Stochastic Processes

Pheneas Newman

October 6, 2025

Contents

Notation

$\mathbb{R}, \mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{C}$ Real, natural, integer, rational, complex numbers.

$\mathcal{B}(\mathbb{R})$ Borel σ-algebra on $\mathbb{R}$.

λ Lebesgue measure.

Ω Sample space.

$\mathcal{F}$ σ-algebra of events.

$\mathcal{F}_t$ Filtration up to time t.

$\mathbb{P}$ Physical (real-world) probability measure.

$\mathbb{Q}$ Risk-neutral (martingale) probability measure.

$\mathbb{E}^{\mathbb{P}}, \mathbb{E}^{\mathbb{Q}}$ Expectation under $\mathbb{P}$ or $\mathbb{Q}$. Assume under $\mathbb{P}$ unless stated otherwise.

ω A generic element of the sample space Ω, i.e. an elementary outcome.

$\mathbb{1}_A$ Indicator of event A.

X_t Generic stochastic process.

M_t Generic martingale.

W_t Standard Brownian motion.

N_t Poisson process.

$\langle X \rangle_t$ Quadratic variation of process X up to time t.

$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$ General stochastic differential equation.

S_t Asset price process.

B_t Bank account / risk-free asset.

r Risk-free rate.

V_t Derivative price process.

Δ Hedge ratio.

$\xrightarrow{a.s.}$ Convergence almost surely.

$\xrightarrow{p}$ Convergence in probability.

$\xrightarrow{L^p}$ Convergence in L^p.

$\xrightarrow{d}$ Convergence in distribution (weak convergence).

Introduction

This aims to explain stochastic processes from measure theory to financial applications. It mainly serves as a way for me to not forget what I've learned, but hopefully I can make something worthwhile and helpful as well.

Chapter 1

Measure Theory

1.1 Measure Spaces

1.1.1 Definition of σ-algebra

Definition 1.1.1 (σ-algebra)**.** A *σ-algebra* on a set Ω is a collection $\mathcal{F}$ of subsets of Ω such that:

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

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

(iii) If $A_1, A_2, \cdots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Example 1.1.2 (Coin tosses)**.** Toss a coin twice, with heads probability $p \in (0,1)$. The sample space is $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, where

ω_1 heads then heads,

ω_2 heads then tails,

ω_3 tails then heads,

ω_4 tails then tails.

The full σ-algebra is the power set $\mathcal{F} = \mathcal{P}(\Omega)$. Equivalently,

$$\mathcal{F} = \sigma(\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}).$$

Example 1.1.3 (Borel σ-algebra on $\mathbb{R}$)**.** The *Borel σ-algebra* on $\mathbb{R}$, denoted $\mathcal{B}(\mathbb{R})$, is the smallest σ-algebra containing all open intervals $(a, b) \subseteq \mathbb{R}$.

- By definition, $(a, b) \in \mathcal{B}(\mathbb{R})$.
- Its complement $(-\infty, a] \cup [b, \infty)$ is also in $\mathcal{B}(\mathbb{R})$.
- By closure under countable unions, $\bigcup_{n=1}^{\infty}(1/n,\, 1 - 1/n) = (0, 1)$ is in $\mathcal{B}(\mathbb{R})$.

Definition 1.1.4 (Generated σ-algebra)**.** For a collection $\mathcal{C} \subseteq 2^{\Omega}$, the *generated σ-algebra* $\sigma(\mathcal{C})$ is the smallest σ-algebra containing $\mathcal{C}$:

$$\sigma(\mathcal{C}) = \bigcap \{\, \mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-algebra and } \mathcal{C} \subseteq \mathcal{G} \,\}.$$

Definition 1.1.5 (Product σ-algebra)**.** If $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are measurable spaces, the *product σ-algebra* on $\Omega_1 \times \Omega_2$ is

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\big(\{A \times B : A \in \mathcal{F}_1,\ B \in \mathcal{F}_2\}\big).$$

Note

A σ-algebra is the collection of events we are allowed to talk about. It is closed under complements and countable unions, so if an event is included, so are "not that event" and "any countable combination of such events."

1.1.2 Measures

Intuitively, a measure is a generalisation of length, area, or volume. Formally, it is a function that assigns a nonnegative number to each set in a σ-algebra, in a way that is consistent with disjoint unions.

Definition 1.1.6 (Measure)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function

$$\mu : \mathcal{F} \to [0, \infty]$$

is called a *measure* if

(i) $\mu(\emptyset) = 0$,

(ii) For any countable collection $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ of pairwise disjoint sets,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

The triple $(\Omega, \mathcal{F}, \mu)$ is then called a *measure space*.

Definition 1.1.7 (σ-finite measure)**.** A measure μ on $(\Omega, \mathcal{F})$ is *σ-finite* if

$$\Omega = \bigcup_{n=1}^{\infty} A_n \quad \text{for sets } A_n \in \mathcal{F} \text{ with } \mu(A_n) < \infty.$$

Definition 1.1.8 (Complete measure space)**.** A measure space $(\Omega, \mathcal{F}, \mu)$ is *complete* if whenever $N \in \mathcal{F}$ with $\mu(N) = 0$ and $A \subseteq N$, then $A \in \mathcal{F}$.

Remark 1.1.9. Intuitively, a measure is a rule for assigning "sizes" or "weights" to sets. - Condition (i) says the empty set has size zero. - Condition (ii) says that the measure is *countably additive*: the size of a disjoint union is the sum of the sizes.

This captures the familiar properties of length, area, or volume, but in a far more general setting.

Some important consequences:

Proposition 1.1.10 (Basic properties of measures). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then:*

(i) ***Monotonicity:*** *If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.*

(ii) ***Finite additivity:*** *If $A_1, \ldots, A_n$ are pairwise disjoint,*

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i).$$

(iii) ***Continuity from below:*** *If $A_1 \subseteq A_2 \subseteq \ldots$, then*

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

(iv) ***Continuity from above:*** *If $A_1 \supseteq A_2 \supseteq \ldots$ and $\mu(A_1) < \infty$, then*

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Proof (sketch). (i) If $A \subseteq B$, then $B = A \cup (B \setminus A)$ disjointly, hence $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$. (ii) Special case of countable additivity. (iii) Define $B = \bigcup_{n=1}^{\infty} A_n$. The sets $B_n = A_n \setminus A_{n-1}$ are disjoint, so $\mu(B) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \mu(A_n)$. (iv) Apply (iii) to complements A_n^c. □

Thus, a measure behaves much like ordinary "volume" but is abstract enough to cover discrete spaces (counting measure), continuous spaces (Lebesgue measure), and probability spaces (where the measure of the whole space is 1).

Example 1.1.11 (Counting measure)**.** On any set Ω, define $\mu(A) = |A|$ if A is finite, and $\mu(A) = \infty$ if A is infinite. This is a measure called the *counting measure*.

Example 1.1.12 (Dirac measure)**.** For a fixed point $\omega_0 \in \Omega$, define

$$\delta_{\omega_0}(A) = \begin{cases} 1 & \text{if } \omega_0 \in A, \\ 0 & \text{if } \omega_0 \notin A. \end{cases}$$

This is a measure concentrated at a single point, called the *Dirac measure*.

Example 1.1.13 (Lebesgue measure)**.** On $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the *Lebesgue measure* λ is defined so that $\lambda((a, b)) = b - a$ for all intervals $a < b$. It extends uniquely to all Borel sets and beyond.

Remark 1.1.14. For an interval $(a, b) \subset \mathbb{R}$, the Lebesgue measure satisfies $\lambda((a, b)) = b - a$. This coincides with the Riemann integral of the constant function 1:

$$\lambda((a, b)) = \int_a^b 1 \, dx.$$

The Lebesgue measure can be viewed as the rigorous extension of this "length of an interval" idea to much more complicated sets.

Note

A measure is a way of assigning "sizes" or "weights" to sets in a consistent way. It generalises length, area, and volume, but can also count discrete points or assign probability mass. The key idea is that disjoint sets add up.

1.1.3 Probability Measures

A probability measure is simply a measure normalised so that the total mass is one.

Definition 1.1.15 (Probability measure)**.** A measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is called a *probability measure* if $\mathbb{P}(\Omega) = 1$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

Example 1.1.16 (Finite probability space)**.** Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ and $\mathcal{F} = 2^\Omega$. If $p_i \geq 0$ with $\sum_{i=1}^n p_i = 1$, define

$$\mathbb{P}(\{\omega_i\}) = p_i, \quad i = 1, \ldots, n.$$

This extends uniquely to a probability measure on all subsets of Ω.

Example 1.1.17 (Coin toss)**.** Let $\Omega = \{H, T\}^2 = \{(H,H), (H,T), (T,H), (T,T)\}$. For a fair coin, assign $\mathbb{P}(\omega) = \frac{1}{4}$ for each $\omega \in \Omega$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space with uniform distribution.

Example 1.1.18 (Gaussian measure)**.** On $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, define for $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

This probability measure corresponds to the standard normal distribution $N(0,1)$.

Proposition 1.1.19 (Basic properties of probability measures)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then for all $A, B \in \mathcal{F}$:*

(i) ***Bounds:*** $0 \leq \mathbb{P}(A) \leq 1$.

(ii) ***Complement rule:*** $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(iii) ***Monotonicity:*** *If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

(iv) ***Finite additivity:*** *If $A \cap B = \emptyset$, then*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

(v) ***Union bound (Boole's inequality):***

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

More generally, for any finite or countable collection $\{A_i\}_{i \geq 1}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

(vi) ***Inclusion–Exclusion (two sets):***

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Proof (sketch). (i) Since $\emptyset \subseteq A \subseteq \Omega$ and $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$, monotonicity gives $0 \leq \mathbb{P}(A) \leq 1$. (ii) Follows because $\Omega = A \cup A^c$ disjointly. (iii) Same as the monotonicity of general measures. (iv) From countable additivity. (v) The sets A and $B \setminus A$ are disjoint, so $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \leq \mathbb{P}(A) + \mathbb{P}(B)$. (vi) Standard rearrangement using $A \cup B = (A \setminus B) \cup B$. □

Note

A probability measure is just a measure with $\mathbb{P}(\Omega) = 1$. This ensures the whole sample space has probability 1, and all other events get a value between 0 and 1. Every probability measure is σ-finite, since Ω itself has finite measure. This forms the rigorous foundation for Kolmogorov's axioms of probability.

1.1.4 Measurable functions

Note

So far, measures are defined only on sets. To assign probabilities to events involving a function $X : \Omega \to \mathbb{R}$ (such as $\{X \leq 1\}$), we need to guarantee that such events belong to $\mathcal{F}$. This requirement is called *measurability*, and it allows us to treat random variables as functions compatible with the underlying probability structure.

Definition 1.1.20 (Measurable function)**.** Let $(\Omega, \mathcal{F})$ be a measurable space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ the real line with its Borel σ-algebra. A function $X : \Omega \to \mathbb{R}$ is called *$\mathcal{F}$-measurable* if

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \qquad \forall B \in \mathcal{B}(\mathbb{R}).$$

Proposition 1.1.21 (Practical criterion for measurability)**.** *A function $X : \Omega \to \mathbb{R}$ is measurable if and only if*

$$\{\, \omega \in \Omega : X(\omega) \leq a \,\} \in \mathcal{F} \qquad \text{for all } a \in \mathbb{R}.$$

Example 1.1.22 (Discrete random variable)**.** Let $\Omega = \{HH, HT, TH, TT\}$, and define $X : \Omega \to \mathbb{R}$ as the number of heads. Then $\{X = 1\} = \{HT, TH\} \in \mathcal{F}$. All preimages of sets of the form $\{0\}, \{1\}, \{2\}$ are measurable, hence X is measurable.

Example 1.1.23 (Identity function)**.** Let $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, and $X(\omega) = \omega$. For any $a \in \mathbb{R}$,

$$\{\omega : X(\omega) \leq a\} = (-\infty, a] \in \mathcal{B}(\mathbb{R}),$$

so X is measurable.

Remark 1.1.24 (Random variables). In probability theory, $\mathcal{F}$-measurable functions are called *random variables*. Thus, a random variable is simply a measurable mapping from the sample space into the real line.

Remark 1.1.25 (Measurability with respect to a smaller σ-algebra). If $\mathcal{G} \subseteq \mathcal{F}$, we say X is *$\mathcal{G}$-measurable* if $X^{-1}(B) \in \mathcal{G}$ for all Borel sets B. Intuitively, measurability depends on the "information" available. For example, if $\mathcal{G} = \{\emptyset, \Omega\}$, the only $\mathcal{G}$-measurable functions are constants.

Remark 1.1.26 (Connection to filtrations). If $(\mathcal{F}_t)_{t \geq 0}$ is a filtration, then X_t is $\mathcal{F}_t$-measurable if its value at time t is determined by the information available up to t. This notion is central for defining *adapted processes* later.

Proposition 1.1.27 (Closure properties)**.** *If X, Y are measurable functions and $f : \mathbb{R} \to \mathbb{R}$ is Borel-measurable, then*

(i) *$X + Y$, $X - Y$, XY, and $\max(X, Y)$ are measurable.*

(ii) *$f \circ X$ is measurable.*

(iii) *If (X_n) is a sequence of measurable functions, then $\sup_n X_n$, $\inf_n X_n$, $\limsup_n X_n$, and $\liminf_n X_n$ are measurable.*

Definition 1.1.28 (Equality almost everywhere)**.** Given a measure space $(\Omega, \mathcal{F}, \mu)$, we say $X = Y$ *almost everywhere* (a.e.) if $\mu(\{\omega : X(\omega) \neq Y(\omega)\}) = 0$.

Remark 1.1.29. In probability theory, random variables that are equal a.e. are considered equivalent, since they induce the same distributions and expectations. Most results hold "a.e." rather than pointwise.

Note

Informally, a function is *measurable* if every event of the form $\{X \in B\}$ has a well-defined probability, i.e. its preimage lies in the σ-algebra.

1.2 Integration

1.2.1 Lebesgue integral

The Riemann integral partitions the *domain* (the x-axis), while the Lebesgue integral partitions the *range* (the y-axis). This shift makes the Lebesgue integral compatible with measure theory, allowing us to integrate functions with many discontinuities or defined on abstract spaces.

Note

Recall: the *Lebesgue measure* λ on $\mathbb{R}$ extends the idea of length, with $\lambda((a,b)) = b-a$. This measure lets us assign "sizes" to complicated sets, forming the foundation of the Lebesgue integral.

Step 1. Simple functions. If $s = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ with $a_i \geq 0$ and A_i measurable, define

$$\int s \, d\mu := \sum_{i=1}^{n} a_i \, \mu(A_i).$$

Step 2. Nonnegative functions. For a measurable $f \geq 0$, define

$$\int f \, d\mu := \sup\{ \int s \, d\mu : 0 \leq s \leq f, \ s \text{ simple}\}.$$

Step 3. General functions. If $f = f^+ - f^-$ with $\int f^+ \, d\mu < \infty$ and $\int f^- \, d\mu < \infty$, set

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

Example 1.2.1 (Riemann vs Lebesgue: $f(x) = x^2$ on $[0,1]$)**.** We compare the two integration approaches:

- **Riemann:** Subdivide $[0,1]$ into n equal subintervals of width $\Delta x = \frac{1}{n}$. The right-endpoint Riemann sum is
$$S_n = \sum_{k=1}^{n} \left(\frac{k}{n}\right)^2 \cdot \frac{1}{n}.$$
Using $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$, one finds
$$\lim_{n\to\infty} S_n = \frac{1}{3}.$$
- **Lebesgue:** Slice the range instead. For $y \in [0,1]$,
$$E_y = \{x \in [0,1] : x^2 > y\} = (\sqrt{y}, 1],$$
with measure $\lambda(E_y) = 1 - \sqrt{y}$. By the layer-cake representation,
$$\int_0^1 x^2 \, d\lambda = \int_0^1 (1 - \sqrt{y}) \, dy = \frac{1}{3}.$$

Both methods give the same result, but the perspective differs: Riemann integration partitions the *domain* into vertical slices, while Lebesgue integration partitions the *range* into horizontal slices.

Example 1.2.2 (Why Lebesgue is stronger)**.** Let $f = \mathbb{1}_{\mathbb{Q}\cap(0,1)}$.

- Riemann: every interval contains rationals and irrationals, so upper sums $= 1$ and lower sums $= 0$. Hence not Riemann-integrable.
- Lebesgue: rationals have measure zero, so $\int f\, d\lambda = 0$.

Proposition 1.2.3 (Basic properties)**.** *If f, g are measurable and integrable, and $\alpha \geq 0$:*

(i) Linearity: $\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu$,

(ii) Positive homogeneity: $\int \alpha f\, d\mu = \alpha \int f\, d\mu$,

(iii) Monotonicity: $f \leq g \implies \int f \leq \int g$,

(iv) Agreement with Riemann when f is Riemann-integrable.

Note

Riemann: slice vertically. Lebesgue: slice horizontally. This distinction allows us to prove the convergence theorems that follow: Monotone Convergence, Fatou's Lemma, and the Dominated Convergence Theorem.

Note

Supremum and Infimum. For a set $A \subseteq \mathbb{R}$: - $\sup A$ = least upper bound, - $\inf A$ = greatest lower bound.
Limsup and Liminf. For a sequence (a_n):

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{k\geq n} a_k, \quad \liminf_{n\to\infty} a_n = \lim_{n\to\infty} \inf_{k\geq n} a_k.$$

If $\limsup = \liminf$, the usual limit exists. These tools are essential for the convergence theorems that follow.

1.2.2 Monotone Convergence Theorem

Theorem 1.2.4 (Monotone Convergence Theorem (MCT))**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $(X_n)_{n\geq 1}$ be an increasing sequence of nonnegative measurable functions, i.e. $0 \leq X_1 \leq X_2 \leq \cdots$ and $X_n(\omega) \uparrow X(\omega)$ for μ-a.e. $\omega \in \Omega$, for some measurable $X : \Omega \to [0, \infty]$. Then

$$\lim_{n\to\infty} \int X_n\, d\mu \;=\; \int X\, d\mu,$$

with the understanding that both sides may be $+\infty$.

Proof. Since $X_n \leq X$ for all n, monotonicity of the integral gives $\int X_n\, d\mu \leq \int X\, d\mu$, hence $\limsup_n \int X_n\, d\mu \leq \int X\, d\mu$.

For the reverse inequality, let s be any simple function with $0 \leq s \leq X$. Write $s = \sum_{j=1}^{m} a_j \mathbf{1}_{A_j}$ with $a_j \geq 0$ and $A_j \in \mathcal{F}$ disjoint. Fix j. On A_j we have $a_j \leq X$. Since $X_n \uparrow X$ a.e., the sets

$$A_{j,n} := \{\omega \in A_j : X_n(\omega) \geq a_j\}$$

increase to A_j (i.e. $A_{j,n} \uparrow A_j$), so by continuity from below of μ, $\mu(A_{j,n}) \uparrow \mu(A_j)$.

For each n,

$$\int X_n \, d\mu \;\geq\; \sum_{j=1}^{m} a_j \, \mu(A_{j,n}).$$

Taking $n \to \infty$ and using $\mu(A_{j,n}) \uparrow \mu(A_j)$,

$$\liminf_{n\to\infty} \int X_n \, d\mu \;\geq\; \sum_{j=1}^{m} a_j \, \mu(A_j) \;=\; \int s \, d\mu.$$

Since this holds for every simple $s \leq X$, taking the supremum over all such s yields $\liminf_n \int X_n \, d\mu \geq \int X \, d\mu$ by the definition of the Lebesgue integral of X. Combining the two inequalities gives $\lim_n \int X_n \, d\mu = \int X \, d\mu$. □

Example 1.2.5 (Indicator functions filling up the interval)**.** Let $\Omega = [0,1]$ with Lebesgue measure λ, and define $X_n(\omega) = \mathbf{1}_{[0,\,1-1/n]}(\omega)$. Then $0 \leq X_1 \leq X_2 \leq \cdots$ and $X_n(\omega) \uparrow 1$ for λ-a.e. $\omega \in [0,1]$. Hence, by MCT,

$$\lim_{n\to\infty} \int_0^1 X_n(\omega) \, d\lambda(\omega) = \int_0^1 1 \, d\lambda = 1.$$

Example 1.2.6 (Truncation of a nonnegative random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to [0,\infty]$ be measurable. Define the increasing sequence $X_n = \min(X, n)$. Then $X_n \uparrow X$ $\mathbb{P}$-a.s., so by MCT

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

This holds whether $\mathbb{E}[X]$ is finite or infinite. In particular, if $\mathbb{E}[X] < \infty$ it shows that expectations can be computed as limits of expectations of the bounded truncations X_n.

Note

MCT justifies interchanging limit and integral for *monotone increasing* nonnegative sequences:

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

1.2.3 Fatou's Lemma

Lemma 1.2.7 (Fatou's Lemma)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of nonnegative measurable random variables on a measure space $(\Omega, \mathcal{F}, \mu)$. Then*

$$\mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big] \;\leq\; \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Proof (sketch). Define

$$Y_k(\omega) := \inf_{n\geq k} X_n(\omega), \qquad k = 1, 2, \ldots$$

so that $Y_1 \leq Y_2 \leq \ldots$ and

$$\lim_{k\to\infty} Y_k(\omega) = \liminf_{n\to\infty} X_n(\omega).$$

By the Monotone Convergence Theorem,

$$\mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big] = \lim_{k\to\infty} \mathbb{E}[Y_k].$$

But $Y_k \le X_n$ for all $n \ge k$, so

$$\mathbb{E}[Y_k] \le \inf_{n \ge k} \mathbb{E}[X_n].$$

Taking limits gives

$$\lim_{k\to\infty} \mathbb{E}[Y_k] \le \lim_{k\to\infty} \inf_{n\ge k} \mathbb{E}[X_n] = \liminf_{n\to\infty} \mathbb{E}[X_n].$$

□

Corollary 1.2.8 (Fatou with integrable bounds)**.** *Let (X_n) be random variables.*

(i) If $X_n \ge Y$ $\mathbb{P}$-a.s. for all n, where $Y \in L^1$, then

$$\mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big] \;\le\; \liminf_{n\to\infty} \mathbb{E}[X_n].$$

(ii) If $X_n \le Y$ $\mathbb{P}$-a.s. for all n, where $Y \in L^1$, then

$$\mathbb{E}\Big[\limsup_{n\to\infty} X_n\Big] \;\ge\; \limsup_{n\to\infty} \mathbb{E}[X_n].$$

(This follows by applying Fatou's Lemma to $-X_n$.)

Note

Informal explanation: Fatou's Lemma says that expectations "preserve inequalities" when passing to limits. - For the *liminf* version, the expectation of the pointwise liminf is at most the liminf of expectations. - For the *limsup* version, the expectation of the pointwise limsup is at least the limsup of expectations.
Intuitively, expectations cannot "overshoot" when you pass to limits. Fatou's Lemma is weaker than the Dominated Convergence Theorem, but it requires fewer assumptions. It is often used as a building block for convergence theorems.

1.2.4 Dominated Convergence Theorem

Theorem 1.2.9 (Dominated Convergence Theorem (DCT))**.** *Let (X_n) be a sequence of random variables that converges to a random variable X $\mathbb{P}$-a.s. Suppose there exists an integrable random variable $Y \in L^1$ such that $|X_n| \le Y$ $\mathbb{P}$-a.s. for all $n \ge 1$. Then*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Proof. Since $|X_n| \le Y$ a.s. and $X_n \to X$ a.s., by continuity of the absolute value we have $|X| \le Y$ a.s., hence $X \in L^1$ and $\mathbb{E}[|X|] \le \mathbb{E}[Y] < \infty$.

Consider the nonnegative random variables

$$U_n := Y + X_n \quad \text{and} \quad V_n := Y - X_n,$$

which converge a.s. to $U := Y + X$ and $V := Y - X$, respectively. By Fatou's lemma applied to (U_n),

$$\mathbb{E}[X] \;\le\; \liminf_{n\to\infty} \mathbb{E}[X_n]. \tag{1}$$

Similarly, applying Fatou's lemma to (V_n),

$$\mathbb{E}[X] \;\ge\; \limsup_{n\to\infty} \mathbb{E}[X_n]. \tag{2}$$

Combining (1) and (2) gives

$$\limsup_{n\to\infty} \mathbb{E}[X_n] \;\le\; \mathbb{E}[X] \;\le\; \liminf_{n\to\infty} \mathbb{E}[X_n],$$

so the limit $\lim_{n\to\infty} \mathbb{E}[X_n]$ exists and equals $\mathbb{E}[X]$. □

Example 1.2.10 (Necessity of domination)**.** Let $\Omega = [0,1]$ with Lebesgue measure λ, and define

$$X_n(x) = n\,\mathbf{1}_{(0,1/n)}(x).$$

Then $X_n(x) \to 0$ for λ-a.e. x, but

$$\int_0^1 X_n(x)\,dx = 1 \quad \text{for all } n.$$

So $\lim \int X_n \neq \int \lim X_n$. Here there is no integrable dominating function Y, so the assumptions of DCT fail. This shows why domination is essential.

Note

The DCT justifies exchanging limit and expectation when the sequence is uniformly dominated by an L^1 random variable. It strengthens Fatou's Lemma by giving equality instead of just inequality, at the cost of requiring a domination condition.

1.2.5 Tonelli's and Fubini's theorems

For these theorems we state them without proof and illustrate their use.

Theorem 1.2.11 (Tonelli's theorem)**.** *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ-finite measure spaces. If $f : \Omega_1 \times \Omega_2 \to [0,\infty]$ is measurable, then*

$$\int_{\Omega_1\times\Omega_2} f\,d(\mu_1\otimes\mu_2) = \int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,d\mu_2(y)\right)d\mu_1(x) = \int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)\,d\mu_1(x)\right)d\mu_2(y).$$

Example 1.2.12 (Tonelli: double integral of a nonnegative function)**.** Let $f(x,y) = e^{-x-y}$ on $(0,\infty)\times(0,\infty)$ with Lebesgue measure. Then

$$\int_0^\infty\int_0^\infty e^{-(x+y)}\,dy\,dx = \int_0^\infty e^{-x}\left(\int_0^\infty e^{-y}\,dy\right)dx = \int_0^\infty e^{-x}(1)\,dx = 1.$$

Example 1.2.13 (Tonelli: indicator function)**.** Let $f(x,y) = \mathbb{1}_{\{x+y\le 1\}}$ on $[0,1]^2$. Then

$$\int_0^1\int_0^1 \mathbb{1}_{\{x+y\le1\}}\,dy\,dx = \int_0^1\int_0^{1-x} 1\,dy\,dx = \int_0^1 (1-x)\,dx = \tfrac{1}{2}.$$

Theorem 1.2.14 (Fubini's theorem)**.** *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ-finite measure spaces. If $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is integrable, i.e.*

$$\int_{\Omega_1\times\Omega_2} |f|\,d(\mu_1\otimes\mu_2) < \infty,$$

then the iterated integrals exist, and

$$\int_{\Omega_1\times\Omega_2} f\,d(\mu_1\otimes\mu_2) = \int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,d\mu_2(y)\right)d\mu_1(x) = \int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)\,d\mu_1(x)\right)d\mu_2(y).$$

Example 1.2.15 (Fubini: alternating sign function)**.** Let $f(x,y) = \sin(x)\cos(y)$ on $[0,\pi]\times[0,\pi]$. Then

$$\int_0^\pi \int_0^\pi \sin(x)\cos(y)\,dy\,dx = \int_0^\pi \sin(x)\left(\int_0^\pi \cos(y)\,dy\right)dx = 0.$$

The same result holds if we swap the order.

Example 1.2.16 (Fubini: absolute integrability required)**.** Let $f(x,y) = \frac{\sin(xy)}{xy}$ on $(0,\infty)\times(0,\infty)$. Then f is integrable, and

$$\int_0^\infty \int_0^\infty \frac{\sin(xy)}{xy}\,dy\,dx = \frac{\pi}{2}.$$

This uses the classical result $\int_0^\infty \frac{\sin(u)}{u}\,du = \frac{\pi}{2}$. Here Fubini ensures we can swap the order of integration safely.

Example 1.2.17 (Failure without absolute integrability)**.** Consider $f(x,y) = \frac{xy}{(x^2+y^2)^2}$ on $(0,\infty)^2$. Both iterated integrals exist, but they are not equal, so the double integral is undefined. This illustrates why absolute integrability is required for Fubini's theorem.

Note

Tonelli vs. Fubini. - Tonelli applies to nonnegative measurable functions, even if the integral is infinite. - Fubini applies to absolutely integrable functions, and guarantees equality of iterated and double integrals.
Why it matters in stochastic calculus. These theorems justify exchanging the order of integration in situations like: - computing expectations of stochastic integrals ($\mathbb{E}\int f\,dW = \int \mathbb{E}[f]\,dW$), - handling double integrals in covariance and quadratic variation computations, - and working with stochastic Fubini theorems when interchanging stochastic and Lebesgue integrals. They are essential whenever integrals over time and probability are combined.

1.2.6 Expectation as a Lebesgue integral

Definition 1.2.18 (Expectation)**.** Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $X:\Omega\to\mathbb{R}$ an $\mathcal{F}$-measurable random variable. If X is integrable, i.e. $\int_\Omega |X|\,d\mathbb{P} < \infty$, the *expectation* of X is

$$\mathbb{E}[X] := \int_\Omega X(\omega)\,d\mathbb{P}(\omega).$$

Equivalently, if $\mu = \mathcal{L}(X)$ denotes the *law* (distribution) of X, then

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\,d\mu(x).$$

Proposition 1.2.19 (Basic properties of expectation)**.** *For integrable random variables X, Y and constants $a, b \in \mathbb{R}$:*

(i) ***Linearity:*** $\mathbb{E}[aX+bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$.

(ii) ***Monotonicity:*** *If $X \le Y$ a.s., then $\mathbb{E}[X] \le \mathbb{E}[Y]$.*

(iii) ***Monotone Convergence:*** *If $X_n \uparrow X$, then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$ (by MCT).*

(iv) ***Dominated Convergence:*** *If $X_n \to X$ a.s. and $|X_n| \le Y \in L^1$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ (by DCT).*

Example 1.2.20 (Discrete random variable)**.** Let $\Omega = \{\omega_1, \omega_2, \dots\}$ with $\mathbb{P}(\{\omega_i\}) = p_i$, and $X(\omega_i) = x_i$. Then

$$\mathbb{E}[X] = \sum_i x_i p_i.$$

Thus the expectation reduces to a weighted average of the values x_i.

Example 1.2.21 (Continuous random variable)**.** Let X have density f with respect to Lebesgue measure λ on $\mathbb{R}$. Then for any integrable X,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x)\, dx.$$

For example, if $X \sim \text{Uniform}[0,1]$, then $\mathbb{E}[X] = \int_0^1 x\, dx = \frac{1}{2}$.

Example 1.2.22 (Mixed random variable)**.** Suppose X is 0 with probability 0.5, and otherwise uniformly distributed on $[0,1]$. Then $\mathcal{L}(X)$ is the mixture measure

$$\mathcal{L}(X) = 0.5\, \delta_0 + 0.5\, \lambda|_{[0,1]},$$

and

$$\mathbb{E}[X] = 0.5 \cdot 0 + 0.5 \int_0^1 x\, dx = \tfrac{1}{4}.$$

This shows how the Lebesgue integral unifies discrete, continuous, and mixed cases.

Note

The expectation is simply the Lebesgue integral with respect to the probability measure. - In the discrete case, the Lebesgue integral becomes a countable sum of values weighted by probabilities. - In the continuous case, it becomes an ordinary integral against the density. - In mixed cases, it naturally combines both.
This unified viewpoint is crucial: there is no need for separate definitions of expectation depending on whether a random variable is discrete or continuous. Everything follows from the Lebesgue integral.

Note

The *law* (or *distribution*) of a random variable $X : \Omega \to \mathbb{R}$ is the probability measure $\mathcal{L}(X)$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

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

It tells us the probability of X falling in any Borel set of $\mathbb{R}$.
A common way to represent the law is via the cumulative distribution function (CDF):

$$F_X(x) := \mathbb{P}(X \le x) = \mathcal{L}(X)((-\infty, x]).$$

Examples:

- If X is discrete with $\mathbb{P}(X = 0) = 0.3$, $\mathbb{P}(X = 1) = 0.7$, then $\mathcal{L}(X)$ is the probability measure assigning mass 0.3 to $\{0\}$ and 0.7 to $\{1\}$.
- If X is standard normal, then $\mathcal{L}(X)(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$, and the CDF is $F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$.

In short: the *law* is the full probability measure; the CDF is one way of describing it.

1.3 Probability Spaces

1.3.1 $(\Omega, \mathcal{F}, \mathbb{P})$

Definition 1.3.1 (Probability space)**.** A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where:

- Ω is the *sample space*, the set of all possible outcomes,
- $\mathcal{F}$ is a σ-algebra of subsets of Ω, called *events*,
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a *probability measure* with $\mathbb{P}(\Omega) = 1$.

Note

This framework, due to Kolmogorov, is the foundation of modern probability theory. It is simply a measure space $(\Omega, \mathcal{F}, \mu)$ with total mass normalised to 1. This structure allows us to rigorously define random variables, expectations, and stochastic processes.

Example 1.3.2 (Coin toss)**.** Let $\Omega = \{H, T\}$, $\mathcal{F} = 2^{\Omega}$, and $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a simple probability space for a fair coin.

Example 1.3.3 (Gaussian distribution)**.** Let $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, and

$$\mathbb{P}(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \qquad A \in \mathcal{B}(\mathbb{R}).$$

This defines a probability space for a standard normal random variable.

1.3.2 Random variables

Definition 1.3.4 (Random variable)**.** A *random variable* is a measurable function

$$X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})),$$

that is,

$$\{\omega \in \Omega : X(\omega) \leq a\} = X^{-1}((-\infty, a]) \in \mathcal{F}, \qquad \forall a \in \mathbb{R}.$$

Note

A random variable is not "random" in itself; it is a deterministic mapping on Ω. Randomness arises from the fact that the outcome ω is unknown. Measurability ensures that events of the form $\{X \leq a\}$ are legitimate events in $\mathcal{F}$.

Example 1.3.5 (Discrete random variable)**.** Toss two coins: $\Omega = \{HH, HT, TH, TT\}$, define $X(\omega) =$ number of heads. Then $X : \Omega \to \{0, 1, 2\}$ is measurable, and $\mathbb{P}(X = 1) = \frac{1}{2}$.

Example 1.3.6 (Continuous random variable)**.** Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, and $\mathbb{P} =$ Lebesgue measure on $[0, 1]$. Define $X(\omega) = \omega$. Then X is measurable and has the *Uniform*[0,1] distribution.

Remark 1.3.7. If $\mathbb{E}[|X|] < \infty$, we say that X is an *integrable random variable*. This distinction is important when working with expectations.

1.3.3 Distribution (Law)

Definition 1.3.8 (Law of a random variable)**.** The *law* (or distribution) of a random variable X is the pushforward measure $\mathcal{L}(X)$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$\mathcal{L}(X)(B) := \mathbb{P}(X \in B), \qquad B \in \mathcal{B}(\mathbb{R}).$$

In other words, the law describes the probabilities of subsets of $\mathbb{R}$ induced by X.

Note

The law is a full probability measure on $\mathbb{R}$. Different representations include:

- **CDF:** $F_X(x) = \mathbb{P}(X \le x) = \mathcal{L}(X)((-\infty, x])$,
- **PMF (discrete case):** $\mathbb{P}(X = x_i)$ for atoms x_i,
- **PDF (continuous case):** a density f such that $\mathcal{L}(X)(A) = \int_A f(x)\,dx$.

All of these are just different ways of describing the same measure $\mathcal{L}(X)$.

Example 1.3.9 (Gaussian distribution)**.** If $X \sim N(0,1)$, then

$$\mathcal{L}(X)(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx, \qquad A \in \mathcal{B}(\mathbb{R}).$$

The CDF is

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt.$$

Example 1.3.10 (Discrete law)**.** If X is the outcome of a fair die, then

$$\mathcal{L}(X)(\{k\}) = \tfrac{1}{6}, \qquad k = 1, \dots, 6.$$

Here the law is described by a probability mass function.

1.3.4 Independence

Definition 1.3.11 (Independence of events)**.** Two events $A, B \in \mathcal{F}$ are *independent* if

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

A family $\{A_i\}$ is independent if for any finite subcollection,

$$\mathbb{P}\left(\bigcap_{j=1}^{n} A_{i_j}\right) = \prod_{j=1}^{n} \mathbb{P}(A_{i_j}).$$

Definition 1.3.12 (Independence of random variables)**.** Random variables X, Y are independent if the σ-algebras they generate are independent. Equivalently,

$$\mathbb{P}(X \in A,\ Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B), \qquad \forall A, B \in \mathcal{B}(\mathbb{R}),$$

where $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$.

Remark 1.3.13. More generally, a family $\{X_i : i \in I\}$ of random variables is independent if the σ-algebras $\sigma(X_i)$ are mutually independent, i.e.

$$\mathbb{P}\left(\bigcap_{j=1}^{n} \{X_{i_j} \in B_j\}\right) = \prod_{j=1}^{n} \mathbb{P}(X_{i_j} \in B_j),$$

for all finite choices $i_1, \ldots, i_n$ and Borel sets $B_1, \ldots, B_n$.

Example 1.3.14 (Independent coin tosses)**.** Let X_1, X_2 be outcomes of two independent fair coins (1 for head, 0 for tail). Then

$$\mathbb{P}(X_1 = 1, X_2 = 1) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2}.$$

Example 1.3.15 (Independent Gaussians)**.** If $X, Y \sim N(0, 1)$ are independent, then the joint law is the product measure:

$$\mathcal{L}(X, Y)(A \times B) = \mathcal{L}(X)(A)\, \mathcal{L}(Y)(B),$$

so the joint density factorises as

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \tfrac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

Note

Independence means that knowing one event (or random variable) provides no information about the other. It is stronger than uncorrelatedness: if $X \sim \text{Uniform}[-1, 1]$ and $Y = X^2$, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$, so X, Y are uncorrelated but not independent.

1.3.5 Markov and Chebyshev inequalities

Theorem 1.3.16 (Markov's inequality)**.** *Let $X \geq 0$ be a random variable with $\mathbb{E}[X] < \infty$. Then*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \qquad \forall a > 0.$$

Proof (sketch). For $X \geq 0$,

$$\mathbb{E}[X] \geq \mathbb{E}[\, X \cdot \mathbf{1}_{\{X \geq a\}}\,] \geq a \cdot \mathbb{P}(X \geq a).$$

Dividing by $a > 0$ gives the result. □

Theorem 1.3.17 (Chebyshev's inequality)**.** *Let X be a random variable with mean μ and variance $\sigma^2 < \infty$. Then*

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \qquad \forall \varepsilon > 0.$$

Proof (sketch). Apply Markov's inequality to the nonnegative random variable $(X - \mu)^2$:

$$\mathbb{P}(|X - \mu| \geq \varepsilon) = \mathbb{P}\big((X - \mu)^2 \geq \varepsilon^2\big) \leq \frac{\mathbb{E}[(X - \mu)^2]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}.$$

□

Note

- Markov provides bounds using only the expectation.
- Chebyshev follows directly from Markov applied to $(X-\mu)^2$.
- These inequalities control the probability of large deviations and are key in proving the Weak Law of Large Numbers and in analysing convergence of stochastic processes.

Example 1.3.18 (Application of Markov)**.** If $X \geq 0$ with $\mathbb{E}[X] = 10$, then $\mathbb{P}(X \geq 100) \leq 0.1$. Even without knowing the distribution, we obtain a useful bound.

Example 1.3.19 (Application of Chebyshev)**.** If $\mathbb{E}[X] = 0$ and $\mathrm{Var}(X) = 1$, then

$$\mathbb{P}(|X| \geq 5) \leq \frac{1}{25} = 0.04.$$

This shows X is very unlikely to deviate far from the mean.

1.4 Conditional Expectation

1.4.1 Definition via the Radon-Nikodym theorem

Motivation. Suppose we want to "average" a random variable X given partial information, represented by a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. We want a $\mathcal{G}$-measurable random variable Y that acts like X when tested against $\mathcal{G}$:

$$\int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P}, \quad \forall G \in \mathcal{G}.$$

The question: does such a Y exist, and is it unique? This is answered by the Radon–Nikodym theorem.

Theorem 1.4.1 (Conditional expectation via Radon–Nikodym)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra, and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists a $\mathcal{G}$-measurable random variable Y, unique up to $\mathbb{P}$-a.s. equality, such that*

$$\int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P}, \quad \forall G \in \mathcal{G}.$$

We call Y the conditional expectation of X given $\mathcal{G}$*, written*

$$Y = \mathbb{E}[X \mid \mathcal{G}].$$

Proof outline. Define a set function ν on $\mathcal{G}$ by

$$\nu(G) := \int_G X \, d\mathbb{P}, \quad G \in \mathcal{G}.$$

- ν is a finite signed measure on $(\Omega, \mathcal{G})$. - Moreover, ν is absolutely continuous with respect to $\mathbb{P}$ restricted to $\mathcal{G}$ (since $\mathbb{P}(G) = 0 \implies \nu(G) = 0$). - By the Radon–Nikodym theorem, there exists a $\mathcal{G}$-measurable function Y such that

$$\nu(G) = \int_G Y \, d\mathbb{P}, \quad \forall G \in \mathcal{G}.$$

This Y is unique $\mathbb{P}$-a.s. and is exactly $\mathbb{E}[X \mid \mathcal{G}]$.

Example 1.4.2 (Trivial σ-algebra)**.** If $\mathcal{G} = \{\emptyset, \Omega\}$, then the only $\mathcal{G}$-measurable random variables are constants. Thus $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$, the unconditional expectation.

Example 1.4.3 (Full σ-algebra)**.** If $\mathcal{G} = \mathcal{F}$, then X itself is $\mathcal{G}$-measurable. Hence $\mathbb{E}[X \mid \mathcal{F}] = X$.

Example 1.4.4 (Discrete case: finite partition)**.** Suppose $\mathcal{G} = \sigma(A_1, \ldots, A_n)$, a finite partition of Ω with $\mathbb{P}(A_i) > 0$. Then

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i=1}^{n} \frac{\int_{A_i} X \, d\mathbb{P}}{\mathbb{P}(A_i)} \mathbf{1}_{A_i},$$

i.e. on each A_i, the conditional expectation is the average of X restricted to A_i.

Example 1.4.5 (Conditional expectation as regression)**.** Let X, Y be square-integrable. If $\mathcal{G} = \sigma(Y)$, then $\mathbb{E}[X \mid \mathcal{G}]$ is the L^2-projection of X onto functions of Y, i.e. the unique $\mathcal{G}$-measurable random variable Z minimising $\mathbb{E}[(X - Z)^2]$. For instance, if (X, Y) are jointly Gaussian with means μ_X, μ_Y, variances σ_X^2, σ_Y^2, and correlation ρ, then

$$\mathbb{E}[X \mid Y] = \mu_X + \rho \frac{\sigma_X}{\sigma_Y}(Y - \mu_Y).$$

Note

The Radon–Nikodym theorem tells us that conditional expectation is just the Radon–Nikodym derivative of one measure with respect to another. Formally, $\nu(G) = \int_G X \, d\mathbb{P}$ is a measure on $\mathcal{G}$, absolutely continuous with respect to $\mathbb{P}|_{\mathcal{G}}$. The conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is its Radon–Nikodym derivative. Intuitively, it is the "best guess" of X given the information encoded in $\mathcal{G}$.

1.4.2 Basic properties

Proposition 1.4.6 (Properties of conditional expectation)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra, and let $X, Y \in L^1$. Then:*

1. ***Linearity:*** *For $a, b \in \mathbb{R}$,*
$$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\,\mathbb{E}[X \mid \mathcal{G}] + b\,\mathbb{E}[Y \mid \mathcal{G}].$$
2. ***Monotonicity:*** *If $X \leq Y$ a.s., then*
$$\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}] \quad \text{a.s.}$$
3. ***Taking out what is known:*** *If Z is bounded and $\mathcal{G}$-measurable (so that $XZ \in L^1$), then*
$$\mathbb{E}[XZ \mid \mathcal{G}] = Z\,\mathbb{E}[X \mid \mathcal{G}].$$
4. ***Tower property:*** *If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, then*
$$\mathbb{E}[\,\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}].$$
5. ***Law of total expectation:***
$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X].$$

Proof sketch. All properties follow directly from the defining identity

$$\int_G \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_G X \, d\mathbb{P}, \quad \forall G \in \mathcal{G}.$$

- (i) Apply the identity to $aX + bY$.
- (ii) If $X \leq Y$, then $\int_G X \leq \int_G Y$ for all $G \in \mathcal{G}$, so the same inequality holds for their conditionals.
- (iii) If Z is $\mathcal{G}$-measurable, pull Z out of the inner integral on each G, giving the result.
- (iv) Both sides are $\mathcal{H}$-measurable and agree on all $H \in \mathcal{H}$, hence must coincide.
- (v) Apply (iv) with $\mathcal{H}$ trivial.

□

Example 1.4.7 (Finite partition)**.** Let $\mathcal{G} = \sigma(A_1, \ldots, A_n)$ with $\mathbb{P}(A_i) > 0$. Then

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i=1}^{n} \frac{1}{\mathbb{P}(A_i)} \int_{A_i} X \, d\mathbb{P} \, \mathbf{1}_{A_i}.$$

On each atom A_i, the conditional expectation is just the average of X over A_i.

Example 1.4.8 (Tower property in action)**.** Suppose X is the outcome of a fair die roll. - Let $\mathcal{G} = \sigma(\{\text{odd}, \text{even}\})$ (parity). - Let $\mathcal{H} = \{\emptyset, \Omega\}$ (trivial). Then $\mathbb{E}[X \mid \mathcal{G}] = 3$ on odd outcomes and 4 on even outcomes. Taking expectation again w.r.t. $\mathcal{H}$ gives 3.5, which is just $\mathbb{E}[X]$.

Example 1.4.9 (Independence and "taking out what is known")**.** If X, Y are independent and integrable, then

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X].$$

Indeed, functions of Y are $\sigma(Y)$-measurable, and independence makes $\mathbb{E}[XZ \mid Y] = \mathbb{E}[X] \cdot Z$ for such Z.

Note

These properties make conditional expectation behave like an "ordinary" expectation, but relative to the information in $\mathcal{G}$. - Linearity and monotonicity mirror those of the usual expectation. - "Taking out what is known" says that $\mathcal{G}$-measurable information can be treated like a constant. - The tower property expresses consistency when conditioning step by step. - The law of total expectation shows that conditional expectation is a refinement of expectation.
These rules form the algebraic toolkit for manipulating conditional expectations in probability and stochastic calculus.

1.4.3 Conditional probability

Definition 1.4.10 (Conditional probability via conditional expectation)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$, and $A \in \mathcal{F}$. The *conditional probability of A given $\mathcal{G}$* is the $\mathcal{G}$-measurable random variable

$$\mathbb{P}(A \mid \mathcal{G}) := \mathbb{E}[\, \mathbf{1}_A \mid \mathcal{G}\,],$$

where $\mathbf{1}_A$ is the indicator of A. It is characterised by

$$\int_G \mathbb{P}(A \mid \mathcal{G}) \, d\mathbb{P} = \mathbb{P}(A \cap G), \quad \forall G \in \mathcal{G}.$$

Example 1.4.11 (Trivial σ-algebra)**.** If $\mathcal{G} = \{\emptyset, \Omega\}$, then

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{P}(A).$$

Here the conditional probability reduces to the unconditional one.

Example 1.4.12 (Full σ-algebra)**.** If $\mathcal{G} = \mathcal{F}$, then

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbf{1}_A.$$

Given complete information, the conditional probability is either 0 or 1 depending on whether A occurs.

Example 1.4.13 (Discrete case: die roll)**.** Let X be the outcome of a fair die and $A = \{X = 6\}$. Take $\mathcal{G} = \sigma(\{\text{odd}, \text{even}\})$. Then

$$\mathbb{P}(A \mid \mathcal{G})(\omega) = \begin{cases} 0 & \text{if } X(\omega) \text{ is odd,} \\ 1/3 & \text{if } X(\omega) \text{ is even.} \end{cases}$$

Thus $\mathbb{P}(A|\mathcal{G})$ is a random variable taking different constant values depending on the parity of the outcome.

Example 1.4.14 (Bayes' rule from conditional expectation)**.** Suppose $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$ and let $\mathcal{G} = \sigma(B)$. Then

$$\mathbb{P}(A \mid \mathcal{G}) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \mathbf{1}_B + \frac{\mathbb{P}(A \cap B^c)}{\mathbb{P}(B^c)} \mathbf{1}_{B^c}.$$

Restricting to B gives the familiar formula

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Note

Conditional probability is just conditional expectation of indicators:

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[1_A|\mathcal{G}].$$

- With no information ($\mathcal{G}$ trivial), we recover the usual probability. - With full information ($\mathcal{G} = \mathcal{F}$), the conditional probability is 0 or 1. - With partial information, it becomes a random variable reflecting what is known.
In stochastic processes we often write $\mathbb{P}(A|\mathcal{F}_t)$: the probability of A given the information available up to time t.

1.4.4 Jensen's inequality

Theorem 1.4.15 (Conditional Jensen's inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$, and let $X \in L^1$. If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex with $X, \varphi(X) \in L^1$, then*

$$\varphi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\varphi(X) \mid \mathcal{G}] \quad a.s.$$

Note

If $\mathcal{G}$ is the trivial σ-algebra, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$, so the theorem reduces to the classical inequality

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Proof sketch. For convex φ, one can find an affine function $a+bx$ (supporting hyperplane) such that

$$\varphi(x) \geq a + bx, \quad \forall x \in \mathbb{R}.$$

Taking conditional expectations gives

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \;\geq\; a + b\,\mathbb{E}[X \mid \mathcal{G}].$$

Since this holds for all supporting affine functions of φ, we conclude

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \;\geq\; \varphi\big(\mathbb{E}[X \mid \mathcal{G}]\big).$$

□

Example 1.4.16 (Quadratic convex function)**.** Let $X \in L^2$ and $\varphi(x) = x^2$. Then

$$\big(\mathbb{E}[X \mid \mathcal{G}]\big)^2 \leq \mathbb{E}[X^2 \mid \mathcal{G}],$$

which reduces in the trivial-σ-algebra case to the variance inequality $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$.

Example 1.4.17 (Martingale to submartingale)**.** If (M_t) is a martingale and φ is convex, then by conditional Jensen,

$$\mathbb{E}[\varphi(M_t) \mid \mathcal{F}_s] \geq \varphi(\mathbb{E}[M_t \mid \mathcal{F}_s]) = \varphi(M_s), \quad s \leq t.$$

Hence $(\varphi(M_t))$ is a submartingale — a fundamental result in martingale theory.

Note

Jensen's inequality shows that convex functions "push expectations upwards". The conditional version says the same holds when averaging relative to partial information $\mathcal{G}$. This is central in stochastic processes: convex transforms of martingales are submartingales, underpinning many inequalities and convergence results in stochastic calculus.

1.5 Convergence Concepts

1.5.1 $\mathbb{P}$-almost surely convergence

Definition 1.5.1 ($\mathbb{P}$-almost sure convergence)**.** Let $(X_n)_{n\geq 1}$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and X another random variable. We say that X_n converges to X $\mathbb{P}$-*almost surely* (or *a.s.*) if

$$\mathbb{P}\big(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\big) = 1.$$

We write

$$X_n \xrightarrow{a.s.} X.$$

Example 1.5.2 (Strong Law of Large Numbers)**.** If (X_i) are i.i.d. with $\mathbb{E}[X_1] < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}[X_1].$$

This is almost sure convergence: the sample average converges pointwise for almost every outcome.

Example 1.5.3 (Failure only on a null set)**.** Let $\Omega = [0,1]$ with Lebesgue measure and define $X_n(\omega) = \mathbf{1}_{[0,1/n]}(\omega)$. Then for $\omega > 0$, eventually $X_n(\omega) = 0$, so $X_n(\omega) \to 0$. At $\omega = 0$, we have $X_n(0) = 1$ for all n, so convergence fails. But $\{0\}$ has measure zero, hence $X_n \to 0$ a.s.

Note

Almost sure convergence is the strongest of the common convergence notions: it requires $X_n(\omega) \to X(\omega)$ for "almost every" ω, except possibly on a set of probability zero.
- It is pointwise convergence, but with tolerance for ignoring null sets. - It implies convergence in probability, but not conversely:

$$X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

- Many limit theorems (e.g. the Strong Law of Large Numbers) are formulated in terms of almost sure convergence.

1.5.2 Convergence in probability

Definition 1.5.4 (Convergence in probability)**.** Let (X_n) and X be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X_n converges to X in probability if

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

We write

$$X_n \xrightarrow{P} X.$$

Example 1.5.5 (Vanishing noise)**.** Let $X_n = X + \xi_n$ where $\xi_n \sim \text{Uniform}(-1/n, 1/n)$ are independent noise terms. Then $X_n \xrightarrow{P} X$, since deviations larger than ε become impossible as $n \to \infty$. However, almost sure convergence need not hold.

Example 1.5.6 (Sample averages: Weak Law of Large Numbers)**.** If (X_i) are i.i.d. with $\mathbb{E}[X_1] = \mu < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

This is the Weak Law of Large Numbers.

Note

Convergence in probability requires that the probability of large deviations goes to zero.
- It is *weaker* than almost sure convergence:

$$X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad X_n \xrightarrow{P} X,$$

but not conversely.
- It is *stronger* than weak convergence:

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{d} X.$$

- In practice, convergence in probability is the most useful notion in statistics, because it captures convergence of estimators to the true parameter.

1.5.3 L^p convergence

Definition 1.5.7 (L^p convergence)**.** Let $p \geq 1$. We say that X_n converges to X in L^p if

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

We write

$$X_n \xrightarrow{L^p} X.$$

Example 1.5.8 (Normal variables with shrinking variance)**.** If $X_n \sim N(0, 1/n)$, then $X_n \xrightarrow{L^2} 0$, since

$$\mathbb{E}[X_n^2] = \mathrm{Var}(X_n) = \tfrac{1}{n} \to 0.$$

Example 1.5.9 (Convergence in L^1 but not almost surely)**.** Let $\Omega = [0, 1]$ with Lebesgue measure and define

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

Then $\mathbb{E}[X_n] = 1/n \to 0$, so $X_n \xrightarrow{L^1} 0$. However, $X_n(0) = 1$ for all n, so X_n does not converge to 0 pointwise at $\omega = 0$. This shows that L^p convergence does not imply a.s. convergence.

Note

L^p convergence is a *norm convergence*: the L^p-distance between X_n and X tends to zero.
- It is stronger than convergence in probability:

$$X_n \xrightarrow{L^p} X \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

- It does *not* imply almost sure convergence in general. - It requires integrability: both X_n and X must belong to L^p.
In applications, L^2 convergence is particularly important because it is tied to variance and Hilbert space structure.

1.5.4 Weak convergence

Definition 1.5.10 (Weak convergence / convergence in distribution)**.** We say that X_n converges in distribution (weakly) to X, written $X_n \xrightarrow{d} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \forall x \in \mathbb{R} \text{ where } F_X \text{ is continuous},$$

where F_X is the cumulative distribution function (CDF) of X.

Example 1.5.11 (Normal variables collapsing to a point mass)**.** If $X_n \sim N(0, 1/n)$, then $X_n \xrightarrow{d} 0$, the degenerate distribution at 0 with $F(x) = \mathbf{1}_{\{x\geq 0\}}$.

Example 1.5.12 (Convergence in probability implies weak convergence)**.** If $X_n \xrightarrow{P} X$, then automatically $X_n \xrightarrow{d} X$. For instance, if $X_n = X + \xi_n$ with $\xi_n \sim \mathrm{Uniform}(-1/n, 1/n)$ as before, then $X_n \xrightarrow{P} X$ and hence also $X_n \xrightarrow{d} X$.

Note

Weak convergence is the weakest of the common convergence modes:
- It refers only to the convergence of *distributions*, not to pointwise behaviour of random variables. - It does not require X_n and X to live on the same probability space. - It is implied by convergence in probability:

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{d} X.$$

- It is the natural framework for many asymptotic results in statistics and probability, most famously the Central Limit Theorem.

1.5.5 Law of Large Numbers

Theorem 1.5.13 (Weak Law of Large Numbers (WLLN)). *Let $(X_i)_{i\geq 1}$ be i.i.d. random variables with mean $\mu = \mathbb{E}[X_1] < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

Theorem 1.5.14 (Strong Law of Large Numbers (SLLN)). *Let $(X_i)_{i\geq 1}$ be i.i.d. random variables with mean $\mu = \mathbb{E}[X_1] < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu.$$

Example 1.5.15 (Sample averages of coin tosses). Let X_i be i.i.d. Bernoulli(p). Then the sample average

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is the proportion of heads in n tosses. By the WLLN, $\overline{X}_n \xrightarrow{P} p$. By the SLLN, $\overline{X}_n \xrightarrow{a.s.} p$. Thus the empirical frequency converges to the true probability both in probability and almost surely.

Note

- The **weak law** ensures that averages converge to the mean in probability. - The **strong law** strengthens this to almost sure convergence, i.e. pointwise convergence for almost every outcome.
Both results formalise the idea that empirical averages stabilise around their expected value, justifying the interpretation of $\mathbb{E}[X]$ as the "long-run average" of repeated trials.

1.5.6 Central Limit Theorem

Theorem 1.5.16 (Central Limit Theorem (CLT)). *Let $(X_i)_{i\geq 1}$ be i.i.d. random variables with mean μ and variance $\sigma^2 \in (0, \infty)$. Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Example 1.5.17 (Coin tosses)**.** Let $X_i \sim \text{Bernoulli}(p)$. Then $\mu = p$, $\sigma^2 = p(1-p)$. By the CLT,

$$\sqrt{n}\left(\overline{X}_n - p\right) \xrightarrow{d} N(0,\, p(1-p))\,.$$

Thus, for large n, the empirical proportion of heads is approximately normal with mean p and variance $p(1-p)/n$.

Note

The CLT is the cornerstone of probability and statistics. It says that suitably normalised sums of i.i.d. random variables converge in distribution to a normal random variable, *regardless of the original distribution* (under mild moment conditions). This explains the ubiquity of the Gaussian distribution in applied probability and statistics.

Summary

Convergence concepts compared:

Almost sure (a.s.): $X_n \xrightarrow{a.s.} X$ if $X_n(\omega) \to X(\omega)$ for almost every ω. *Strongest form:* pointwise convergence except on a null set.

In probability: $X_n \xrightarrow{P} X$ if $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ for all $\varepsilon > 0$. *Weaker than a.s., stronger than in distribution.*

L^p: $X_n \xrightarrow{L^p} X$ if $\mathbb{E}[|X_n - X|^p] \to 0$. *Norm convergence, implies convergence in probability.*

Weak (in distribution): $X_n \xrightarrow{d} X$ if $F_{X_n}(x) \to F_X(x)$ at continuity points of F_X. *Weakest form:* distributional convergence only.

1.6 Filtrations and Adapted Processes

1.6.1 Filtration $(\mathcal{F}_t)_{t\geq 0}$

Definition 1.6.1 (Filtration)**.** On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a *filtration* is a family of sub-σ-algebras

$$(\mathcal{F}_t)_{t\geq 0} \subseteq \mathcal{F} \quad \text{such that} \quad \mathcal{F}_s \subseteq \mathcal{F}_t \text{ whenever } s \leq t.$$

We interpret $\mathcal{F}_t$ as the information available up to time t.

Definition 1.6.2 (Usual conditions / augmentation)**.** A filtration $(\mathcal{F}_t)$ satisfies the *usual conditions* if:

1. **Completeness:** Every $\mathbb{P}$-null set in $\mathcal{F}$ belongs to $\mathcal{F}_0$ (and hence to all $\mathcal{F}_t$).
2. **Right-continuity:** $\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$ for all $t \geq 0$.

Given any filtration, one can construct its $\mathbb{P}$-*augmentation* that satisfies these usual conditions.

Example 1.6.3 (Natural filtration of a process)**.** Let $(X_t)_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. Its *natural filtration* is

$$\mathcal{F}_t^X := \sigma\big(X_s : 0 \leq s \leq t\big),$$

the smallest σ-algebra making the path segment $(X_s)_{s\leq t}$ observable.

Example 1.6.4 (Brownian filtration)**.** If $(W_t)_{t\geq 0}$ is a Brownian motion, the canonical choice is the (completed, right-continuous) natural filtration

$$\mathbb{F}^W = (\mathcal{F}_t^W)_{t\geq 0}, \qquad \mathcal{F}_t^W := \sigma(W_s : s \leq t) \text{ augmented to satisfy the usual conditions.}$$

This is the standard setup for martingales and Itô calculus.

Example 1.6.5 (Discrete-time filtration from observations)**.** At times $0 = t_0 < t_1 < \cdots < t_n$, define

$$\mathcal{F}_{t_k} = \sigma(X_{t_0}, \ldots, X_{t_k}).$$

Then $\mathcal{F}_{t_0} \subseteq \cdots \subseteq \mathcal{F}_{t_n}$ captures information from progressively more observations.

Proposition 1.6.6 (Basic facts)**.** *Let $(\mathcal{F}_t)$ be a filtration and $t \mapsto X_t$ a process.*

1. *If Y is $\mathcal{F}_s$-measurable and $s \leq t$, then Y is also $\mathcal{F}_t$-measurable.*
2. *If X is adapted to $(\mathcal{F}_t)$ (see next subsection), then for any Borel g, $g(X_t)$ is $\mathcal{F}_t$-measurable.*
3. *If $(\mathcal{F}_t)$ is right-continuous, then conditioning at t or just after t coincides:*

$$\mathbb{E}[\,\cdot\,|\mathcal{F}_t] = \mathbb{E}[\,\cdot\,|\bigcap_{u>t} \mathcal{F}_u].$$

Summary

Filtration = information growth. $\mathcal{F}_t$ encodes what can be observed by time t. The usual conditions (complete & right-continuous) provide the technical framework needed for martingale theory and stochastic integration.

1.6.2 Adapted processes

Definition 1.6.7 (Adapted process)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $(\mathcal{F}_t)_{t\geq 0}$. A stochastic process $(X_t)_{t\geq 0}$ is said to be *adapted* to $(\mathcal{F}_t)$ if

$$X_t \text{ is } \mathcal{F}_t\text{-measurable for each } t \geq 0.$$

Example 1.6.8 (Natural filtration)**.** If (X_t) is any process, then it is automatically adapted to its *natural filtration*

$$\mathcal{F}_t^X = \sigma(X_s : 0 \leq s \leq t).$$

Example 1.6.9 (Brownian motion)**.** Let (W_t) be a Brownian motion with its completed, right-continuous natural filtration $(\mathcal{F}_t^W)$. Then (W_t) is adapted to $(\mathcal{F}_t^W)$. This is the canonical setup in stochastic calculus.

Example 1.6.10 (Non-adapted process)**.** Suppose (W_t) is a Brownian motion with filtration $(\mathcal{F}_t^W)$. Define $X_t = W_T - W_t$ for some fixed $T > t$. Then X_t depends on the future increment $W_T - W_t$, which is not $\mathcal{F}_t^W$-measurable. Thus (X_t) is *not* adapted to $(\mathcal{F}_t^W)$.

Proposition 1.6.11 (Basic facts)**.** *Let (X_t) and (Y_t) be processes adapted to $(\mathcal{F}_t)$.*

1. *For any Borel function g, the process $(g(X_t))$ is adapted.*
2. *If both (X_t) and (Y_t) are adapted, then so are $(X_t + Y_t)$ and $(X_t Y_t)$.*

3. *If (X_t) is adapted and $(\mathcal{F}_t)$ is right-continuous, then the left-limit process (X_{t-}) is also adapted.*

Summary

Adapted process = no anticipation. At each time t, the value X_t is measurable with respect to the information $\mathcal{F}_t$ available up to that time. Thus, an adapted process cannot "see the future". This non-anticipativity condition is essential for martingale theory and stochastic integration.

1.6.3 Stopping times

Definition 1.6.12 (Stopping time)**.** Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. A random time $\tau : \Omega \to [0, \infty]$ is called a *stopping time* (with respect to $(\mathcal{F}_t)$) if

$$\{\tau \leq t\} \in \mathcal{F}_t, \quad \forall t \geq 0.$$

Note

Intuitively, whether $\tau \leq t$ has occurred must be decidable using only the information available by time t. In other words, a stopping time is a random time that does not anticipate the future.

Example 1.6.13 (Deterministic times)**.** Any fixed time t_0 defines a stopping time $\tau(\omega) = t_0$. Indeed, $\{\tau \leq t\}$ is either $\emptyset$ or Ω, both of which are measurable.

Example 1.6.14 (First hitting time)**.** For a process (X_t) adapted to $(\mathcal{F}_t)$, define

$$\tau = \inf\{t \geq 0 : X_t \geq a\}.$$

Then τ is a stopping time: the event $\{\tau \leq t\}$ depends only on the trajectory $(X_s)_{s\leq t}$, not on the future.

Example 1.6.15 (Non-example)**.** Let $\tau = \inf\{t \geq 0 : X_t \geq a\} - 1$. This is generally *not* a stopping time, because to decide if $\tau \leq t$ one needs to know whether X_{t+1} has crossed the threshold, i.e. information from the future.

Proposition 1.6.16 (Basic properties)**.**

1. *If τ and σ are stopping times, then $\tau \wedge \sigma$ and $\tau \vee \sigma$ are stopping times.*
2. *If τ is a stopping time and $c \geq 0$ is deterministic, then $\tau + c$ is a stopping time* provided the filtration is right-continuous.
3. *If (τ_n) is a sequence of stopping times, then $\sup_n \tau_n$ and $\inf_n \tau_n$ are stopping times.*

Example 1.6.17 (Exit times for Brownian motion)**.** For Brownian motion (W_t), define

$$\tau = \inf\{t \geq 0 : |W_t| \geq 1\}.$$

This is a stopping time, representing the first exit from $(-1, 1)$. Moreover, $\tau < \infty$ almost surely.

Summary

Stopping times = random but observable decision times. They mark the random instants at which a process hits a condition, in a way consistent with available information. Stopping times are essential for martingale theory, optimal stopping problems, and stochastic control.

Chapter 2

Discrete-Time Processes

2.1 Basic Definitions

2.1.1 Stochastic process in discrete time

Note

A stochastic process is just a collection of random variables indexed by time. In the discrete-time setting, time takes integer values (e.g. $n = 0, 1, 2, \dots$). Think of this as observing the random system only at discrete snapshots rather than continuously.

Definition 2.1.1 (Discrete-time stochastic process)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *discrete-time stochastic process* is a sequence of random variables

$$(X_n)_{n\geq 0}, \quad X_n : \Omega \to \mathbb{R}^d,$$

indexed by the non-negative integers. Each X_n represents the state of the system at time n.

Example 2.1.2 (Simple random walk)**.** Let $(\xi_n)_{n\geq 1}$ be i.i.d. random variables with

$$\mathbb{P}(\xi_n = 1) = \mathbb{P}(\xi_n = -1) = \tfrac{1}{2}.$$

Define $X_0 = 0$ and

$$X_n = \sum_{k=1}^{n} \xi_k, \quad n \geq 1.$$

Then $(X_n)_{n\geq 0}$ is called the *simple symmetric random walk*. It models a particle on $\mathbb{Z}$ taking independent ± 1 steps.

Example 2.1.3 (Markov chain on a finite state space)**.** Let $\mathcal{S} = \{1, 2, \dots, m\}$ be a finite set. A process $(X_n)_{n\geq 0}$ taking values in $\mathcal{S}$ is specified by an initial distribution π_0 and a transition matrix $P = (p_{ij})_{i,j\in\mathcal{S}}$, where

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

This is a discrete-time stochastic process with the *Markov property*.

Theorem 2.1.4 (Kolmogorov extension theorem, discrete case)**.** *Given consistent finite-dimensional distributions*

$$\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n), \quad A_i \in \mathcal{B}(\mathbb{R}^d),$$

there exists a discrete-time stochastic process $(X_n)_{n\geq 0}$ realising these distributions.

Note

Finite-dimensional distributions are the joint laws of finitely many coordinates $(X_{n_1}, \ldots, X_{n_k})$. The extension theorem guarantees that specifying these consistently is enough to define a full process.

Example 2.1.5 (IID process)**.** If $X_n \sim$ i.i.d. $\mathcal{N}(0,1)$, then $(X_n)_{n\geq 0}$ is a discrete-time process with independent standard Gaussian increments. This is the discrete analogue of white noise.

Note

Key mental models:

- Think of (X_n) as "random sequences" instead of a single random variable.
- Examples include coin tosses, dice rolls, daily stock returns, or random walks.
- The structure we impose later (filtrations, martingales, Markov property) will tell us how the randomness unfolds over time.

2.1.2 Adaptedness to a filtration

Note

In probability theory, a filtration $(\mathcal{F}_n)_{n\geq 0}$ models the growth of information over time. A process (X_n) is *adapted* if its value at time n can be determined from the information available up to time n. In other words, the process does not "see into the future".

Definition 2.1.6 (Adapted process)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $(\mathcal{F}_n)_{n\geq 0}$. A discrete-time stochastic process $(X_n)_{n\geq 0}$ is said to be *adapted* to $(\mathcal{F}_n)$ if

$$X_n \text{ is } \mathcal{F}_n\text{-measurable for each } n \geq 0.$$

Example 2.1.7 (Natural filtration)**.** Given any process (X_n), its *natural filtration* is

$$\mathcal{F}_n^X := \sigma(X_0, X_1, \ldots, X_n).$$

This is the smallest filtration making (X_n) adapted. Intuitively, it records exactly the information generated by the process up to time n.

Example 2.1.8 (Random walk)**.** Let (X_n) be a simple random walk, $X_n = \sum_{k=1}^n \xi_k$ with i.i.d. steps (ξ_k). If we define $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, then (X_n) is adapted to $(\mathcal{F}_n)$ because X_n is a measurable function of the increments up to time n.

Example 2.1.9 (A non-adapted process)**.** Suppose (W_n) is a random walk with its natural filtration $(\mathcal{F}_n^W)$. Define $Y_n = W_{n+1} - W_n$, i.e. the *future* increment at step n. Then Y_n is not $\mathcal{F}_n^W$-measurable, so (Y_n) is not adapted. This illustrates the "no peeking into the future" rule.

Proposition 2.1.10 (Basic properties)**.** *Let (X_n) and (Y_n) be processes adapted to a filtration $(\mathcal{F}_n)$. Then:*

1. *For any Borel function g, the process $(g(X_n))$ is also adapted.*
2. *Linear combinations and products of adapted processes are adapted.*
3. *If (X_n) is adapted, then $(\max_{k\leq n} X_k)$ is also adapted.*

Note

Adaptedness is a minimal requirement for most constructions in stochastic processes. It ensures that decisions or events at time n depend only on the information available up to n, not on future outcomes. This is crucial for defining martingales, stopping times, and stochastic integrals.

2.1.3 Independence vs. Markov property

Note

Independence and the Markov property are two different ways of describing how random variables relate over time.

- **Independence:** future variables are completely unrelated to the past.
- **Markov property:** the future may depend on the present, but *only* through the most recent state, not the full history.

In practice, independence is stronger, while the Markov property is weaker but often more realistic.

Definition 2.1.11 (Independence)**.** A family of random variables $(X_n)_{n\geq 0}$ is said to be *independent* if, for any finite set of indices $n_1 < \cdots < n_k$,

$$\mathbb{P}(X_{n_1} \in A_1, \ldots, X_{n_k} \in A_k) = \prod_{j=1}^{k} \mathbb{P}(X_{n_j} \in A_j), \quad A_j \in \mathcal{B}(\mathbb{R}).$$

Definition 2.1.12 (Markov property)**.** A process $(X_n)_{n\geq 0}$ with values in a measurable space $(S, \mathcal{S})$ satisfies the *Markov property* with respect to a filtration $(\mathcal{F}_n)$ if

$$\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A \mid X_n), \quad A \in \mathcal{S},\ n \geq 0.$$

In words: given the present state X_n, the future X_{n+1} is conditionally independent of the past $(X_0, \ldots, X_{n-1})$.

Example 2.1.13 (IID sequence)**.** Let (ξ_n) be i.i.d. random variables. Then the process $X_n = \xi_n$ is independent. It also satisfies the Markov property (trivially), since knowing X_n gives no information about X_{n+1}.

Example 2.1.14 (Random walk)**.** Let $X_n = \sum_{k=1}^{n} \xi_k$ with i.i.d. increments (ξ_k). The increments are independent, but the process (X_n) itself is *not* independent: X_{n+1} and X_n are strongly related. However, (X_n) does satisfy the Markov property, because the distribution of X_{n+1} depends only on X_n and not on the full history.

Example 2.1.15 (Dependent but Markov)**.** Consider a two-state Markov chain (X_n) with transition matrix

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix}.$$

Clearly, X_{n+1} depends on X_n, so the sequence is not independent. But it is Markov: the distribution of the next state depends only on the current state.

Note

- Independence $\implies$ Markov property, but not vice versa.
- Independence is rare in time series (most real data have memory).
- The Markov property is the natural compromise: the process has memory, but only of the most recent state.
- Financial models like random walks, binomial trees, and Markov chains rely on this property.

2.2 Markov Chains

2.2.1 Definition and transition matrices

Note

A *Markov chain* is the most fundamental discrete-time model of dependence. It assumes that the next state of the system depends only on the current state, not on the full history. This makes it both mathematically tractable and widely applicable (queues, population models, finance, etc.).

Definition 2.2.1 (Markov chain)**.** Let $(X_n)_{n\geq 0}$ be a stochastic process on a countable state space $\mathcal{S}$. We say (X_n) is a *Markov chain* with respect to a filtration $(\mathcal{F}_n)$ if

$$\mathbb{P}(X_{n+1} = j \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} = j \mid X_n), \quad \forall j \in \mathcal{S},\ n \geq 0.$$

The conditional probabilities

$$p_{ij} := \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad i, j \in \mathcal{S},$$

are called the *transition probabilities*.

Definition 2.2.2 (Transition matrix)**.** For a finite or countable state space $\mathcal{S}$, the transition probabilities can be arranged into a matrix

$$P = (p_{ij})_{i,j\in\mathcal{S}}, \quad p_{ij} \geq 0, \quad \sum_{j\in\mathcal{S}} p_{ij} = 1 \ \ \forall i.$$

P is called the *transition matrix*. It describes the full dynamics of the Markov chain.

Example 2.2.3 (Two-state chain)**.** Let $\mathcal{S} = \{0, 1\}$ with transition matrix

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

If $X_n = 0$, then $X_{n+1} = 0$ with probability 0.7 and $X_{n+1} = 1$ with probability 0.3. Similarly, from state 1, the chain moves to 0 with probability 0.4 and remains at 1 with probability 0.6.

Example 2.2.4 (Random walk on $\mathbb{Z}$)**.** Let (S_n) be a simple symmetric random walk:

$$\mathbb{P}(S_{n+1} = S_n + 1 \mid S_n) = \tfrac{1}{2}, \quad \mathbb{P}(S_{n+1} = S_n - 1 \mid S_n) = \tfrac{1}{2}.$$

Here $\mathcal{S} = \mathbb{Z}$ and the transition matrix is infinite, with $p_{i,i+1} = p_{i,i-1} = \frac{1}{2}$.

Proposition 2.2.5 (Distribution update)**.** *If π_n is the distribution of X_n written as a row vector, then*

$$\pi_{n+1} = \pi_n P.$$

Thus, starting from initial distribution π_0, we have $\pi_n = \pi_0 P^n$.

Note

Key ideas to keep in mind:

- The Markov property means *"memoryless" beyond the present.*
- The transition matrix P encodes all the dynamics.
- Computing future distributions reduces to repeated matrix multiplication.
- For infinite state spaces, P is not a literal matrix but an operator with the same interpretation.

2.2.2 Chapman-Kolmogorov equations

Note

The Chapman-Kolmogorov equations formalise the idea that transitions over multiple steps can be decomposed into successive one-step transitions. They provide the link between short-term and long-term behaviour of a Markov chain.

Theorem 2.2.6 (Chapman-Kolmogorov equations)**.** *Let (X_n) be a Markov chain on state space $\mathcal{S}$ with transition matrix $P = (p_{ij})$. For all $i, j \in \mathcal{S}$ and integers $m, n \geq 0$,*

$$\mathbb{P}(X_{m+n} = j \mid X_0 = i) = \sum_{k \in \mathcal{S}} \mathbb{P}(X_m = k \mid X_0 = i)\, \mathbb{P}(X_n = j \mid X_0 = k).$$

Equivalently, in terms of P,

$$P^{m+n} = P^m P^n.$$

Idea of proof. Condition on the intermediate state X_m:

$$\mathbb{P}(X_{m+n} = j \mid X_0 = i) = \sum_{k \in \mathcal{S}} \mathbb{P}(X_m = k \mid X_0 = i)\, \mathbb{P}(X_{m+n} = j \mid X_m = k).$$

By the Markov property, the second factor reduces to an n-step transition probability from k to j. This gives the formula above. □

Example 2.2.7 (Two-step transitions)**.** Suppose $\mathcal{S} = \{0, 1\}$ with transition matrix

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

Then the two-step transition matrix is

$$P^2 = P \cdot P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}.$$

For example, starting at state 0, the probability of being in state 1 after two steps is 0.39.

Example 2.2.8 (Random walk on $\mathbb{Z}$)**.** For the simple symmetric random walk, the probability of moving from i to j in n steps is

$$p_{ij}^{(n)} = \mathbb{P}(S_n = j \mid S_0 = i) = \binom{n}{\frac{n+j-i}{2}} 2^{-n},$$

whenever $n + j - i$ is even; otherwise $p_{ij}^{(n)} = 0$. These probabilities arise by repeatedly applying the Chapman-Kolmogorov relation.

Note

Key consequences:

- The entire multi-step behaviour of a Markov chain is determined by the one-step matrix P.
- Transition probabilities over n steps are given by the nth power P^n.
- This allows us to study long-run behaviour using matrix analysis (eigenvalues, eigenvectors, convergence).

2.2.3 Classification of states

Note

Not all states of a Markov chain behave the same way. Some states can be left forever (transient), others are revisited infinitely often (recurrent). Among recurrent states, some are periodic, others aperiodic. Classifying states is the first step in understanding the long-run behaviour of a Markov chain.

Definition 2.2.9 (Communicating states)**.** In a Markov chain (X_n) with state space $\mathcal{S}$:

- We say i *leads to* j (written $i \to j$) if there exists $n \geq 0$ such that $p_{ij}^{(n)} > 0$.
- States i and j *communicate* if $i \to j$ and $j \to i$.

Communication is an equivalence relation. The equivalence classes are called *communicating classes.*

Definition 2.2.10 (Irreducibility)**.** A Markov chain is called *irreducible* if all states communicate with each other, i.e. the chain consists of a single communicating class.

Definition 2.2.11 (Recurrence and transience)**.** A state $i \in \mathcal{S}$ is

- *recurrent* if $\mathbb{P}_i(X_n = i \text{ infinitely often}) = 1$, equivalently $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$.
- *transient* if $\mathbb{P}_i(X_n = i \text{ infinitely often}) < 1$, equivalently $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$.

Definition 2.2.12 (Periodicity)**.** The *period* of a state i is

$$d(i) := \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}.$$

If $d(i) = 1$, then i is *aperiodic*; otherwise i is *periodic* with period $d(i)$.

Example 2.2.13 (Random walk on $\mathbb{Z}$)**.** For the simple symmetric random walk on $\mathbb{Z}$:

- All states communicate, so the chain is irreducible.
- Each state is recurrent in dimension 1, but transient in higher dimensions (classical result).
- Each state has period 2, since returns are only possible after an even number of steps.

Example 2.2.14 (Two-state chain)**.** For the chain with transition matrix

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix},$$

the states communicate, so the chain is irreducible. Both states are recurrent (finite irreducible chains have only recurrent states). They are also aperiodic, since $p_{ii} > 0$ implies possible returns in both even and odd numbers of steps.

Proposition 2.2.15 (Finite irreducible chains)**.** *If a Markov chain has a finite state space and is irreducible, then all states are positive recurrent. That is, the expected return time to any state is finite.*

Note

Key insights:

- Transient states may be visited, but eventually are left behind forever.
- Recurrent states are visited infinitely often; if the chain is finite and irreducible, every state is recurrent.
- Periodicity matters for convergence: a chain with period $d > 1$ oscillates between classes of states, while aperiodic chains "mix" smoothly.

These classifications pave the way for the study of *stationary distributions*.

2.2.4 Stationary distributions

Note

A stationary distribution describes the long-run behaviour of a Markov chain. It is a probability distribution on the state space that remains unchanged as the chain evolves. If the chain is irreducible and aperiodic (under mild conditions), it converges to its stationary distribution regardless of the starting state.

Definition 2.2.16 (Stationary distribution)**.** Let (X_n) be a Markov chain with transition matrix P on state space $\mathcal{S}$. A probability vector $\pi = (\pi_i)_{i \in \mathcal{S}}$ is called a *stationary distribution* if

$$\pi = \pi P, \qquad \sum_{i \in \mathcal{S}} \pi_i = 1, \quad \pi_i \geq 0.$$

Equivalently, if $X_0 \sim \pi$, then $X_n \sim \pi$ for all $n \geq 0$.

Example 2.2.17 (Two-state chain)**.** Consider the chain with transition matrix

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

To find the stationary distribution, solve $\pi = \pi P$ with $\pi_0 + \pi_1 = 1$. This gives

$$\pi = \left(\tfrac{4}{7}, \tfrac{3}{7}\right).$$

Thus in the long run, the chain spends about 57% of the time in state 0 and 43% in state 1.

Example 2.2.18 (Random walk on a finite cycle)**.** Let $\mathcal{S} = \{0, 1, \ldots, m-1\}$ and define a symmetric random walk with

$$p_{i,i+1 \,(\mathrm{mod}\ m)} = p_{i,i-1 \,(\mathrm{mod}\ m)} = \tfrac{1}{2}.$$

This chain is irreducible and symmetric. The stationary distribution is uniform:

$$\pi_i = \tfrac{1}{m}, \quad i = 0, \ldots, m-1.$$

Proposition 2.2.19 (Existence and uniqueness)**.** *If a Markov chain is finite, irreducible, and aperiodic, then:*

- *It admits a unique stationary distribution π.*
- *For any initial distribution μ, the distribution of X_n converges to π as $n \to \infty$:*

$$\mu P^n \to \pi.$$

Note

Key insights:

- The stationary distribution is the long-run equilibrium of the chain.
- In finite irreducible aperiodic chains, every trajectory "forgets" its starting point and settles into π.
- For reducible chains, multiple stationary distributions may exist, each supported on a closed communicating class.
- In applications, stationary distributions often describe steady-state behaviour: e.g. long-run market share, equilibrium queue lengths, or genetic distributions.

Summary

- **Definition:** A Markov chain is a discrete-time process where the future depends only on the present, not on the full past. Transition probabilities p_{ij} form the transition matrix P, which encodes the dynamics.
- **Chapman–Kolmogorov equations:** Multi-step transition probabilities are obtained by matrix powers:

$$P^{m+n} = P^m P^n, \quad \pi_n = \pi_0 P^n.$$

- **Classification of states:**
  - States communicate if transitions are possible in both directions.
  - Chains are *irreducible* if all states communicate.
  - States can be *recurrent* (visited infinitely often) or *transient* (eventually abandoned).
  - The period of a state i is $d(i) = \gcd\{n : p_{ii}^{(n)} > 0\}$. Aperiodic states mix smoothly.
- **Stationary distributions:**
  - A stationary distribution π satisfies $\pi = \pi P$.
  - If the chain is finite, irreducible, and aperiodic, then there exists a unique stationary distribution, and

$$\mu P^n \to \pi \quad \text{for any initial distribution } \mu.$$

  - Stationary distributions describe the long-run equilibrium behaviour of the chain.

2.3 Discrete-Time Martingales

2.3.1 Martingales, submartingales, supermartingales

2.3.2 Symmetric random walk

2.3.3 Doob martingale

2.3.4 Properties

2.3.5 Doob's decomposition

2.3.6 Doob's maximal inequality

2.3.7 Optional stopping theorem

2.3.8 Martingale convergence theorem

2.3.9 Strong law of large numbers

Chapter 3

Continuous-Time Processes

3.1 Basic Concepts

3.1.1 Definition of a continuous-time process

3.1.2 Filtration and adaptedness

3.1.3 Right-continuous (càdlàg) sample paths

3.1.4 Kolmogorov continuity theorem

3.2 Poisson Process

3.2.1 Definition and construction

3.2.2 Inter-arrival times

3.2.3 Properties

3.2.4 Distribution

3.2.5 Applications

3.3 Brownian Motion

3.3.1 Definition and properties

3.3.2 Scaling and time-homogeneity

3.3.3 Quadratic variation

3.3.4 Quadratic covariation

3.3.5 Lévy's characterisation

3.3.6 Strong Markov property

3.3.7 Reflection principle

3.4 Continuous-Time Martingales

Chapter 4

Stochastic Calculus

4.1 Motivation

4.1.1 Why ordinary calculus fails for Brownian motion

4.1.2 Itô formula vs. Taylor expansion

4.2 Quadratic Variation and Covariation

4.2.1 Quadratic variation

4.2.2 Quadratic covariation

4.3 Stochastic Integrals

4.3.1 Definition of the Itô integral

4.3.2 Extension to square-integrable processes

4.3.3 Itô isometry

4.3.4 Key properties

4.4 Itô's Lemma

4.4.1 Statement for one-dimensional Brownian motion

4.4.2 Multidimensional Itô's lemma

4.4.3 Examples

4.5 Stochastic Differential Equations

4.5.1 General form

4.5.2 Existence and uniqueness

4.5.3 Weak vs. strong solutions

4.5.4 Examples

Chapter 5

Financial Applications

5.1 Risk-Neutral Valuation

5.1.1 Fundamental theorem of asset pricing

5.1.2 Equivalent martingale measure

5.1.3 Risk-neutral pricing formula

5.1.4 Change of numéraire

5.1.5 Incomplete markets

5.2 Black-Scholes Model

5.2.1 Market assumptions

5.2.2 Black–Scholes PDE

5.2.3 Closed-form option pricing

5.2.4 Greeks

5.2.5 Hedging strategies

5.3 Numerical Methods

5.3.1 Monte Carlo simulation

5.3.2 Variance reduction methods

5.3.3 Binomial and trinomial trees

5.3.4 Finite difference methods

5.4 Exotic Options

5.4.1 Barrier options

5.4.2 Asian options

Bibliography

[1] Krishna B. Athreya and Soumendra N. Lahiri. *Measure Theory and Probability Theory*. Springer, 2006.

Appendix A

Code

Content [1]