a.s. Conditionally on $X_0 \in A$, the successive gaps $\tau_k = T_k - T_{k-1}$ between visits to A form a stationary sequence and

$$\mathbb{E}(T_1 \mid X_0 \in A) = \frac{1}{\mathbb{P}(X_0 \in A)}.$$

This identity follows from stationarity of the successive return gaps.

6.4 A Subadditive Ergodic Theorem

Definition 6.19 (Subadditive process). For a family $(X_{m,n})_{0 \leq m < n}$, the subadditivity condition used here is

$$X_{0,n} \leq X_{0,m} + X_{m,n}.$$

The inequality compares the value on $[0, n]$ with the sum of the values on $[0, m]$ and $[m, n]$. For minimum passage times, it follows by joining a path on the first interval to a path on the second. Superadditive quantities are handled by changing signs.

Theorem 6.20 (Liggett–Kingman subadditive theorem; Book 6.4.1). Suppose $(X_{m,n})$ satisfies:

1. $X_{0,n} \le X_{0,m} + X_{m,n}$ for $0 < m < n$;
2. for each k, $(X_{nk,(n+1)k})_{n\ge 0}$ is stationary;
3. the law of $(X_{m,m+k})_{k\ge 1}$ does not depend on m;
4. $\mathbb{E}X_{0,1}^{+} < \infty$ and $\mathbb{E}X_{0,n} \ge \gamma_0 n$ for a finite γ_0.

Then

$$\frac{\mathbb{E}X_{0,n}}{n} \longrightarrow \gamma := \inf_{m\ge 1} \frac{\mathbb{E}X_{0,m}}{m}, \qquad \frac{X_{0,n}}{n} \longrightarrow X \quad \text{a.s. and in } L^1,$$

with $\mathbb{E}X = \gamma$. If every block sequence in item 2 is ergodic, then $X = \gamma$ a.s.

Proof sketch. Expectations are subadditive, so the division algorithm proves Fekete's formula $\mathbb{E}X_{0,n}/n \to \inf_m \mathbb{E}X_{0,m}/m$. Decompose $[0,n]$ into intervals of fixed length m and a remainder. Birkhoff gives convergence of the average over these intervals, while the normalized remainder tends to zero. Hence $\limsup_n X_{0,n}/n \le m^{-1}\mathbb{E}(X_{0,m} \mid \mathcal{I}_m)$.

For the lower bound, truncate the liminf and subtract $\varepsilon > 0$. Partition the interval into pieces of bounded length on which this truncated value approximates the normalized process. Stationarity bounds the expected contribution from the remaining pieces and gives

$$\mathbb{E}\left(\liminf_n \frac{X_{0,n}}{n}\right) \ge \gamma.$$

The upper and lower bounds imply a.s. convergence. Estimates for the positive parts and the averages over fixed intervals then give convergence in L^1.

Example 6.21 (Additive sums, range, and longest common subsequences). For a stationary sequence ξ_k, taking $X_{m,n} = \sum_{j=m+1}^{n} \xi_j$ recovers an additive ergodic limit. The number of distinct sites visited by a walk with stationary increments is subadditive and therefore has a linear growth rate. The longest common subsequence length $L_{m,n}$ is superadditive under concatenation, so $L_{0,n}/n$ has an a.s. limit after applying the theorem to $-L$.

Remark 6.22 (Stationarity and the limiting rate). The stationarity assumption applies to the entire sequence $(X_{nk,(n+1)k})_{n\ge 0}$ for each fixed k. Together with the lower bound on expectations, it gives an integrable limiting rate. Ergodicity makes that rate constant; its numerical value generally requires a separate argument.

6.5 Applications

Example 6.23 (Products of random matrices). Let (A_n) be a stationary sequence of positive $k \times k$ matrices and assume the relevant logarithms are integrable. Submultiplicativity of an operator norm gives

$$\log \|A_1 \cdots A_n\| \le \log \|A_1 \cdots A_m\| + \log \|A_{m+1} \cdots A_n\| .$$

Hence $n^{-1} \log \|A_1 \cdots A_n\|$ converges a.s.; under ergodicity its limit is a constant top Lyapunov exponent. Strict positivity also makes the exponential growth rates of individual matrix entries agree.

Example 6.24 (Longest increasing subsequence). Place a Poisson process of intensity one in the plane and let $L_{s,t}$ be the longest northeast chain in the square $(s,t]^2$. Concatenation along the diagonal gives $L_{0,m} + L_{m,n} \le L_{0,n}$, so

$$\frac{L_{0,n}}{n} \to \gamma \quad \text{a.s.}$$

If $\tau(N)$ is the side length needed to contain N points, the strong law gives $\tau(N)/\sqrt{N} \to 1$. Conditioned on the point count, coordinate ranks form a uniform permutation, hence

$$\frac{\ell(\pi_N)}{\sqrt{N}} \to \gamma.$$

Subadditivity proves existence; separate arguments give the exact constant $\gamma = 2$.

Remark 6.25 (The Poisson construction). The almost sure convergence above uses the permutations obtained from one Poisson process, so all sizes are represented on a common probability space. In particular, the distribution of the longest increasing subsequence of a uniform permutation satisfies $\ell(\pi_N)/\sqrt{N} \to 2$ in probability.

Example 6.26 (First-passage percolation). Assign i.i.d. nonnegative passage times to nearest-neighbor edges of $\mathbb{Z}^d$ and let $t(x,y)$ be the minimum path time. For $u \in \mathbb{Z}^d$,

$$t(0,nu) \le t(0,mu) + t(mu,nu),$$

so $t(0,nu)/n$ converges a.s. and in L^1 to a deterministic time constant under the usual integrability and ergodicity assumptions. Finite mean of the minimum of the $2d$ incident edge weights is the natural threshold for a finite limit. These limits describe passage times in each fixed direction. A shape theorem follows when convergence can also be controlled uniformly over directions.

Example 6.27 (Age-dependent branching). Start with one individual in generation 0, born at time 0. Each individual has a nonnegative lifetime with common distribution F and, at death, produces k offspring with probability p_k. The lifetimes and offspring counts are independent across individuals and independent of each other; the offspring begin independent copies of the same process. Assume $p_0 = 0$ and that F has finite mean.

Let $X_{0,n}$ be the earliest birth time in generation n, and choose an individual i_m born at time $X_{0,m}$, resolving ties independently of the descendant processes. For $m < n$, define $X_{m,n}$ as the time from the birth of i_m until the first of its descendants in generation n is born. Following those descendants gives

$$X_{0,n} \le X_{0,m} + X_{m,n}.$$

For each fixed k, the sequence $(X_{nk,(n+1)k})_{n\ge 0}$ is i.i.d., and the law of $(X_{m,m+k})_{k\ge 1}$ does not depend on m. Since $X_{0,n} \ge 0$ and $X_{0,1}$ has distribution F, Liggett's theorem gives $X_{0,n}/n \to \gamma$ a.s. and in L^1 for a deterministic γ. Subadditivity for arbitrary triples can fail: i_m need not descend from i_ℓ when $0 < \ell < m$.

7 Brownian Motion

Brownian motion is a continuous Gaussian process with independent increments. Its Markov and martingale properties determine hitting probabilities and exit times, while its quadratic variation gives the second derivative term in Itô's formula.

7.1 Definition and Construction

Definition 7.1 (Standard Brownian motion). A real-valued process $(B_t)_{t\ge 0}$ is a standard Brownian motion started at x if

1. $B_0 = x$ almost surely;
2. increments over disjoint time intervals are independent;
3. $B_{s+t} - B_s \sim N(0, t)$ for $s, t \ge 0$; and
4. with probability one, $t \mapsto B_t$ is continuous.

When $x = 0$, the finite-dimensional laws are equivalently characterized by Gaussianity, zero mean, and covariance

$$\mathbb{E}[B_s B_t] = s \wedge t.$$

For a Gaussian process, zero covariance of disjoint increments implies their independence. This implication is special to joint Gaussian laws.

Proposition 7.2 (Translation and scaling). Under $\mathbb{P}^x$, the process $B_t - x$ is standard Brownian motion started at zero. If $c > 0$, then

$$\{B_{ct} : t \ge 0\} \overset{d}{=} \{\sqrt{c}\, B_t : t \ge 0\}$$

for Brownian motion started at zero.

The scaling identity holds in distribution. In particular, an increment over a time interval of length h has standard deviation $h^{1/2}$.

Constructing the canonical process. The finite-dimensional densities are specified by the Gaussian transition kernel

$$p_t(x,y) = (2\pi t)^{-1/2} \exp\left(-\frac{(y-x)^2}{2t}\right).$$

The semigroup identity $p_s * p_t = p_{s+t}$ gives consistency, so Kolmogorov's extension theorem constructs a process on the dyadic rational times. Moment estimates imply uniform continuity on each compact interval almost surely. These paths extend uniquely to continuous functions, giving a probability measure on $C([0,\infty),\mathbb{R})$.

Finite-dimensional distributions determine the law on the coordinate σ-field. The continuity argument supplies a version supported on continuous paths.

Theorem 7.3 (Kolmogorov continuity criterion; Book 7.1.3). Suppose a process indexed by dyadic times in $[0,1]$ satisfies

$$\mathbb{E}\,|X_t - X_s|^{\beta} \le K\,|t-s|^{1+\alpha}$$

for some $\alpha,\beta>0$. For every $\gamma<\alpha/\beta$, it has a version whose paths satisfy

$$|X_t - X_s| \le C(\omega)\,|t-s|^{\gamma}$$

on the dyadic set and hence extend continuously.

Proof sketch. At level n, the moment estimate and a union bound show that the probabilities of an adjacent dyadic increment exceeding $2^{-\gamma n}$ are summable. Borel–Cantelli therefore bounds all such increments for sufficiently large n. A chaining argument expresses the difference between any two dyadic values as a sum of adjacent increments whose bounds form a geometric series.

For Brownian motion, $\mathbb{E}\,|B_t - B_s|^{2m} = C_m\,|t-s|^m$. Applying the criterion with arbitrarily large m gives every Hölder exponent below $1/2$.

Theorem 7.4 (Brownian path regularity; Book 7.1.5–7.1.6). On every compact interval, Brownian paths are Hölder continuous of every order $\gamma<1/2$. Almost surely they are not Lipschitz, and in fact are differentiable at no time.

Remark 7.5 (Modulus of continuity). Brownian paths are almost surely not Hölder continuous of order $1/2$ on a nontrivial interval. Their sharp modulus of continuity includes a logarithmic factor in addition to the scale $h^{1/2}$ of an increment.

Along nested dyadic partitions of $[0,t]$,

$$\sum_i (B_{t_{i+1}} - B_{t_i})^2 \longrightarrow t \quad \text{a.s.}$$

This quadratic variation contributes the second derivative term in Itô's formula.

Definition 7.6 (Multidimensional Brownian motion). A d-dimensional Brownian motion has independent one-dimensional Brownian coordinates. Its transition density is

$$p_t(x,y) = (2\pi t)^{-d/2} \exp\left(-\frac{|y-x|^2}{2t}\right).$$

Its increments are centered Gaussian vectors with covariance $\mathbb{E}[(B_t^i - B_s^i)(B_t^j - B_s^j)] = (t-s)\mathbf{1}_{\{i=j\}}$.

Remark 7.7 (Diffusive scaling). A centered random walk with finite variance has displacement of order $\sqrt{n}$ after n steps. Donsker's theorem gives Brownian motion as the limit after space is divided by $\sqrt{n}$ and time by n.

7.2 Markov Property and Blumenthal's Zero–One Law

Let

$$\mathcal{F}_t^0 = \sigma(B_s : s \le t), \qquad \mathcal{F}_t^+ = \bigcap_{u>t} \mathcal{F}_u^0.$$

Right continuity ensures that decreasing limits of stopping times are stopping times. Completing $\mathcal{F}_t^+$ by null sets gives the usual Brownian filtration.

For a path ω, let $(\theta_s\omega)(t) = \omega(s+t)$.

Theorem 7.8 (Markov property; Book 7.2.1). If Y is a bounded measurable functional of a path, then

$$\mathbb{E}^x[Y \circ \theta_s \mid \mathcal{F}_s^+] = \mathbb{E}^{B_s} Y.$$

Thus $(B_{s+t} - B_s)_{t\ge 0}$ is a Brownian motion independent of $\mathcal{F}_s^+$.

Proof sketch. Independent increments and the Gaussian transition density give the identity for cylinder functionals depending on finitely many future times. A monotone class argument extends it to bounded measurable path functionals. Letting deterministic times decrease to s and using dominated convergence establishes the identity for $\mathcal{F}_s^+$.

The Markov identity implies that conditioning on $\mathcal{F}_s^+$ gives the same result as conditioning on $\mathcal{F}_s^0$. At $s = 0$ this becomes a zero–one law.

Theorem 7.9 (Blumenthal's zero–one law; Book 7.2.3). For an event A in the germ field

$$\mathcal{F}_{0+} = \bigcap_{t>0} \mathcal{F}_t,$$

$\mathbb{P}^x(A)$ is either zero or one.

The germ field contains events determined by the path in every interval $[0,\varepsilon]$. Any such event with positive probability consequently has probability one.

Example 7.10 (Immediate oscillation at the origin). Starting from zero, Brownian motion enters $(0,\infty)$ immediately with probability one. Symmetry gives the same statement for $(-\infty,0)$. Continuity therefore implies that zero is revisited at arbitrarily small positive times. The initial zero is not an isolated zero.

Theorem 7.11 (Time inversion; Book 7.2.6). If B starts at zero, define $X_0 = 0$ and

$$X_t = tB_{1/t}, \qquad t > 0.$$

Then X is again Brownian motion.

The covariance calculation gives the finite-dimensional laws. Continuity at zero follows from $B_u/u \to 0$ as $u \to \infty$, obtained from a strong law plus control of the oscillation between integer times.

Time inversion turns a tail event at infinity into a germ event at zero.

Theorem 7.12 (Brownian tail zero–one law; Book 7.2.7). If $A \in \bigcap_{t\geq 0} \sigma(B_s : s \geq t)$, then either $\mathbb{P}^x(A) = 0$ for every x or $\mathbb{P}^x(A) = 1$ for every x.

An application is

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{t}} = \infty, \qquad \liminf_{t\to\infty} \frac{B_t}{\sqrt{t}} = -\infty \quad \text{a.s.}$$

Continuity then shows that one-dimensional Brownian motion hits every level repeatedly and returns to zero at unbounded times.

Remark 7.13 (Recurrence in higher dimensions). Two-dimensional Brownian motion returns to every neighborhood of its starting point but does not hit a prescribed point after time zero. In dimensions three and higher, Brownian motion is transient. These properties follow from the radial harmonic functions used in potential theory.

7.3 Stopping Times and the Strong Markov Property

Definition 7.14 (Stopping time in continuous time). A random variable $T : \Omega \to [0,\infty]$ is a stopping time if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. With a right-continuous filtration this is equivalent to requiring $\{T < t\} \in \mathcal{F}_t$.

For a continuous adapted process, the first entrance time into an open set is a stopping time because

$$\{T_G < t\} = \bigcup_{q\in\mathbb{Q},\, q<t} \{B_q \in G\}.$$

The union is countable and belongs to $\mathcal{F}_t$. Hitting times of closed sets follow by approximating the sets with decreasing open neighborhoods and using path continuity.

Proposition 7.15 (Closure operations). Increasing or decreasing limits of stopping times are stopping times. The minimum, maximum, sum, supremum, infimum, limsup, and liminf of countably many stopping times are stopping times whenever the corresponding expressions take values in $[0,\infty]$.

Definition 7.16 (Information at a stopping time). For a stopping time T,

$$\mathcal{F}_T = \{A : A \cap \{T \le t\} \in \mathcal{F}_t \text{ for every } t \ge 0\}.$$

If $S \le T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$, and B_T is $\mathcal{F}_T$-measurable.

Approximate T from above by dyadic stopping times

$$T_n = 2^{-n}(\lfloor 2^n T \rfloor + 1).$$

Then $T_n \downarrow T$, $B_{T_n} \to B_T$, and $\mathcal{F}_T = \bigcap_n \mathcal{F}_{T_n}$. Identities at deterministic times can then be extended to stopping times by passing to the limit.

Theorem 7.17 (Strong Markov property; Book 7.3.9). Let T be a stopping time. On $\{T < \infty\}$, conditional on $\mathcal{F}_T$, the process

$$\{B_{T+t} - B_T : t \ge 0\}$$

is a Brownian motion started at zero and independent of $\mathcal{F}_T$. Equivalently, for an appropriate bounded jointly measurable functional Y_t,

$$\mathbb{E}^x[Y_T \circ \theta_T \mid \mathcal{F}_T] = \mathbb{E}^{B_T} Y_T.$$

Proof sketch. For a stopping time taking countably many values, the identity follows by partitioning according to those values. The dyadic approximations $T_n \downarrow T$, path continuity, and the Feller continuity of $x \mapsto \mathbb{E}^x f(B_t)$ give the result for bounded continuous cylinder functions by bounded convergence. A monotone class argument extends it to bounded measurable path functionals.

Remark 7.18 (Regeneration). At an almost surely finite hitting time T, the increments after T have the law of Brownian motion and are independent of $\mathcal{F}_T$. Successive hitting times can therefore be used to construct independent path segments, just as returns to a recurrent state produce independent excursions for a Markov chain.

7.4 Path Properties

7.4.1 Zeros of Brownian Motion

Let $Z = \{t \ge 0 : B_t = 0\}$. Path continuity makes Z closed. Strong Markov at the first zero after a rational time, combined with immediate return to zero, shows that Z has no isolated points. Hence it is uncountable.

On the other hand, for each $T < \infty$, Tonelli's theorem gives

$$\mathbb{E}\lambda(Z \cap [0,T]) = \int_0^T \mathbb{P}(B_t = 0)\, \mathrm{d}t = 0.$$

Here λ denotes Lebesgue measure. Thus the zero set is uncountable but has Lebesgue measure zero. Its Hausdorff dimension is $1/2$.

7.4.2 Hitting Times

For $a \geq 0$, let $T_a = \inf\{t \geq 0 : B_t = a\}$ under $\mathbb{P}^0$.

Theorem 7.19 (Process of hitting times; Book 7.4.1). The process $(T_a)_{a\geq 0}$ has stationary independent increments and scaling

$$T_a \stackrel{d}{=} a^2 T_1.$$

Its finite-dimensional laws are those of a stable subordinator of index $1/2$. A right-continuous version is

$$\widehat{T}_a = \inf\{t \geq 0 : B_t > a\}, \qquad a \geq 0.$$

For $0 \leq a < b$, the strong Markov property at T_a shows that $T_b - T_a$ has the law of T_{b-a} and is independent of $\mathcal{F}_{T_a}$.

Theorem 7.20 (Reflection principle; Book 7.4.2). For $a > 0$ and $t > 0$,

$$\mathbb{P}^0(T_a \leq t) = 2\mathbb{P}^0(B_t \geq a).$$

Consequently T_a has density

$$f_{T_a}(s) = \frac{a}{\sqrt{2\pi s^3}} \exp\left(-\frac{a^2}{2s}\right), \qquad s > 0.$$

Proof sketch. Reflecting the path about a after T_a preserves its law by the strong Markov property and symmetry. Among paths that hit a before t, the endpoint is therefore equally likely to lie above or below a. Since $\{B_t > a\} \subseteq \{T_a < t\}$, this gives the factor two. Differentiating the Gaussian tail gives the density.

The density is asymptotic to $a(2\pi)^{-1/2}s^{-3/2}$ as $s \to \infty$. Thus $\mathbb{E}T_a = \infty$, although $T_a < \infty$ almost surely.

Example 7.21 (Arcsine law for the last zero). Let $L = \sup\{t \leq 1 : B_t = 0\}$. Then

$$\mathbb{P}(L \leq s) = \frac{2}{\pi} \arcsin\sqrt{s}, \qquad 0 \leq s \leq 1,$$

with density $\{\pi\sqrt{s(1-s)}\}^{-1}$. The density is largest near the endpoints: the last zero is more likely to be very early or very late than near $1/2$.

Remark 7.22 (Discrete reflection and arcsine limits). Under diffusive rescaling, simple random walk converges to Brownian motion. Applying the continuous mapping theorem to the last zero gives the corresponding arcsine limit for the walk.

7.5 Brownian Martingales

Theorem 7.23 (Optional sampling at bounded stopping times). If M_t is a right-continuous martingale adapted to a right-continuous filtration and T is bounded, then $\mathbb{E}M_T = \mathbb{E}M_0$.

Dyadic approximation of T reduces the claim to discrete optional sampling. Reverse martingale convergence passes from $T_n \downarrow T$ to T.

Proposition 7.24 (Basic Brownian martingales). For fixed $\theta \in \mathbb{R}$,

$$B_t, \qquad B_t^2 - t, \qquad \exp\left(\theta B_t - \frac{\theta^2 t}{2}\right)$$

are martingales. More generally, if a sufficiently integrable smooth function u solves the backward heat equation

$$\partial_t u + \frac{1}{2}\partial_{xx} u = 0,$$

then $u(t, B_t)$ is a martingale.

Example 7.25 (Two-sided exit probability). For $a < x < b$ and $T_c = \inf\{t : B_t = c\}$,

$$\mathbb{P}^x(T_b < T_a) = \frac{x - a}{b - a}.$$

Stop B_t at $T_a \wedge T_b \wedge n$, then use bounded convergence. The affine answer is also the unique harmonic function on (a, b) with boundary values 0 and 1.

Example 7.26 (Expected exit time). If $T = \inf\{t : B_t \notin (a, b)\}$ with $a < 0 < b$, then

$$\mathbb{E}^0 T = -ab.$$

Stop $B_t^2 - t$ at $T \wedge n$ and use monotone convergence for $T \wedge n$ and bounded convergence for $B_{T\wedge n}^2$.

Theorem 7.27 (Laplace transform of a hitting time; Book 7.5.7). For $a > 0$ and $\lambda > 0$,

$$\mathbb{E}^0 e^{-\lambda T_a} = e^{-a\sqrt{2\lambda}}.$$

Stopping the exponential martingale with $\theta = \sqrt{2\lambda}$ at $T_a \wedge n$ gives an integrand bounded by $e^{\theta a}$. Bounded convergence yields the transform, which also identifies the stable index $1/2$.

Polynomial solutions of the backward heat equation give the martingales

$$B_t^3 - 3tB_t, \qquad B_t^4 - 6tB_t^2 + 3t^2,$$

which give identities for higher moments of exit times after optional stopping.

Theorem 7.28 (Second moment of an exit time; Book 7.5.9). For $a > 0$, let $T = \inf\{t \geq 0 : B_t \notin (-a, a)\}$ under $\mathbb{P}^0$. Then

$$\mathbb{E}^0 T^2 = \frac{5a^4}{3}.$$

Proof sketch. Stopping $B_t^4 - 6tB_t^2 + 3t^2$ at $T \wedge n$ gives

$$\mathbb{E}^0[B_{T\wedge n}^4 - 6(T \wedge n)B_{T\wedge n}^2] = -3\mathbb{E}^0(T \wedge n)^2.$$

Since $\mathbb{E}^0 T = a^2$ and $|B_{T\wedge n}| \leq a$, dominated convergence on the left and monotone convergence on the right yield $a^4 - 6a^2\mathbb{E}^0 T = -3\mathbb{E}^0 T^2$, which gives the formula.

7.6 Itô's Formula*

In a Taylor expansion along a continuous path of finite variation, the sum of squared increments tends to zero as the partition is refined. For Brownian motion this sum tends to elapsed time, so the second derivative contributes to the limiting formula.

Definition 7.29 (Itô integral along Brownian motion). For a simple adapted process

$$H_s = \sum_{k=0}^{m-1} H_k \mathbf{1}_{(t_k, t_{k+1}]}(s), \qquad H_k \in \mathcal{F}_{t_k},$$

define

$$\int_0^t H_s \, \mathrm{d}B_s = \sum_{k=0}^{m-1} H_k (B_{t_{k+1}\wedge t} - B_{t_k \wedge t}).$$

Completion in L^2 defines the integral for predictable square-integrable integrands.

The Itô isometry states that

$$\mathbb{E}\left(\int_0^t H_s \, \mathrm{d}B_s\right)^2 = \mathbb{E}\int_0^t H_s^2 \, \mathrm{d}s.$$

It follows first by expanding the square: cross terms vanish because each future increment has conditional mean zero. The isometry makes the approximating sums Cauchy in L^2.

Theorem 7.30 (One-dimensional Itô formula; Book 7.6.1). If $f \in C^2(\mathbb{R})$, then almost surely, for every $t \geq 0$,

$$f(B_t) - f(B_0) = \int_0^t f'(B_s) \, \mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s) \, \mathrm{d}s.$$

Proof sketch. For bounded derivatives, Taylor expansion on a refining sequence of partitions gives a linear sum converging in L^2 to the Itô integral. Quadratic variation makes the sum involving squared increments converge to the time integral of $f''(B_s)$, while continuity of f'' controls the remainders. Stopping at the exit time from $[-M, M]$ and letting $M \to \infty$ extends the formula to every $f \in C^2$.

Taking $f(x) = x^2$ yields

$$\int_0^t 2B_s \, \mathrm{d}B_s = B_t^2 - B_0^2 - t.$$

The term t is the quadratic variation of Brownian motion on $[0, t]$.

Theorem 7.31 (Martingale property of the stochastic integral). If g is continuous and $\mathbb{E}\int_0^t |g(B_s)|^2 \, \mathrm{d}s < \infty$, then

$$M_t = \int_0^t g(B_s) \, \mathrm{d}B_s$$

is a continuous martingale. For a continuous g without this integrability assumption, the integral is a continuous local martingale.

Theorem 7.32 (Space–time Itô formula; Book 7.6.5). For $f \in C^{1,2}([0,\infty) \times \mathbb{R})$,

$$f(t, B_t) - f(0, B_0) = \int_0^t \partial_s f(s, B_s) \, \mathrm{d}s + \int_0^t \partial_x f(s, B_s) \, \mathrm{d}B_s + \frac{1}{2}\int_0^t \partial_{xx} f(s, B_s) \, \mathrm{d}s.$$

Thus $f(t, B_t)$ has drift $(\partial_t + \frac{1}{2}\partial_{xx})f$. Vanishing drift produces a local martingale and, with suitable integrability, a true martingale.

Example 7.33 (Exponential Brownian motion). For $X_t = e^{\mu t + \sigma B_t}$,

$$\mathrm{d}X_t = \left(\mu + \frac{\sigma^2}{2}\right) X_t \, \mathrm{d}t + \sigma X_t \, \mathrm{d}B_t.$$

If the exponential is written as $e^{\sigma B_t - \sigma^2 t/2}$, the drift vanishes and the process is a martingale.

Theorem 7.34 (Multidimensional Itô formula; Book 7.6.7). For d-dimensional Brownian motion and $f \in C^{1,2}([0,\infty) \times \mathbb{R}^d)$,

$$\mathrm{d}f(t, B_t) = \left(\partial_t f + \frac{1}{2}\Delta f\right)(t, B_t) \, \mathrm{d}t + \sum_{j=1}^{d} \partial_j f(t, B_t) \, \mathrm{d}B_t^j.$$

Quadratic covariations vanish between independent coordinates, while $\mathrm{d}B_t^j \, \mathrm{d}B_t^j = \mathrm{d}t$.

Remark 7.35 (Generator). The generator of Brownian motion is $L = \frac{1}{2}\Delta$. For a smooth test function f, Itô's formula identifies $Lf(B_t)$ as the drift of $f(B_t)$. Consequently, harmonic functions and solutions of the heat equation give Brownian local martingales.

8 Applications to Random Walk

Functional central limit theorems describe the convergence of rescaled random paths. Embedding partial sums in Brownian motion reduces these limits to the convergence of the associated stopping times. The same construction relates the almost sure fluctuations of partial sums to the Brownian law of the iterated logarithm.

8.1 Donsker's Theorem

Definition 8.1 (Interpolated partial sums). Let $X_1, X_2, \dots$ be real random variables and $S_k = \sum_{j=1}^{k} X_j$, with $S_0 = 0$. Define $S(u)$ for $u \geq 0$ by $S(k) = S_k$ at integers and by linear interpolation on every interval $[k, k+1]$. For $n \geq 1$, define the rescaled path by

$$W_n(t) = \frac{S(nt)}{\sqrt{n}}, \qquad 0 \leq t \leq 1.$$

It is a random element of $C[0,1]$, equipped with the uniform norm $\|f\|_\infty = \sup_{0\leq t\leq 1} |f(t)|$.

Definition 8.2 (Weak convergence of paths). Random paths W_n converge weakly to W in $C[0,1]$, written $W_n \Rightarrow W$, if

$$\mathbb{E}\Phi(W_n) \longrightarrow \mathbb{E}\Phi(W)$$

for every bounded continuous $\Phi : C[0,1] \to \mathbb{R}$. This assertion includes convergence of all finite-dimensional distributions and the additional tightness needed to control oscillations between fixed times.

Theorem 8.3 (Skorokhod embedding; Book 8.1.1). If X has $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$, then one can construct a Brownian motion B and a Brownian stopping time T such that

$$B_T \overset{d}{=} X, \qquad \mathbb{E}T = \mathbb{E}X^2.$$

Proof sketch. If X takes two values $a < 0 < b$, the exit time from (a, b) has the required law: the exit probabilities follow from $\mathbb{E}X = 0$, and optional stopping for $B_t^2 - t$ gives $\mathbb{E}T = -ab = \mathbb{E}X^2$. A general centered law is a mixture of centered laws supported on two points, together with any mass at zero. Choosing the endpoints according to this mixture and then taking the corresponding exit time gives the stated terminal law and mean time. The filtration includes the auxiliary choice of endpoints, which is independent of the Brownian motion.

Remark 8.4 (Choice of embedding). A centered law can admit several embeddings with different distributions of T. The finite second moment determines the mean stopping time in this construction; T and B_T are generally dependent.

Theorem 8.5 (Embedding an i.i.d. random walk; Book 8.1.2). Let $X_1, X_2, \dots$ be i.i.d. with mean 0 and variance 1. There are Brownian stopping times

$$0 = T_0 \le T_1 \le T_2 \le \cdots$$

such that

$$(S_0, S_1, \dots, S_k) \overset{d}{=} (B_{T_0}, B_{T_1}, \dots, B_{T_k})$$

for every k, and the increments $T_k - T_{k-1}$ are i.i.d. with mean 1. Consequently $T_n/n \to 1$ a.s.

Proof sketch. By the strong Markov property, the process after T_{k-1}, translated to start at zero, is Brownian motion independent of the past. Embedding X_k in this process defines $T_k - T_{k-1}$. Induction gives the joint distribution of all partial sums, and the strong law for the stopping time increments yields $T_n/n \to 1$.

Corollary 8.6 (Scalar central limit theorem; Book 8.1.3). Under the preceding assumptions,

$$\frac{S_n}{\sqrt{n}} \Rightarrow N(0,1).$$

Indeed, $\widetilde{B}_t^{(n)} = B_{nt}/\sqrt{n}$ is Brownian motion and $B_{T_n}/\sqrt{n} = \widetilde{B}^{(n)}_{T_n/n}$. Since $T_n/n \to 1$, Brownian continuity implies that this differs from $\widetilde{B}_1^{(n)}$ by a quantity tending to zero in probability.

Lemma 8.7 (Convergence of embedded paths; Book 8.1.9). Suppose a triangular array of partial sums is embedded as $S_{n,m} = B_{\tau_{n,m}}$, $0 \le m \le n$, with $m \mapsto \tau_{n,m}$ increasing. If, for every $s \in [0,1]$,

$$\tau_{n,\lfloor ns \rfloor} \longrightarrow s \quad \text{in probability,}$$

then the linear interpolation satisfies

$$\sup_{0 \le t \le 1} |S_n(nt) - B_t| \longrightarrow 0 \quad \text{in probability.}$$

Proof sketch. A finite deterministic partition reduces the estimate to the stopping times at its endpoints. Their monotonicity bounds all intermediate values between the values at adjacent endpoints. Uniform continuity of Brownian paths then bounds the difference between the embedded path and B, including the linear interpolation. Refining the partition makes the supremum of this difference tend to zero in probability.

Lemma 8.8 (Bounded continuous functionals; Book 8.1.10). If $\|W_n - B\|_\infty \to 0$ in probability under a coupling, then every bounded continuous $\Phi : C[0,1] \to \mathbb{R}$ satisfies $\mathbb{E}\Phi(W_n) \to \mathbb{E}\Phi(B)$.

Proof sketch. Every subsequence has a further subsequence along which $\|W_n - B\|_\infty \to 0$ almost surely. Continuity of Φ and bounded convergence give convergence of expectations along that further subsequence. The subsequence criterion then proves convergence along the full sequence.

Theorem 8.9 (Donsker invariance principle; Book 8.1.4). If $X_1, X_2, \ldots$ are i.i.d. with mean 0 and variance 1, then

$$\left\{ \frac{S(nt)}{\sqrt{n}} : 0 \le t \le 1 \right\} \Rightarrow \{B_t : 0 \le t \le 1\}$$

in $C[0,1]$. More generally, if $\mathbb{E}X_1 = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 \in (0,\infty)$, then $\{(S(nt) - nt\mu)/\sqrt{n} : 0 \le t \le 1\} \Rightarrow \sigma B$, with the same interpolation convention.

Proof sketch. The embedding of the scaled increments $X_m/\sqrt{n}$ has stopping times equal in law to T_m/n by Brownian scaling. The strong law gives $T_{\lfloor nt \rfloor}/n \to t$. The lemma on embedded paths therefore gives a coupling in which the interpolated partial sums converge uniformly to Brownian motion in probability. Convergence of expectations of bounded continuous functionals follows.

Theorem 8.10 (Continuous mapping theorem; Book 8.1.5). Let $\Psi : C[0,1] \to \mathbb{R}$ be measurable and continuous at Brownian paths with probability one. Then

$$\Psi(W_n) \Rightarrow \Psi(B).$$

Example 8.11 (Running maximum). The map $f \mapsto \max_{0\le t\le 1} f(t)$ is Lipschitz in the uniform norm. Hence

$$\frac{1}{\sqrt{n}} \max_{0\le k\le n} S_k \Rightarrow M_1 := \max_{0\le t\le 1} B_t.$$

The reflection principle identifies the limit:

$$\mathbb{P}(M_1 \ge a) = 2\mathbb{P}(B_1 \ge a), \qquad a > 0.$$

Thus $M_1 \overset{d}{=} |N(0,1)|$.

Example 8.12 (Last zero and the arcsine law). The functional giving the last zero is continuous at a path that is nonzero at time 1 and takes both positive and negative values arbitrarily close to its last zero from the left. Brownian motion has these properties almost surely. Therefore

$$\frac{1}{n}\sup\{k \le n : S_{k-1}S_k \le 0\} \Rightarrow \sup\{t \le 1 : B_t = 0\}.$$

The sign crossings record the zeros of the interpolated path, including crossings between integer times. The limiting law is the arcsine distribution.

Example 8.13 (Occupation times). For fixed $a \in \mathbb{R}$, the functional

$$\Psi_a(f) = \int_0^1 \mathbf{1}_{\{f(t)>a\}}\, \mathrm{d}t$$

is continuous at every path whose level set $\{t : f(t) = a\}$ has Lebesgue measure zero. Fubini's theorem shows that Brownian level sets have zero Lebesgue measure a.s. Consequently,

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{S_k>a\sqrt{n}\}} \Rightarrow \int_0^1 \mathbf{1}_{\{B_t>a\}}\, \mathrm{d}t.$$

For $a = 0$, the limit again has the arcsine law.

Remark 8.14 (Discrete occupation counts). Finite second moment gives

$$\max_{k\le n} |X_k| / \sqrt{n} \to 0 \quad \text{in probability.}$$

The interpolated path and the process observed at integer times are therefore uniformly close after rescaling. Together with the zero Lebesgue measure of Brownian level sets, this proves the same limit for the discrete occupation count.

Theorem 8.15 (Donsker on the half-line; Book 8.1.11). The same rescaled process converges to Brownian motion in $C[0,\infty)$ with the topology of uniform convergence on compact time intervals. Equivalently, its restriction to $[0, M]$ converges in $C[0, M]$ for every finite M.

Example 8.16 (First passage times). Let

$$N_n = \inf\{m : S_m \ge \sqrt{n}\}, \qquad T_1 = \inf\{t : B_t \ge 1\}.$$

The truncated hitting time is a continuous functional at almost every Brownian path, and the law of T_1 has no atoms. Applying Donsker first on $[0, M]$ and then letting $M \to \infty$ gives

$$\mathbb{P}(N_n \le nt) \longrightarrow \mathbb{P}(T_1 \le t), \qquad t > 0.$$

8.2 Central Limit Theorems for Martingales

Theorem 8.17 (Embedding a square-integrable martingale; Book 8.2.1). Let $(S_n, \mathcal{F}_n)$ be a square-integrable martingale with $S_0 = 0$. There are increasing Brownian stopping times T_n such that, for every k,

$$(S_0, S_1, \dots, S_k) \overset{d}{=} (B_{T_0}, B_{T_1}, \dots, B_{T_k}).$$

If $\Delta S_k = S_k - S_{k-1}$ and $\tau_k = T_k - T_{k-1}$, then the construction can be chosen so that

$$\mathbb{E}(\tau_k \mid \mathcal{F}_{k-1}) = \mathbb{E}((\Delta S_k)^2 \mid \mathcal{F}_{k-1})$$

under the embedded identification.

Proof sketch. Given $\mathcal{F}_{k-1}$, the regular conditional law of ΔS_k has mean zero. The strong Markov property permits this law to be embedded after T_{k-1}. Induction preserves the joint distribution of the martingale values, and optional stopping for $B_t^2 - t$ identifies the conditional mean of $T_k - T_{k-1}$ with the conditional second moment of ΔS_k.

Theorem 8.18 (Finite predictable quadratic variation; Book 8.2.2). Let

$$A_\infty = \sum_{k=1}^{\infty} \mathbb{E}((S_k - S_{k-1})^2 \mid \mathcal{F}_{k-1}).$$

Then S_n converges to a finite limit on $\{A_\infty < \infty\}$.

Proof sketch. Stopping before the predictable quadratic variation exceeds a bounds the expected total embedded time by a. The stopped sequence of embedding times consequently has a finite limit almost surely, so Brownian continuity gives convergence of the stopped martingale. Letting $a \uparrow \infty$ proves convergence on $\{A_\infty < \infty\}$.

Definition 8.19 (Martingale difference array). For each n, let

$$X_{n,m} \in \mathcal{F}_{n,m}, \qquad \mathbb{E}(X_{n,m} \mid \mathcal{F}_{n,m-1}) = 0, \qquad 1 \le m \le n.$$

This is a martingale difference array. Write

$$S_{n,k} = \sum_{m=1}^{k} X_{n,m}, \qquad V_{n,k} = \sum_{m=1}^{k} \mathbb{E}(X_{n,m}^2 \mid \mathcal{F}_{n,m-1}).$$

$V_{n,k}$ is the predictable quadratic variation of the nth row.

Theorem 8.20 (Functional CLT with uniformly small jumps; Book 8.2.3). Suppose a martingale difference array satisfies, for every $t \in [0,1]$,

$$V_{n,\lfloor nt \rfloor} \longrightarrow t \quad \text{in probability,}$$

and there are deterministic $\varepsilon_n \downarrow 0$ such that

$$|X_{n,m}| \le \varepsilon_n \quad \text{for every } m \le n.$$

Then the linearly interpolated process $S_n(n\,\cdot)$ converges weakly to standard Brownian motion in $C[0,1]$.

Proof sketch. For each embedded row, the conditional mean of an increment of the stopping times is the corresponding increment of $V_{n,k}$. The bound on $X_{n,m}$ controls the second moments of the embedded durations. After stopping where $V_{n,k}$ is bounded, an L^2 estimate gives $T_{n,\lfloor nt \rfloor} - V_{n,\lfloor nt \rfloor} \to 0$ in probability. The lemma on embedded paths then gives the functional central limit theorem.

Definition 8.21 (Conditional Lindeberg quantity). For $\varepsilon > 0$, define

$$L_n(\varepsilon) = \sum_{m=1}^{n} \mathbb{E}\left(X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| > \varepsilon\}} \mid \mathcal{F}_{n,m-1} \right).$$

The conditional Lindeberg condition is $L_n(\varepsilon) \to 0$ in probability for every $\varepsilon > 0$. It makes the contribution of increments larger than any fixed ε to the conditional variance tend to zero.

Theorem 8.22 (Martingale Lindeberg–Feller theorem; Book 8.2.4). If

$$V_{n,\lfloor nt \rfloor} \to t \quad \text{in probability for every } t \in [0,1],$$

and the conditional Lindeberg condition holds, then

$$S_n(n\,\cdot) \Rightarrow B \quad \text{in } C[0,1].$$

Lemma 8.23 (Diagonal truncation level; Book 8.2.5). If $L_n(\varepsilon) \to 0$ in probability for every fixed $\varepsilon > 0$, then one can choose $\varepsilon_n \downarrow 0$ slowly enough that

$$\varepsilon_n^{-2} L_n(\varepsilon_n) \longrightarrow 0 \quad \text{in probability.}$$

Since $\sum_m \mathbb{P}(|X_{n,m}| > \varepsilon_n \mid \mathcal{F}_{n,m-1}) \le \varepsilon_n^{-2} L_n(\varepsilon_n)$, the sum of the conditional probabilities of removed increments also tends to zero.

Lemma 8.24 (Centered truncated array; Book 8.2.6). Set

$$\bar{X}_{n,m} = X_{n,m}\mathbf{1}_{\{|X_{n,m}|\le\varepsilon_n\}}, \qquad \widetilde{X}_{n,m} = \bar{X}_{n,m} - \mathbb{E}(\bar{X}_{n,m} \mid \mathcal{F}_{n,m-1}).$$

Then $\left|\widetilde{X}_{n,m}\right| \le 2\varepsilon_n$, its predictable quadratic variation has the same limit as $V_{n,k}$, and the theorem for uniformly small jumps applies to its partial sums.

Proof sketch. The preceding diagonal choice of ε_n makes the conditional variance removed by truncation tend to zero. Conditional centering preserves the martingale difference property and changes the predictable quadratic variation by a negligible amount. The theorem for uniformly small jumps applies to $\widetilde{X}_{n,m}$. The probability of a removed increment and the sum of the centering corrections both tend to zero, so the original partial sums have the same limit.

Lemma 8.25 (Controlling an adapted union; Book 8.2.7). If $A_m \in \mathcal{G}_m$ and $\delta \ge 0$ is $\mathcal{G}_0$-measurable, then

$$\mathbb{P}\left(\bigcup_{m=1}^{n} A_m \mid \mathcal{G}_0\right) \le \delta + \mathbb{P}\left(\sum_{m=1}^{n} \mathbb{P}(A_m \mid \mathcal{G}_{m-1}) > \delta \;\middle|\; \mathcal{G}_0\right).$$

This converts control of predictable conditional probabilities into control of the probability that any exceptional event occurs.

Remark 8.26 (Application to truncation). With $A_m = \{|X_{n,m}| > \varepsilon_n\}$, the preceding lemma bounds the probability of any removed increment using the sum of its conditional probabilities. This sum converges in probability by the diagonal truncation estimate, which is sufficient for the bound to tend to zero.

Theorem 8.27 (Functional CLT for a martingale difference sequence; Book 8.2.8). Let $(X_k, \mathcal{F}_k)$ be a martingale difference sequence, $S_n = \sum_{k=1}^{n} X_k$, and

$$V_n = \sum_{k=1}^{n} \mathbb{E}(X_k^2 \mid \mathcal{F}_{k-1}).$$

Suppose

$$\frac{V_n}{n} \longrightarrow \sigma^2 > 0 \quad \text{in probability}$$

and, for every $\varepsilon > 0$,

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \mathbf{1}_{\{|X_k| > \varepsilon\sqrt{n}\}}\right] \longrightarrow 0.$$

Then

$$\frac{S(n\,\cdot)}{\sqrt{n}} \Rightarrow \sigma B \quad \text{in } C[0,1].$$

8.3 Central Limit Theorems for Stationary Sequences

Theorem 8.28 (Stationary martingale differences; Book 8.3.1). Let $(X_n)_{n\in\mathbb{Z}}$ be stationary and ergodic, with $\mathbb{E}X_0^2 = \sigma^2 < \infty$. Set $\mathcal{F}_n = \sigma(X_m : m \le n)$ and assume

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = 0.$$

Then, for $S_n = \sum_{k=1}^n X_k$,

$$\frac{S(n\,\cdot)}{\sqrt{n}} \Rightarrow \sigma B.$$

Proof sketch. The conditional variances $U_n = \mathbb{E}(X_n^2 \mid \mathcal{F}_{n-1})$ form a stationary ergodic sequence. Birkhoff's theorem gives $n^{-1}\sum_{k=1}^n U_k \to \mathbb{E}U_0 = \sigma^2$ a.s. Stationarity reduces the Lindeberg average to $\mathbb{E}[X_0^2 \mathbf{1}_{\{|X_0|>\varepsilon\sqrt{n}\}}]$, which vanishes by dominated convergence. The functional central limit theorem for martingale differences now applies.

Definition 8.29 (Filtration compatible with the shift). Represent a two-sided stationary sequence as $X_n = X_0 \circ \theta^n$ under a measure-preserving shift θ. An increasing filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}}$ is shift compatible when $X_n \in \mathcal{F}_n$ and

$$\theta^{-1}\mathcal{F}_n = \mathcal{F}_{n+1}.$$

The filtration may contain more information than the natural past of X.

Theorem 8.30 (Projective criterion for a stationary sequence; Book 8.3.2). Let $(X_n)_{n\in\mathbb{Z}}$ be centered, square integrable, stationary, and ergodic, and let $(\mathcal{F}_n)$ be shift compatible. If

$$\sum_{n=1}^{\infty} \|\mathbb{E}(X_0 \mid \mathcal{F}_{-n})\|_2 < \infty,$$

then

$$\frac{S(n\,\cdot)}{\sqrt{n}} \Rightarrow \sigma B,$$

where

$$\sigma^2 = \mathbb{E}X_0^2 + 2\sum_{n=1}^{\infty} \mathbb{E}(X_0 X_n).$$

The covariance series converges absolutely, and σ^2 is allowed to be zero.

Martingale–coboundary decomposition. The projective condition gives $Y_0, Z_0 \in L^2$ satisfying

$$X_0 = Y_0 + Z_0 - Z_0 \circ \theta, \qquad \mathbb{E}(Y_0 \mid \mathcal{F}_{-1}) = 0.$$

They can be defined by

$$Z_0 = \sum_{j \geq 0} \mathbb{E}(X_j \mid \mathcal{F}_{-1}), \qquad Y_0 = \sum_{j \geq 0} \{\mathbb{E}(X_j \mid \mathcal{F}_0) - \mathbb{E}(X_j \mid \mathcal{F}_{-1})\}.$$

The projective summability assumption makes both series converge in L^2. Summing the decomposition gives

$$S_n = \sum_{k=1}^{n} Y_0 \circ \theta^k + Z_0 \circ \theta - Z_0 \circ \theta^{n+1}.$$

The first term is a sum of stationary ergodic martingale differences. The coboundary telescopes, and the maximum of its partial sums divided by $\sqrt{n}$ tends to zero in probability by the L^2 tail estimate. The stationary martingale theorem therefore gives the functional limit, and $n^{-1}\mathbb{E}S_n^2$ identifies its variance.

Example 8.31 (Zero limiting variance). If $X_k = Z \circ \theta^k - Z \circ \theta^{k+1}$ with $Z \in L^2$, then $S_n = Z \circ \theta - Z \circ \theta^{n+1}$ is bounded in L^2. Consequently $n^{-1}\mathbb{E}S_n^2 \to 0$, even when $\mathrm{Var}(X_0) > 0$.

Example 8.32 (M-dependent sequences). Suppose the stationary sequence is centered and square integrable and $\sigma(X_j : j \leq 0)$ is independent of $\sigma(X_j : j > M)$. Then $\mathbb{E}(X_0 \mid \mathcal{F}_{-n}) = 0$ for $n > M$, so the projective condition is automatic. The limiting variance is

$$\sigma^2 = \mathbb{E}X_0^2 + 2\sum_{k=1}^{M} \mathbb{E}(X_0 X_k).$$

The dependence contributes the finite sum of covariances to the limiting variance.

Example 8.33 (Additive functionals of Markov chains). Let $(\zeta_n)_{n \in \mathbb{Z}}$ be an irreducible stationary Markov chain with law π and let $X_n = f(\zeta_n)$, where $\pi f = 0$ and $f \in L^2(\pi)$. Use the chain filtration $\mathcal{F}_m = \sigma(\zeta_j : j \leq m)$. Then

$$\mathbb{E}(X_0 \mid \mathcal{F}_{-n}) = (P^n f)(\zeta_{-n}), \qquad \|\mathbb{E}(X_0 \mid \mathcal{F}_{-n})\|_2 = \|P^n f\|_{L^2(\pi)}.$$

Hence the functional CLT follows whenever

$$\sum_{n=1}^{\infty} \|P^n f\|_{L^2(\pi)} < \infty.$$

For a finite irreducible aperiodic chain, exponential convergence to equilibrium verifies this condition for every centered f.

Example 8.34 (Moving averages). Let

$$X_m = \sum_{k\geq 0} c_k \xi_{m-k}, \qquad \sum_{k\geq 0} c_k^2 < \infty,$$

where the ξ_j are two-sided i.i.d., centered, and have variance 1. With the innovation filtration,

$$\|\mathbb{E}(X_0 \mid \mathcal{F}_{-n})\|_2 = \left(\sum_{k\geq n} c_k^2\right)^{1/2}.$$

Thus the projective theorem applies if the sum of these tail norms is finite. For $c_k = (1+k)^{-p}$, this direct criterion requires $p > 3/2$. Sharper martingale approximations reduce the sufficient condition to absolute summability of (c_k).

Example 8.35 (Long memory). If $c_k \asymp k^{-p}$ with $1/2 < p < 1$, correlations decay too slowly and $\mathrm{Var}(S_n)$ has order n^{3-2p} rather than n. Thus the natural standard deviation is of order $n^{3/2-p}$. The coefficients are square summable, so each X_m remains in L^2.

8.3.1 Mixing Properties

Definition 8.36 (Mixing coefficient between sigma fields). For two σ-fields $\mathcal{G}$ and $\mathcal{H}$, define

$$\alpha(\mathcal{G},\mathcal{H}) = \sup_{A\in\mathcal{G},\, B\in\mathcal{H}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

It is zero exactly when the two fields are independent. For a stationary sequence, these coefficients measure dependence between sigma fields separated by a given number of time steps.

Lemma 8.37 (Davydov covariance inequality; Book 8.3.6). Let $p, q, r \in (1,\infty]$ satisfy

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1.$$

If X is $\mathcal{G}$-measurable, Y is $\mathcal{H}$-measurable, and $X \in L^p$, $Y \in L^q$, then

$$|\mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y| \leq 8\,\|X\|_p\,\|Y\|_q\,\alpha(\mathcal{G},\mathcal{H})^{1/r}.$$

Proof sketch. When $r = \infty$, Hölder's inequality and the triangle inequality give the bound. For bounded variables, integral representations in terms of indicators reduce the covariance estimate to the definition of α. Truncating general X and Y gives a bound for their bounded parts and tail estimates in terms of the L^p and L^q norms. Choosing the truncation levels as powers of α^{-1} gives the exponent $1/r$.

Theorem 8.38 (One-sided mixing FCLT; Book 8.3.7). Let $(X_n)_{n\in\mathbb{Z}}$ be centered, stationary, and ergodic, and suppose

$$\mathbb{E}\,|X_0|^{2+\delta} < \infty$$

for some $\delta > 0$. Set

$$\alpha(n) = \alpha\big(\sigma(X_m : m \le -n), \sigma(X_0)\big).$$

If

$$\sum_{n=1}^{\infty} \alpha(n)^{\delta/[2(2+\delta)]} < \infty,$$

then

$$\frac{S(n\,\cdot)}{\sqrt{n}} \Rightarrow \sigma B, \qquad \sigma^2 = \mathbb{E}X_0^2 + 2\sum_{n=1}^{\infty} \mathbb{E}(X_0 X_n).$$

Proof sketch. The L^2 duality identity gives

$$\|\mathbb{E}(X_0 \mid \mathcal{F}_{-n})\|_2 = \sup\{\mathbb{E}(X_0 Y) : Y \in L^2(\mathcal{F}_{-n}),\ \|Y\|_2 = 1\}.$$

Applying the covariance inequality with $p = 2+\delta$ and $q = 2$ yields

$$\|\mathbb{E}(X_0 \mid \mathcal{F}_{-n})\|_2 \le 8\,\|X_0\|_{2+\delta}\,\alpha(n)^{\delta/[2(2+\delta)]}.$$

The assumed summability therefore implies the projective criterion.

Definition 8.39 (Strong mixing of a sequence). Let

$$\bar{\alpha}(n) = \alpha\big(\sigma(X_m : m \le -n), \sigma(X_m : m \ge 0)\big).$$

The sequence is strongly mixing if $\bar{\alpha}(n) \to 0$. A classical scalar CLT follows under the stronger summability condition

$$\sum_{n=1}^{\infty} \bar{\alpha}(n)^{\delta/(2+\delta)} < \infty$$

together with a $(2+\delta)$ moment.

Example 8.40 (Checking mixing through a Markov kernel). For a stationary Markov chain, total variation convergence of $P^n(x,\cdot)$ to π bounds the mixing coefficient. An estimate of the form

$$\alpha(n) \le C \sum_x \pi(x)\,\|P^n(x,\cdot) - \pi\|_{d_{\mathrm{TV}}}$$

gives a sufficient mixing rate for the central limit theorem when the right side is summable to the required power.

Remark 8.41 (Comparison of mixing coefficients). Since $\sigma(X_0) \subseteq \sigma(X_m : m \geq 0)$, one has $\alpha(n) \leq \bar{\alpha}(n)$. The two central limit criteria use different powers because the projective criterion estimates conditional expectations in L^2, whereas the scalar mixing theorem controls sums through covariances.

8.4 Empirical Distributions and Brownian Bridge

Definition 8.42 (Empirical distribution function). For i.i.d. observations $X_1, \ldots, X_n$ with distribution function F, define

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}}.$$

It is a random distribution function placing mass $1/n$ at every observation.

Theorem 8.43 (Glivenko–Cantelli; Book 8.4.1). For every distribution function F,

$$\sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right| \longrightarrow 0 \quad \text{a.s.}$$

Thus the empirical distribution converges uniformly to F almost surely.

Definition 8.44 (Kolmogorov–Smirnov statistic). For a fully specified continuous distribution function F, set

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right|.$$

The probability integral transform reduces its distribution to the case of i.i.d. uniform observations $U_1, \ldots, U_n$ on $(0, 1)$.

Remark 8.45 (Reduction to uniform observations). When F is continuous, $F(X_k)$ is uniform and intervals on which F is constant carry no observations. Thus

$$\sup_{x} \left| \widehat{F}_n(x) - F(x) \right| = \sup_{0<u<1} \left| \widehat{G}_n(u) - u \right|.$$

Hence the distribution of the statistic is the same for every continuous F.

Proposition 8.46 (Extrema occur at order statistics). Let $U_{1:n} < \cdots < U_{n:n}$ be the order statistics of a continuous uniform sample. Then

$$\sup_{0<u<1} \{\widehat{G}_n(u) - u\} = \max_{1\le k\le n} \left(\frac{k}{n} - U_{k:n} \right),$$

while

$$\inf_{0<u<1} \{\widehat{G}_n(u) - u\} = \min_{1\le k\le n} \left(\frac{k-1}{n} - U_{k:n} \right).$$

The difference between k/n and $(k-1)/n$ is negligible after multiplication by $\sqrt{n}$.

Lemma 8.47 (Exponential representation of order statistics; Book 8.4.2). Let $W_1, W_2, \ldots$ be i.i.d. exponential random variables of mean 1 and $Z_k = W_1 + \cdots + W_k$. Then

$$(U_{1:n}, \ldots, U_{n:n}) \overset{d}{=} \left(\frac{Z_1}{Z_{n+1}}, \ldots, \frac{Z_n}{Z_{n+1}} \right).$$

Proof sketch. The joint density of uniform order statistics is $n!$ on $0 < u_1 < \cdots < u_n < 1$. In the joint density of the exponential partial sums, the change of variables $u_k = Z_k/Z_{n+1}$, with Z_{n+1} as the final coordinate, has Jacobian Z_{n+1}^n. Integrating over Z_{n+1} gives $n!$ on the ordered simplex.

Definition 8.48 (Brownian bridge). For a Brownian motion B, define

$$B_t^0 = B_t - tB_1, \qquad 0 \le t \le 1.$$

This centered Gaussian process has covariance

$$\operatorname{Cov}(B_s^0, B_t^0) = \min(s,t) - st.$$

Its finite-dimensional distributions are the regular conditional Gaussian distributions of Brownian motion given $B_1 = 0$.

Theorem 8.49 (Kolmogorov–Smirnov limit; Book 8.4.3). For i.i.d. observations from a continuous F,

$$D_n \Rightarrow \sup_{0\le t\le 1} \left| B_t^0 \right| = \sup_{0\le t\le 1} |B_t - tB_1|.$$

The limit does not depend on F.

Proof sketch. Substituting the exponential representation into the formulas for the order statistics and multiplying by $\sqrt{n}$ gives the centered exponential walk $Z_{\lfloor nt\rfloor} - \lfloor nt\rfloor$ and a correction proportional to $t(Z_{n+1}-n)$. The strong law gives $n/Z_{n+1} \to 1$, and the last exponential increment divided by $\sqrt{n}$ tends to zero in probability. Donsker's theorem gives convergence of the centered walk to B. Applying the continuous map $f \mapsto \sup_t |f(t) - tf(1)|$ yields the supremum of the Brownian bridge.

Example 8.50 (Bridge transition law). For $0 \le s < t < 1$, conditional on $B^0_s = x$,

$$B^0_t \sim N\left(x\frac{1-t}{1-s}, \frac{(t-s)(1-t)}{1-s}\right).$$

Thus the bridge is a Markov process with transitions that depend on both s and t. For fixed s and x, the conditional mean and variance tend to zero as $t \uparrow 1$. Another useful representation is

$$\{B^0_t : 0 \le t < 1\} \overset{d}{=} \left\{(1-t)B_{t/(1-t)} : 0 \le t < 1\right\}.$$

Theorem 8.51 (Brownian motion killed between two barriers; Book 8.4.4). Let $a < x, y < b$, let $T_c = \inf\{t : B_t = c\}$, and write

$$p_t(x,z) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(z-x)^2}{2t}\right).$$

The density of B_t at y on survival inside (a,b) is

$$p^{(a,b)}_t(x,y) = \sum_{k\in\mathbb{Z}} \left\{p_t(x, y+2k(b-a)) - p_t(x, 2b-y+2k(b-a))\right\}.$$

Equivalently, $\mathbb{P}_x(T_a \wedge T_b > t, B_t \in \mathrm{d}y) = p^{(a,b)}_t(x,y)\,\mathrm{d}y$.

Proof sketch. Reflection after the first hit of a or b expresses the density of paths that exit through that boundary as a Gaussian density at a reflected point. Repeated reflections account for paths crossing the opposite boundary and give the alternating sum of image points spaced by $2(b-a)$. Gaussian tail estimates make the remainder tend to zero and ensure convergence of the series.

Theorem 8.52 (Kolmogorov distribution). For $b > 0$,

$$\mathbb{P}\left(\sup_{0\le t\le 1}\left|B_t^0\right| < b\right) = \sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2b^2}.$$

Equivalently,

$$\mathbb{P}\left(\sup_{0\le t\le 1}\left|B_t^0\right| \ge b\right) = 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2b^2}.$$

For the one-sided maximum,

$$\mathbb{P}\left(\sup_{0\le t\le 1} B_t^0 > b\right) = e^{-2b^2}.$$

Proof sketch. The transition density for Brownian motion killed on leaving $(-b, b)$, evaluated at $x = y = 0$ and $t = 1$, divided by the unrestricted density at zero gives the conditional probability for the bridge. The two families of image points combine into the displayed alternating series.

8.5 Laws of the Iterated Logarithm

Definition 8.53 (Normalization for the law of the iterated logarithm). For $t > e$, define

$$a(t) = \sqrt{2t\log\log t}.$$

This grows faster than $\sqrt{t}$ but slower than $t^{1/2+\varepsilon}$ for every $\varepsilon > 0$. It describes the almost sure upper and lower limits of Brownian motion and centered random walks with finite variance.

Theorem 8.54 (Brownian law of the iterated logarithm; Book 8.5.1). For standard Brownian motion,

$$\limsup_{t\to\infty}\frac{B_t}{\sqrt{2t\log\log t}} = 1 \quad \text{a.s.}$$

By symmetry,

$$\liminf_{t\to\infty}\frac{B_t}{\sqrt{2t\log\log t}} = -1 \quad \text{a.s.}$$

Upper bound. At geometric times $t_n = \alpha^n$, one has $\log\log t_n \sim \log n$. Gaussian tail estimates and the reflection principle show that the probabilities of exceeding a suitable multiple of $\sqrt{2t\log\log t}$ on successive intervals $[t_n, t_{n+1}]$ are summable. Borel–Cantelli gives an eventual upper bound on every such interval. Letting the multiplicative constant decrease and then $\alpha \downarrow 1$ gives the upper limit 1.

Lower bound. For geometric times with α large, the increments $B_{t_{n+1}} - B_{t_n}$ are independent. Gaussian lower bounds show that the sum of the probabilities of a sufficiently large positive increment diverges. The second Borel–Cantelli lemma therefore gives infinitely many such increments. Subtracting the upper bound for $|B_{t_n}|$ and letting $\alpha \to \infty$ gives the lower limit 1.

Theorem 8.55 (Hartman–Wintner LIL; Book 8.5.2). Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_1 = 0$ and $\mathrm{Var}(X_1) = 1$. Then

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \quad \text{a.s.},$$

and the corresponding liminf is -1. For variance $\sigma^2 \in (0, \infty)$, divide by $\sigma\sqrt{2n \log\log n}$.

Proof sketch. The Skorokhod embedding gives $S_n = B_{T_n}$ with $T_n/n \to 1$ almost surely. On geometric time intervals, the reflection principle and Gaussian tail estimates bound the Brownian oscillation between n and T_n. Borel–Cantelli then gives

$$B_{T_n} - B_n = o(\sqrt{n \log\log n}) \quad \text{a.s.}$$

The Brownian law of the iterated logarithm consequently gives the same upper and lower limits for S_n.

Example 8.56 (Eventual envelope and repeated crossings). For every $\varepsilon > 0$, a.s. there is a random N such that for all $n \geq N$,

$$|S_n| \leq (1+\varepsilon)\sqrt{2\sigma^2 n \log\log n}.$$

At the same time,

$$S_n \geq (1-\varepsilon)\sqrt{2\sigma^2 n \log\log n}$$

and its negative counterpart each occur infinitely often. Thus the upper bound holds eventually, while levels arbitrarily close to both limiting constants are reached infinitely often.

Theorem 8.57 (Strassen's functional law of the iterated logarithm; Book 8.5.3). Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_1 = 0$ and $\mathrm{Var}(X_1) = 1$, and let $S(u)$ be the usual linear interpolation of their partial sums. For $n \geq 3$, define

$$Z_n(t) = \frac{S(nt)}{\sqrt{2n \log\log n}}, \qquad 0 \leq t \leq 1.$$

Almost surely, (Z_n) is relatively compact in $C[0,1]$ with the uniform norm, and its set of subsequential limits is exactly

$$K = \left\{ f : f(t) = \int_0^t g(s)\, ds, \quad g \in L^2[0,1], \quad \int_0^1 g(s)^2\, ds \leq 1 \right\}.$$

For $f \in K$, Cauchy–Schwarz gives $|f(1)| \le 1$, and every value in $[-1,1]$ is attained by $f(t) = ct$ with $|c| \le 1$. Thus the scalar normalized sums have limit set $[-1,1]$, recovering the upper and lower limits in the Hartman–Wintner law. Equality at either endpoint forces $f(t) = t$ or $f(t) = -t$, respectively.

9 Multidimensional Brownian Motion

The generator $\frac{1}{2}\Delta$ relates Brownian martingales to harmonic functions and solutions of the heat equation. Optional stopping then gives representations in terms of exit distributions and occupation times. Throughout, B_t is standard Brownian motion in $\mathbb{R}^d$, and $\Delta f = \sum_{i=1}^{d} \partial_{ii} f$ is the Laplacian.

9.1 Martingales

Theorem 9.1 (Martingale associated with the Laplacian). Book Theorem 9.1.1. If $v \in C^2$ and $\mathbb{E}\int_0^t \|\nabla v(B_s)\|^2 \, \mathrm{d}s < \infty$ for every t, then

$$v(B_t) - \frac{1}{2}\int_0^t \Delta v(B_s) \, \mathrm{d}s$$

is a continuous martingale. In particular, $v(B_t)$ is a local martingale whenever v is harmonic.

Proof sketch. Itô's formula expresses $v(B_t) - v(B_0)$ as $\frac{1}{2}\int_0^t \Delta v(B_s)\,\mathrm{d}s$ plus the stochastic integral of $\nabla v(B_s)$. The integrability assumption makes this stochastic integral a martingale.

Let $S_r = \inf\{t : |B_t| = r\}$. When $|x| < R$, $S_R < \infty$ almost surely.

Theorem 9.2 (Mean exit time from a ball, Theorem 9.1.2). For $|x| < R$,

$$\mathbb{E}_x S_R = \frac{R^2 - |x|^2}{d}.$$

Optional stopping for $|B_t|^2 - dt$ at $S_R \wedge t$ gives the identity as $t \to \infty$ by bounded and monotone convergence.

Definition 9.3 (Radial fundamental harmonic function). Away from the origin set

$$\phi(x) = \begin{cases} \log|x|, & d = 2, \\ |x|^{2-d}, & d \ge 3. \end{cases}$$

Then $\Delta\phi = 0$ on $\mathbb{R}^d \setminus \{0\}$.

Lemma 9.4 (Annulus stopping identity, Lemma 9.1.3). For $r < |x| < R$ and $\tau = S_r \wedge S_R$, $\phi(x) = \mathbb{E}_x \phi(B_\tau)$. Consequently,

$$\mathbb{P}_x(S_r < S_R) = \frac{\phi(R) - \phi(x)}{\phi(R) - \phi(r)}.$$

Proof sketch. A compactly supported C^2 function agreeing with ϕ on the annulus gives the same stopped martingale up to τ. Bounded convergence as $t \to \infty$ gives $\phi(x) = \mathbb{E}_x \phi(B_\tau)$. The exit point lies on one of the two spheres, so expanding this expectation and solving for $\mathbb{P}_x(S_r < S_R)$ gives the formula.

Corollary 9.5 (Recurrence and polarity). In $d = 2$, letting $R \to \infty$ gives $\mathbb{P}_x(S_r < \infty) = 1$: every nonempty open set is visited infinitely often. Letting $r \downarrow 0$ instead shows that a fixed point is never hit at positive time. Thus points are polar even though every neighborhood is recurrent. For $d \geq 3$,

$$\mathbb{P}_x(S_r < \infty) = \left(\frac{r}{|x|}\right)^{d-2}, \qquad |x| > r,$$

so Brownian motion is transient.

Theorem 9.6 (Escape to infinity, Theorem 9.1.4). For $d \geq 3$, $|B_t| \to \infty$ almost surely. After first hitting radius n, the strong Markov property and the probability of hitting a sphere show that the probability of ever returning to radius $n^{1-\varepsilon}$ is $n^{-\varepsilon(d-2)}$. The probability of permanent escape at these scales tends to one.

Theorem 9.7 (Dvoretsky–Erdős lower envelope, Theorem 9.1.5). Let $d \geq 3$ and let g be positive and decreasing. Then

$$\mathbb{P}_0\Big(|B_t| \leq g(t)\sqrt{t} \text{ infinitely often as } t \to \infty\Big)$$

is 1 or 0 according as $\int^\infty g(t)^{d-2}\, \mathrm{d}t/t$ diverges or converges. For example, the logarithmic threshold comes from the integral test for $t^{-1}(\log t)^{-\alpha}$.

9.2 Heat Equation

Consider

$$u_t = \frac{1}{2}\Delta u \quad \text{on } (0,\infty) \times \mathbb{R}^d, \qquad u(0,x) = f(x),$$

with bounded continuous f. Here $C^{1,2}$ means one continuous time derivative and two continuous space derivatives.

Theorem 9.8 (Martingale for the backward heat equation, Theorem 9.2.1). If $u \in C^{1,2}$ solves the heat equation, then for fixed t, $M_s = u(t - s, B_s)$ is a local martingale on $0 \leq s \leq t$.

Proof sketch. Itô's formula gives drift $(-u_t + \frac{1}{2}\Delta u)\,\mathrm{d}s = 0$. Stopping on compact subsets bounds the derivatives and identifies the remaining stochastic integral as a local martingale.

Theorem 9.9 (Uniqueness representation, Theorem 9.2.2). Every bounded classical solution equals

$$v(t,x) = \mathbb{E}_x f(B_t).$$

The local martingale is bounded, hence uniformly integrable, and its terminal value as $s \uparrow t$ is $f(B_t)$.

Theorem 9.10 (Heat semigroup and initial values, Theorems 9.2.3–9.2.4). If v is $C^{1,2}$, then the Markov identity $v(t+h,x) = \mathbb{E}_x v(t, B_h)$ and a second-order Taylor expansion imply $v_t = \frac{1}{2}\Delta v$. Brownian scaling gives $v(t_n, x_n) = \mathbb{E} f(x_n + \sqrt{t_n} Z) \to f(x)$ whenever $(t_n, x_n) \to (0, x)$, so v attains the initial value f continuously.

Theorem 9.11 (Regularity of the heat kernel, Theorem 9.2.5). The probabilistic formula can be written

$$v(t,x) = \int_{\mathbb{R}^d} (2\pi t)^{-d/2} \exp\left(-\frac{|x-y|^2}{2t}\right) f(y)\ \mathrm{d}y.$$

For bounded continuous f, differentiation of the Gaussian kernel shows $v \in C^{1,2}$ for $t > 0$. Hence v is the unique bounded classical solution.

9.3 Inhomogeneous Heat Equation

For bounded continuous g, consider

$$u_t = \frac{1}{2}\Delta u + g, \qquad u(0,x) = f(x).$$

By linearity, the solution is the sum of the homogeneous solution with initial value f and the solution with source g and zero initial value.

Theorem 9.12 (Compensated martingale, Theorem 9.3.1). If u solves the PDE, then

$$M_s = u(t-s, B_s) + \int_0^s g(B_r)\ \mathrm{d}r$$

is a local martingale on $[0,t]$. The additive functional cancels the drift $-g(B_s)$.

Theorem 9.13 (Duhamel representation, Theorem 9.3.2). For zero initial data, every solution bounded on each strip $[0,T]\times\mathbb{R}^d$ equals

$$v(t,x)=\mathbb{E}_x\int_0^t g(B_s)\,\mathrm{d}s.$$

Boundedness of u and g makes the local martingale uniformly integrable on $[0,t]$. For general initial data, add $\mathbb{E}_x f(B_t)$.

Theorem 9.14 (Inhomogeneous heat semigroup, Theorems 9.3.3–9.3.4). If $v\in C^{1,2}$, the Markov decomposition over the first h units of time gives

$$v(t+h,x)=\mathbb{E}_x\left[v(t,B_h)+\int_0^h g(B_s)\,\mathrm{d}s\right].$$

Taylor expansion and division by h yield $v_t=\frac{1}{2}\Delta v+g$. Also $|v(t,x)|\le t\,\|g\|_\infty$, so the zero initial condition holds uniformly.

Theorem 9.15 (Regularity with a source, Theorem 9.3.5). If g is Hölder continuous, then $v(t,x)=\mathbb{E}_x\int_0^t g(B_s)\,\mathrm{d}s$ belongs to $C^{1,2}$ for $t>0$. For second spatial derivatives, subtracting $g(x)$ from $g(y)$ in the kernel integral gives a Hölder bound that makes the singularity near $s=0$ integrable.

9.4 Feynman–Kac Formula

Let c be bounded and continuous and consider

$$u_t=\frac{1}{2}\Delta u+cu,\qquad u(0,x)=f(x).$$

Put $C_s=\int_0^s c(B_r)\,\mathrm{d}r$.

Theorem 9.16 (Multiplicative martingale, Theorem 9.4.1). If u solves the PDE, then $M_s=u(t-s,B_s)e^{C_s}$ is a local martingale on $[0,t]$. The product rule adds the drift cu, which cancels the PDE drift.

Theorem 9.17 (Feynman–Kac uniqueness, Theorem 9.4.2). Every solution bounded on finite time strips equals

$$v(t,x)=\mathbb{E}_x\left[f(B_t)\exp\left(\int_0^t c(B_s)\,\mathrm{d}s\right)\right].$$

Since c is bounded, the exponential weight is bounded on each fixed time interval, so the stopped local martingale is uniformly integrable.

Theorem 9.18 (Equation and initial values, Theorems 9.4.3–9.4.4). If $v \in C^{1,2}$, the Markov identity

$$v(t+h,x) = \mathbb{E}_x\left[v(t,B_h)\exp\left(\int_0^h c(B_s)\,\mathrm{d}s\right)\right]$$

and Taylor expansion yield $v_t = \frac{1}{2}\Delta v + cv$. Boundedness of c makes the exponential weight tend uniformly to 1 as $t \downarrow 0$. Continuity of f and Brownian scaling then give $v(t,x) \to f(x)$.

Theorem 9.19 (Feynman–Kac smoothness, Theorem 9.4.5). If f and c are Hölder continuous as well as bounded, then $v \in C^{1,2}$ and is the classical solution.

Proof sketch. Expanding the exponential weight gives the integral equation

$$v(t,x) = \mathbb{E}_x f(B_t) + \int_0^t \mathbb{E}_x[c(B_s)v(t-s,B_s)]\,\mathrm{d}s.$$

The first term is smooth for $t > 0$ by differentiation of the heat kernel. Local Hölder estimates for v and the Hölder continuity of c give the required regularity of the source $c(x)v(r,x)$. The theorem for the inhomogeneous heat equation then applies to the second term.

Remark 9.20 (Killing). If $c \le 0$, the factor $\exp(\int_0^t c(B_s)\,\mathrm{d}s)$ is the conditional survival probability for Brownian motion killed at rate $-c(B_s)$. For bounded c of either sign, its absolute value is at most $e^{t\|c\|_\infty}$ on a fixed time interval.

9.5 Dirichlet Problem

Let $G \subset \mathbb{R}^d$ be open, $\tau = \inf\{t > 0 : B_t \notin G\}$, and assume in this section that $\mathbb{P}_x(\tau < \infty) = 1$ for $x \in G$. The Dirichlet problem asks for

$$\Delta u = 0 \quad \text{in } G, \qquad u = f \quad \text{on } \partial G,$$

where f is bounded. The exit distribution gives the candidate solution $\mathbb{E}_x f(B_\tau)$.

Lemma 9.21 (Time change of a local martingale, Lemma 9.5.1). Suppose $u(B_t)$ is a local martingale on $[0,\tau)$. Inserting deterministic holding intervals at its localizing times gives a time change γ for which $u(B_{\gamma(t)})$ is a martingale and $\gamma(t) \uparrow \tau$ as $t \to \infty$. Each bounded interval in the new time variable involves only finitely many localized segments.

Theorem 9.22 (Dirichlet uniqueness, Theorem 9.5.2). Every bounded classical solution is

$$v(x) = \mathbb{E}_x f(B_\tau).$$

The time-changed martingale is bounded, so it is uniformly integrable; its terminal value is $f(B_\tau)$ by the boundary condition and path continuity.

Theorem 9.23 (Harmonicity from the mean value property, Theorem 9.5.3). If $v \in C^2$, then $\Delta v = 0$ in G. For a ball $D(x,r) \Subset G$ with exit time τ_r, strong Markov gives $v(x) = \mathbb{E}_x v(B_{\tau_r})$. Taylor expansion, stopped coordinate martingales, and $\mathbb{E}_x \tau_r \asymp r^2$ yield $0 = \frac{1}{2}\Delta v(x)\mathbb{E}_x\tau_r + o(r^2)$.

Theorem 9.24 (Interior smoothness, Theorem 9.5.4). The function $v(x) = \mathbb{E}_x f(B_\tau)$ is C^∞ inside G. It has the spherical mean value property on every ball compactly contained in G. Convolving locally with a smooth radial kernel preserves that mean and expresses a constant multiple of v as a smooth function.

Example 9.25 (Irregular boundaries, Examples 9.5.5–9.5.6). In a three-dimensional ball with a line segment removed, Brownian motion almost surely exits through the outer sphere because the segment is polar. Thus boundary data supported on the segment give zero in the exit representation. Lebesgue's thorn removes increasingly thin boxes accumulating at the origin. If their widths shrink sufficiently fast, Brownian motion started at the origin has positive probability to avoid all of them initially, so the origin is not regular.

Definition 9.26 (Regular boundary point). A point $y \in \partial G$ is regular if $\mathbb{P}_y(\tau = 0) = 1$. Regularity says Brownian motion started at y leaves G immediately. At such a point, the exit representation attains continuous boundary data.

Theorem 9.27 (Boundary convergence, Theorem 9.5.7). If f is bounded and continuous, y is regular, and $x_n \in G$ tends to y, then $v(x_n) \to f(y)$.

Lemma 9.28 (Semicontinuity of exit probabilities, Lemma 9.5.8). For each $t > 0$, $x \mapsto \mathbb{P}_x(\tau \le t)$ is lower semicontinuous. Delaying the process by $\varepsilon > 0$ gives a continuous approximation through the Gaussian transition density and dominated convergence. Letting $\varepsilon \downarrow 0$ gives the semicontinuity statement.

Lemma 9.29 (Concentration of exit locations, Lemma 9.5.9). If y is regular and $x_n \to y$ from inside G, then for every $\delta > 0$,

$$\mathbb{P}_{x_n}(\tau < \infty,\ B_\tau \in D(y,\delta)) \to 1.$$

Regularity and lower semicontinuity make τ small with high probability; Brownian continuity makes the path displacement before that small time unlikely to exceed δ.

Theorem 9.30 (Cone condition, Theorem 9.5.10). If a cone with vertex y lies in G^c near y, then y is regular. Spherical symmetry gives a positive probability, independent of scale, that B_t lies in the cone, while path continuity keeps it in the local neighborhood. Hence $\mathbb{P}_y(\tau = 0) > 0$, and Blumenthal's zero–one law makes this probability one.

9.5.1 Exit Distributions

For the half-space $H = \{(x, y) : x \in \mathbb{R}^{d-1}, y > 0\}$, define

$$h_\theta(x, y) = \frac{C_d y}{(|x - \theta|^2 + y^2)^{d/2}}, \qquad \theta \in \mathbb{R}^{d-1},$$

where C_d normalizes the integral over θ to one.

Lemma 9.31 (Half-space Poisson kernel, Lemma 9.5.11). For fixed θ, $\Delta h_\theta = 0$ in H. Direct differentiation shows cancellation between the tangential and normal second derivatives.

Theorem 9.32 (Half-space exit law, Theorem 9.5.12). For bounded continuous boundary data,

$$u(x, y) = \int_{\mathbb{R}^{d-1}} h_\theta(x, y) f(\theta, 0)\, \mathrm{d}\theta$$

solves the Dirichlet problem in H. Thus $h_\theta(x, y)\, \mathrm{d}\theta$ is the law of the exit location. Rescaling proves normalization and concentration near x as $y \downarrow 0$. In $d = 2$ this is the Cauchy density $y/[\pi((\theta - x)^2 + y^2)]$.

Theorem 9.33 (Ball Poisson kernel, Theorem 9.5.13). For the unit ball D, normalized surface measure π, and $k_y(x) = (1 - |x|^2)/\,|x - y|^d$,

$$\mathbb{E}_x f(B_{\tau_D}) = \int_{\partial D} k_y(x) f(y)\, \pi(\,\mathrm{d}y).$$

The kernel is harmonic in x, integrates to one, and concentrates at the nearest boundary point as x approaches ∂D.

9.6 Green's Functions and Potential Kernels

For a ball D, recurrence gives infinite total occupation time in $d \le 2$, while transience gives finite expected occupation time in $d \ge 3$:

$$\mathbb{P}_x\left(\int_0^\infty \mathbf{1}_D(B_t)\, \mathrm{d}t = \infty\right) = 1 \quad (d \le 2), \qquad \mathbb{E}_x \int_0^\infty \mathbf{1}_D(B_t)\, \mathrm{d}t < \infty \quad (d \ge 3).$$

In recurrent dimensions, successive excursions give independent contributions to occupation time. In transient dimensions, the finite expectation follows by integrating the heat kernel.

Definition 9.34 (Potential kernel). For $d \ge 3$,

$$G(x, y) = \int_0^\infty p_t(x, y)\, \mathrm{d}t = \frac{\Gamma(d/2 - 1)}{2\pi^{d/2}} |x - y|^{2-d}.$$

Then for nonnegative f, $\mathbb{E}_x \int_0^\infty f(B_t)\, \mathrm{d}t = \int G(x, y) f(y)\, \mathrm{d}y$.

Proof sketch. Tonelli's theorem interchanges expectation and the integrals in time and space. The substitution $t = |x-y|^2/(2s)$ in the heat kernel integral gives a gamma integral, yielding the stated constant and the power $|x-y|^{2-d}$.

Definition 9.35 (Renormalized recurrent kernels). When $d \le 2$, subtract a time-dependent reference value before integrating:

$$G(x,y) = \int_0^\infty [p_t(x,y) - a_t]\, \mathrm{d}t.$$

Choosing $a_t = p_t(0,0)$ in $d=1$ and $a_t = p_t(0,e_1)$ in $d=2$ gives

$$G(x,y) = -|x-y| \quad (d=1), \qquad G(x,y) = -\pi^{-1}\log|x-y| \quad (d=2).$$

The term involving a_t vanishes when integrated against a function whose integral is zero.

Remark 9.36 (Normalization of the fundamental solution). In every dimension, $G(x,y)$ is a constant multiple of the radial harmonic function from Section 9.1. The constants are chosen so that $\frac{1}{2}\Delta_x G(x,y) = -\delta_y$ in the distributional sense. Adding a harmonic constant to a recurrent potential kernel leaves its integrals against functions of integral zero unchanged.

9.7 Poisson's Equation

Let G now be a bounded open set and g bounded and continuous. Consider

$$\frac{1}{2}\Delta u = -g \quad \text{in } G, \qquad u = 0 \quad \text{on } \partial G.$$

Nonzero boundary data can be added through the Dirichlet solution.

Theorem 9.37 (Poisson martingale, Theorem 9.7.1). If u solves the equation and τ is the exit time from G, then $M_t = u(B_t) + \int_0^t g(B_s)\,\mathrm{d}s$ is a local martingale on $[0,\tau)$.

Theorem 9.38 (Occupation representation, Theorem 9.7.2). Every bounded solution equals

$$v(x) = \mathbb{E}_x \int_0^\tau g(B_t)\, \mathrm{d}t.$$

Because G is bounded, $\sup_x \mathbb{E}_x\tau < \infty$; bounded u and g therefore give an integrable dominating variable at the exit time. For $g \equiv 1$, $v(x) = \mathbb{E}_x\tau$.

Theorem 9.39 (Poisson equation in the interior, Theorem 9.7.3). If $v \in C^2$, then $\frac{1}{2}\Delta v = -g$. Stopping on a small interior ball gives

$$v(x) = \mathbb{E}_x \int_0^{\tau_r} g(B_s)\,\mathrm{d}s + \mathbb{E}_x v(B_{\tau_r}).$$

Taylor expansion gives the Laplacian term; continuity of g makes the occupation integral $[g(x) + o(1)]\mathbb{E}_x\tau_r$.

Theorem 9.40 (Zero boundary values, Theorem 9.7.4). If G and g are bounded and y is a regular boundary point, then $v(x_n) \to 0$ whenever $x_n \to y$ from inside G. Splitting the occupation integral at a small deterministic ε bounds its initial part by $\varepsilon \|g\|_\infty$, and the remaining conditional expectation is controlled by $\mathbb{P}_{x_n}(\tau > \varepsilon) \to 0$.

For compactly supported g on $\mathbb{R}^d$, define $w(x) = \int G(x,y)g(y)\,\mathrm{d}y$ using the ordinary or renormalized potential kernel. Strong Markov gives $v(x) = w(x) - \mathbb{E}_x w(B_\tau)$; the second term is harmonic in G.

Theorem 9.41 (Potential regularity, Theorem 9.7.5). If $d \geq 2$ and g is Hölder continuous, then $w \in C^2$. In $d = 1$, continuity of g suffices. Hence the occupation representation satisfies the Poisson equation classically in G. It attains zero at regular boundary points, so it solves the full boundary value problem if every point of ∂G is regular.

9.7.1 Occupation Times

Let $H = \{x : x_d > 0\}$ and let $\bar{y} = (y_1, \ldots, y_{d-1}, -y_d)$ be reflection across ∂H.

Theorem 9.42 (Killed half-space kernel, Theorem 9.7.6). For nonnegative compactly supported f inside H,

$$\mathbb{E}_x \int_0^{\tau_H} f(B_t)\,\mathrm{d}t = \int_H [G(x,y) - G(x,\bar{y})] f(y)\,\mathrm{d}y.$$

The reflection principle replaces the killed transition density by $p_t(x,y) - p_t(x,\bar{y})$. Subtracting the same renormalization constant from both terms makes the formula valid in recurrent and transient dimensions.

Example 9.43 (One-dimensional half-line). For $x, y > 0$, $G_H(x,y) = 2(x \wedge y)$. Although Brownian motion is killed at zero, the expected occupation density at every level $y \geq x$ is the same, namely $2x$.

Theorem 9.44 (Green function of the unit ball, Theorem 9.7.7). For bounded measurable g and exit time τ_D,

$$\mathbb{E}_x \int_0^{\tau_D} g(B_t)\,\mathrm{d}t = \int_D G_D(x,y)g(y)\,\mathrm{d}y,$$

where $G_D(x,y) = G(x,y) - \int_{\partial D} k_z(x)G(z,y)\pi(\,\mathrm{d}z)$. The subtracted harmonic extension forces the boundary value to vanish.

Theorem 9.45 (Explicit ball kernels, Theorems 9.7.8–9.7.10). For $d \geq 3$ and $0 < |y| < 1$,

$$G_D(x,y) = G(x,y) - |y|^{2-d}\, G\left(x, \frac{y}{|y|^2}\right).$$

For $d = 2$ and $0 < |y| < 1$, the analogous image formula is

$$G_D(x,y) = -\pi^{-1}\left[\log|x-y| - \log\left||y|\,x - y/|y|\right|\right].$$

In $d = 1$, $D = (-1,1)$ and

$$G_D(x,y) = \begin{cases} (1-x)(1+y), & y \leq x, \\ (1-y)(1+x), & x \leq y. \end{cases}$$

These formulas follow from the method of images applied to the kernel on $\mathbb{R}^d$.

Lemma 9.46 (Inversion geometry, Lemma 9.7.9). If $|x| = 1$ and $y \neq 0$, then

$$\left||y|\,x - y/|y|\right| = |x-y|.$$

This identity makes the image term agree with $G(x,y)$ on the sphere and hence enforces the zero boundary condition.

9.8 Schrödinger Equation

For bounded connected G, consider

$$\frac{1}{2}\Delta u + cu = 0 \quad \text{in } G, \qquad u = f \quad \text{on } \partial G,$$

with bounded continuous c and f. Let $C_t = \int_0^t c(B_s)\,\mathrm{d}s$.

Theorem 9.47 (Multiplicative martingale before exit, Theorem 9.8.1). If u solves the equation, then $M_t = u(B_t)e^{C_t}$ is a local martingale on $[0,\tau)$. This is the elliptic counterpart of Feynman–Kac.

Example 9.48 (Positive constant potential on an interval, Example 9.8.2). For $G = (-a, a)$, $c \equiv \gamma > 0$, and $f \equiv 1$, the ODE has candidate $u(x) = \cos(x\sqrt{2\gamma})/\cos(a\sqrt{2\gamma})$. The denominator first vanishes at $\gamma = \pi^2/(8a^2)$, the first Dirichlet eigenvalue, and the expression may change sign beyond this value. The expectation $\mathbb{E}_x e^{\gamma\tau}$ is infinite at and above the first eigenvalue.

Define the gauge

$$w(x) = \mathbb{E}_x \exp\left(\int_0^\tau c(B_s)\,\mathrm{d}s\right).$$

Lemma 9.49 (Exponential exit moments, Lemma 9.8.3). For every $\theta > 0$, there is $\mu > 0$ such that if an open H has Lebesgue measure at most μ, then $\sup_x \mathbb{E}_x e^{\theta\tau_H} \le 2$. A heat kernel estimate gives $\mathbb{P}_x(\tau_H > \gamma) \le 1/4$; the Markov property iterates this geometric tail, which dominates the exponential moment.

Lemma 9.50 (Local gauge comparison, Lemma 9.8.4). For a sufficiently small interior scale, if $y \in D(x, \delta)$ and $D(x, 2\delta) \subset G$, then $w(y) \le 2^{d+2} w(x)$. Exit from small balls, independence of exit time and uniform exit location, Jensen's inequality, and averaging over radii compare the local integrals of w around x and y.

Theorem 9.51 (Gauge dichotomy, Theorems 9.8.5–9.8.6). If G is connected and w is finite at one point, it is finite everywhere: local comparison shows that the set $\{x \in G : w(x) < \infty\}$ is both open and closed in G. If G also has finite volume, then $\sup_{x\in G} w(x) < \infty$. A compact subset is covered by finitely many balls on which local comparison applies, and the remaining set has sufficiently small measure for the exponential moment estimate.

For the rest of the section assume G is bounded, connected, and open; f, c are bounded and continuous; and w is not identically infinite.

Theorem 9.52 (Schrödinger uniqueness, Theorem 9.8.7). Every bounded classical solution equals

$$v(x) = \mathbb{E}_x\left[f(B_\tau)\exp\left(\int_0^\tau c(B_s)\,\mathrm{d}s\right)\right].$$

Stopping the local martingale at $t \wedge \tau$ gives a term on $\{\tau \le t\}$ and a remainder on $\{\tau > t\}$. Finiteness of the gauge permits dominated convergence in the first term. A positive uniform lower bound on w bounds the remainder by the tail of an integrable random variable, which tends to zero as $t \to \infty$.

Theorem 9.53 (Interior equation, Theorem 9.8.8). If $v \in C^2$, then $\frac{1}{2}\Delta v + cv = 0$ in G. Stopping at a small ball gives $v(x) = \mathbb{E}_x[v(B_\sigma)e^{C_\sigma}]$. Comparing the second order spatial expansion with $e^{C_\sigma} = 1 + [c(x) + o(1)]\sigma$ gives the differential equation.

Theorem 9.54 (Boundary values, Theorem 9.8.9). At every regular $y \in \partial G$, $v(x_n) \to f(y)$ as $x_n \to y$ from G. On $\{\tau \le 1\}$, concentration of the exit time and location gives the limit by bounded convergence. On $\{\tau > 1\}$, the Markov property and bounded gauge give a bound tending to zero because $\mathbb{P}_{x_n}(\tau > 1) \to 0$.

Theorem 9.55 (Smoothness, Theorem 9.8.10). If, in addition, c is Hölder continuous, then $v \in C^2$ and hence solves the interior Schrödinger equation. At regular boundary points it also attains the prescribed boundary values.

Theorem 9.56 (Interval spectral threshold, Theorem 9.8.11). For exit time τ from $(-a, a)$,

$$\mathbb{E}_x e^{\gamma\tau} = \frac{\cos(x\sqrt{2\gamma})}{\cos(a\sqrt{2\gamma})} \quad \text{if } 0 < \gamma < \frac{\pi^2}{8a^2},$$

while $\mathbb{E}_x e^{\gamma\tau} = \infty$ for $\gamma \ge \pi^2/(8a^2)$. The threshold is the first eigenvalue of $-\frac{1}{2}\Delta$ with zero boundary values on $(-a, a)$.

A Measure Theory Details

A.1 Carathéodory's Extension Theorem

Definition A.1 (Semialgebra and generated algebra). A semialgebra $\mathcal{S}$ contains $\varnothing$, is closed under intersections, and has the property that the complement of each member is a finite disjoint union of members of $\mathcal{S}$. The algebra $\overline{\mathcal{S}}$ generated by $\mathcal{S}$ consists exactly of finite disjoint unions of semialgebra sets.

Theorem A.2 (Extension from a semialgebra, Theorem A.1.1). Let μ be defined on a semialgebra $\mathcal{S}$, with $\mu(\varnothing) = 0$. Assume:

$$S = \dot{\bigcup_{i=1}^{n}} S_i \quad \Longrightarrow \quad \mu(S) = \sum_{i=1}^{n} \mu(S_i),$$

and, whenever $S \subseteq \bigcup_{i \ge 1} S_i$ with all sets in $\mathcal{S}$, $\mu(S) \le \sum_i \mu(S_i)$. Then μ extends uniquely to a measure $\bar{\mu}$ on $\overline{\mathcal{S}}$. If this extension is sigma-finite, it has a unique extension to $\sigma(\mathcal{S})$.

Proof sketch. For $A = \dot{\bigcup}_{i=1}^{n} S_i$, define $\bar{\mu}(A) = \sum_i \mu(S_i)$. Refine two representations by all intersections $S_i \cap T_j$ to prove the value is well defined. Finite additivity on the generated algebra is immediate; countable additivity follows by combining the assumed covering inequality with finite additivity of partial unions. The remaining existence and uniqueness steps are Carathéodory's theorem below.

Lemma A.3 (Finite decompositions and covers, Lemma A.1.2). Under finite additivity on $\mathcal{S}$, the induced $\bar{\mu}$ is finitely additive on $\overline{\mathcal{S}}$. If $A \subseteq \bigcup_{i=1}^{n} B_i$ with all sets in the algebra, then $\bar{\mu}(A) \leq \sum_i \bar{\mu}(B_i)$.

Theorem A.4 (Carathéodory extension, Theorem A.1.3). Every sigma-finite measure on an algebra $\mathcal{A}$ has a unique extension to $\sigma(\mathcal{A})$.

Definition A.5 (π- and λ-systems). A π-system is closed under finite intersections. A λ-system contains the whole space, is closed under $B \setminus A$ when $A \subseteq B$, and is closed under increasing countable unions.

Theorem A.6 (π–λ theorem, Theorem A.1.4). If a λ-system $\mathcal{L}$ contains a π-system $\mathcal{P}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.

Proof sketch. Let $\lambda(\mathcal{P})$ be the smallest λ-system containing $\mathcal{P}$. It is enough to show closure under intersections, since a λ-system with this extra property is a sigma-field. For fixed A, consider $\mathcal{G}_A = \{B : A \cap B \in \lambda(\mathcal{P})\}$. First take $A \in \mathcal{P}$, then $A \in \lambda(\mathcal{P})$, applying minimality twice.

Theorem A.7 (Uniqueness on a generator, Theorem A.1.5). Let ν_1, ν_2 agree on a π-system $\mathcal{P}$. If some $A_n \in \mathcal{P}$ satisfy $A_n \uparrow \Omega$ and $\nu_i(A_n) < \infty$, then the two measures agree on $\sigma(\mathcal{P})$.

Proof sketch. For $A \in \mathcal{P}$ with finite measure, the class of B satisfying $\nu_1(A \cap B) = \nu_2(A \cap B)$ is a λ-system. The generator theorem extends equality to all $B \in \sigma(\mathcal{P})$. Let $A = A_n$ and use continuity from below. Finiteness allows subtraction when verifying closure under differences of nested sets.

Definition A.8 (Outer measure). For an algebra measure μ, define for every $E \subseteq \Omega$

$$\mu^*(E) = \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) : A_i \in \mathcal{A},\ E \subseteq \bigcup_{i=1}^{\infty} A_i\right\}.$$

An outer measure vanishes at $\varnothing$, is monotone, and is countably subadditive. A set E is Carathéodory measurable when, for every $F \subseteq \Omega$,

$$\mu^*(F) = \mu^*(F \cap E) + \mu^*(F \cap E^c).$$

Remark A.9 (Verifying measurability). Subadditivity gives $\mu^*(F) \leq \mu^*(F \cap E) + \mu^*(F \cap E^c)$ for every E and F. Thus measurability follows from the reverse inequality.

Lemma A.10 (Agreement on the original algebra, Lemma A.1.6). For every $A \in \mathcal{A}$, $\mu^*(A) = \mu(A)$ and A is Carathéodory measurable. The first statement uses countable subadditivity of the original algebra measure; the second splits the sets in an approximating cover into their parts inside and outside A.

Theorem A.11 (Carathéodory measurable sets, Lemma A.1.7). The class $\mathcal{A}^*$ of Carathéodory measurable sets is a sigma-field, and $\mu^*|_{\mathcal{A}^*}$ is a measure. This conclusion holds for every outer measure.

Proof sketch. The criterion is symmetric under complements. Prove closure under finite unions by successively splitting an arbitrary test set; intersections follow by De Morgan. Disjointify a countable union and iterate the splitting identity for finite partial unions. Monotonicity and subadditivity then give measurability of the full union and countable additivity of μ^* on measurable sets.

Remark A.12 (Sigma-finiteness). Carathéodory's construction gives an extension for any pre-measure. Sigma-finiteness ensures uniqueness on the generated sigma-field and the representation of completed measurable sets in Theorem A.2.2.

A.2 Which Sets Are Measurable?

Write $\mathcal{A}_\sigma$ for countable unions of algebra sets and $\mathcal{A}_{\sigma\delta}$ for countable intersections of such unions. Let $A \triangle B = (A \setminus B) \cup (B \setminus A)$ denote symmetric difference.

Lemma A.13 (Approximation by algebraic sets, Lemma A.2.1). If $\mu^*(E) < \infty$, then:

1. for every $\varepsilon > 0$, some $A \in \mathcal{A}_\sigma$ contains E and has $\mu^*(A) \le \mu^*(E) + \varepsilon$;
2. when E is measurable, for every $\varepsilon > 0$ there is an algebra set A with $\mu^*(E \triangle A) < \varepsilon$;
3. some $C \in \mathcal{A}_{\sigma\delta}$ contains E and satisfies $\mu^*(C) = \mu^*(E)$.

Proof sketch. Choose a nearly optimal countable algebra cover. A sufficiently large finite partial union approximates that cover, proving the assertion about symmetric difference for measurable E. For the final claim, intersect outer approximations whose errors decrease to zero.

Theorem A.14 (Structure of the Carathéodory sigma-field, Theorem A.2.2). If μ is σ-finite on $\mathcal{A}$, then $B \in \mathcal{A}^*$ if and only if

$$B = A \setminus N, \qquad A \in \mathcal{A}_{\sigma\delta}, \qquad \mu^*(N) = 0.$$

Thus every Carathéodory measurable set differs from a set in the generated sigma-field by a set of outer measure zero.

Proof sketch. The reverse implication follows because sets in the generated sigma-field and sets of outer measure zero are measurable. For the forward implication, partition Ω into algebra sets of finite measure, approximate B on each piece with total error at most $1/n$, unite those approximations, and intersect over n. The excess set is contained in errors of arbitrarily small outer measure.

Definition A.15 (Complete measure space). $(\Omega, \mathcal{F}, \mu)$ is complete if every subset of every μ-null measurable set belongs to $\mathcal{F}$. Completeness is a property of the sigma-field as well as of the numerical measure.

Theorem A.16 (Completion, Theorem A.2.3). Every measure space has a completion $(\Omega, \overline{\mathcal{F}}, \bar{\mu})$, where $E \in \overline{\mathcal{F}}$ if and only if $E = A \cup B$ for some $A \in \mathcal{F}$ and $B \subseteq N \in \mathcal{F}$ with $\mu(N) = 0$. Set $\bar{\mu}(E) = \mu(A)$. This is well defined, countably additive, complete, and extends μ.

Proof sketch. Unions and complements preserve the stated representation, so $\overline{\mathcal{F}}$ is a sigma-field. Two representations differ only inside a measurable null set, which proves that $\bar{\mu}$ is well defined. For disjoint completed sets, the measurable sets A in their representations are also disjoint, so countable additivity follows from that of μ.

Example A.17 (Vitali's nonmeasurable set, Theorem A.2.4). On $[0, 1)$ declare $x \sim y$ when $x - y \in \mathbb{Q}$ and choose one representative from each equivalence class. The resulting set B is not Lebesgue measurable. Its rational rotations $q +^* B$, indexed by $q \in \mathbb{Q} \cap [0, 1)$, are disjoint, cover $[0, 1)$, and would all have the same measure: positive measure would make the union infinite, while zero measure would make it zero.

Lemma A.18 (Rotation modulo one preserves measure, Lemma A.2.5). If measurable $E \subseteq [0, 1)$ and $x \in (0, 1)$, then $x +^* E = \{(x + y) \bmod 1 : y \in E\}$ is measurable and has the same Lebesgue measure as E. Split E at $1 - x$, translate the two pieces separately, and use disjoint additivity.

Remark A.19 (Choice and nonmeasurability). The choice of one representative from each equivalence class uses the axiom of choice. The resulting set shows that Lebesgue measure cannot extend to every subset of $\mathbb{R}$ while retaining countable additivity and translation invariance.

A.3 Kolmogorov's Extension Theorem

Equip $\mathbb{R}^{\mathbb{N}}$ with the product sigma-field generated by finite-dimensional cylinder sets. A family μ_n of laws on $\mathbb{R}^n$ is consistent when projecting μ_{n+1} onto its first n coordinates gives μ_n.

Theorem A.20 (Kolmogorov extension, Theorem A.3.1). Every consistent family of probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ determines a unique probability measure $\mathbb{P}$ on $\mathbb{R}^{\mathbb{N}}$ whose first n coordinates have law μ_n for every n.

Example A.21 (Independent coordinates, Example A.3.2). Given distribution functions $F_1, F_2, \ldots$, take μ_n to be the product of their first n Stieltjes measures. Consistency is automatic. Under the extension, the coordinate maps are independent and the nth coordinate has distribution F_n.

Proof sketch. Define $\mathbb{P}$ on finite-dimensional rectangles from the prescribed marginals; these rectangles form a semialgebra. Finite additivity follows inside one sufficiently large $\mathbb{R}^n$. To obtain countable additivity, it is enough to show continuity at the empty set: if algebra sets $B_n \downarrow \varnothing$, then $\mathbb{P}(B_n) \downarrow 0$. Approximate B_n from within by finite unions of compact rectangles. If their probabilities stayed positive, nested compactness and a diagonal subsequence would produce a point in every B_n, a contradiction. Then apply the extension theorem and generator uniqueness.

Lemma A.22 (Continuity at the empty set, Lemma A.3.3). For the cylinder premeasure, if B_n belong to the cylinder algebra and $B_n \downarrow \varnothing$, then $\mathbb{P}(B_n) \downarrow 0$. The diagonal compactness step is the essential tightness argument specific to standard coordinate spaces.

Remark A.23 (Consistency and dependence). Consistency requires the finite-dimensional laws to agree under coordinate projections. The joint laws μ_n specify the dependence among coordinates; product laws give independence. The extension theorem also holds for standard Borel state spaces.

A.4 Radon–Nikodym Theorem

Definition A.24 (Signed measure). A signed measure on $(\Omega, \mathcal{F})$ is a map $\alpha : \mathcal{F} \to (-\infty, \infty]$ with $\alpha(\varnothing) = 0$. For every disjoint decomposition $E = \bigcup_i E_i$, $\alpha(E) = \sum_i \alpha(E_i)$. If $\alpha(E) < \infty$, the series converges absolutely. If $\alpha(E) = \infty$, then $\sum_i \alpha(E_i)^- < \infty$ and $\sum_i \alpha(E_i)^+ = \infty$. A measurable set is positive if every measurable subset has nonnegative α-mass, and negative if every subset has nonpositive mass.

Example A.25 (Basic signed measures, Examples A.4.1–A.4.2). If f^- is integrable, then $\alpha(A) = \int_A f \, d\mu$ is a signed measure. If $\mu_2(\Omega) < \infty$, then $\alpha = \mu_1 - \mu_2$ is another example. The Jordan theorem shows that every signed measure has the latter form with singular parts.

Lemma A.26 (Unions of positive or negative sets, Lemma A.4.3). Measurable subsets of positive sets are positive, and countable unions of positive sets are positive. The same statements hold for negative sets. Disjointify the union before using countable additivity.

Lemma A.27 (Existence of a negative subset, Lemma A.4.4). If $\alpha(E) < 0$, then E contains a negative measurable F with $\alpha(F) < 0$. Repeatedly remove subsets whose positive measures approach the supremum over the current set. Either the procedure stops, or absolute convergence implies that every measurable subset of the remaining intersection has nonpositive measure. Since only sets of positive measure were removed, the remaining set F satisfies $\alpha(F) \le \alpha(E) < 0$.

Theorem A.28 (Hahn decomposition, Theorem A.4.5). There are disjoint measurable sets P, N with $P \cup N = \Omega$, where P is positive and N is negative for α. If P', N' is another such decomposition, every measurable subset of $P \triangle P' = N \triangle N'$ has α-measure zero.

Proof sketch. Minimize $\alpha(B)$ over negative sets. A union of a minimizing sequence is negative and attains the infimum. If its complement were not positive, Lemma A.4.4 would produce an additional negative piece and lower the infimum, a contradiction.

Definition A.29 (Mutual singularity and total variation). Measures μ_1, μ_2 are mutually singular, written $\mu_1 \perp \mu_2$, if some measurable A satisfies $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$. For a signed measure with Jordan parts, $|\alpha| = \alpha^+ + \alpha^-$ is its total variation measure.

Theorem A.30 (Jordan decomposition, Theorem A.4.6). Every signed measure has a unique representation $\alpha = \alpha^+ - \alpha^-$ in which α^+ and α^- are mutually singular positive measures. Given a Hahn decomposition $P \cup N$, $\alpha^+(E) = \alpha(E \cap P)$ and $\alpha^-(E) = -\alpha(E \cap N)$.

Theorem A.31 (Lebesgue decomposition, Theorem A.4.7). For sigma-finite measures μ, ν, there are unique measures ν_r, ν_s such that

$$\nu = \nu_r + \nu_s, \qquad \nu_s \perp \mu, \qquad \nu_r(E) = \int_E g \, d\mu$$

for some measurable $g \ge 0$.

Proof sketch. After reducing to finite measures, consider all $g \ge 0$ whose induced measure is dominated by ν. This class is closed under maxima. Choose a maximizing sequence, take successive maxima, and use monotone convergence to obtain an extremal h. The remainder $\nu_s = \nu - h\mu$ must be singular: otherwise a Hahn decomposition of $\nu_s - \varepsilon\mu$ would reveal a positive set on which h could be increased.

Definition A.32 (Absolute continuity). ν is absolutely continuous with respect to μ, written $\nu \ll \mu$, when $\mu(A) = 0$ implies $\nu(A) = 0$.

Theorem A.33 (Radon–Nikodym, Theorem A.4.8). If μ and ν are sigma-finite and $\nu \ll \mu$, then there is measurable $g \geq 0$ such that $\nu(E) = \int_E g \, \mathrm{d}\mu$ for all measurable E. The density $g = \mathrm{d}\nu/\mathrm{d}\mu$ is unique μ-almost everywhere.

Proof sketch. Take the Lebesgue decomposition of ν relative to μ. The singular part is supported on a μ-null set and therefore must vanish by $\nu \ll \mu$. For uniqueness, integrate two candidate densities over subsets of $\{g > h\}$ and $\{g < h\}$, localizing to finite measure and bounded level sets under sigma-finiteness.

Example A.34 (Failure without sigma-finiteness, Example A.4.9). On $\mathbb{R}$, Lebesgue measure is absolutely continuous with respect to counting measure, because counting measure has no nonempty null set. Yet no density with respect to counting measure can reproduce Lebesgue measure. The dominating measure is not sigma-finite.

Remark A.35 (Calculus of densities). For $\nu_1, \nu_2 \ll \mu$ and $\pi \ll \nu \ll \mu$, respectively,

$$\frac{\mathrm{d}(\nu_1 + \nu_2)}{\mathrm{d}\mu} = \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu} + \frac{\mathrm{d}\nu_2}{\mathrm{d}\mu}, \qquad \frac{\mathrm{d}\pi}{\mathrm{d}\mu} = \frac{\mathrm{d}\pi}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

Also

$$\int f \, \mathrm{d}\nu = \int f \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\mu,$$

and when $\nu \ll \mu \ll \nu$, the two Radon–Nikodym densities are reciprocals almost everywhere.

A.5 Differentiating under the Integral

Theorem A.36 (Integrability of the derivative). Book Theorem A.5.1. Let $u(x) = \int_S f(x,s)\,\mu(\mathrm{d}s)$ near y. Suppose $f(x,\cdot)$ is integrable, $\partial_x f(x,s)$ exists and is continuous in x for fixed s, $v(x) = \int_S \partial_x f(x,s)\,\mu(\mathrm{d}s)$ is continuous at y, and

$$\int_S \int_{-\delta}^{\delta} |\partial_x f(y+\theta, s)| \; \mathrm{d}\theta \, \mu(\mathrm{d}s) < \infty.$$

Then $u'(y) = v(y)$.

Proof sketch. Use the fundamental theorem of calculus in x and Fubini to obtain

$$\frac{u(y+h)-u(y)}{h} = \frac{1}{h}\int_0^h v(y+\theta)\,\mathrm{d}\theta.$$

Continuity of v at y makes this average converge to $v(y)$. The assumed integrability of the derivative allows the use of Fubini's theorem.

Example A.37 (Gaussian Fourier integral, Example A.5.2). For

$$u(x) = \int_{\mathbb{R}} \cos(xs)e^{-s^2/2}\,\mathrm{d}s,$$

the derivative is dominated by $|s|\,e^{-s^2/2}$, an integrable function independent of x. Thus differentiation under the integral is justified.

Theorem A.38 (Dominating supremum criterion, Theorem A.5.3). In Theorem A.5.1, it is enough to replace the last two hypotheses by

$$\int_S \sup_{|\theta|\le\delta} |\partial_x f(y+\theta,s)|\;\mu(\mathrm{d}s) < \infty.$$

This bound implies both continuity of the integrated derivative by dominated convergence and integrability over the parameter rectangle.

Example A.39 (Derivative of an mgf, Example A.5.4). If $\phi(\theta) = \mathbb{E}e^{\theta Z}$ is finite for $\theta \in [-\varepsilon,\varepsilon]$, then $\phi'(0) = \mathbb{E}Z$. On the smaller interval $|\theta| \le \varepsilon/2$, the derivative $|Z|\,e^{\theta Z}$ is bounded by a constant times $e^{\varepsilon Z} + e^{-\varepsilon Z}$.

Theorem A.40 (Termwise differentiation, Theorem A.5.5). Suppose $u(x) = \sum_{n\ge1} f_n(x)$ converges absolutely near y, every f_n' exists and is continuous, and $\sum_n \sup_{|\theta|<\delta} |f_n'(y+\theta)| < \infty$. Then $u'(y) = \sum_n f_n'(y)$. This is Theorem A.5.3 with counting measure.

Example A.41 (Geometric series, Example A.5.6). For $p \in (0,1)$, termwise differentiation is valid in

$$\left(\sum_{n=1}^{\infty}(1-p)^n\right)' = -\sum_{n=1}^{\infty} n(1-p)^{n-1}.$$

Choose a compact interval inside $(0,1)$; the derivative terms there are dominated by a summable geometric sequence after absorbing the factor n into a slightly slower rate.