People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

Mohamed El Bachir El Ibrahimi University, Bordj Bou-Arreridj
Faculty of Mathematics and Computer Science
Department of Mathematics

Course Notes – Stochastic Processes

Course & Exercises

Prepared by:

Dr.REBIHA ZEGHDANE

Destined for Master 1 students- Mathematical Analysis and Applications (LMD)

First Year, Semester 2

Academic Year: 2024–2025

Contents

Preface

This course on stochastic processes is designed for Master 1 students specializing in mathematics analysis and applications. It offers a structured and rigorous introduction to the mathematical modeling of random phenomena evolving over time, a field that has become essential in probability theory, statistics, and their applications across the sciences and engineering. The text is organized into six chapters: Chapter 1 provides a review of probability essentials and fundamental concepts in probability theory, including probability spaces, random variables, expectations. These form the theoretical foundation for the study of stochastic processes. Chapter 2 offers an overview of stochastic processes as collections of random variables indexed by time. Several motivating examples are presented, illustrating key notions such as random vectors, conditional expectation, stationarity, Filtrations, independence structures and martingales which are an important class of stochastic processes characterized by conditional expectation properties. Chapter 3 introduces Markov chains, focusing on discrete-time Markov chains. Topics include transition probabilities, classification of states, recurrence, ergodic theorems, and convergence to equilibrium distributions. Chapter 4 presents the Poisson process, the canonical model for random counting and the occurrence of events over time. Its construction, properties, and applications are discussed in detail. Chapter 5 is devoted to Brownian motion, developed rigorously as the central continuous-time stochastic process. Topics include its definition, path properties, Markov and martingale characteristics, and its role as a fundamental building block of stochastic calculus.

The primary goal of this course is to equip students with a rigorous theoretical foundation in stochastic processes while fostering the ability to apply these concepts across diverse fields. Particular emphasis is placed on clarity, mathematical rigor, and the use of illustrative examples to highlight key ideas and techniques.

It is my intention that these notes serve not only as a reliable reference but also as an entry point to deeper study and research in probability theory, stochastic analysis, and their broad range of applications.

Dr. Rebiha Zeghdane, June 2025

Introduction

Mathematical analysis is a cornerstone of higher mathematics, providing the rigorous framework upon which many branches of both pure and applied mathematics are built. Its tools and methods are indispensable not only in classical analysis but also in modern areas such as functional analysis, numerical methods, and mathematical modeling. Among the most significant fields where mathematical analysis plays a decisive role is the theory of stochastic processes, which lies at the interface of probability theory, differential equations, and applications in diverse scientific domains. Stochastic processes are fundamental tools in modeling systems that evolve randomly over time. From the unpredictable movement of particles in physics to the modeling of financial markets, these processes offer a rich mathematical framework to describe uncertainty and temporal dynamics. In this course, we concentrate on a set of fundamental topics that form the backbone of stochastic analysis and are essential for introducing more advanced concepts such as stochastic differential equations and stochastic optimal control. These topics include probability spaces, stochastic processes, Markov chains, martingales, the Poisson process, and Brownian motion. The manuscript is structured into five chapters, each building upon the previous ones in order to provide a coherent and progressive understanding of the subject.

Chapter 1 provides a comprehensive review of the essentials of probability theory, covering probability spaces, random variables, expectation. These elements constitute the theoretical foundation upon which the study of stochastic processes is based.

Chapter 2 offers an introduction to stochastic processes, regarded as families of random variables indexed by time. Several motivating examples are presented, highlighting the notions of distributions, stationarity, and dependence structures, conditional expectation and martingales which are fundamental in understanding both discrete and continuous-time models.

Chapter 3 is dedicated to the study of Markov chains, with particular attention to the discrete-time case. The chapter explores transition probabilities, the classification of states, recurrence and transience, ergodic theorems, and convergence to equilibrium distributions. These concepts are central in probability theory and have wide applications in modeling phenomena that evolve in steps.

Chapter 4 presents the Poisson process, which serves as the canonical model for random counting and the occurrence of events over time. The chapter discusses its construction, fundamental properties, and a wide range of applications in queuing theory, telecommunications, and reliability modeling.

Chapter 5 is devoted to Brownian motion, developed rigorously as the prototypical continuous-time stochastic process. Its definition, sample path properties, Markov and martingale characteristics, and its pivotal role in the construction of stochastic calculus are discussed in detail. As a fundamental building block, Brownian motion connects probability theory with analysis, partial differential equations, and mathematical finance.

The primary objective of this course is to establish a comprehensive and rigorous understanding of stochastic processes, providing students with both the theoretical framework and the methodological tools necessary to analyze random phenomena. Beyond the development of abstract concepts, the course aims to cultivate the ability to apply stochastic techniques in diverse domains, including physics, engineering, quantitative finance, and data science. Throughout, particular emphasis is placed on mathematical precision, conceptual clarity, and pedagogical structure, with carefully selected examples and applications designed to illustrate the interplay between theory and practice.

These lecture notes are intended not only to serve as a dependable reference for students during and after the course, but also to act as a foundation for deeper exploration into advanced topics in probability theory, stochastic analysis, and their many applications in contemporary scientific and technological research. It is my hope that this material will stimulate curiosity, encourage independent study, and provide a stepping stone for those interested in pursuing further work or research in this rich and evolving field.

Chapter 1
Probability Essentials

It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge. Probability is simply how likely something is to happen. Whenever we're unsure about the outcome of an event, we can talk about the probabilities of certain outcomes how likely they are. The analysis of events governed by probability is called statistics. Probability theory, a branch of mathematics concerned with the analysis of random phenomena. The outcome of a random event cannot be determined before it occurs, but it may be any one of several possible outcomes. The actual outcome is considered to be determined by chance. Applications probability theory is applied in everyday life in risk assessment and in trade on financial markets. Governments apply probabilistic methods in environmental regulation, where it is called pathway analysis. Another significant application of probability theory in everyday life is reliability. Many consumer products, such as automobiles and consumer electronics, use reliability theory in product design to reduce the probability of failure. Failure probability may influence a manufacturer's decisions on a product's warranty. The range of applications extends beyond games into business decisions, insurance, law, medical tests, and the social sciences The telephone network, call centers, and airline companies with their randomly fluctuating loads could not have been economically designed without probability theory. Some uses of Probability in real Life are as follows:

- Sports – In games such as basketball, football, or cricket, a coin toss is often used to decide which team starts. Each team has an equal 50/50 chance of winning the toss.
- Board Games – In a dice game, the probability of rolling an even number on a fair six-sided die is 50
- Medical Decisions – When patients are advised to undergo surgery, they usually inquire about the success rate of the operation. This rate is essentially a probability, and patients use it to make informed decisions about whether or not to proceed.
- Life Expectancy – Predictions of life expectancy are based on statistical data, such as the average number of years that similar groups of people have lived in the past.
- Weather Forecasting – When planning outdoor activities, people often check the probability of rain. Meteorologists use historical data, temperature patterns, and past natural events to make predictions. These forecasts are never given as certainties but rather as probabilities and approximations.

1.1 Probability theory

Probability theory centers on three main objects: random events, random variables, and stochastic processes. These provide a systematic mathematical framework for studying uncertainty and randomness [1, 2, 8, 9, 11].

1.1.1 Probability triple

We begin by introducing the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which serves as the fundamental framework for modern probability theory.

Definition 1.1 (Probability Triple)**.** *The triple* $(\Omega, \mathcal{F}, P)$ *is a probability triple if:*

- $\mathcal{F}$ *is a* σ*-algebra:*
  - $\Omega \in \mathcal{F}$, $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,
  - $A_n \in \mathcal{F} \Rightarrow \bigcup_n A_n \in \mathcal{F}$.
- P *is* σ*-additive:*
$$P\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} P(A_n),$$
  for disjoint events $A_n \in \mathcal{F}$*, and* $P(\Omega) = 1$, $P(A^c) = 1 - P(A)$.

Examples (σ-algebra)

- $\mathcal{F}_0 = \{\Omega, \emptyset\}$,
- $\mathcal{F}^* =$ all subsets of Ω.
- Let A be a subset of Ω; the σ-algebra generated by A is
$$\mathcal{F}_A = \{\Omega, \emptyset, A, A^c\}.$$
- Let A, B be subsets of Ω; the σ-algebra generated by A and B is
$$\mathcal{F}_{A,B} = \{\Omega, \emptyset, A, A^c, B, B^c, A \cap B, A \cup B, A^c \cap B^c, A^c \cup B^c, A^c \cap B, A^c \cup B, A \cap B^c, A \cup B^c\}.$$
- A finite set of subsets $A_1, A_2, \ldots, A_n$ of Ω which are pairwise disjoint and whose union is Ω is called a **partition** of Ω. It generates the σ-algebra
$$\mathcal{A} = \left\{ A = \bigcup_{j \in J} A_j \; : \; J \subset \{1, \ldots, n\} \right\}.$$
  This σ-algebra has 2^n elements. Every finite σ-algebra is of this form. The smallest nonempty elements $\{A_1, \ldots, A_n\}$ of this algebra are called **atoms**.

Theorem 1.1 (Monotone Convergence)**.** *The measure* P *is* σ*-additive if and only if for all sequences* $\{A_k\}$ *of nondecreasing events and* $A = \bigcup_{k \geq 1} A_k$*,*
$$\lim_{n \to \infty} P(A_n) = P(A).$$

Proof. Suppose that P is σ-additive. Let $\{A_n\}$ be a nondecreasing sequence of events:

$$A_1 \subseteq A_2 \subseteq \cdots, \qquad A = \bigcup_{n=1}^{\infty} A_n.$$

Define the disjoint differences

$$B_1 = A_1, \qquad B_n = A_n \setminus A_{n-1}, \quad n \geq 2.$$

Then

$$A_n = \bigcup_{k=1}^{n} B_k, \qquad A = \bigcup_{k=1}^{\infty} B_k,$$

and the B_k are disjoint. By finite additivity,

$$P(A_n) = \sum_{k=1}^{n} P(B_k).$$

By countable additivity,

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

Hence,

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) = \sum_{k=1}^{\infty} P(B_k) = P(A).$$

Conversely, assume the monotone convergence property holds. Let $\{C_n\}$ be a sequence of disjoint sets, and define

$$C = \bigcup_{n=1}^{\infty} C_n, \qquad D_n = \bigcup_{k=1}^{n} C_k.$$

Then $\{D_n\}$ is an increasing sequence with $\bigcup_{n=1}^{\infty} D_n = C$. By the monotone convergence assumption,

$$P(C) = \lim_{n\to\infty} P(D_n).$$

Since the C_k are disjoint,

$$P(D_n) = \sum_{k=1}^{n} P(C_k).$$

Therefore,

$$P(C) = \lim_{n\to\infty} \sum_{k=1}^{n} P(C_k) = \sum_{k=1}^{\infty} P(C_k).$$

Thus, P is σ-additive. □

1.1.2 Construction of probability measure

Let $(E, \mathcal{O})$ be a topological space, where $\mathcal{O}$ denotes the collection of open sets in E. The σ-algebra $\mathcal{B}(E)$ generated by $\mathcal{O}$ is called the **Borel σ-algebra** of E. Hence, $(E, \mathcal{B}(E))$ forms a measurable space, and any set $B \in \mathcal{B}(E)$ is referred to as a **Borel set**. As a fundamental example, $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is a measurable space, where $\mathcal{B}(\mathbb{R}^n)$ is the σ-algebra generated by the collection of open balls in $\mathbb{R}^n$.

Theorem 1.2 (Carathéodory). *Let $\mathcal{B} = \sigma(\mathcal{A})$, the smallest σ-algebra containing an algebra $\mathcal{A}$ of subsets of E. Let μ_0 be a σ-additive measure on $(E, \mathcal{A})$. Then there exists a unique measure μ on $(E, \mathcal{B})$ which extends μ_0, that is,*

$$\mu(A) = \mu_0(A), \quad \forall A \in \mathcal{A}.$$

Definition 1.2 (Measurable Functions). *A map f from a measure space $(X, \mathcal{A})$ to another measure space $(Y, \mathcal{B})$ is called **measurable** if for every $B \in \mathcal{B}$,*

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

*If $f : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ is measurable, we say f is a **Borel function**.*

Example 1.1. *For example, for $f(x) = x^2$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, one has*

$$f^{-1}([1, 4]) = [1, 2] \cup [-2, -1].$$

*In general, every continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$ is a **Borel function** since the inverse image of open sets in $\mathbb{R}^n$ are open in $\mathbb{R}^n$.*

Definition 1.3. *(Random Variable) A function $X : \Omega \to \mathbb{R}$ is called a **random variable** if it is a measurable map from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, that is,*

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

for all Borel sets $B \in \mathcal{B}(\mathbb{R})$. Every random variable X defines a σ-algebra

$$\mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\},$$

*called the **σ-algebra generated by** X.*

Definition 1.4 (Induced Measure and Distribution). *Let X be a random variable. We define the **induced measure** on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by*

$$\mu(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

*The **distribution function** F is defined by*

$$F(x) = P(X(\omega) \leq x), \quad x \in \mathbb{R}.$$

The function $F : \mathbb{R} \to [0, 1]$ is nondecreasing, right-continuous, with left limits everywhere, and satisfies

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

*Such a function F is called a **distribution function** on $\mathbb{R}$.*

Example 1.2. *(Random Variable and Probability Measure) Let $\Omega = \mathbb{R}$ and let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra. Note that*

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + \tfrac{1}{n}), \quad [a, b] = \bigcap_{n=1}^{\infty} (a - \tfrac{1}{n}, b + \tfrac{1}{n}),$$

thus $\mathcal{B}(\mathbb{R})$ coincides with the σ-algebra generated by semi-closed intervals. Let $\mathcal{A}$ be the algebra of finite disjoint unions of semi-closed intervals $(a_i, b_i]$. Define P_0 on $\mathcal{A}$ by

$$P_0 \left(\bigcup_{k=1}^{n} (a_k, b_k] \right) = \sum_{k=1}^{n} \left(F(b_k) - F(a_k) \right),$$

where F is a distribution function on $\mathbb{R}$. By the Carathéodory Theorem, there exists a measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ extending P_0. Thus, the random variable $X(\omega) = \omega$ on $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is uniquely identified with its distribution function F. For example, if $F(x) = x$ (on $[0, 1]$), then P_0 is the Lebesgue measure dx on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

1.2 Random vectors

In many situations, it is important to examine the joint behavior of several random variables rather than studying them in isolation. Assuming that all variables are defined on the same probability space $(\Omega, \mathcal{F}, P)$, we can construct their joint distribution in a natural way. As in the case of a single random variable, the treatment depends on whether the variables are discrete or continuous.

1. **Joint probability mass function.**

   If the random variables $X_1, \ldots, X_n$ are discrete, their *joint probability mass function* is defined by
   $$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n),$$
   where each x_i ranges over the possible values of the corresponding random variable X_i.

The function $f(x)$ is often referred to simply as the *distribution* of the random vector $X = (X_1, \ldots, X_n)$. It satisfies the same fundamental property as the probability mass function of a single random variable: if C is a subset of the possible values D of X, then

$$P(X \in C) = \sum_{x \in C} f(x).$$

For practical purposes, having introduced the general definition, we shall in the sequel restrict our attention to the case $n = 2$. The extension of definitions and results to higher dimensions is usually straightforward, though notationally cumbersome.

1.2.1 Marginal and joint density function

One very important consequence of the key property above is that we can recover the separate (marginal) distributions of X and Y from a knowledge of their joint distribution $f(x, y)$. The event $\{X = x\}$ can be written as

$$\{X = x\} = \bigcup_{y \in D} \{X = x, Y = y\} = \bigcup_{y \in D} \left(\{X = x\} \cap \{Y = y\}\right),$$

thus we have

$$f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f(x, y).$$

Likewise,

$$f_Y(y) = \sum_x f(x, y).$$

If you visualize $f(x, y)$ as a matrix and place its row sums and column sums at the appropriate sides, you will appreciate why $f_X(x)$ and $f_Y(y)$ are sometimes called *marginal distributions*.

Example 1.3. *(Trinomial distribution)*
Let X and Y have the joint distribution

$$f(x, y) = \frac{n!}{x!y!(n - x - y)!} p^x q^y (1 - p - q)^{n-x-y}.$$

where $x \geq 0$, $y \geq 0$, $0 \leq p, q \leq 1$, and $x + y \leq n$. Then this is called the ***trinomial distribution****, and the marginal is*

$$f_X(x) = \sum_{y=0}^{n-x} f(x,y) = \binom{n}{x} p^x (1-p)^{n-x},$$

which is the Binomial distribution $\mathcal{B}(n,p)$.

Now we turn from the discrete to the continuous case: The pair of random variables (X,Y) is said to be *jointly continuous* with density $f(x,y)$ if, for all a, b, c, d,

$$P(\{a < X < b\} \cap \{c < Y \leq d\}) = \int_a^b \int_c^d f(x,y)\,dy\,dx.$$

In particular, this supplies the joint distribution function of X and Y, denoted by $F(x,y)$, as

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du. \tag{1.1}$$

It follows that, analogously to the case of a single random variable, we may obtain the density $f(x,y)$ from $F(x,y)$ by differentiation:

$$f(x,y) = \begin{cases} \dfrac{\partial^2 F}{\partial x \partial y}, & \text{where the derivative exists,} \\ 0, & \text{elsewhere.} \end{cases}$$

The density has the same key property as in the discrete case: if C is some subset of the domain D of (X,Y),

$$P((X,Y) \in C) = \iint_{(x,y)\in C} f(x,y)\,dx\,dy.$$

Just as in the discrete case, the key property enables us to recover the separate distributions of X and Y from a knowledge of their joint density. We write:

$$P(a < X \leq b) = P(\{a \leq X \leq b\} \cap \{-\infty < Y < \infty\}) = \int_a^b \int_{-\infty}^{\infty} f(x,y)\,dy\,dx = \int_a^b f_X(x)\,dx,$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and similarly} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

are seen to be the **marginal density** of X and Y. Since distributions of random variables supply probabilities, it is not surprising that the concept of independence is important here also.

1.2.2 Independence of random variables

The pair of random variables (X,Y) are said to be *independent* if, for all x and y,

$$P(X \leq x, Y \leq y) = F_{(X,Y)}(x,y) = F_X(x)\,F_Y(y) = P(X \leq x)\,P(Y \leq y).$$

Independence of larger collections of random variables, as well as pairwise independence, may be defined analogously to independence of events.

Equivalently, the above definition holds for random variables of any type, and it follows that X and Y are independent if, for any two sets $A, B \subseteq \mathbb{R}$,

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B).$$

In the special case where X and Y are jointly continuous, differentiating the joint distribution function yields that they are independent if, for all $x, y \in \mathbb{R}$,

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y).$$

Example 1.4. *Let X and Y be independent exponential random variables with parameters λ and μ, respectively. Find $P(X < Y)$.*
Solution. *The joint density is*

$$f(x, y) = \lambda\mu e^{-\lambda x - \mu y}, \quad \textit{for} \quad x > 0,\ y > 0.$$

Hence,

$$P(X < Y) = \iint_{0<x<y} \lambda\mu e^{-\lambda x - \mu y}\, dx\, dy$$

which can be rewritten as

$$P(X < Y) = \int_0^\infty \left(\int_0^y \lambda\mu e^{-\lambda x - \mu y}\, dx \right) dy.$$

Evaluating the integral, we get

$$P(X < Y) = \int_0^\infty \lambda\mu e^{-\mu y} \left(\int_0^y e^{-\lambda x}\, dx \right) dy = \int_0^\infty \lambda\mu e^{-\mu y} \left(\frac{1 - e^{-\lambda y}}{\lambda} \right) dy$$

$$P(X < Y) = \mu \int_0^\infty \left(e^{-\mu y} - e^{-(\lambda+\mu) y} \right) dy.$$

Now integrating term by term, we obtain

$$P(X < Y) = \mu \left(\frac{1}{\mu} - \frac{1}{\lambda + \mu} \right) = 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu}.$$

1.3 Transformations of random variables

Just as with individual random variables, we are often interested in functions of jointly distributed random variables. Let X and Y be random variables with joint density $f_{(X,Y)}(x, y)$. Suppose we define new random variables U and V by

$$U = u(X, Y), \quad V = v(X, Y),$$

where the transformation

$$u = u(x, y), \quad v = v(x, y)$$

is specified by given functions $u(\cdot,\cdot)$ and $v(\cdot,\cdot)$. is one-to-one, and hence invertible as

$$x = x(u, v), \quad y = y(u, v).$$

Assume also that $x(u, v)$ and $y(u, v)$ have continuous derivatives in both arguments and define the Jacobian

$$J(u, v) = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v} \neq 0.$$

Then it is a basic result, using the calculus of several variables, that U and V have joint density

$$f_{U,V}(u, v) = f(x(u, v), y(u, v))\, |J(u, v)|\,. \tag{1.2}$$

Example 1.5. *Let*

$$U = X + Y, \quad V = \frac{X}{X+Y},$$

where X *and* Y *have joint density* $f(x, y)$*. Then*

$$x = uv, \quad y = u(1 - v).$$

Hence,

$$J(u, v) = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}.$$

Computing the derivatives, we obtain

$$\frac{\partial x}{\partial u} = v, \quad \frac{\partial x}{\partial v} = u,$$

$$\frac{\partial y}{\partial u} = 1 - v, \quad \frac{\partial y}{\partial v} = -u,$$

thus

$$J(u, v) = v \times (-u) - (1 - v) \times u = -uv - u(1 - v) = -u(v + 1 - v) = -u.$$

Then, U *and* V *have joint density*

$$f_{U,V}(u, v) = f(uv, u(1 - v))|J(u, v)| = f(uv, u(1 - v))u.$$

Let us note at this point that if we integrate with respect to V*, we obtain the marginal density of the sum* $U = X + Y$*,*

$$f_U(u) = \int f(uv, u(1 - v))u \, dv = \int f(z, u - z)dz.$$

1.4 Caracteristic function

To verify convergence in distribution of random variables $(X_n)_{n\in\mathbb{N}}$, one requires convergence of $\mathbb{E}[f(X_n)]$ for all bounded continuous f. In one dimension, this reduces to checking convergence of cdfs $F_n(x)$ at the continuity points of the limit F.

For random vectors in $(\mathbb{R}^p, \mathcal{B}^p)$, it suffices to restrict to the special class of bounded continuous functions

$$\{\exp(it^\top x) : t \in \mathbb{R}^p\},$$

which characterize the distribution via characteristic functions. Here, i denotes the imaginary unit.

Definition 1.5 (Characteristic Function)**.** *The function*

$$\varphi_X(t) = \mathbb{E}\Big[\exp(it^\top X)\Big]$$

is called the characteristic function (cf) *of* X*.*

The characteristic function may be viewed as the Fourier transform of the density f_X when the latter exists. A key advantage is that every distribution on $\mathbb{R}^p$ admits a characteristic function, even in the absence of moments. Since $\exp(iu) = \cos(u) + i\sin(u)$, it follows that the function $x \mapsto \exp(it^\top x)$ is bounded for each fixed t.

Example 1.6 (Normal distribution). *Let*

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

be the density of X.

Then

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(itx - \frac{x^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\tfrac{1}{2}(x-it)^2 - \tfrac{t^2}{2}\right) dx = \exp\left(-\tfrac{t^2}{2}\right).$$

Example 1.7 (Uniform distribution). *Let $f(x) = \frac{1}{2}$ for $-1 < x < 1$. Then*

$$\varphi(t) = \frac{1}{2} \int_{-1}^{1} \exp(itx)\, dx = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin(t)}{t}.$$

Example 1.8 (Cauchy distribution). *Let*

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

Then

$$\varphi_X(t) = e^{-|t|}.$$

Remark 1.1 (Continuity). *Of course all characteristic functions are continuous by the dominated convergence theorem. Since*

$$\left| e^{it^\top x} - e^{iu^\top x} \right| \leq 2$$

for all t, u, x, we can pass the limit as $u \to t$ under the integral

$$\int \left(e^{it^\top x} - e^{iu^\top x} \right) d\mu_X(x)$$

to get 0 for the limit.

Remark 1.2 (Smoothness). *The smoothness of the characteristic function is related to the existence of moments. Suppose that X is a random variable with finite mean. We can write*

$$|e^{ix} - 1|^2 = |\cos(x) + i\sin(x) - 1|^2 = 2 - 2\cos(x) = 2\int_0^x \sin(t)\, dt \leq 2\int_0^x t\, dt = x^2.$$

This implies that $|e^{ix} - 1| \leq |x|$ for all x. Clearly, $|e^{ix} - 1| \leq 2$ for all x also. So

$$|e^{ix} - 1| \leq \min\{2, |x|\}. \tag{1.3}$$

This implies that $\frac{e^{ixt}-1}{t}$ is bounded by a μ_X-integrable function $|x|$. By the dominated convergence theorem, we can pass the limit as $t \to 0$ under the integral to get that $\varphi'(0)$ exists and equals $i\mathbb{E}[X]$. With a bit more effort, similar results hold if higher moments exist:

$$\varphi^{(k)}(0) = i^k \mathbb{E}[X^k].$$

Some basic properties of characteristic functions are summarized below.

Proposition 1.1 (Basic properties of cf)**.** *All characteristic functions φ have the following properties:*

1. $\varphi(0) = 1, \quad |\varphi(t)| \leq 1,$
2. $\varphi(-t) = \overline{\varphi(t)}$ *(complex conjugate),*
3. $|\varphi(t+h) - \varphi(t)| \leq \mathbb{E}[|e^{ihX} - 1|]$ *(uniform continuity),*
4. $\varphi_{aX+b}(t) = e^{itb}\,\varphi_X(at)$.

The next result gives a sufficient condition for $\varphi(t)$ to be a characteristic function.

Theorem 1.3 (Pólya's Criterion)**.** *Let φ be continuous, real, nonnegative, symmetric, decreasing and convex on $[0, \infty)$, such that $\varphi(0) = 1$ and $\lim_{t\to\infty} \varphi(t) = 0$. Then φ is a characteristic function.*

Proposition 1.2 (Characteristic function of sum of independent r.v.'s)**.** *If X and Y are independent, then*

$$\varphi_{X+Y}(t) = \varphi_X(t)\,\varphi_Y(t).$$

The main remaining results concerning convergence in distribution are as follows:

- **Inversion (uniqueness) theorem:** every characteristic function uniquely determines a distribution.
- **Continuity theorem:** $X_n \xrightarrow{D} X$ if and only if $\varphi_{X_n}(t) \to \varphi_X(t)$ for all $t \in \mathbb{R}^p$ (the "only if" direction being immediate).
- **Central limit theorem:** suitably normalized sums of independent (not necessarily identically distributed) random variables with finite variance converge in distribution to a standard normal random variable.

Lemma 1.1 (Cramér–Wold)**.** *Let X and Y be $\mathbb{R}^p$-valued random vectors. Then X and Y have the same distribution if and only if $\alpha^\top X$ and $\alpha^\top Y$ have the same distribution for every $\alpha \in \mathbb{R}^p$.*

Proof. By definition, X and Y have the same distribution if and only if their characteristic functions agree, i.e.,

$$\varphi_X(t) = \varphi_Y(t), \quad \forall\, t \in \mathbb{R}^p.$$

This condition holds if and only if

$$\varphi_X(s\alpha) = \varphi_Y(s\alpha), \quad \forall\, \alpha \in \mathbb{R}^p,\ \forall\, s \in \mathbb{R}.$$

Now observe that $\varphi_X(s\alpha)$ is exactly the characteristic function of the scalar random variable $\alpha^\top X$ (as a function of s), and similarly $\varphi_Y(s\alpha)$ is the characteristic function of $\alpha^\top Y$.

Hence, $\varphi_X(s\alpha) = \varphi_Y(s\alpha)$ for all $\alpha \in \mathbb{R}^p$ and all $s \in \mathbb{R}$ if and only if $\alpha^\top X$ and $\alpha^\top Y$ have the same distribution for every $\alpha \in \mathbb{R}^p$. □

Proposition 1.3. *If φ is the characteristic function of the cdf F on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and if φ is integrable, then F has a density*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\, \varphi(t)\, dt, \tag{1.4}$$

which is continuous.

The link between characteristic functions and convergence in distribution is summarized in the following result.

Theorem 1.4 (Continuity theorem)**.** *Let $\{P_n\}_{n=1}^{\infty}$ be a sequence of probability measures on $(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p})$, and let P be another probability measure. Let φ_n be the characteristic function of P_n, and let φ be the characteristic function of P. Then*

$$P_n \xrightarrow{D} P \quad \text{if and only if} \quad \lim_{n\to\infty} \varphi_n(t) = \varphi(t) \quad \text{for all } t \in \mathbb{R}^p.$$

Example 1.9. *For each j, let Y_j have a uniform distribution on the interval $[-1, 1]$ and let*

$$X_n = \sqrt{\frac{3}{n}} \sum_{j=1}^{n} Y_j.$$

Then the characteristic function of X_n is

$$\varphi_n(t) = \left(\frac{\sin(t\sqrt{\frac{3}{n}})}{t\sqrt{\frac{3}{n}}} \right)^n .$$

We can write $\sin(t) = t - t^3/6 + o(t^3)$ *so that, for each t,*

$$\frac{\sin(t\sqrt{\frac{3}{n}})}{t\sqrt{\frac{3}{n}}} = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \quad \text{as } n \to \infty.$$

It follows easily that $\lim_{n\to\infty} \varphi_n(t) = e^{-t^2/2}$. *This is the characteristic function of the standard normal distribution.*

The following two results are useful in proving convergence in distribution.

Corollary 1.1. *If* $\lim_{n\to\infty} \varphi_n(t)$ *exists for all t and is continuous at* 0, *then the limit is a characteristic function, and the distributions converge to the distribution with that characteristic function.*

Corollary 1.2 (Cramér–Wold Theorem)**.** *If $\{X_n\}_{n=1}^{\infty}$ is a sequence of p-dimensional random vectors and X is a random vector, then*

$$X_n \xrightarrow{D} X \quad \text{if and only if} \quad \alpha^\top X_n \xrightarrow{D} \alpha^\top X \quad \text{for all } \alpha \in \mathbb{R}^p.$$

1.5 Expectation

In this section, we define the expectation of a random variable X on $(\Omega, \mathcal{F}, P)$.

Definition 1.6. *A simple random variable X is defined by*

$$X(\omega) = \sum_i x_k \mathbb{I}_{A_k(\omega)},$$

where $\{A_k\}$ is a partition of Ω, i.e., $A_k \in \mathcal{F}$ are disjoint and $\bigcup_k A_k = \Omega$. Then, the expectation of X is given by

$$E[X] = \sum_k x_k P(A_k).$$

Theorem 1.5. *(Approximation): For every random variable $X(\omega) \geq 0$, there exists a sequence of simple random variables $\{X_n\}$ such that $0 \leq X_n(\omega) \leq X(\omega)$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$.*

Proof: For $n \geq 1$, define a sequence of simple random variables by

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{I}_{I_{k,n}(\omega)} + n\mathbb{I}_{\{X(\omega)>n\}},$$

where $I_{k,n} = \left\{\omega : \frac{k-1}{2^n} < X(\omega) \leq \frac{k}{2^n}\right\}$ is the indicator function of the set. It is easy to verify that $X_n(\omega)$ is monotonically non-decreasing, $X_n(\omega) \leq X(\omega)$, and thus $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$.

Definition 1.7. *(Expectation): For a nonnegative random variable X, we define the expectation by*

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

where the limit exists since $E[X_n]$ is an increasing sequence of numbers.

Note that $X = X^+ - X^-$ with $X^+(\omega) = \max(0, X(\omega))$ and $X^-(\omega) = \max(0, -X(\omega))$. Therefore, we can apply the Theorem and Definition to X^+ and X^-. We then have:

$$E[X] = E[X^+] - E[X^-].$$

If $E[X^+], E[X^-] < \infty$, then X is integrable and

$$E[|X|] = E[X^+] + E[X^-].$$

Corollary 1.3. *Let μ_X be the induced distribution of a random variable X, i.e.,*

$$\mu_X(x) = P(\{X(\omega) \leq x\}).$$

Then, for a Borel function $f : \mathbb{R} \to \mathbb{R}$, we have as $n \to \infty$,

$$E[f(X_n)] = \sum_{k=1}^{n2^n} f\left(\frac{k-1}{2^n}\right)\left(\mu_X\left(\frac{k}{2^n}\right) - \mu_X\left(\frac{k-1}{2^n}\right)\right) + f(n)\left(1 - \mu_X(n)\right) \to \int_0^1 f(x)\, d\mu_X(x).$$

which is the Lebesgue-Stieltjes integral with respect to measure $d\mu_X$. Thus,

$$E[f(X)] = \int_\Omega f(X(\omega))P(d\omega) = \int_{-\infty}^{\infty} f(x)\, d\mu_X(x).$$

Definition 1.8. *(Absolutely continuous): The measure Q is absolutely continuous with respect to the measure P if*

$$Q(A) = 0 \text{ if } P(A) = 0 \quad \text{for all } A \in \mathcal{F}.$$

Theorem 1.6. *(Radon-Nykodým derivative): If Q is absolutely continuous with respect to the measure P, then there exists a nonnegative random variable f such that*

$$\int_A dQ(\omega) = \int_A f(\omega) dP(\omega)$$

for all $A \in \mathcal{F}$, i.e.,

$$\frac{dQ}{dP}(\omega) = f(\omega) \text{ for almost every } \omega.$$

If μ_X is absolutely continuous with respect to the Lebesgue measure dx, then there exists a probability density function $p(x) \geq 0$ of X, i.e.,

$$\frac{d\mu_X}{dx} = p(x) \quad \text{(Radon-Nykodým derivative)},$$

and

$$E[f(X)] = \int_\Omega f(X(\omega))P(d\omega) = \int_{-\infty}^{\infty} f(x)p(x)\, dx.$$

Theorem 1.7. *(Independent Random Variables): For independent random variables X and Y and Borel functions f and g, if $f(X)$ and $g(Y)$ are integrable, then the product $f(X)g(Y)$ is integrable and*

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

Proof: If X and Y are simple random variables, i.e.,

$$X = \sum_k x_k \mathbb{I}_{A_k}(\omega), \quad Y = \sum_j y_j \mathbb{I}_{B_j}(\omega),$$

where $\{A_k\}$ and $\{B_j\}$ are partitions of $(\Omega, \mathcal{F})$, then

$$E[f(X)g(Y)] = \sum_k f(x_k) \sum_j g(y_j) P(X^{-1}(A_k) \cap (Y^{-1}(B_j)).$$

Since $P(X^{-1}(A_k) \cap Y^{-1}(B_j)) = P(X^{-1}(A_k)P(Y^{-1}(B_j))$, we have

$$E[f(X)g(Y)] = \left(\sum_k f(x_k)P(X^{-1}(A_k))\right)\left(\sum_j g(y_j)P(Y^{-1}(B_j))\right) = E[f(X)]E[g(Y)].$$

For the general case, we approximate X and Y by a sequence of simple random variables by the approximation.

Theorem 1.8. *The variables X snd Y are independent if and only the Borel functions f and g are independent and we have $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$.*

Theorem 1.9. *(Convolution): If X and Y are independent random variables, and let $F(x) = P(X \leq x)$, and $G(y) = P(Y \leq y)$, then*

$$P(X + Y \leq z) = \int_{-\infty}^{z} F(z - y)\, dG(y).$$

The integral on the right-hand side is called the convolution of F and G and is denoted by $(F * G)(z)$.

Proof: Let $h(x,y) = \mathbb{I}_{x+y\le z}$. Let μ and ν be the probability measures with distribution functions F and G, respectively. Since for fixed y,

$$\int h(x,y)\,\mu(dx) = \int I_{(-\infty,z-y]}(x)\mu(dx) = F(z-y),$$

Thus,

$$P(X+Y\le z) = \int \mathbb{I}_{x+y\le z}\,\mu(dx)\,\nu(dy) = \int F(z-y)\,\nu(dy) = \int F(z-y)\,dG(y).$$

For a random vector $X(\omega) = (X_1(\omega),\ldots,X_n(\omega))$ in $\mathbb{R}^n$, we define the mean m and variance matrix R by

$$m = E[X] = \int_{\mathbb{R}^n} x\,d\mu_X(x),$$

and

$$R_{ij} = \operatorname{Var}(X) = E[(X_i-m_i)(X_j-m_j)] = \int_{\mathbb{R}^n} (x_i-m_i)(x_j-m_j)\,d\mu_X(x).$$

The variance matrix R is symmetric and nonnegative definite since

$$\xi^T R\xi = E\left[\Big|\sum_k \xi_k(X_k-m_k)\Big|^2\right] \ge 0$$

for all $\xi\in\mathbb{R}^n$. For $m\in\mathbb{R}^n$ and R a symmetric positive definite matrix on $\mathbb{R}^n$, and

$$p(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(R)}}\exp\left(-\frac{1}{2}(x-m)^T R^{-1}(x-m)\right),$$

is the probability density function of a random vector X, where $\det(R)$ is the determinant of R, i.e.,

$$\mu(B) = \int_B p(x)\,dx, \quad B\in\mathcal{B}(\mathbb{R}^n),$$

defines the Gaussian distribution on $\mathbb{R}^n$. Thus, a Gaussian random variable is completely determined by its statistics (m,R).
If X and Y are discrete random variables, we have

$$E[g(X,Y)] = \sum g(x,y)f(x,y),$$

and if X and Y are jointly continuous, so

$$E[g(X,Y)] = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)f(x,y)\,dx\,dy,$$

provided that the sum and the integral converge absolutely. These results have many important consequences; we highlight the following:

- **Linearity of E.** If $E[X]$ and $E[Y]$ exist, then for constants a and b,

$$E[aX+bY] = aE[X]+bE[Y].$$

- $$X\le Y \Longrightarrow EX\le EY.$$

1.5.1 Covariance and correlation

The covariance of X and Y is

$$\operatorname{cov}(X,Y) = E[(X - E[X])(Y - E[Y])].$$

The correlation coefficient (or correlation) is

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}}.$$

It is an exercise for you to show that if X and Y are independent, then $\rho(X,Y) = 0$. The converse is not true in general.

Definition 1.9. *(Square Integrable Random Variables:) Define*

$$L^2(\Omega; P) = \Big\{ \textit{The space of square integrable random variables, } E[|X|^2] < \infty. \Big\}$$

The space $L^2(\Omega; P)$ is a vector space. For $\alpha, \beta \in \mathbb{R}$, we have

$$E[|\alpha X + \beta Y|^2] \le 2(\alpha^2 E[|X|^2] + \beta^2 E[|Y|^2]),$$

since

$$|E[XY]| \le \sqrt{E[|X|^2]} \cdot \sqrt{E[|Y|^2]}.$$

In fact, if we define the inner product by

$$(X,Y)_{L^2} = E[XY],$$

then $L^2(\Omega; P)$ is a complete inner product space, i.e., a Hilbert space.

1.6 Exercises

Exercise 1.1. *Two people agree to meet between 2:00 p.m. and 3:00 p.m., with the condition that neither of them will wait more than 15 minutes. What is the probability that they actually meet? We model the arrival time of each person as a random variable uniformly distributed over the interval $[0,1]$, where the origin of time corresponds to 2:00 p.m. We also assume that these two random variables are independent.*

Solution. *Let X denote the arrival time of the first person and Y the arrival time of the second person. Both random variables are independent and uniformly distributed over $[0,1]$. Define the random variable Z as the absolute difference between the two arrival times:*

$$Z = |X - Y|.$$

Clearly, Z takes values in $[0,1]$. Since X and Y are independent, their joint density is given by

$$f_{X,Y}(x,y) = \mathbf{1}_{[0,1]}(x)\,\mathbf{1}_{[0,1]}(y).$$

For $z \in [0,1]$, we compute

$$P(Z \le z) = P(|X - Y| \le z) = \iint_{\{(x,y)\in\mathbb{R}^2 \,:\, |x-y|\le z\}} \mathbf{1}_{[0,1]\times[0,1]}(x,y)\,dx\,dy = 1 - (1-z)^2.$$

Exercise 1.2. 1. *Show that* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

2. *Show that if* X *is a random variable, then* $\{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ *is a* σ*-algebra.*

3. *Show that* X *is a random variable if and only if* $\{\omega : X(\omega) \leq a\} \in \mathcal{F}$ *for all* $a \in \mathbb{R}$.

4. *Show that a distribution function has at most countably many discontinuities.*

5. *Show that if* f *is a Borel function and* X *is a random variable, then* $f(X)$ *is a random variable.*

Chapter 2
Introduction to stochastic processes

When working with measured data, it is common to view it as a realization of a *stochastic process*. Before proceeding, it is important to understand what a stochastic process is and why it is useful. Why introduce a statistical model for data? One may ask: why model the data statistically instead of working directly with the observed values? This question deserves careful consideration, and we offer two main reasons:

1. **Measurement Noise and Uncertainty:** Measured data are often affected by various disturbances errors from measuring devices, environmental influences, and other noise sources. Accurately modeling all such disturbances is usually infeasible. However, by assuming the data are realizations of a stochastic process with known properties, we can design algorithms that exploit these statistical characteristics without needing a full deterministic description.

2. **Simplified Representation:** Even when the underlying signal is well understood, it is often advantageous to model it as a realization of a process characterized by a few parameters. This way, we estimate a small set of parameters rather than a large number of unknowns. Though this may introduce approximation error, it is often preferable to attempting infeasible exact recovery from limited data.

Examples of stochastic processes include:

- **Binomial process:** sequences of “heads” and “tails” from repeated coin tosses;
- **Brownian motion:** the random movement of a particle suspended in a fluid;
- **Poisson process:** counting random events such as arrivals in a queue;
- **Epidemic spread:** tracking the number of infected individuals over time;
- **Meteorological processes:** measuring daily sunshine hours at a location;
- **Financial processes:** observing stock price fluctuations over days.

The apparent randomness in such processes typically originates from the combined effect of numerous unpredictable “agents.” For example, although the microscopic dynamics governing molecular collisions in Brownian motion are well understood, it is practically impossible to trace every single collision, thereby necessitating a stochastic description. Likewise, in financial markets, predicting price movements directly from the mental states of traders is unfeasible, and thus a probabilistic framework becomes indispensable. Although randomness is intrinsic at the microscopic level, the collective or macroscopic behavior of a stochastic process often exhibits simple and reproducible statistical patterns, such as the mean and variance, when observed over appropriate time scales. The theory of stochastic processes [7,11–13,16,24] seeks to formalize these macroscopic laws of evolution by substituting detailed microscopic mechanisms with well-chosen probabilistic assumptions.

2.1 Conditional expectation

This section is devoted to the introduction of the concept of *conditional expectation*, a fundamental notion in probability theory and stochastic processes.

2.1.1 Conditioning

When dealing with a random vector, it may happen that the value of one or more of its components is known, or that certain constraints are imposed on their values. Alternatively, we may wish to investigate how such information affects the distribution of the remaining components. This naturally leads us to the notion of *conditional distributions*. As usual, we begin with the simplest case.

In the discrete setting, the situation is particularly straightforward.

(1) Conditional Probability Mass Function. If X and Y are jointly discrete random variables, the conditional probability mass function of X given $Y = y$ is defined, for all y such that $f_Y(y) > 0$, by

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

In the continuous case, the conditional density of X given $Y = y$ is defined, for all y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Similarly, the conditional distribution of Y given X is

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad \text{when } f_X(x) > 0.$$

Note that the conditional distribution $f_{X|Y}(x|y)$ is indeed a probability distribution in the sense that it is nonnegative and sums to unity:

$$\sum_x f_{X|Y}(x|y) = \sum_x \frac{f(x, y)}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1.$$

It can therefore have an expectation as defined in the previous section; we now introduce this important concept.

(2) Conditional Expectation. The conditional expectation of X given that $Y = y$ is denoted $E(X \mid Y = y)$ and defined as

$$E(X \mid Y = y) = \sum_x x f_{X|Y}(x|y),$$

when the sum is absolutely convergent. Any other conditional mean is defined in a similar and obvious way.

Just as the conditional distribution is in fact a probability distribution, the conditional expectation possesses all the properties of the ordinary expectation, such as:

$$E(X+Y \mid Z=z) = E(X \mid Z=z) + E(Y \mid Z=z),$$

as well as conditional moments and conditional correlation. The case where X and Y are jointly continuous is not as elementary or straightforward. We can proceed by analogy:

(3) Conditional Density and Expectation. If X and Y are jointly continuous, then the conditional density of X given Y is defined by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad \text{when } f_Y(y) > 0.$$

The conditional expectation of X given $Y = y$ is

$$E(X \mid Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx,$$

provided that the integral converges absolutely. Conditional moments are defined in a similarly obvious manner.

The conditional distribution function of X given $Y = y$ is

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(u|y)\,du.$$

However, note that $f_{X|Y}(x|y)$ is indeed a probability density, in the sense that it is nonnegative and that

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,dx = 1.$$

Moreover, it shares all the other useful properties of the conditional mass function, such as:

(4) Partition Lemma. Let X and Y be jointly distributed. (a) If X and Y are discrete, then

$$f_X(x) = \sum_y f_{X|Y}(x|y) f_Y(y).$$

If X and Y are continuous, then

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\,dy.$$

Throughout, we work on a fixed probability space $(\Omega, \mathcal{F}, P)$, and all random variables considered in this course are assumed to be integrable. Unless stated otherwise, inequalities are always understood in the almost-sure sense.

Let $\mathcal{G}$ denote a sub-σ-algebra of $\mathcal{F}$; that is, a σ-algebra of subsets of Ω satisfying $\mathcal{G} \subseteq \mathcal{F}$. Moreover, let $X : \Omega \to \mathbb{R}$ be an integrable random variable.

Definition 2.1 (Conditional Expectation w.r.t. a Sub-σ-Algebra)**.** *A conditional expectation of X with respect to $\mathcal{G}$ is a random variable Y such that:*

1. *Y is $\mathcal{G}$-measurable;*

2. For any $A \in \mathcal{G}$, we have

$$\mathbb{E}[1_A X] = \mathbb{E}[1_A Y].$$

While for integrable random variables the conditional expectation always exists, it might be nonunique, i.e., more than one random variable may satisfy the properties from Definition 2.1. Nevertheless, it is defined uniquely up to a set of measure zero.

For clarity, we denote the conditional expectation by $\mathbb{E}[X \mid \mathcal{G}]$. It should be kept in mind that we are working in the space $L^1(\Omega, \mathcal{F}, \mathbb{P})$ of equivalence classes of integrable random variables, rather than on the set of all random variables. Next, we define a conditional expectation of X with respect to another random variable Y.

Definition 2.2 (Conditional Expectation w.r.t. a Random Variable)**.** *Let X and Y be random variables. We denote by $\mathbb{E}[X \mid Y]$ the conditional expectation of X with respect to Y, defined as*

$$\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)],$$

where $\sigma(Y)$ is the σ-algebra generated by Y.

From Definition 2.1 we know that $\mathbb{E}[X \mid Y]$ is $\sigma(Y)$-measurable. Thus, from the Doob–Dynkin lemma it follows that

$$\mathbb{E}[X \mid Y] = g(Y),$$

for some Borel measurable function $g : \mathbb{R} \to \mathbb{R}$.

From now on, we use the convention

$$\mathbb{E}[X \mid Y = y] := g(y), \quad y \in \mathbb{R}.$$

Before we outline basic properties of conditional expectation, let us present the definition of conditional expectations.

Theorem 2.1. *Let $\mathcal{G} = \sigma(\{A_1, A_2, \ldots, A_n\})$, where $\{A_i\}_{i=1}^n$ is a finite partition of Ω. Then, for any integrable random variable X we get the representation*

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \begin{cases} 0, & \text{if } \mathbb{P}[A_i] = 0, \\ \dfrac{\mathbb{E}[1_{A_i} X]}{\mathbb{P}[A_i]}, & \text{if } \omega \in A_i \text{ with } \mathbb{P}[A_i] > 0, \end{cases}$$

Note that in in the previous equation we can substitute 0 with any real number, for any $i \in \{1, \ldots, n\}$ such that $\mathbb{P}[A_i] = 0$. Also, using the convention $0/0 = 0$, we can rewrite

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i=1}^{n} \frac{\mathbb{E}[1_{A_i} X]}{\mathbb{P}[A_i]} 1_{A_i}.$$

Proof. We divide the proof into a few steps.

Step 1. First of all, let us show that

$$\mathcal{G} = \Big\{ \bigcup_{i \in I} A_i : I \subseteq \{1, 2, \ldots, n\} \Big\}. \tag{2.1}$$

Let $\widetilde{\mathcal{G}}$ denote the right-hand side of equation (2.1). Because $\widetilde{\mathcal{G}}$ contains $\{A_1, \ldots, A_n\}$ we get that $\mathcal{G} \subseteq \widetilde{\mathcal{G}}$. On the other hand, for any $I \subseteq \{1, \ldots, n\}$ we get $\bigcup_{i \in I} A_i \in \mathcal{G}$, so $\widetilde{\mathcal{G}} \subseteq \mathcal{G}$. Thusn we obtain the results.

Step 2. Next, we show that a random variable Y is $\mathcal{G}$-measurable if and only if Y is constant on each A_i.

Suppose, for the sake of contradiction, that there exists Y and some i_0 such that Y is $\mathcal{G}$-measurable but not constant on A_{i_0}. Then, there exists $x \in \mathbb{R}$ such that for $B = (-\infty, x]$ we have

$$Y^{-1}(B) \cap A_{i_0} \neq \varnothing \quad \text{and} \quad Y^{-1}(B^c) \cap A_{i_0} \neq \varnothing.$$

This contradicts yhe definition of $\mathcal{G}$, because $Y^{-1}(B) \cap A_{i_0}$ is a (proper) subset of A_{i_0}, which belongs to $\mathcal{G}$.

Step 3. Finally, from the previous steps we know that

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i=1}^{n} a_i 1_{A_i},$$

for some sequence of real numbers $a_1, \ldots, a_n$. Now, using Definition 2.1 and noting that for $i \neq j$ we have $1_{A_i} 1_{A_j} = 1_{A_i \cap A_j} = 0$, we get

$$\mathbb{E}[1_{A_i} X] = \mathbb{E}\left[1_{A_i} \sum_{j=1}^{n} a_j 1_{A_j}\right] = \mathbb{E}[1_{A_i} a_i] = \mathbb{P}[A_i]\, a_i,$$

which concludes the proof. □

Theorem 2.2. *Let Y be a random variable that takes finitely many values $\{y_1, \ldots, y_n\}$, all with positive probability. Then, for any integrable random variable X we get*

$$\mathbb{E}[X \mid Y] = \sum_{i=1}^{n} \frac{\mathbb{E}\left[1_{\{Y=y_i\}} X\right]}{\mathbb{P}[Y = y_i]} 1_{\{Y=y_i\}}. \tag{2.2}$$

Alternatively, we can rewrite (2.2) as

$$\mathbb{E}[X \mid Y = y] = \begin{cases} 0, & \text{if } \mathbb{P}[Y = y] = 0, \\ \dfrac{\mathbb{E}[\, 1_{\{Y=y\}} X\,]}{\mathbb{P}[Y = y]}, & \text{if } y = y_i \text{ for some } i = 1, \ldots, n. \end{cases}$$

Proof. We know that

$$\sigma(Y) = \sigma(\{A_1, \ldots, A_n\}),$$

where $A_i := \{Y = y_i\}$ for $i = 1, \ldots, n$. Moreover, it is easy to note that $\{A_1, \ldots, A_n\}$ is a partition of Ω. Thus, the proof follows from the proof of Theorem 2.1.

For conciseness, all properties and theorems in this section are stated without proofs. We begin with the elementary properties of conditional expectation.

Theorem 2.3 (Basic Properties)**.** *Let X and Y be integrable random variables and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then:*

1. *If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$.*

2. *If X is $\mathcal{G}$-measurable and XY is integrable, then*

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

3. *If X is independent of $\mathcal{G}$, then* $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.
4. *If X is independent of $\mathcal{G}$, and $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{F}$, then*

$$\mathbb{E}[X \mid \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \mid \mathcal{H}].$$

5. *If* $\mathcal{G} = \{\Omega, \varnothing\}$, *then* $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.
6. *For any $a, b \in \mathbb{R}$ we have*

$$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\,\mathbb{E}[X \mid \mathcal{G}] + b\,\mathbb{E}[Y \mid \mathcal{G}].$$

7. *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then*

$$\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}\Big[\mathbb{E}[X \mid \mathcal{G}] \,\Big|\, \mathcal{H}\Big].$$

8. *If $X \geq 0$, then* $\mathbb{E}[X \mid \mathcal{G}] \geq 0$.
9. *If $X \geq Y$, then* $\mathbb{E}[X \mid \mathcal{G}] \geq \mathbb{E}[Y \mid \mathcal{G}]$.

2.1.2 Useful classical theorems

Next, we present a few classical results extended from the static to the conditional case. For brevity, most proofs are omitted, as they closely parallel the unconditional versions, which can be found in any standard text on differential and integral calculus.

Theorem 2.4 (Conditional Monotone Convergence Theorem)**.** *Let (X_n) be a non-negative non-decreasing sequence of random variables that converges (almost surely) to X. Then, for any sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ we have*

$$\mathbb{E}[X_n \mid \mathcal{G}] \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[X \mid \mathcal{G}].$$

Theorem 2.5 (Conditional Dominated Convergence Theorem)**.** *Let (X_n) be a sequence of random variables that converges (almost surely) to X, and let Y be an integrable random variable such that for any $n \in \mathbb{N}$ we have $|X_n| \leq Y$. Then, the random variable X is integrable and for any sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ we get*

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

Theorem 2.6 (Conditional Fatou Lemma)**.** *Let (X_n) be a non-negative sequence of integrable random variables. Then, for any sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ we get*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \,\middle|\, \mathcal{G}\right] \;\leq\; \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}].$$

Finally, observe that if X is square-integrable, i.e. $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then the conditional expectation with respect to a sub-σ-algebra $\mathcal{G}$ may be interpreted as the *orthogonal projection* of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. This means that $\mathbb{E}[X \mid \mathcal{G}]$ is the least-squares optimal $\mathcal{G}$-measurable approximation of X. For example, if $\mathcal{G}$ captures information about a system (such as the stock market) and X denotes a related random variable (such as a stock price), then $\mathbb{E}[X \mid \mathcal{G}]$ serves as the best prediction of X given $\mathcal{G}$.

Theorem 2.7. *(Conditional expectation as the least-squares predictor) Let X be a square integrable random variable and let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then, for any square integrable $\mathcal{G}$-measurable random variable Z we get*

$$\mathbb{E}\Big[(X - \mathbb{E}[X \mid \mathcal{G}])^2\Big] \;\le\; \mathbb{E}\Big[(X - Z)^2\Big]. \tag{2.3}$$

Proof. Let Z be square integrable and $\mathcal{G}$-measurable. Using the tower property, we get

$$\mathbb{E}[(X - Z)^2] = \mathbb{E}\Big[\big((X - \mathbb{E}[X \mid \mathcal{G}]) + (\mathbb{E}[X \mid \mathcal{G}] - Z)\big)^2\Big].$$

Expanding the square and using linearity of expectation gives

$$\mathbb{E}[(X-Z)^2] = \mathbb{E}\Big[(X - \mathbb{E}[X \mid \mathcal{G}])^2\Big] + 2\,\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])(\mathbb{E}[X \mid \mathcal{G}] - Z)] + \mathbb{E}\Big[(\mathbb{E}[X \mid \mathcal{G}] - Z)^2\Big].$$

Now observe that

$$\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])(\mathbb{E}[X \mid \mathcal{G}] - Z)] = \mathbb{E}\Big[\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])(\mathbb{E}[X \mid \mathcal{G}] - Z) \mid \mathcal{G}]\Big].$$

Since Z is $\mathcal{G}$-measurable, this becomes

$$\mathbb{E}\Big[(\mathbb{E}[X \mid \mathcal{G}] - Z)\,\mathbb{E}[X - \mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}]\Big] = \mathbb{E}\Big[(\mathbb{E}[X \mid \mathcal{G}] - Z)\cdot 0\Big] = 0.$$

Thus,

$$\mathbb{E}[(X - Z)^2] = \mathbb{E}\Big[(X - \mathbb{E}[X \mid \mathcal{G}])^2\Big] + \mathbb{E}\Big[(\mathbb{E}[X \mid \mathcal{G}] - Z)^2\Big] \;\ge\; \mathbb{E}\Big[(X - \mathbb{E}[X \mid \mathcal{G}])^2\Big].$$

This concludes the proof. □

2.2 Introduction on stochastic processes

Let $(\Omega, \mathcal{F}, P)$ denote the underlying probability space. All random variables introduced in this course are assumed to be integrable, and unless stated otherwise, all inequalities are understood in the almost-sure sense.

As a starting point, we consider a classical example: the **simple random walk**.

2.2.1 Simple random walk

A simple random walk can be interpreted as a model for repeated gambling. Suppose you start with an initial fortune of \$$a$, and repeatedly place bets of \$1. At each bet, you win \$1 with probability p and lose \$1 with probability q, where $p + q = 1$.

If we let X_n denote your fortune at time n, then $X_0 = a$. After the first bet, we have

$$X_1 = \begin{cases} a + 1, & \text{with probability } p, \\ a - 1, & \text{with probability } q. \end{cases}$$

Similarly, after two bets,

$$X_2 \in \{a + 2,\; a,\; a - 2\},$$

depending on whether you win both bets, win once and lose once, or lose both bets. Continuing in this way, we obtain a sequence $(X_n)_{n \ge 0}$ of random variables representing the gambler's fortune over time. This stochastic process is called the **simple random walk**.

An equivalent and more formal construction is as follows. Let $(Z_i)_{i\geq 1}$ be i.i.d. random variables such that

$$P(Z_i = 1) = p, \qquad P(Z_i = -1) = q,$$

with $0 < p < 1$. Here, $Z_i = 1$ corresponds to a win on the i-th bet, and $Z_i = -1$ to a loss. Then the random walk is given by

$$X_0 = a, \qquad X_n = a + \sum_{i=1}^{n} Z_i, \quad n \geq 1.$$

Example 2.1. *Consider simple random walk with $a = 8$ and $p = \frac{1}{3}$, so you start with \$8 and have probability $\frac{1}{3}$ of winning each bet. Then the probability that you have \$9 after one bet is given by*

$$P(X_1 = 9) = P(8 + Z_1 = 9) = P(Z_1 = 1) = \frac{1}{3},$$

as it should be. Also, the probability that you have \$7 after one bet is given by

$$P(X_1 = 7) = P(8 + Z_1 = 7) = P(Z_1 = -1) = \frac{2}{3}.$$

On the other hand, the probability that you have \$10 after two bets is given by

$$P(X_2 = 10) = P(8 + Z_1 + Z_2 = 10) = P(Z_1 + Z_2 = 2) = P(Z_1 = 1, Z_2 = 1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}.$$

In this section, we use t to denote time. In the discrete case, t is typically associated with the set of days or years, e.g.,

$$T = \{1, 2, \dots, t\} \quad \text{for some fixed } t \in \mathbb{N}, \quad \text{or } T = \mathbb{N}, \quad \text{or } T = \mathbb{Z}.$$

In the continuous case, T is usually linked to some fixed interval of $\mathbb{R}$, e.g.,

$$T = [0, t] \quad \text{for some fixed } t \in \mathbb{R}_+, \quad \text{or } T = \mathbb{R}_+, \quad \text{or } T = \mathbb{R}.$$

2.2.2 Basic definitions and Properties

Stochastic processes describe dynamical systems whose evolution over time is probabilistic in nature. The formal definition is as follows.

Definition 2.3 (Stochastic process)**.** *Let T be an ordered index set, $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, and $(E, \mathcal{G})$ a measurable space. A* stochastic process *is a family of random variables*

$$X = \{X_t : t \in T\},$$

where, for each fixed $t \in T$, the mapping $X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{G})$ is a random variable. The set Ω is called the sample space*, and E is the* state space *of the process.*

The index set T may be discrete (e.g., $\mathbb{Z}_+$) or continuous (e.g., $\mathbb{R}_+$). In most cases, the state space E is $\mathbb{R}^d$ equipped with the Borel σ-algebra.

A stochastic process X can be regarded as a function of both time $t \in T$ and outcome $\omega \in \Omega$. Thus, one may write $X(t)$, $X(t, \omega)$, or $X_t(\omega)$ to denote its value. For each fixed sample point $\omega \in \Omega$, the map

$$t \mapsto X_t(\omega), \quad T \to E,$$

is called a *realization*, *sample path*, or *trajectory* of the process.

Definition 2.4. *(Finite-dimensional distributions). The* finite-dimensional distributions (fdd) *of a stochastic process are the joint distributions of the E^k-valued random variables*

$$(X(t_1), X(t_2), \ldots, X(t_k))$$

for arbitrary positive integer k and arbitrary times $t_i \in T$, $i \in \{1, \ldots, k\}$. That is,

$$F(x) = \mathbb{P}\left(X(t_i) \leq x_i,\, i = 1, \ldots, k\right).$$

From experiments or numerical simulations we can only obtain information about the *finite-dimensional distributions* of a process. This raises a natural question: are the finite-dimensional distributions of a stochastic process sufficient to determine the process uniquely. For processes with continuous paths,[1] the answer is affirmative. These are precisely the class of stochastic processes that we will focus on in these notes. The family of finite-dimensional distributions determines the statistical properties of the process $(X_t)_{t \in T}$. Conversely, given a family

$$\{\nu_{t_1, \ldots, t_n} : t_k \in T,\, n \in \mathbb{N}\}$$

of probability measures on E^n satisfying the two natural consistency conditions, it follows from Kolmogorov's extension theorem that we can construct a stochastic process.

Theorem 2.8 (Kolmogorov's Extension Theorem). *For all $t_1, \ldots, t_n \in T$, let $\nu_{t_1, \ldots, t_n}$ be probability measures on E^n satisfying:*

1. ***Permutation invariance:*** *for any permutation π of $\{1, \ldots, n\}$,*

$$\nu_{t_{\pi(1)}, \ldots, t_{\pi(n)}}(B_1 \times \cdots \times B_n) = \nu_{t_1, \ldots, t_n}(B_{\pi^{-1}(1)} \times \cdots \times B_{\pi^{-1}(n)}),$$

for all Borel sets $B_1, \ldots, B_n \subseteq E$.

2. ***Consistency under marginalization:***

$$\nu_{t_1, \ldots, t_n}(B_1 \times \cdots \times B_n) = \nu_{t_1, \ldots, t_{n+m}}(B_1 \times \cdots \times B_n \times E \times \cdots \times E),$$

for all $m \geq 1$.

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $(X_t)_{t \in T}$ on Ω such that

$$\nu_{t_1, \ldots, t_n}(B_1 \times \cdots \times B_n) = P(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n),$$

for all $t_k \in T$, $n \in \mathbb{N}$, and Borel sets $B_k \subseteq E$.

Definition 2.5. *We say that two processes X_t and Y_t are* equivalent *if they have the same finite-dimensional distributions.*

[1] A precise version of this statement relies on Kolmogorov's extension theorem together with path regularity results.

2.2.3 Gaussian stochastic processes

An especially important class of continuous-time processes is that of *Gaussian processes*, which play a central role in both theory and applications.

Definition 2.6. *A one-dimensional continuous-time* Gaussian process *is a stochastic process for which* $E = \mathbb{R}$ *and all the finite-dimensional distributions are Gaussian, i.e., every finite-dimensional vector*

$$(X_{t_1}, X_{t_2}, \dots, X_{t_k})$$

is a multivariate Gaussian random variable $\mathcal{N}(\mu_k, K_k)$ *for some vector* $\mu_k \in \mathbb{R}^k$ *and a symmetric nonnegative definite matrix* $K_k \in \mathbb{R}^{k \times k}$*, for all* $k \in \mathbb{N}$ *and all* $t_1, \dots, t_k \in \mathbb{R}_+$*.*

From the above definition, we conclude that the finite-dimensional distributions of a Gaussian continuous-time stochastic process are multivariate normal with probability density function:

$$\gamma_{\mu_k, K_k}(x) = \frac{1}{(2\pi)^{k/2}(\det K_k)^{1/2}} \exp\left(-\frac{1}{2}\langle K_k^{-1}(x - \mu_k), x - \mu_k\rangle\right),$$

where $x = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$. It is straightforward to extend the above definition to arbitrary dimensions. A Gaussian process $x(t)$ is characterized by its mean and covariance functions:

- **Mean function:** $m(t) := \mathbb{E}[x(t)]$
- **Covariance function:**

$$C(t, s) := \mathbb{E}\left[(x(t) - m(t))^T (x(s) - m(s))\right]$$

Thus, the first two moments of a Gaussian process are sufficient for a complete characterization of the process.

Simulation of Gaussian Processes

Simulating Gaussian stochastic processes on a computer is straightforward. Given a random number generator that produces independent $\mathcal{N}(0, 1)$ (pseudo)random numbers, one can sample from a Gaussian process by computing the square root of its covariance matrix (for instance, using a Cholesky decomposition).

A simple algorithm to construct a skeleton of a continuous-time Gaussian process is as follows:

1. Fix Δt and define $t_j = (j - 1)\Delta t$, for $j = 1, \dots, N$.
2. Set $X_j := X(t_j)$ and define the Gaussian random vector

$$X_N := \left\{X_j^N\right\}_{j=1}^N.$$

   Then $X_N \sim \mathcal{N}(\mu_N, \Gamma_N)$, with:

$$\mu_N = (\mu(t_1), \dots, \mu(t_N)), \quad \Gamma_N^{ij} = C(t_i, t_j).$$

Then $X_N = \mu_N + \Lambda\, Z$, where $Z \sim \mathcal{N}(0, I)$ and $\Gamma_N = \Lambda\Lambda^T$. We can calculate the square root of the covariance matrix Γ_N either using the *Cholesky factorization*, via the *spectral decomposition* of Γ_N, or by using the *singular value decomposition (SVD)*.

Examples of Gaussian Stochastic Processes

- **Random Fourier series:** Let $\xi_j, \zeta_j \sim \mathcal{N}(0,1)$, for $j = 1, \dots, N$, and define

$$X(t) = \sum_{j=1}^{N} \left(\xi_j \cos(2\pi j t) + \zeta_j \sin(2\pi j t)\right).$$

- **Brownian motion:** A Gaussian process with

$$m(t) = 0, \quad C(t,s) = \min(t,s).$$

- **Brownian bridge:** A Gaussian process with

$$m(t) = 0, \quad C(t,s) = \min(t,s) - ts.$$

- **Ornstein–Uhlenbeck process:** A Gaussian process with

$$m(t) = 0, \quad C(t,s) = \lambda e^{-\alpha|t-s|}, \quad \alpha, \lambda > 0.$$

2.2.4 Stationary processes

In many applied settings, stochastic processes exhibit statistical properties that do not change over time. Such processes are referred to as *stationary*. We distinguish between processes for which *all* finite-dimensional distributions are invariant under time shifts (*strictly stationary*) and those for which only the *first two moments* remain constant (*weakly stationary*).

Strictly Stationary Processes

Definition 1.5. A stochastic process is called *(strictly) stationary* if all its finite-dimensional distributions are invariant under time translation: for any integer k and times $t_1, \dots, t_k \in T$, the distribution of

$$(X(t_1), X(t_2), \dots, X(t_k))$$

is equal to that of

$$(X(s+t_1), X(s+t_2), \dots, X(s+t_k))$$

for all s such that $s + t_i \in T$ for all $i = 1, \dots, k$. In other words, for any measurable sets $A_1, \dots, A_k \subseteq \mathbb{R}$,

$$\mathbb{P}(X_{t_1+s} \in A_1, \dots, X_{t_k+s} \in A_k) = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_k} \in A_k), \quad \forall s \in T.$$

Example 2.2. *Let $Y_0, Y_1, \dots$ be a sequence of independent, identically distributed random variables and consider the stochastic process $X_n = Y_n$. Then X_n is a strictly stationary process.*

2.2.5 Second-order stationary processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X_t, $t \in T$ (with $T = \mathbb{R}$ or $\mathbb{Z}$), be a real-valued stochastic process on this space with finite second moment, i.e., $\mathbb{E}[|X_t|^2] < +\infty$ (so $X_t \in L^2(\Omega, \mathbb{P})$ for all $t \in T$). Assume that X_t is strictly stationary. Then,

$$\mathbb{E}[X_{t+s}] = \mathbb{E}[X_t], \quad \forall s \in T, \tag{2.4}$$

from which we conclude that $\mathbb{E}[X_t]$ is constant. Also,

$$\mathbb{E}[(X_{t_1+s} - \mu)(X_{t_2+s} - \mu)] = \mathbb{E}[(X_{t_1} - \mu)(X_{t_2} - \mu)], \quad \forall s \in T, \tag{2.5}$$

which implies that the covariance function depends only on the time difference:

$$C(t, s) = C(t - s).$$

This motivates the following definition.

Definition 2.7. *A stochastic process $X_t \in L^2$ is called* second-order stationary*,* wide-sense stationary*, or* weakly stationary *if the first moment $\mathbb{E}[X_t]$ is constant, and the covariance function $\mathbb{E}[(X_t - \mu)(X_s - \mu)]$ depends only on the difference $t - s$:*

$$\mathbb{E}[X_t] = \mu, \quad \mathbb{E}[(X_t - \mu)(X_s - \mu)] = C(t - s).$$

The constant μ is called the *mean* of the process X_t. Without loss of generality, we can assume $\mu = 0$, since if $\mathbb{E}[X_t] = \mu$, then the process $Y_t := X_t - \mu$ is mean-zero. A mean-zero process is called a *centered process*.

The function $C(t)$ is called the *covariance function* of X_t, and is sometimes referred to as the *autocovariance* or *autocorrelation function*. Note that $C(t) = \mathbb{E}[X_t X_0]$, while $C(0) = \mathbb{E}[X_t^2]$, which is finite by assumption. Since X_t is real-valued, we have

$$C(t) = C(-t), \quad t \in \mathbb{R}.$$

Now, let X_t be a strictly stationary process with finite second moment. By definition, $\mathbb{E}[X_t] = \mu$ and

$$\mathbb{E}[(X_t - \mu)(X_s - \mu)] = C(t - s).$$

Hence, any strictly stationary process with finite second moment is also *weakly stationary* (stationary in the wide sense). The converse is not true in general.

However, for **Gaussian processes**, the first two moments fully determine the process. Consequently, a Gaussian process is strictly stationary if and only if it is weakly stationary.

Example 2.3. *Let $Y_0, Y_1, \dots$ be a sequence of independent, identically distributed random variables and consider the stochastic process $X_n = Y_n$. From Example 2.2, we know that this is a strictly stationary process, irrespective of whether $\mathbb{E}[Y_0^2] < +\infty$. Assume now that $\mathbb{E}[Y_0] = 0$ and $\mathbb{E}[Y_0^2] = \sigma^2 < +\infty$. Then X_n is a second-order stationary process with mean zero and correlation function*

$$R(k) = \sigma^2 \delta_{k0}.$$

Notice that in this case, there is no correlation between the values of the stochastic process at different times n and k.

Example 2.4. *Let Z be a single random variable and consider the stochastic process $X_n = Z$, $n = 0, 1, 2, \ldots$. From Example 2.3, we know that this is a strictly stationary process irrespective of whether $\mathbb{E}[|Z|^2] < +\infty$ or not. Assume now that $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^2] = \sigma^2$. Then X_n becomes a second-order stationary process with*

$$R(k) = \sigma^2.$$

Notice that in this case, the values of our stochastic process at different times are strongly correlated.

Continuity in the L^2-sense. Continuity properties of the covariance function are equivalent to continuity properties of the paths of X_t in the L^2-sense, i.e.,

$$\lim_{h\to 0} \mathbb{E}[|X_{t+h} - X_t|^2] = 0. \tag{2.6}$$

Lemma 2.1. *Assume that the covariance function $C(t)$ of a second-order stationary process is continuous at $t = 0$. Then it is continuous for all $t \in \mathbb{R}$. Furthermore, the continuity of $C(t)$ is equivalent to the continuity of the process X_t in the L^2-sense.*

Proof. Fix $t \in \mathbb{R}$ and, without loss of generality, assume $\mathbb{E}[X_t] = 0$. Then

$$\begin{aligned} |C(t+h) - C(t)|^2 &= |\mathbb{E}[X_{t+h}X_0] - \mathbb{E}[X_t X_0]|^2 = |\mathbb{E}[(X_{t+h} - X_t)X_0]|^2 \\ &\leq \mathbb{E}[X_0^2] \cdot \mathbb{E}[(X_{t+h} - X_t)^2] = C(0) \cdot \mathbb{E}[(X_{t+h} - X_t)^2]. \end{aligned}$$

Since $\lim_{h\to 0} \mathbb{E}[(X_{t+h} - X_t)^2] = 0$, it follows that

$$\lim_{h\to 0} |C(t+h) - C(t)| = 0.$$

Hence, continuity of $C(\cdot)$ at 0 implies continuity for all $t \in \mathbb{R}$.

Conversely, assume that $C(t)$ is continuous. From the above calculation, we have

$$\mathbb{E}[|X_{t+h} - X_t|^2] = 2\Big(C(0) - C(h)\Big), \tag{2.7}$$

which converges to 0 as $h \to 0$. Thus, X_t is L^2-continuous. Similarly, if X_t is L^2-continuous, then (2.7) implies

$$\lim_{h\to 0} C(h) = C(0).$$

Note also from (2.7) that $C(0) \geq C(h)$ for all $h \in \mathbb{R}$.

The Fourier transform of the covariance function of a second-order stationary process always exists. This fact allows the study of second-order stationary processes using Fourier analysis. To make the connection precise, we will invoke *Bochner's theorem*, which applies to all nonnegative definite functions.

Definition 2.8. *A function $f(x) : \mathbb{R} \to \mathbb{R}$ is called nonnegative definite if*

$$\sum_{i,j=1}^{n} f(t_i - t_j) c_i \overline{c_j} \geq 0 \quad \textit{for all } n \in \mathbb{N}, t_1, \ldots, t_n \in \mathbb{R}, c_1, \ldots, c_n \in \mathbb{C}.$$

Lemma 2.2. *The covariance function of a second-order stationary process is a nonnegative definite function.*

Proof. We will use the notation $X_t^c := \sum_{i=1}^n X_{t_i} c_i$. We have

$$\sum_{i,j=1}^n C(t_i - t_j) c_i \overline{c_j} = \sum_{i,j=1}^n \mathbb{E}[X_{t_i} X_{t_j}] c_i \overline{c_j} = \mathbb{E}\left[\left(\sum_{i=1}^n X_{t_i} c_i\right)\left(\sum_{j=1}^n X_{t_j} \overline{c_j}\right)\right] = \mathbb{E}[|X_t^c|^2] \geq 0.$$

In many applications, it is desirable to impose additional properties on stochastic processes, such as continuity of sample paths. Some of these properties are summarized in the following definition. For simplicity, we present them in the continuous-time setting. While certain properties, such as integrability, extend naturally to discrete-time processes, others, such as continuity, do not.

Definition 2.9 (Path Properties)**.** *Let X be a continuous-time stochastic process. We say that X is:*

1. Measurable *if the map $X : T \times \Omega \to \mathbb{R}$ is $\mathcal{F} \otimes \mathcal{B}(T)$-measurable.*
2. Continuous *(resp. left-continuous, right-continuous) if all sample paths of X are continuous (resp. left-continuous, right-continuous).*
3. Continuous in probability *(or* stochastically continuous*) if for any $t \in T$, we have $X_s \to X_t$ in probability whenever $s \to t$, i.e.,*

$$\forall \varepsilon > 0 \; \exists \delta > 0 : \quad \text{for all } s \in T \text{ such that } |t - s| \leq \delta, \quad \mathbb{P}(|X_t - X_s| \geq \varepsilon) \leq \varepsilon.$$

4. Càdlàg *(continue à droite et limites à gauche) if all sample paths are right-continuous and have left limits for every $t \in T$.*
5. Integrable *if $\mathbb{E}[|X_t|] < \infty$ for all $t \in T$.*
6. Square-integrable *if $\mathbb{E}[|X_t|^2] < \infty$ for all $t \in T$.*
7. p-power integrable *if $\mathbb{E}[|X_t|^p] < \infty$ for some $p \in \mathbb{N}$ and all $t \in T$.*

Now, we want to identify processes for which almost all trajectories are the same

Definition 2.10 (Equivalence Notions)**.** *Let X and Y be two stochastic processes defined on the same probability space. We say that:*

1. *Y is* indistinguishable *from X if*

$$\mathbb{P}(X_t = Y_t \text{ for all } t \in T) = 1.$$

2. *Y is a* modification *of X if for every $t \in T$,*

$$\mathbb{P}(X_t = Y_t) = 1.$$

3. *Y has the same* finite-dimensional distributions *as X if for every $n \in \mathbb{N}$, for all time points $(t_1, \ldots, t_n) \in T^n$, and for all $A \in \mathcal{B}(\mathbb{R}^n)$,*

$$\mathbb{P}((X_{t_1}, \ldots, X_{t_n}) \in A) = \mathbb{P}((Y_{t_1}, \ldots, Y_{t_n}) \in A).$$

Sometimes, instead of saying "modification," we say that X is a version *of Y. Note that if X and Y are modifications of each other and are (almost surely) continuous, then they are indistinguishable.*

Example 2.5. *Let $T = [0, 1]$ and let $\Omega = [0, 1]$ be a standard probability space. Let X be such that*

Let $X_t \equiv 0$ for all $t \in T$, and let Y given by

$$Y_t(\omega) = \begin{cases} 1 & \text{si } t \neq \omega, \\ 0 & \text{si } t = \omega. \end{cases}$$

Hence, for all $t \in T$, we have

$$\mathbb{P}[X_t = Y_t] = \mathbb{P}[\Omega \setminus \{t\}] = 1, \quad \text{mais} \quad \mathbb{P}[X_t = Y_t \; \forall t \in T] = 0.$$

Definition 2.11. *Let $X = (X_t)_{t \in T}$ be a continuous-time stochastic process. We say that X has* independent increments *if for any finite set of time-points $t_1 \leq t_2 \leq \cdots \leq t_n$ (from T), the incremental random variables*

$$X_{t_2} - X_{t_1}, \quad X_{t_3} - X_{t_2}, \quad \ldots, \quad X_{t_n} - X_{t_{n-1}}$$

are independent. Moreover, if the distribution of the increment $X_t - X_s$ depends only on $t - s$, then we say that X has independent and stationary increments*.*

2.2.6 Kolmogorov's extension theorem

We now present Kolmogorov's *extension theorem*, which addresses the existence of a stochastic process corresponding to a given collection of finite-dimensional distributions. Before stating the theorem, we first define finite-dimensional distributions precisely.

Definition 2.12. *Let X be a stochastic process. The mapping $\mathbb{P}_X : \mathcal{B}_T \to [0, 1]$, which characterises the finite-dimensional distributions of X, is given by*

$$\mathbb{P}_X[A] := \mathbb{P}\left[\{\omega \in \Omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in \Gamma\}\right],$$

where $A = \{x \in \mathbb{R}^T : (x_{t_1}, \ldots, x_{t_n}) \in \Gamma\}$ is a cylinder set, and $\mathcal{B}_T$ is the collection of all cylinder sets on T.

(a) We call $A \subset \mathbb{R}^T$ a cylinder set if $A = \{x \in \mathbb{R}^T : (x_{t_1}, \ldots, x_{t_n}) \in \Gamma\}$, where $n \in \mathbb{N}$, $(t_i)_{i=1}^n \subset T$ is a finite sequence of time points, and $\Gamma \subset \mathbb{R}^n$ is a Borel-measurable set. Note that while $\mathcal{B}_T$ is always an algebra, it is not necessarily a σ-algebra. Given the auxiliary mapping $\mathbb{P}_X$ defined on the cylinder sets $\mathcal{B}_T$, we want to (uniquely) extend it to a probability measure on $\mathcal{B}(\mathbb{R}^T)$ in order to get the distribution of the corresponding stochastic process. In other words, we want to check if the mapping $\mathbb{P}_X$ can be used to characterise X. This can be done if the mapping preserves the consistency property. Before we state the main result, we introduce some additional notation. Given T, let $\mathcal{T}$ denote the set of all finite subsets of T, partially ordered by inclusion. Assume we are given a family $(\mathbb{P}_I)_{I \in \mathcal{T}}$ of probability measures $\mathbb{P}_I$ defined on $(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I))$. Then, for $I_1 \subset I_2 \subset T$ and $A \in \mathcal{B}(\mathbb{R}^{I_1})$, we define the projection sets:

$$C_{I_2, I_1}(A) := \left\{\omega \in \mathbb{R}^{I_2} : (\omega(t))_{t \in I_1} \in A\right\}, \quad C_{I_1}(A) := \left\{\omega \in \mathbb{R}^T : (\omega(t))_{t \in I_1} \in A\right\}.$$

Note that $C_{I_2,I_1} : \mathcal{B}(\mathbb{R}^{I_1}) \to \mathcal{B}(\mathbb{R}^{I_2})$ and $C_{I_1} : \mathcal{B}(\mathbb{R}^{I_1}) \to \mathcal{B}(\mathbb{R}^{T})$.

We say that a family $(\mathbb{P}_I)_{I\in\mathcal{T}}$ is *consistent* if for any $I_1 \subset I_2 \subset T$ and for all $A \in \mathcal{B}(\mathbb{R}^{I_1})$, we have

$$\mathbb{P}_{I_1}[A] = \mathbb{P}_{I_2}[C_{I_2,I_1}(A)].$$

Now, we present the Kolmogorov's extension theorem.

Theorem 2.9 (Kolmogorov's Extension Theorem). *Let $(\mathbb{P}_I)_{I\in\mathcal{T}}$ be a consistent family of measures. Then, there exists a unique probability measure $\mathbb{P}$ defined on $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$ such that*

$$\mathbb{P}[C_I(A)] = \mathbb{P}_I[A],$$

for any $I \in \mathcal{T}$ and $A \in \mathcal{B}(\mathbb{R}^I)$.

Note that the consistency property is very natural: we simply want the distribution to remain unchanged if we exclude some time points from a larger set I_2 and restrict ourselves to a subset $I_1 \subset I_2$.

The theorem states that, given a stochastic process X and the corresponding mapping $\mathbb{P}_X$, we can extend this mapping to a full probability measure on $\mathbb{R}^T$ (with its Borel σ-algebra).

2.2.7 Continuous modifications and Kolmogorov's theorems

We now show that any stochastic process admitting a continuous modification is necessarily continuous in probability.

Proposition 2.1. *Let X be a stochastic process that possesses a continuous modification. Then X is continuous in probability.*

Proof. Let us fix $t \in T$ and $\varepsilon > 0$. Let

$$A_n := \left\{\omega \in \Omega : \exists s \in T \text{ such that } |t-s| \leq \frac{1}{n} \text{ and } |\widetilde{X}_t(\omega) - \widetilde{X}_s(\omega)| \geq \varepsilon\right\}.$$

From the continuity of $\widetilde{X}$, we know that A_n is measurable since it can be expressed as

$$A_n := \left\{\omega \in \Omega : \exists s \in T \cap \mathbb{Q} \text{ such that } |t-s| \leq \frac{1}{n} \text{ and } |\widetilde{X}_t(\omega) - \widetilde{X}_s(\omega)| \geq \varepsilon\right\}.$$

□

Also, we know that $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence such that

$$\mathbb{P}\left[\bigcap_{n\in\mathbb{N}} A_n\right] = 0.$$

Thus, there exists $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}[A_{n_0}] \leq \varepsilon. \tag{2.8}$$

Next, because $\widetilde{X}$ is a modification of X and

$$\widetilde{X}_t - \widetilde{X}_s = (\widetilde{X}_t - X_t) + (X_t - X_s) + (X_s - \widetilde{X}_s),$$

we get

$$\mathbb{P}\left[|\widetilde{X}_t - \widetilde{X}_s| \geq \varepsilon\right] = \mathbb{P}\left[|X_t - X_s| \geq \varepsilon\right]. \tag{2.9}$$

Combining (2.8) with (2.9), for any $\delta < \frac{1}{n_0}$ and any $s \in T$ such that $|t-s| \leq \delta$, we get

$$\mathbb{P}\left[|X_t - X_s| \geq \varepsilon\right] = \mathbb{P}\left[|\widetilde{X}_t - \widetilde{X}_s| \geq \varepsilon\right] \leq \mathbb{P}[A_{n_0}] \leq \varepsilon,$$

which concludes the proof.
From Proposition 2.1, we know that any continuous process is continuous in probability. Moreover, any modification of a process that is continuous in probability is itself continuous in probability. We now present an example of a stochastic process that is continuous in probability but does not admit a continuous modification.

Example 2.6. *Let $T = [0,1]$ and let $Z \sim \mathcal{U}[0,1]$. Let $X = (X_t)_{t\in T}$ be given by*

$$X_t(\omega) = \mathbf{1}_{[0,Z(\omega))}(t),$$

for $t \in [0,1]$ and $\omega \in \Omega$. One can easily check that there exists no continuous modification of X, but X is continuous in probability.

2.2.8 Kolmogorov's continuity theorem and Hölder continuity

Before presenting (without proof) two simplified versions of Kolmogorov's continuity theorems which is a fundamental tool in the theory of stochastic processes. Let us first recall the notion of Hölder continuity.

Definition 2.13. *We say that a function $f : I \to \mathbb{R}$ is* Hölder continuous *on $I = [a,b]$ with exponent $\alpha \in (0,1)$ if there exists a constant $C > 0$ such that for any $x, y \in I$,*

$$|f(x) - f(y)| \leq C|x-y|^{\alpha}.$$

Of course, any Hölder continuous function is continuous. Also, if f is Hölder continuous with exponent $\alpha \in (0,1)$, then it is Hölder continuous with any exponent $\gamma \leq \alpha$.

Theorem 2.10 (Kolmogorov's Continuity)**.** *Let $T = [a,b]$ and let X be a stochastic process. If there exist constants $p > 0$, $K > 0$, and $\varepsilon > 0$ such that for any $t, s \in T$,*

$$\mathbb{E}|X_t - X_s|^p \leq K|t-s|^{1+\varepsilon}, \tag{2.10}$$

then X has a modification which is Hölder continuous for any exponent $\alpha \in \left(0, \frac{\varepsilon}{p}\right)$.

While the condition $T = [a,b]$ might look restrictive, it is often used to prove the existence of continuous modifications. The following is a direct consequence of Theorem 2.10.

Theorem 2.11 (Kolmogorov's Continuity)**.** *Let $T = \mathbb{R}_+$ and let X be a stochastic process. Assume that for any $T \in \mathbb{R}_+$, there exist constants $p > 0$, $K > 0$, and $\varepsilon > 0$ such that inequality (2.10) holds for all $s, t \leq T$. Then, there exists a continuous modification of X.*

2.3 Filtrations and adaptiveness

We start with basic definitions.

Definition 2.14. *A* filtration *is a non-decreasing family of sub-σ-algebras of $\mathcal{F}$ indexed by time, i.e., a family $\mathbb{F} := (\mathcal{F}_t)_{t\in T}$ such that*

$$\mathcal{F}_s \subseteq \mathcal{F}_t, \quad \text{for all } s \le t,\ s,t \in T.$$

Typically, $\mathcal{F}_t$ represents the information available about the system up to time t. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F}$ is called a filtered probability space.

Example 2.7. *Consider two successive tosses of a fair coin. The sample space is*

$$\Omega = \{HH, HT, TH, TT\},$$

where H denotes heads *and T denotes* tails. *Each outcome has probability $\frac{1}{4}$.*

We observe the tosses at times $t = 0, 1, 2$ and define the natural filtration $(\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2)$ by

$$\mathcal{F}_0 = \{\varnothing, \Omega\}, \qquad \mathcal{F}_1 = \sigma(\textit{first toss}), \qquad \mathcal{F}_2 = \mathcal{P}(\Omega).$$

Thus,

$$\mathcal{F}_1 = \{\varnothing, \Omega, \{HH, HT\}, \{TH, TT\}\}.$$

Define a random variable (payoff) R by

$$R(HH) = 2, \quad R(HT) = 1, \quad R(TH) = 1, \quad R(TT) = 0.$$

$$\mathbb{E}[R] = 2\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 0\cdot\frac{1}{4} = \frac{2+1+1+0}{4} = 1.$$

The σ-algebra $\mathcal{F}_1$ distinguishes two events:

$$A_H = \{HH, HT\} \quad \textit{and} \quad A_T = \{TH, TT\}.$$

- *If the first toss is H, then*

$$\mathbb{E}[R \mid \mathcal{F}_1](\omega \in A_H) = 2 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{3}{2} = 1.5.$$

- *If the first toss is T, then*

$$\mathbb{E}[R \mid \mathcal{F}_1](\omega \in A_T) = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2} = 0.5.$$

Hence:

$$\mathbb{E}[R \mid \mathcal{F}_1] = \begin{cases} 1.5, & \textit{if the first toss is } H, \\ 0.5, & \textit{if the first toss is } T. \end{cases}$$

$$\mathbb{E}[R \mid \mathcal{F}_0] = \mathbb{E}[R] = 1,$$

$$\mathbb{E}[R \mid \mathcal{F}_1] = \begin{cases} 1.5 & \textit{on } \{HH, HT\}, \\ 0.5 & \textit{on } \{TH, TT\}, \end{cases}$$

$$\mathbb{E}[R \mid \mathcal{F}_2] = R.$$

For continuous-time filtrations, a right-continuity condition is often imposed for regularity.

Definition 2.15. *A filtration* $\mathbb{F}$ *is* right-continuous *if, for any* $t \in T$,

$$\mathcal{F}_t = \mathcal{F}_{t+}, \quad \text{where } \mathcal{F}_{t+} := \bigcap_{s>t,\, s\in T} \mathcal{F}_s.$$

Filtered probability spaces are often assumed to satisfy the *usual conditions*: the filtration is right-continuous and complete.[2]

The filtration encodes the evolving information about a system, while the stochastic process represents its evolution. This relation is formalized through the notion of adaptiveness.

Definition 2.16. *A stochastic process* $X = (X_t)_{t\in T}$ *is* adapted *to the filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\in T}$, *or* $\mathbb{F}$-adapted, *if* X_t *is* $\mathcal{F}_t$*-measurable for every* $t \in T$.

Definition 2.17. *For a process* X, *the* filtration generated by X, *denoted* $\mathbb{F}^X = (\mathcal{F}_t^X)_{t\in T}$, *is defined by*

$$\mathcal{F}_t^X := \sigma(X_s : s \le t,\ s \in T).$$

In other words, $\mathbb{F}^X$ is the smallest filtration such that X_s is $\mathcal{F}_t^X$-measurable for all $s \le t$. It is important to note that a filtration generated by a right-continuous process is not necessarily right-continuous itself.[3] Conversely, a filtration generated by a process that is not right-continuous may still be right-continuous.[4] Under certain conditions, such as the Feller property, the generated filtration may indeed be right-continuous.

Now, having defined filtrations, we are ready to introduce the notion of progressively measurable processes.

Definition 2.18. *A process* $X = (X_t)_{t\in T}$, *defined on a filtered probability space, is called* progressively measurable *if it is* $\mathbb{F}$*-adapted and for any* $t \in T$, *the map*

$$(T \cap (-\infty, t]) \times \Omega \to \mathbb{R}, \quad (s, \omega) \mapsto X_s(\omega)$$

is measurable with respect to $\mathcal{B}(T \cap (-\infty, t]) \otimes \mathcal{F}_t$.

Intuitively, we require that the stochastic process truncated at time $t \in T$ remains measurable with respect to the available information up to time t. As the next example shows, measurability and progressive measurability are not equivalent.

Example 2.8. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a standard probability space with* $\Omega = [0, 1]$.[5] *Define*

$$\mathcal{A} := \sigma(\{N \subset [0, 1] : \#N < \infty\}),$$

i.e., the σ*-algebra generated by all countable sets (and their complements).*

[2]A filtration is complete if $\mathcal{F}$ and $\mathcal{F}_t$, for all $t \in T$, contain all $\mathbb{P}$-null sets. A probability space $(\Omega, \Sigma, \mathbb{P})$ is complete if, for any $S \subseteq N$ with $N \in \Sigma$ and $\mathbb{P}[N] = 0$, we have $S \in \Sigma$.

[3]For example, let $X = (X_t)_{t\in[0,1]}$ be defined by $X_t = tZ$, where $Z \sim \mathcal{U}[0, 1]$.

[4]Let Z be a strictly positive random variable and define $X = (X_t)_{t\in\mathbb{R}_+}$ by $X_t = tZ$ for $t \in [0, 1] \cup (2, \infty)$ and $X_t = 0$ otherwise.

[5]That is, Ω is equipped with the Borel σ-algebra and Lebesgue measure.

Let the time horizon be $T = [0, +\infty)$, and define the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in T}$ by

$$\mathcal{F}_t := \begin{cases} \mathcal{A} & \text{for } t \in [0,1), \\ \mathcal{F} & \text{for } t \geq 1. \end{cases}$$

Next, define the stochastic process $X = (X_t)_{t\in T}$ by

$$X_t(\omega) := \mathbf{1}_\Delta(t,\omega) = \begin{cases} 1 & \text{if } t = \omega \text{ and } t \leq \frac{1}{2}, \\ 0 & \text{otherwise}, \end{cases} \quad \text{where } \Delta := \{(t,t) : t \in [0, \tfrac{1}{2}]\} \subset T \times \Omega.$$

It is straightforward to verify that X is a measurable process, since $\Delta \in \mathcal{B}([0,1]) \otimes \mathcal{F}$. Moreover, X is $\mathbb{F}$-adapted, since for any $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R})$, we have

$$X_t^{-1}(A) = \begin{cases} \emptyset & \text{if } t \in [0,\infty),\ 0 \notin A,\ 1 \notin A, \\ \{t\} & \text{if } t \in [0,\frac{1}{2}],\ 0 \notin A,\ 1 \in A, \\ [0,1] \setminus \{t\} & \text{if } t \in [0,\frac{1}{2}],\ 0 \in A,\ 1 \notin A, \\ [0,1] & \text{if } t \in [0,\frac{1}{2}],\ 0 \in A,\ 1 \in A, \\ [0,1] & \text{if } t \in [\frac{1}{2},\infty),\ 0 \in A. \end{cases}$$

Hence, $X_t^{-1}(A) \in \mathcal{F}_t$, and thus X is $\mathbb{F}$-adapted.

Now we demonstrate that X is not *progressively measurable. Fix $T = \frac{1}{2}$, and we aim to show that*

$$\Delta \notin \mathcal{B}([0, \tfrac{1}{2}]) \otimes \mathcal{F}_{1/2}, \tag{2.4}$$

i.e., $\Delta \notin \mathcal{B}([0, \frac{1}{2}]) \otimes \mathcal{A}$.

Suppose the contrary, i.e., $\Delta \in \sigma(A_n : n \in \mathbb{N}) \otimes \sigma(D_n : n \in \mathbb{N})$ for some sequences $(A_n) \subset \mathcal{B}([0, \frac{1}{2}])$ and $(D_n) \subset \mathcal{A}$. Then by the Fubini theorem and the definition of Δ, we would get that for each $t \in [0, \frac{1}{2}]$,

$$\{t\} = \{\omega \in \Omega : (\omega, t) \in \Delta\} \in \sigma(D_n : n \in \mathbb{N}),$$

which is a contradiction because the singleton $\{t\}$ does not belong to $\mathcal{A}$ (since $\mathcal{A}$ does not contain all singletons of uncountable sets like $[0,1]$). Therefore, (2.4) holds and X is not progressively measurable.

et $D := \bigcup_{n\in\mathbb{N}} D_n$. Noting that

$$\sigma(D_n : n \in \mathbb{N}) \subset \{A \subset \Omega : A = \Gamma \text{ or } A = \Gamma \cup (\Omega \setminus D),\ \text{where } \Gamma \subset D\},$$

we conclude that for any $t \in [0, \frac{1}{2}]$, we have $\{t\} \subset \bigcup_{n\in\mathbb{N}} D_n$, which implies

$$[0, \tfrac{1}{2}] \subset D.$$

This contradicts the fact that the union of the D_n must be countable.

It is interesting to note that the stochastic process $Y \equiv 0$ is a progressively measurable modification of X, because:

$$\mathbb{P}[X_t = 0] = \mathbb{P}[\{\omega \in \Omega : X_t(\omega) = 0\}] = \mathbb{P}[\Omega \setminus \{t\}] = 1.$$

2.3.1 Predictability

Now, we define the concept of *predictability*.

Definition 2.19. *Let $\mathcal{P}_T$ be the σ-algebra on $T \times \Omega$ generated by left-continuous and adapted stochastic processes, i.e., the smallest σ-algebra containing sets of the form*

$$A \times (s,t], \quad \text{with } s,t \in T,\ s < t,\ A \in \mathcal{F}_s.$$

Then:

- *$\mathcal{P}_T$ is called the* predictable σ-algebra*.*
- *A stochastic process X is said to be* predictable *if it is $\mathcal{P}_T$-measurable.*

Sometimes, it is also useful to define the optional σ-algebra*, denoted $\mathcal{O}_T$, i.e., the σ-algebra on $T \times \Omega$ generated by càdlàg adapted stochastic processes. A stochastic process is called* optional *if it is $\mathcal{O}_T$-measurable.*

2.3.2 Stopping times

Suppose a stochastic process represents the value of a game or the price of a stock. We may then be interested in defining a *stopping condition*, for example, the time when a player wins or loses a certain amount, or when the stock price crosses a given threshold. This intuitive notion is formalized in the concept of *stopping times*.

Definition 2.20. *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. A random variable τ, taking values in the time set $T \cup \{+\infty\}$, that is*

$$\tau : \Omega \to T \cup \{+\infty\},$$

is called a random time*. We say that τ is a* stopping time *(or* $\mathbb{F}$-stopping time*, or* Markov moment*) if*

$$\{\tau \leq t\} \in \mathcal{F}_t$$

for all $t \in T$.

Intuitively, a random time τ is a *stopping time* if, at any time $t \in T$, we can determine whether the event $\{\tau \leq t\}$ has occurred using only the information available up to time t.

Stopping times formalize *stopping rules*, which dictate how we control a stochastic process (e.g., stop a game if your reward exceeds a pre-specified threshold).

Unless stated otherwise, we assume all stopping times are finite:

$$\mathbb{P}[\tau = +\infty] = 0.$$

This assumption is mainly technical and simplifies the presentation of results. We also adopt the convention

$$\inf \emptyset := +\infty,$$

so that if a stopping condition is never satisfied, the corresponding stopping time takes the value $+\infty$. We now illustrate the concept of stopping times with a few examples.

Example 2.9 (First entry time)**.** *Let $T = \mathbb{N}$ and let X be an $\mathbb{F}$-adapted stochastic process. Then, for any $B \in \mathcal{B}(\mathbb{R})$, the random variable*

$$\tau_B := \inf\{t \in T : X_t \in B\}$$

is a stopping time. It can be interpreted as the first time the process X enters the set B.

In fact, for any $t \in T$, we have

$$\{\tau_B \leq t\} = \bigcup_{s \in T,\, s \leq t} \{X_s \in B\}.$$

Since X is $\mathbb{F}$-adapted, $\{X_s \in B\} \in \mathcal{F}_s \subseteq \mathcal{F}_t$, and because the union is countable, we conclude that $\{\tau_B \leq t\} \in \mathcal{F}_t$, hence τ_B is a stopping time. One may also ask whether the last hitting time

$$\rho = \sup\{t \in T : X_t \in B\}$$

is a stopping time. Unfortunately, this is not generally the case (except for some degenerate cases). Intuitively, one would need to know the future in order to determine if the process X returns to the set B. (The proof is left as an exercise.) Another natural question is whether Example 2.9 can be extended to continuous-time filtrations. We will show that this is in fact possible, under additional technical conditions. Before presenting such an example, let us first recall some basic facts about stopping times.

Exercise 2.1 (True or False)**.** *Let (S_n) be a simple symmetric random walk on $\mathbb{Z}$ and $\mathcal{F}_n = \sigma(S_0, S_1, \ldots, S_n)$. Which of the following variables are stopping times with respect to $(\mathcal{F}_n)$?*

1. $T_1 = \min\{n \geq 0 \mid S_n = 2017\}$,
2. $T_2 = \min\{n \geq 2017 \mid S_n = S_{n-2017}\}$,
3. $T_3 = \min\{n \geq 0 \mid S_n = S_{n+2017}\}$,
4. $T_4 = \min\{n \geq T_1 \mid S_n = 0\}$,
5. $T_5 = \max\{n \in [0, 2017] \mid S_n = 0\}$,
6. $T_6 = \min\{n \in [0, 2017] \mid \forall m \in [0, 2017],\ S_m \leq S_n\}$.

Solution
The times T_1, T_2, and T_4 are stopping times, since in each case the event $\{T \leq n\}$ depends only on $(S_0, S_1, \ldots, S_n)$. On the other hand, T_3, T_5, and T_6 are not stopping times, because the events $\{T_3 = 0\}$, $\{T_5 = 0\}$, and $\{T_6 = 0\}$ are not $\mathcal{F}_0$-measurable.

Lemma 2.3. *Let $\mathbb{F}$ be a filtration and let τ be a stopping time. Then, for any $t \in T$, the events $\{\tau > t\}$, $\{\tau < t\}$, and $\{\tau = t\}$ all belong to $\mathcal{F}_t$.*

Proof. Since $\{\tau \leq t\} \in \mathcal{F}_t$ and $\mathcal{F}_t$ is a σ-algebra, we have

$$\{\tau > t\} = \{\tau \leq t\}^c \in \mathcal{F}_t.$$

Next, we can write

$$\{\tau < t\} = \bigcup_{s<t,\, s \in \mathbb{Q} \cap T} \{\tau \leq s\} \in \mathcal{F}_t,$$

since $\mathbb{Q} \cap T$ is countable and each $\{\tau \leq s\} \in \mathcal{F}_t$. Finally,

$$\{\tau = t\} = \{\tau \leq t\} \setminus \{\tau < t\} \in \mathcal{F}_t.$$

□

Unfortunately, in continuous time we usually cannot replace the condition $\{\tau \le t\} \in \mathcal{F}_t$ with $\{\tau < t\} \in \mathcal{F}_t$ or $\{\tau = t\} \in \mathcal{F}_t$. Nevertheless, if we impose some additional technical assumptions, then the definitions might become equivalent.

Lemma 2.4. *Let $T = \mathbb{R}_+$ and let $\mathbb{F}$ be right-continuous. Then, a random time τ is a stopping time if and only if for any $t \in T$,*

$$\{\tau < t\} \in \mathcal{F}_t. \tag{2.11}$$

Proof. The first implication (i.e., that stopping time implies 2.11) is already given in previous Lemma. Now assume that τ satisfies condition 2.11. Then, for any $t \in T$,

$$\{\tau \le t\} = \bigcap_{s>t,\, s\in\mathbb{Q}\cap T} \{\tau < s\} \in \mathcal{F}_{t+}.$$

Since $\mathbb{F}$ is right-continuous, we have $\mathcal{F}_{t+} = \mathcal{F}_t$, and the result follows. □

Next, note that if $(\tau_n)_{n\in\mathbb{N}}$ is a sequence of stopping times, then their supremum $\sup \tau_n$ is also a stopping time. Under suitable continuity conditions, the same holds for the infimum, limit superior, and limit inferior. Here, $\sup \tau_n$ denotes the *pointwise supremum* with respect to n, i.e., the supremum is taken for each $\omega \in \Omega$ rather than over n globally.

Lemma 2.5. *Let $(\tau_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{F}$-stopping times. Then, $\sup \tau_n$ is a stopping time. Moreover, if $\mathbb{F}$ is right-continuous, then $\inf \tau_n$, $\limsup \tau_n$, and $\liminf \tau_n$ are also stopping times.*

Proof. It is clear that $\sup \tau_n$ is a random time. Moreover, for any $t \in T$,

$$\{\sup \tau_n \le t\} = \bigcap_{n\in\mathbb{N}} \{\tau_n \le t\} \in \mathcal{F}_t.$$

Now assume $\mathbb{F}$ is right-continuous. Then, using Lemma 2.3, it suffices to show the strict inequalities.

Taking the complement of $\{\inf \tau_n < t\}$, we have

$$\{\inf \tau_n \ge t\} = \bigcap_{n\in\mathbb{N}} \{\tau_n \ge t\} = \bigcap_{n\in\mathbb{N}} \{\tau_n < t\}^c \in \mathcal{F}_t,$$

so that $\{\inf \tau_n < t\} \in \mathcal{F}_t$.

Next,

$$\{\limsup \tau_n < t\} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\tau_m < t - \tfrac{1}{k}\} \in \mathcal{F}_t,$$

$$\{\liminf \tau_n > t\} = \bigcup_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{\tau_m > t + \tfrac{1}{k}\} \in \mathcal{F}_t,$$

which concludes the proof. □

Example 2.10. *Let $T = \mathbb{R}_+$, and let X be an $\mathbb{F}$-adapted stochastic process. Assume that both X and $\mathbb{F}$ are right-continuous. Then, for any open (or closed) set $B \in \mathcal{B}(\mathbb{R})$, the random variable*

$$\tau_B := \inf\{t \in T : X_t \in B\}$$

is a stopping time.

Proof. Let B be an open subset of $\mathbb{R}$. Due to previous Lemma, it is enough to work with strict inequalities (and their complements). We know that for any $t \in T$,

$$\{\tau_B \geq t\} = \{X_s \in B^c,\ s \in T,\ s < t\}.$$

Now, because X is right-continuous, we get

$$\{X_s \in B^c,\ s \in T,\ s < t\} = \bigcap_{\substack{s<t \\ s \in T \setminus \mathbb{Q}}} \{X_s \in B^c\} \in \mathcal{F}_t,$$

which concludes the proof for open subsets. Now, let us assume that B is a closed subset of $\mathbb{R}$. For $\varepsilon > 0$, let B_ε denote the ε-hull of B, i.e., the set of points whose distance from B is strictly less than ε. Since B_ε is open, we know that τ_{B_ε} is a stopping time. Noting that

$$\tau_B = \lim_{\varepsilon \to 0^+} \tau_{B_\varepsilon}$$

and using Lemma 2.5, we get that τ_B is a stopping time, which concludes the proof. □

If time is discrete, we can reformulate the definition of a stopping time using equality instead of inequality.

Proposition 2.2. *Let $\mathbb{F}$ be a discrete-time filtration. Then τ is an $\mathbb{F}$-stopping time if and only if*

$$\{\tau = t\} \in \mathcal{F}_t$$

for any $t \in T$.

Proof. If τ is a stopping time, then

$$\{\tau = t\} = \{\tau \leq t\} \setminus \bigcup_{\substack{s \in T \\ s<t}} \{\tau \leq s\},$$

which proves the claim. On the other hand, we have

$$\{\tau \leq t\} = \bigcup_{\substack{s \leq t \\ s \in T}} \{\tau = s\},$$

so the converse implication also holds. □

Finally, let us show that one can perform some basic operations on stopping times.

Proposition 2.3. *Let $\mathbb{F}$ be a filtration and let τ, ρ be two $\mathbb{F}$-stopping times. Then:*

1. *If τ and ρ are non-negative, then $\tau + \rho$ is a stopping time.*
2. *$\tau \wedge \rho := \min\{\tau, \rho\}$ is a stopping time.*
3. *$\tau \vee \rho := \max\{\tau, \rho\}$ is a stopping time.*

Proof. To prove 1), it is enough to note that

$$\{\tau + \rho > t\} = \{\tau = 0\} \setminus \{\rho > t\} \cup \bigcup_{\substack{s \in T \setminus \mathbb{Q} \\ s \in (0, +\infty)}} \{\tau > s\} \cap \{\rho > t - s\}.$$

Next, to prove 2) and 3), it is enough to note that

$$\{\tau \wedge \rho \leq t\} = \{\tau \leq t\} \cap \{\rho \leq t\}$$

and

$$\{\tau \vee \rho \leq t\} = \{\tau \leq t\} \cup \{\rho \leq t\}.$$

□

We now introduce key concepts and notation related to stopping times. Assuming the stochastic processes are progressively measurable, stopping times can be used to halt a process or define a random sample. Additionally, one can define the σ-algebra representing the information available up to a stopping time.

Definition 2.21. *Let τ be a (finite) $\mathbb{F}$-stopping time and let X be a progressively $\mathbb{F}$-measurable stochastic process.*

1. *A random time sample from the stochastic process X picked at τ, and denoted by X_τ, is given by*
$$X_\tau(\omega) := X_{\tau(\omega)}(\omega) \quad \text{for } \omega \in \Omega.$$
2. *A process stopped at τ, denoted $X_\tau = (X^t_\tau)_{t\in T}$, is given by*
$$X^t_\tau := X^{\wedge\tau}_t \quad \text{for } t \in T.$$
3. *The σ-algebra at a stopping time τ, denoted by $\mathcal{F}_\tau$, is given by*
$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, \forall t \in T\}.$$

Note that, in general, the random variable X_τ may not be measurable, and hence not $\mathcal{F}_\tau$-measurable. However, if the process is progressively measurable, X_τ is indeed measurable. We also need to verify that $\mathcal{F}_\tau$ itself forms a σ-algebra.

Proposition 2.4. *Let τ be a (finite) $\mathbb{F}$-stopping time. Then, the family of sets $\mathcal{F}_\tau$ is a σ-algebra, and τ is $\mathcal{F}_\tau$-measurable. Moreover, if the stochastic process X is progressively $\mathbb{F}$-measurable, then X_τ is $\mathcal{F}_\tau$-measurable (and X^t_τ is $\mathcal{F}_t$-measurable).*

Proof. From the definition of τ, we get $\Omega \in \mathcal{F}_\tau$, as $\{\tau \leq t\} \in \mathcal{F}_t$ for any $t \in T$. Now, let us assume that $A \in \mathcal{F}_\tau$. Then, for any $t \in T$, we get

$$A^c \cap \{\tau \leq t\} = \{\tau \leq t\} \setminus (A \cap \{\tau \leq t\}) \in \mathcal{F}_t;$$

and consequently $A^c \in \mathcal{F}_\tau$. Next, we know that for a sequence (A_n), we have that

$$\bigcup_n A_n \in \mathcal{F}_\tau$$

since $\mathcal{F}_\tau$ is closed under countable unions, and similar results hold for intersections, concluding the proof. Next, we know that for a sequence $(A_n)_{n\in\mathbb{N}}$, where $A_n \in \mathcal{F}_\tau$, we get

$$\left(\bigcup_{n\in\mathbb{N}} A_n\right) \cap \{\tau \leq t\} = \bigcup_{n\in\mathbb{N}} (A_n \cap \{\tau \leq t\}) \in \mathcal{F}_t$$

for any $t \in T$, which implies $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}_\tau$. This implies that $\mathcal{F}_\tau$ is indeed a σ-algebra. Now, to prove that τ is $\mathcal{F}_\tau$-measurable, it is enough to show that for any $s \in \mathbb{R}$, the event $\{\tau \leq s\}$ belongs to $\mathcal{F}_\tau$. We can assume that $s \in T$ (since τ has values in T). Then, for $A_s = \{\tau \leq s\}$, we get

$$A_s \cap \{\tau \leq t\} = \{\tau \leq t \wedge s\} \in \mathcal{F}_{t\wedge s} \subset \mathcal{F}_t$$

for any $t \in T$, which concludes this part of the proof. Finally, we need to show that for $\Gamma \in \mathcal{B}(\mathbb{R})$ and any $t \in T$, we get

$$\{X_\tau \in \Gamma\} \cap \{\tau \leq t\} \in \mathcal{F}_t.$$

Let us fix $t \in T$. Noting that $\tau \wedge t$ is a stopping time, we get that the map $Z : \Omega \to \Omega \times [0, t]$ given by

$$Z(\omega) = (\omega, \tau(\omega) \wedge t)$$

is $\mathcal{F}$-measurable (as its margins are measurable). From the progressive measurability of X, we know that the map $W : \Omega \times [0, t] \to \mathbb{R}$ given by

$$W(\omega, s) = X_s(\omega)$$

is $\mathcal{F}_t \times \mathcal{B}(\mathbb{R})$-measurable. Consequently, the map $V : \Omega \to \mathbb{R}$ given by

$$V(\omega) = W(Z(\omega)) = X_{\tau(\omega)\wedge t}(\omega)$$

is $\mathcal{F}_t$-measurable. Now, we note that

$$\{X_\tau \in \Gamma\} \cap \{\tau \leq t\} = \{X_{\tau\wedge t} \in \Gamma\} \cap \{\tau \leq t\} = V^{-1}(\Gamma) \cap \{\tau \leq t\} \in \mathcal{F}_t,$$

for any $t \in T$ (as the union of two measurable events), which concludes the proof. □

2.4 Martingales

Martingales are discrete stochastic processes that generalize the process of summing up IID random variables. They are powerful tools with many applications.

2.4.1 Definition and examples

In this subsection, we introduce and study an important class of stochastic processes called *martingales*. Martingales naturally arise in many areas of stochastic process theory and are particularly useful in the analysis of Brownian motion (BM). Here, the index set T may be any interval in $\mathbb{Z}_+$ or $\mathbb{R}_+$.

Definition 2.22 (Martingale). *An $(\mathcal{F}_t)$-adapted, real-valued process $M = (M_t)_{t\in T}$ is called a* martingale *(with respect to the filtration $(\mathcal{F}_t)$) if:*

1. $\mathbb{E}[\,|M_t|\,] < \infty$ *for all* $t \in T$*;*
2. $\mathbb{E}(M_t \mid \mathcal{F}_s) = M_s \quad$ *a.s. for all* $s \leq t$.

If property 2. holds with "$\geq$" (resp. "$\leq$") instead of "$=$", then M is called a submartingale *(resp.* supermartingale*).*

Intuitively, a martingale is a stochastic process that is "constant on average." Given all information up to time s, the best prediction for the process at a later time $t \geq s$ is the current value M_s. In particular, it satisfies the property 2. implies that

$$\mathbb{E}[M_t] = \mathbb{E}[M_0], \qquad \forall t \in T.$$

Similarly, a *submartingale* is a process that tends to increase on average, while a *supermartingale* tends to decrease on average. Clearly, M is a submartingale if and only if $-M$ is a supermartingale, and M is a martingale if it is both a submartingale and a supermartingale. Using the basic properties of conditional expectation, we obtain the following results and examples.

Example 2.11. *Let $(X_n,\, n = 1, \dots)$ be a sequence of i.i.d. real-valued integrable random variables. Take, e.g., the filtration $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then*

$$M_n = \sum_{k=1}^{n} X_k$$

is a martingale if $\mathbb{E}[X_1] = 0$, a submartingale if $\mathbb{E}[X_1] > 0$, and a supermartingale if $\mathbb{E}[X_1] < 0$. The process $M = (M_n)_n$ can be viewed as a random walk on the real line.

If $\mathbb{E}[X_1] = 0$, but X_1 is square integrable, then

$$M_n^0 = M_n^2 - n\,\mathbb{E}[X_1^2]$$

is a martingale.

Example 2.12. *(Doob martingale) Suppose that X is an integrable random variable and $(\mathcal{F}_t)_{t\in T}$ a filtration. For $t \in T$, define*

$$M_t = \mathbb{E}[X \mid \mathcal{F}_t],$$

or, more precisely, let M_t be a version of $\mathbb{E}[X \mid \mathcal{F}_t]$. Then $M = (M_t)_{t\in T}$ is an $(\mathcal{F}_t)$-martingale, and M is uniformly integrable.

Exercise 2.2. *Let U be a random variable uniformly distributed on $[0,1]$. Let $f : [0,1] \to \mathbb{R}$ be a uniformly Lipschitz function. For $n \in \mathbb{N}$ and $k = 0, \dots, 2^n - 1$, set*

$$I_{k,n} = \Big[k2^{-n}, (k+1)2^{-n}\Big[, \qquad \mathcal{F}_n = \sigma\Big\{I_{k,n} : 0 \leq k \leq 2^n - 1\Big\}.$$

Define

$$X_n = \sum_{k=0}^{2^n-1} \frac{f\Big((k+1)2^{-n}\Big) - f\Big(k2^{-n}\Big)}{2^{-n}} \mathbf{1}_{\{U\in I_{k,n}\}}.$$

1. ***Show that (X_n) is a martingale with respect to $(\mathcal{F}_n)$.***
   Since f is K-Lipschitz,

$$\Big|f\Big((k+1)2^{-n}\Big) - f\Big(k2^{-n}\Big)\Big| \leq K2^{-n},$$

hence

$$\mathbb{E}\Big[|X_n|\Big] \leq \sum_{k=0}^{2^n-1} K\,P\{U \in I_{k,n}\} = K.$$

Thus X_n is integrable.

Each interval $I_{\ell,n}$ is the disjoint union of $I_{2\ell,n+1}$ and $I_{2\ell+1,n+1}$, of equal length, and

$$\mathbb{E}\Big[\mathbf{1}_{\{U\in I_{2\ell,n+1}\}} \mid \mathcal{F}_n\Big] = \mathbb{E}\Big[\mathbf{1}_{\{U\in I_{2\ell+1,n+1}\}} \mid \mathcal{F}_n\Big] = \tfrac{1}{2}\mathbf{1}_{\{U\in I_{\ell,n}\}}.$$

Separating the even and odd terms, we obtain

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \sum_{\ell=0}^{2^n-1} \frac{f\Big((2\ell+2)2^{-(n+1)}\Big) - f\Big(2\ell 2^{-(n+1)}\Big)}{2^{-n}} \mathbf{1}_{\{U\in I_{\ell,n}\}} = X_n.$$

Hence $(X_n, \mathcal{F}_n)$ is a martingale.

2. ***Almost sure convergence.***
We have $\mathbb{E}[|X_n|] \le K$, so (X_n) is bounded in L^1. By the martingale convergence theorem, $X_n \to X_\infty$ almost surely.

3. ***Convergence in L^1.***
Since $|X_n(\omega)| \le K$ for all ω, the sequence (X_n) is uniformly integrable, and therefore $X_n \to X_\infty$ in L^1.

4. ***For all $0 \le a < b \le 1$, compute $\mathbb{E}\Big[X_\infty \mathbf{1}_{\{U\in[a,b]\}}\Big]$.***
For all n,

$$X_n \mathbf{1}_{\{U\in[a,b]\}} = \sum_{\substack{0\le k\le 2^n-1 \\ I_{k,n}\cap[a,b]\neq\emptyset}} \frac{f((k+1)2^{-n}) - f(k2^{-n})}{2^{-n}} \mathbf{1}_{\{U\in I_{k,n}\}}.$$

Let $k_-(n)$ and $k_+(n)$ be the smallest and largest indices such that $I_{k,n} \cap [a,b] \neq \emptyset$. Taking expectations gives

$$\mathbb{E}\Big[X_n \mathbf{1}_{\{U\in[a,b]\}}\Big] = f\Big((k_+(n)+1)2^{-n}\Big) - f\Big(k_-(n)2^{-n}\Big).$$

As $n\to\infty$, we have $k_-(n)2^{-n} \to a$ and $(k_+(n)+1)2^{-n} \to b$, so

$$\mathbb{E}\Big[X_\infty \mathbf{1}_{\{U\in[a,b]\}}\Big] = f(b) - f(a).$$

5. ***Case $f \in C^1$.***
By the fundamental theorem of calculus,

$$f(b) - f(a) = \int_a^b f'(u)\,du = \int_a^b X_\infty(u)\,du \qquad (\textit{with } U(\omega)=\omega).$$

It follows that

$$X_\infty(\omega) = f'(\omega) \quad \textit{for almost every } \omega.$$

In the next section we give the theory for discrete-time martingales.

2.4.2 Discrete-time martingales

In this section we restrict ourselves to martingales (and filtrations) that are indexed by (a subinterval of) $\mathbb{Z}_+$. We will assume the underlying filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ to be fixed. Note that as a consequence, it only makes sense to consider $\mathbb{Z}_+$-valued stopping times. In discrete time, τ is a stopping time with respect to the filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ if

$$\{\tau \le n\} \in \mathcal{F}_n \quad \text{for all } n \in \mathbb{Z}_+.$$

Martingale transforms

If the value of a process at time n is already known at time $n-1$, we call the process *predictable*. The precise definition is as follows.

Definition 2.23 (Predictable process)**.** *We call a discrete-time process X predictable with respect to the filtration $(\mathcal{F}_n)_n$ if X_n is $\mathcal{F}_{n-1}$-measurable for every n.*

In the following definition we introduce discrete-time 'integrals'. This is a useful tool in martingale theory.

Definition 2.24. *Let M and X be two discrete-time processes. We define the process $X \cdot M$ by*

$$(X \cdot M)_0 = 0, \qquad (X \cdot M)_n = \sum_{k=1}^{n} X_k\,(M_k - M_{k-1}), \quad n \geq 1.$$

We call $X \cdot M$ the discrete integral *of X with respect to M. If M is a (sub-, super-)martingale, it is often called the* martingale transform *of M by X.*

Martingale transforms can be regarded as a discrete analogue of the Itô integral. Predictability is crucial in defining the Itô integral. The following lemma illustrates their usefulness: the integral of a predictable process with respect to a martingale is itself a martingale.

Lemma 2.6. *Let X be a predictable process, such that for all n there exists a constant K_n with $|X_1|, \ldots, |X_n| \leq K_n$. If M is a martingale, then $X \cdot M$ is a martingale. If M is a submartingale (resp. a supermartingale) and X is non-negative, then $X \cdot M$ is a submartingale (resp. supermartingale) as well.*

Proof. Put $Y = X \cdot M$. Clearly Y is adapted. Since X is bounded, say $|X_n| \leq K$ a.s. for all n, we have

$$\mathbb{E}[|Y_n|] \leq 2K_n \sum_{k \leq n} \mathbb{E}[|M_k|] < \infty.$$

Now suppose first that M is a submartingale and X is non-negative. Then a.s.

$$\mathbb{E}(Y_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(Y_{n-1} + X_n(M_n - M_{n-1}) \mid \mathcal{F}_{n-1}) = Y_{n-1} + X_n\,\mathbb{E}(M_n - M_{n-1} \mid \mathcal{F}_{n-1}) \geq Y_{n-1}.$$

Consequently, Y is a submartingale. If M is a martingale, the last inequality becomes an equality, regardless of the sign of X_n, which implies that Y is a martingale as well. □

Using this lemma, it is easy to see that a stopped (sub-, super-)martingale is again a (sub-, super-)martingale.

Theorem 2.12. *Let M be a $(\mathcal{F}_n)_n$ (sub-, super-)martingale and τ an $(\mathcal{F}_n)_n$-stopping time. Then the stopped process M^τ is an $(\mathcal{F}_n)_n$ (sub-, super-)martingale as well.*

Proof. Define the process X by $X_n = \mathbf{1}_{\{\tau \geq n\}}$. Verify that

$$M^\tau = M_0 + X \cdot M.$$

Since τ is a stopping time, we have $\{\tau \geq n\} = \{\tau \leq n-1\}^c \in \mathcal{F}_{n-1}$. Hence the process X is predictable. It is also a bounded process, and so the statement follows from the preceding lemma.

We also give a direct proof. First note that

$$\mathbb{E}[|M_t^\tau|] = \mathbb{E}[|M_{t\wedge\tau}|] \le \sum_{n=0}^{t} \mathbb{E}[|M_n|] < \infty, \qquad t \in T.$$

Write

$$M_t^\tau = M_{t\wedge\tau} = \sum_{n=0}^{t-1} \mathbf{1}_{\{\tau=n\}} M_{t\wedge\tau} + \mathbf{1}_{\{\tau\ge t\}} M_t = \sum_{n=0}^{t-1} M_n \mathbf{1}_{\{\tau=n\}} + M_t \mathbf{1}_{\{\tau\ge t\}}.$$

□

Taking conditional expectations yields

$$\mathbb{E}(M_{t\wedge\tau} \mid \mathcal{F}_{t-1}) = \sum_{n=0}^{t-1} M_n \mathbf{1}_{\{\tau=n\}} + \mathbf{1}_{\{\tau\ge t\}}\, \mathbb{E}(M_t \mid \mathcal{F}_{t-1}),$$

since $\{\tau \ge t\} \in \mathcal{F}_{t-1}$. The rest follows immediately. □

The following result can be viewed as a first version of the so-called *optional sampling theorem.*

Theorem 2.13 (Optional sampling theorem, simple version)**.** *Let M be a (sub)martingale and let σ, τ be two stopping times such that $\sigma \le \tau \le K$, for some constant $K > 0$. Then*

$$\mathbb{E}(M_\tau \mid \mathcal{F}_\sigma)\,(\ge)\, M_\sigma, \quad a.s. \tag{2.12}$$

An adapted integrable process M is a martingale if and only if

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_\sigma],$$

for any pairs of bounded stopping times $\sigma \le \tau$.

Proof. Suppose first that M is a martingale. Define the predictable process

$$X_n = \mathbf{1}_{\{\tau\ge n\}} - \mathbf{1}_{\{\sigma\ge n\}}.$$

Note that $X_n \ge 0$ a.s. Hence, $X \cdot M = M_\tau - M_\sigma$. By Lemma 2.6 the process $X \cdot M$ is a martingale, hence

$$\mathbb{E}(M_{\tau\wedge n} - M_{\sigma\wedge n}) = \mathbb{E}[(X\cdot M)_n] = 0, \quad \forall n.$$

Since $\sigma \le \tau \le K$ a.s., it follows that

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_{\tau\wedge K}] = \mathbb{E}[M_{\sigma\wedge K}] = \mathbb{E}[M_\sigma].$$

Now we take $A \in \mathcal{F}_\sigma$ and define the 'truncated' random times

$$\sigma_A = \sigma \mathbf{1}_A + K\mathbf{1}_{A^c}, \qquad \tau_A = \tau\mathbf{1}_A + K\mathbf{1}_{A^c}. \tag{2.13}$$

By definition of $\mathcal{F}_\sigma$ it holds for every n that

$$\{\sigma_A \le n\} = (A \cap \{\sigma \le n\}) \cup (A^c \cap \{K \le n\}) \in \mathcal{F}_n,$$

and so σ_A is a stopping time. Similarly, τ_A is a stopping time and clearly $\sigma_A \le \tau_A \le K$ a.s.

By the first part of the proof, it follows that $\mathbb{E}[M_{\sigma_A}] = \mathbb{E}[M_{\tau_A}]$, in other words

$$\int_A M_\sigma \, d\mathbb{P} + \int_{A^c} M_K \, d\mathbb{P} = \int_A M_\tau \, d\mathbb{P} + \int_{A^c} M_K \, d\mathbb{P}. \tag{2.14}$$

Hence

$$\int_A M_\sigma \, d\mathbb{P} = \int_A M_\tau \, d\mathbb{P}.$$

Since $A \in \mathcal{F}_\sigma$ is arbitrary, this implies $\mathbb{E}(M_\tau \mid \mathcal{F}_\sigma) = M_\sigma$ a.s. (Recall that M_σ is $\mathcal{F}_\sigma$-measurable. Let M be an adapted process with $\mathbb{E}[M_\sigma] = \mathbb{E}[M_\tau]$ for each bounded pair $\sigma \le \tau$ of stopping times. Take $\sigma = n-1$ and $\tau = n$ in the preceding and use truncated stopping times σ_A and τ_A for $A \in \mathcal{F}_{n-1}$. Then, for $A \in \mathcal{F}_{n-1}$ and stopping times σ_A and τ_A implies that

$$\mathbb{E}(M_n \,|\, \mathcal{F}_{n-1}) = M_{n-1} \quad \text{a.s.}$$

In other words, M is a martingale.

If M is a submartingale, the same reasoning applies, but with inequalities instead of equalities.

As in the previous lemma, we will also give a direct proof First note that

$$\mathbb{E}(M_K \,|\, \mathcal{F}_n) \ge M_n \quad \text{a.s.} \iff \mathbb{E}(M_K \mathbf{1}_F) \ge \mathbb{E}(M_n \mathbf{1}_F)\,, \quad \forall\, F \in \mathcal{F}_n.$$

We will first show that

$$\mathbb{E}(M_K \,|\, \mathcal{F}_\sigma) \ge M_\sigma \quad \text{a.s.}$$

Similarly, it is sufficient to show that

$$\mathbb{E}[\mathbf{1}_F M_\sigma] \le \mathbb{E}[\mathbf{1}_F M_K] \quad \text{for all } F \in \mathcal{F}_\sigma.$$

Now,

$$\mathbb{E}[\mathbf{1}_F M_\sigma] = \mathbb{E}\Big[\mathbf{1}_F \Big(\sum_{n=0}^{K} \mathbf{1}_{\{\sigma=n\}} + \mathbf{1}_{\{\sigma>K\}}\Big) M_\sigma\Big] = \sum_{n=0}^{K} \mathbb{E}[\mathbf{1}_{F\cap\{\sigma=n\}} M_n].$$

Hence

$$\mathbb{E}[\mathbf{1}_F M_\sigma] \le \sum_{n=0}^{K} \mathbb{E}[\mathbf{1}_{F\cap\{\sigma=n\}} M_K] = \mathbb{E}\Big[\mathbf{1}_F \Big(\sum_{n=0}^{K} \mathbf{1}_{\{\sigma=n\}} + \mathbf{1}_{\{\sigma>K\}}\Big) M_K\Big] = \mathbb{E}[\mathbf{1}_F M_K].$$

In the second and fourth equalities we have used that

$$\mathbb{E}[\mathbf{1}_{\{\sigma>K\}} M_\sigma] = \mathbb{E}[\mathbf{1}_{\{\sigma>K\}} M_K] = 0,$$

since $\mathbb{P}(\sigma > K) = 0$. The fact that $F \cap \{\sigma = n\} \in \mathcal{F}_n$ (why?). This gives the resiults.

We have

$$\mathbb{E}(M_K^\tau \mid \mathcal{F}_\sigma) \ge M_\sigma^\tau.$$

Now, note that $M_K^\tau = M_\tau$ a.s. and $M_\sigma^\tau = M_\sigma$ a.s. (why?). Note that we may in fact allow that $\sigma \le \tau \le K$ a.s. Later on we need $\sigma \le \tau$ everywhere.

Inequalities

Markov's inequality implies that if M is a discrete-time process, then

$$\lambda\,\mathbb{P}(M_n \geq \lambda) \leq \mathbb{E}[|M_n|], \quad \forall n \in \mathbb{Z}_+,\ \lambda > 0.$$

Doob's classical submartingale inequality states that for submartingales we have a much stronger result.

Theorem 2.14 (Doob's submartingale inequality)**.** *Let M be a submartingale. For all $\lambda > 0$ and $n \in \mathbb{N}$,*

$$\lambda\,\mathbb{P}\Big(\max_{k\leq n} M_k \geq \lambda\Big) \leq \mathbb{E}[M_n\,\mathbf{1}_{\{\max_{k\leq n} M_k \geq \lambda\}}] \leq \mathbb{E}[|M_n|].$$

Proof. Define $\tau = n \wedge \inf\{k : M_k \geq \lambda\}$. This is a stopping time with $\tau \leq n$. We have $\mathbb{E}[M_n] \geq \mathbb{E}[M_\tau]$. It follows that

$$\mathbb{E}[M_n] \geq \mathbb{E}[M_\tau \mathbf{1}_{\{\max_{k\leq n} M_k \geq \lambda\}}] + \mathbb{E}[M_\tau \mathbf{1}_{\{\max_{k\leq n} M_k < \lambda\}}].$$

Since $M_\tau \geq \lambda$ on $\{\max_{k\leq n} M_k \geq \lambda\}$, this gives

$$\mathbb{E}[M_n] \geq \lambda\,\mathbb{P}\Big(\max_{k\leq n} M_k \geq \lambda\Big) + \mathbb{E}[M_n\,\mathbf{1}_{\{\max_{k\leq n} M_k < \lambda\}}].$$

This yields the first inequality. The second one is obvious. □

Theorem 2.15 (Doob's L^p inequality)**.** *If M is a martingale or a non-negative submartingale and $p > 1$, then for all $n \in \mathbb{N}$ we have*

$$\mathbb{E}\Big[\max_{k\leq n} |M_k|^p\Big] \leq \Big(\frac{p}{p-1}\Big)^p\,\mathbb{E}[|M_n|^p],$$

provided $M \in L^p$.

Proof. Define $M^* = \max_{k\leq n} |M_k|$. Assume that M is defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any $m \in \mathbb{N}$ we have

$$\begin{aligned}
\mathbb{E}[(M^* \wedge m)^p] &= \int_\Omega (M^*(\omega) \wedge m)^p\,d\mathbb{P}(\omega) \\
&= \int_\Omega \int_0^{M^*(\omega)\wedge m} p x^{p-1}\,dx\,d\mathbb{P}(\omega) \\
&= \int_\Omega \int_0^m p x^{p-1} \mathbf{1}_{\{M^*(\omega)\geq x\}}\,dx\,d\mathbb{P}(\omega) \\
&= \int_0^m p x^{p-1}\,\mathbb{P}(M^* \geq x)\,dx.
\end{aligned}$$

where we have used Fubini's theorem in the last equality (non-negative integrand).

By conditional Jensen's inequality, $|M|$ is a submartingale, and so we can apply Doob's submartingale inequality to estimate $\mathbb{P}(M^* \geq x)$. Thus

$$\mathbb{P}(M^* \geq x) \leq \frac{\mathbb{E}\big[|M_n| \mathbf{1}_{\{M^* \geq x\}}\big]}{x}.$$

Hence

$$\begin{aligned}\mathbb{E}[(M^* \wedge m)^p] &\le \int_0^m p x^{p-2}\, \mathbb{E}\Big[|M_n| \mathbf{1}_{\{M^* \ge x\}}\Big]\, dx \\ &= \int_\Omega |M_n(\omega)| \int_0^{M^*(\omega)\wedge m} p x^{p-2}\, dx\, d\mathbb{P}(\omega) \\ &= \frac{p}{p-1}\, \mathbb{E}\Big[|M_n|(M^* \wedge m)^{p-1}\Big].\end{aligned}$$

By Hölder's inequality, it follows that with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E}[(M^* \wedge m)^p] \le \frac{p}{p-1} \Big(\mathbb{E}[|M_n|^p]\Big)^{1/p} \Big(\mathbb{E}[(M^* \wedge m)^{(p-1)q}]\Big)^{1/q}.$$

Since $p > 1$, we have $q = \frac{p}{p-1}$, hence $(p-1)q = p$. Therefore

$$\mathbb{E}[(M^* \wedge m)^p] \le \frac{p}{p-1} \Big(\mathbb{E}[|M_n|^p]\Big)^{1/p} \Big(\mathbb{E}[(M^* \wedge m)^p]\Big)^{(p-1)/p}.$$

Now take the p-th power of both sides and cancel common factors. We obtain

$$\mathbb{E}[(M^* \wedge m)^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|M_n|^p].$$

Finally, letting $m \to \infty$ and using monotone convergence completes the proof. □

2.4.3 Doob decomposition

Any adapted, integrable process X can be decomposed into the sum of a martingale and a predictable process. This is known as the *Doob decomposition* of X.

Theorem 2.16 (Doob decomposition)**.** *Let X be an adapted, integrable process. There exists a martingale M and a predictable process A, such that $A_0 = M_0 = 0$ and*

$$X = X_0 + M + A.$$

The processes M and A are a.s. unique. Moreover, X is a submartingale if and only if A is a.s. increasing (i.e. $\mathbb{P}(A_n \le A_{n+1}) = 1$).

Proof. Suppose first that there exist a martingale M and a predictable process A such that $A_0 = M_0 = 0$ and $X = X_0 + M + A$. The martingale property of M and predictability of A show that a.s.

$$\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = A_n - A_{n-1}, \quad \text{a.s.} \tag{2.15}$$

Since $A_0 = 0$ it follows that

$$A_n = \sum_{k=1}^{n} \mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}), \qquad n \ge 1, \tag{2.16}$$

and hence $M_n = X_n - A_n - X_0$. This shows that M and A are a.s. unique.

Conversely, given a process X, equation (2.16) defines a predictable process A. It is easily seen that the process M defined by $M = X - A - X_0$ is a martingale. This proves the existence of the decomposition.

Finally, equation (2.15) shows that X is a submartingale if and only if A is increasing. □

An important application of the Doob decomposition is the following.

Corollary 2.1. *Let M be a martingale with $\mathbb{E}[M_n^2] < \infty$ for all n. Then there exists an a.s. unique predictable, increasing process A with $A_0 = 0$ such that*

$$M^2 - A \quad \text{is a martingale.}$$

Moreover, the random variable $A_n - A_{n-1}$ is a version of the conditional variance of M_n given $\mathcal{F}_{n-1}$, i.e.

$$A_n - A_{n-1} = \mathbb{E}\Big[(M_n - \mathbb{E}(M_n \mid \mathcal{F}_{n-1}))^2 \,\Big|\, \mathcal{F}_{n-1}\Big] = \mathbb{E}\Big[(M_n - M_{n-1})^2 \mid \mathcal{F}_{n-1}\Big], \quad a.s.$$

It follows that Pythagoras' theorem holds for square integrable martingales:

$$\mathbb{E}[M_n^2] = \mathbb{E}[M_0^2] + \sum_{k=1}^{n} \mathbb{E}\Big[(M_k - M_{k-1})^2\Big].$$

The process A is called the predictable quadratic variation process *of M and is often denoted by $\langle M \rangle$.*

Proof. By conditional Jensen, it follows that M^2 is a submartingale. Hence Theorem 2.16 applies. The only thing left to prove is the statement about conditional variance. Since M is a martingale, we have a.s.

$$\begin{aligned}
\mathbb{E}\Big[(M_n - M_{n-1})^2 \mid \mathcal{F}_{n-1}\Big] &= \mathbb{E}\Big(M_n^2 - 2M_nM_{n-1} + M_{n-1}^2 \,\Big|\, \mathcal{F}_{n-1}\Big) \\
&= \mathbb{E}[M_n^2 \mid \mathcal{F}_{n-1}] - 2M_{n-1}\mathbb{E}[M_n \mid \mathcal{F}_{n-1}] + M_{n-1}^2 \\
&= \mathbb{E}[M_n^2 \mid \mathcal{F}_{n-1}] - M_{n-1}^2 \\
&= \mathbb{E}[M_n^2 - M_{n-1}^2 \mid \mathcal{F}_{n-1}] \\
&= A_n - A_{n-1}.
\end{aligned}$$

□

Using the Doob decomposition in combination with the submartingale inequality yields the following result.

Theorem 2.17. *Let X be a sub- or supermartingale. For all $\lambda > 0$ and $n \in \mathbb{Z}_+$,*

$$\lambda\, \mathbb{P}\Big(\max_{k\le n} |X_k| \ge 3\lambda\Big) \le 4\,\mathbb{E}[|X_0|] + 3\,\mathbb{E}[|X_n|].$$

Proof. Suppose that X is a submartingale. By the Doob decomposition theorem there exist a martingale M and an increasing, predictable process A such that $M_0 = A_0 = 0$ and

$$X = X_0 + M + A.$$

By the triangle inequality and the fact that A is increasing,

$$\mathbb{P}\Big(\max_{k\le n} |X_k| \ge 3\lambda\Big) \le \mathbb{P}(|X_0| \ge \lambda) + \mathbb{P}\Big(\max_{k\le n} |M_k| \ge \lambda\Big) + \mathbb{P}(A_n \ge \lambda).$$

Hence, by Markov's inequality and the submartingale inequality (note that $|M_n|$ is a submartingale!),

$$\lambda\, \mathbb{P}\Big(\max_{k\le n} |X_k| \ge 3\lambda\Big) \le \mathbb{E}[|X_0|] + \mathbb{E}[|M_n|] + \mathbb{E}[A_n].$$

Since $M_n = X_n - X_0 - A_n$, the right-hand side is bounded by

$$2\,\mathbb{E}[|X_0|] + \mathbb{E}[|X_n|] + 2\,\mathbb{E}[A_n].$$

We know that A_n is given by $(E(X_n - X_{n-1}|\mathcal{F}_{n-1}) = A_n - A_{n-1} \quad a.s.$ Taking expectations in the latter expression shows that

$$\mathbb{E}[A_n] = \mathbb{E}[X_n] - \mathbb{E}[X_0] \leq \mathbb{E}[|X_n|] + \mathbb{E}[|X_0|].$$

This completes the proof. □

Convergence theorems

Let M be a supermartingale, and consider a compact interval $[a, b] \subset \mathbb{R}$. The *number of upcrossings* of $[a, b]$ by the process up to time n is the number of times the process moves from a level below a to a level above b. A precise definition is given below.

Definition 2.25 (Number of upcrossings)**.** *The number $U_n[a, b]$ is the largest value $k \in \mathbb{Z}_+$ such that there exist*

$$0 \leq s_1 < t_1 < s_2 < \cdots < s_k < t_k \leq n$$

with $M_{s_i} < a$ and $M_{t_i} > b$, for $i = 1, \ldots, k$.

First we define the *limit σ-algebra*

$$\mathcal{F}_\infty = \sigma\left(\bigcup_n \mathcal{F}_n\right).$$

Lemma 2.7 (Doob's upcrossing lemma)**.** *Let M be a supermartingale. Then for all $a < b$, the number of upcrossings $U_n[a, b]$ of the interval $[a, b]$ by M up to time n is an $\mathcal{F}_n$-measurable random variable and satisfies*

$$(b - a)\,\mathbb{E}[U_n[a, b]] \leq \mathbb{E}[(M_n - a)^-].$$

The total number of upcrossings $U_\infty[a, b]$ is $\mathcal{F}_\infty$-measurable.

Theorem 2.18 (Doob's martingale convergence theorem)**.** *If M is a supermartingale that is bounded in L^1, then M_n converges a.s. to a finite $\mathcal{F}_\infty$-measurable limit M_∞ as $n \to \infty$, with $\mathbb{E}|M_\infty| < \infty$.*

Proof. Assume that M is defined on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that $M(\omega)$ does not converge to a limit in $[-\infty, \infty]$. Then there exist two rationals $a < b$ such that

$$\liminf_n M_n(\omega) < a < b < \limsup_n M_n(\omega).$$

In particular, we must have $U_1[a, b](\omega) = \infty$. By Doob's upcrossing lemma $\mathbb{P}\{U_1[a, b] = \infty\} = 0$. Now note that

$$A := \{\omega \mid M(\omega) \text{ does not converge in } [-\infty, \infty]\} \subset \bigcup_{\substack{a,b \in \mathbb{Q} \\ a<b}} \{\omega \mid U_1[a, b](\omega) = \infty\}.$$

Hence

$$\mathbb{P}(A) \leq \sum_{\substack{a,b \in \mathbb{Q} \\ a<b}} \mathbb{P}\{U_1[a, b] = \infty\} = 0.$$

This implies that M_n a.s. converges to a limit $M_\infty \in [-\infty, \infty]$.

Moreover, by Fatou's lemma,

$$\mathbb{E}|M_\infty| = \mathbb{E}\Big(\liminf_n |M_n|\Big) \leq \liminf_n \mathbb{E}|M_n| \leq \sup_n \mathbb{E}|M_n| < \infty.$$

It follows that M_∞ is a.s. finite and integrable. Since M_n is $\mathcal{F}_n$-measurable, it is also $\mathcal{F}_\infty$-measurable. Finally, $M_\infty = \lim_{n\to\infty} M_n$ is the limit of $\mathcal{F}_\infty$-measurable functions, hence $\mathcal{F}_\infty$-measurable as well. □ □

If the supermartingale M is not only bounded in L^1 but also uniformly integrable, then in addition to a.s. convergence we have convergence in L^1. Moreover, in this case, the whole sequence $(M_n)_{n\geq 1}$ is a supermartingale.

Theorem 2.19. *Let M be a supermartingale that is bounded in L^1. Then $M_n \to M_\infty$ in L^1 as $n \to \infty$ if and only if $\{M_n \mid n \in \mathbb{Z}_+\}$ is uniformly integrable, where M_∞ is integrable and $\mathcal{F}_\infty$-measurable. In that case,*

$$\mathbb{E}(M_\infty \mid \mathcal{F}_n) \leq M_n, \quad a.s. \tag{2.17}$$

If in addition M is a martingale, then there is equality in (2.17)*, in other words, M is a Doob martingale.*

Theorem 2.20 (Lévy's upward theorem)**.** *Let X be an integrable random variable, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $(\mathcal{F}_n)_n$ be a filtration, $\mathcal{F}_n \subset \mathcal{F}$ for all n. Then as $n \to \infty$*

$$\mathbb{E}(X \mid \mathcal{F}_n) \longrightarrow \mathbb{E}(X \mid \mathcal{F}_\infty),$$

a.s. and in L^1.

Finally, we consider the case of general $\mathcal{F}$-measurable X. Then $X = X^+ - X^-$, is the difference of two non-negative $\mathcal{F}$-measurable functions X^+ and X^-. Use the linearity of conditional expectation. □

The message here is that one cannot know more than what one can observe. We will also need the corresponding result for decreasing families of σ-algebras. If we have a filtration of the form $(\mathcal{F}_{-n})_{n\in\mathbb{Z}_+}$, i.e. a collection of σ-algebras such that $\mathcal{F}_{-(n+1)} \subseteq \mathcal{F}_{-n}$, then we define

$$\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_{-n}.$$

Theorem 2.21 (Lévy-Doob downward theorem)**.** *Let $(\mathcal{F}_{-n},\, n \in \mathbb{Z}_+)$ be a collection of σ-algebras, such that $\mathcal{F}_{-(n+1)} \subseteq \mathcal{F}_{-n}$ for every n, and let $M = (\dots, M_{-2}, M_{-1})$ be a supermartingale, i.e.*

$$\mathbb{E}(M_{-m} \,|\, \mathcal{F}_{-n}) \leq M_{-n} \quad a.s., \qquad \textit{for all } -n \leq -m \leq -1.$$

If $\sup_n \mathbb{E} M_{-n} < \infty$, then the process M is uniformly integrable and the limit

$$M_{-\infty} = \lim_{n\to\infty} M_{-n}$$

exists a.s. and in L^1. Moreover,

$$\mathbb{E}(M_{-n} \,|\, \mathcal{F}_{-\infty}) \leq M_{-\infty} \quad a.s. \tag{2.18}$$

If M is a martingale, we have equality in (2.18) *and in particular*

$$M_{-\infty} = \mathbb{E}(M_{-1} \,|\, \mathcal{F}_{-\infty}).$$

Corollary 2.2. *Suppose that $X_n \to X$ a.s., and that $|X_n| \leq Y$ a.s. for all n, where Y is an integrable random variable. Moreover, suppose that*

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \quad (\textit{resp. } \mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \cdots)$$

is an increasing (resp. decreasing) sequence of σ-algebras. Then

$$\mathbb{E}(X_n \mid \mathcal{F}_n) \to \mathbb{E}(X \mid \mathcal{F}_1) \quad a.s.,$$

where $\mathcal{F}_1 = \sigma(\bigcup_n \mathcal{F}_n)$ (resp. $\mathcal{F}_1 = \bigcap_n \mathcal{F}_n$).

In case of an increasing sequence of σ-algebras, the corollary is known as *Hunt's lemma.*

Proof. For $m \in \mathbb{Z}_+$, put

$$U_m = \inf_{n \geq m} X_n, \qquad V_m = \sup_{n \geq m} X_n.$$

Since $X_n \to X$ a.s., necessarily $V_m - U_m \to 0$ a.s., as $m \to \infty$. Furthermore, $|V_m - U_m| \leq 2Y$. Dominated convergence then implies that $\mathbb{E}(V_m - U_m) \to 0$ as $m \to \infty$.

Fix $\varepsilon > 0$ and choose m so large that $\mathbb{E}(V_m - U_m) < \varepsilon$. For $n \geq m$ we have

$$U_m \leq X_n \leq V_m \quad \text{a.s.} \tag{2.19}$$

Consequently,

$$\mathbb{E}(U_m \mid \mathcal{F}_n) \leq \mathbb{E}(X_n \mid \mathcal{F}_n) \leq \mathbb{E}(V_m \mid \mathcal{F}_n) \quad \text{a.s.}$$

The processes on the left and right are martingales that satisfy the conditions of the upward (resp. downward) theorem. Letting $n \to \infty$ we obtain

$$\mathbb{E}(U_m \mid \mathcal{F}_1) \leq \liminf_n \mathbb{E}(X_n \mid \mathcal{F}_n) \leq \limsup_n \mathbb{E}(X_n \mid \mathcal{F}_n) \leq \mathbb{E}(V_m \mid \mathcal{F}_1) \quad \text{a.s.} \tag{2.20}$$

It follows that

$$0 \leq \mathbb{E}\Big(\limsup_n \mathbb{E}(X_n \mid \mathcal{F}_n) - \liminf_n \mathbb{E}(X_n \mid \mathcal{F}_n)\Big) \leq \mathbb{E}(\mathbb{E}(V_m \mid \mathcal{F}_1) - \mathbb{E}(U_m \mid \mathcal{F}_1)) \leq \mathbb{E}(V_m - U_m) < \varepsilon.$$

Letting $\varepsilon \downarrow 0$ yields that

$$\limsup_n \mathbb{E}(X_n \mid \mathcal{F}_n) = \liminf_n \mathbb{E}(X_n \mid \mathcal{F}_n) \quad \text{a.s.}$$

and so $\mathbb{E}(X_n \mid \mathcal{F}_n)$ converges a.s.

We wish to identify the limit. Let $n \to \infty$ in (2.19). Then $U_m \leq X \leq V_m$ a.s. Hence

$$\mathbb{E}(U_m \mid \mathcal{F}_1) \leq \mathbb{E}(X \mid \mathcal{F}_1) \leq \mathbb{E}(V_m \mid \mathcal{F}_1) \quad \text{a.s.}$$

□

Previous equations imply that both $\lim E(X_n \mid \mathcal{F}_n)$ and $E(X \mid \mathcal{F}_1)$ are a.s. between V_m and U_m. Consequently,

$$\mathbb{E}\Big|\lim E(X_n \mid \mathcal{F}_n) - E(X \mid \mathcal{F}_1)\Big| \leq \mathbb{E}(V_m - U_m) < \varepsilon.$$

By letting $\varepsilon \downarrow 0$ we obtain that

$$\lim_{n\to\infty} E(X_n \mid \mathcal{F}_n) = E(X \mid \mathcal{F}_1) \quad \text{a.s.} \quad \square$$

2.4.4 Continuous-time martingales

In this section, we consider martingales indexed by $T \subset \mathbb{R}_+$. If $M = (M_t)_{t\geq 0}$ has sufficiently regular paths (e.g., right-continuous), it can be well-approximated by a discrete-time martingale $(M_{t_n})_n$ on a countable dense subset $\{t_n\} \subset T$. This allows many discrete-time results to extend to the continuous-time setting.

Example 2.13. *If $\mathcal{F}$ is a filtration and Y a random variable, then*

$$M_t = \mathbb{E}[Y \mid \mathcal{F}_t]$$

defines a martingale. A martingale of this form is called a closed martingale.

Doob's inequality extends the Bienaymé–Chebyshev inequality to martingale paths:

Theorem 2.22 (Doob's Theorem). *Let $M = (M_t)_{t\geq 0}$ be a martingale with almost surely continuous paths. Then, for any $T \geq 0$, $p \geq 1$, and $\lambda > 0$, we have*

$$\mathbb{P}\left(\sup_{0\leq t\leq T} |M_t| \geq \lambda\right) \leq \frac{\mathbb{E}[|M_T|^p]}{\lambda^p}. \tag{2.21}$$

Note that if M is a martingale and φ a convex function, then for $s \leq t$:

$$\varphi(M_s) = \varphi\big(\mathbb{E}[M_t \mid \mathcal{F}_s]\big) \leq \mathbb{E}[\varphi(M_t) \mid \mathcal{F}_s].$$

The process $(\varphi(M_t))_{t\geq 0}$ is called a **submartingale**.

Example 2.14. *Let $\Omega = \{-1, 1\}$, $\mathcal{F}_t = \{\Omega, \varnothing\}$ for $t \leq 1$ and $\mathcal{F}_t = \{\Omega, \varnothing, \{1\}, \{-1\}\}$ for $t > 1$. Let $\mathbb{P}(\{1\}) = \mathbb{P}(\{-1\}) = \frac{1}{2}$.*

Note that $(\mathcal{F}_t)$ is not right-continuous, since $\mathcal{F}_1 \neq \mathcal{F}_{1+}$. Define

$$Y_t(\omega) = \begin{cases} 0, & t \leq 1, \\ 1, & t > 1, \end{cases} \qquad X_t(\omega) = \begin{cases} 0, & t < 1, \\ 1, & t \geq 1. \end{cases}$$

Consider $Y = (Y_t)_t$, which is a martingale but not right-continuous, while $X = (X_t)_t$ is right-continuous. Although $\mathbb{E}[Y_t] = 0$ is right-continuous in t, we have $Y_{t+} = X_t$ and $\mathbb{P}(X_1 = Y_1) = 0$, so X is not a càdlàg modification of Y. Consequently, Y has no càdlàg modification.

It follows that $\mathbb{E}[X_t \mid \mathcal{F}_t] = Y_t$, implying that X is not a martingale with respect to $(\mathcal{F}_t)_t$. However, by the same lemma, X is a right-continuous martingale with respect to $(\mathcal{F}_{t+})_t$, whereas Y is not.

Example 2.15. *Let $\Omega = \{-1, 1\}$, $\mathcal{F}_t = \{\Omega, \varnothing\}$ for $t < 1$ and $\mathcal{F}_t = \{\Omega, \varnothing, \{1\}, \{-1\}\}$ for $t \geq 1$. Let $\mathbb{P}(\{1\}) = \mathbb{P}(\{-1\}) = \frac{1}{2}$. Define*

$$Y_t(\omega) = \begin{cases} 0, & t \leq 1, \\ 1 - t, & \omega = 1,\ t > 1, \\ -1, & \omega = -1,\ t > 1, \end{cases} \qquad X_t(\omega) = \begin{cases} 0, & t < 1, \\ 1 - t, & \omega = 1,\ t \geq 1, \\ -1, & \omega = -1,\ t \geq 1. \end{cases}$$

In this case the filtration $(\mathcal{F}_t)_t$ is right-continuous and Y and X are both supermartingales with respect to $(\mathcal{F}_t)_t$. Furthermore, $X_t = \lim_{q\downarrow t} Y_q$ for $t \geq 0$, but $\mathbb{P}(X_1 = Y_1) = 0$ and hence X is not a modification of Y.

Convergence theorems

Based on the results of the previous section, we henceforth restrict our attention to martingales that are right-continuous everywhere. Under this assumption, many discrete-time theorems can be extended to the continuous-time setting.

Theorem 2.23. *Let M be a right-continuous supermartingale that is bounded in L^1. Then M_t converges a.s. to a finite $\mathcal{F}_\infty$-measurable limit M_∞, as $t \to \infty$, with $\mathbb{E}|M_\infty| < \infty$.*

Proof. The first step to show is that we can restrict to take a limit along rational time sequences. In other words, that $M_t \to M_\infty$ a.s. as $t \to \infty$ if and only if

$$\lim_{q \to \infty} M_q = M_\infty \quad \text{a.s.} \tag{2.22}$$

To prove the non-trivial implication in this assertion, assume that (2.22) holds. Fix $\varepsilon > 0$ and $\omega \in \Omega$ for which $M_q(\omega) \to M_\infty(\omega)$. Then there exists a number $a = a(\omega, \varepsilon) > 0$ such that

$$|M_q(\omega) - M_\infty(\omega)| < \varepsilon \quad \text{for all } q > a.$$

Now let $t > a$ be arbitrary. Since M is right-continuous, there exists $q_0 > t$ such that

$$|M_{q_0}(\omega) - M_t(\omega)| < \varepsilon.$$

By the triangle inequality, it follows that

$$|M_t(\omega) - M_\infty(\omega)| \le |M_{q_0}(\omega) - M_\infty(\omega)| + |M_t(\omega) - M_{q_0}(\omega)| < 2\varepsilon.$$

This proves that $M_t(\omega) \to M_\infty(\omega)$ as $t \to \infty$.

To prove convergence to a finite $\mathcal{F}_\infty$-measurable, integrable limit, we may assume that M is indexed by the countable set $\mathbb{Q}_+$. The proof can now be finished by arguing as in the proof of Theorem 2.18. □

Corollary 2.3. *A non-negative, right-continuous supermartingale M converges a.s. as $t \to \infty$, to a finite, integrable, $\mathcal{F}_\infty$-measurable random variable.*

Proof. Simple consequence of previous theorem. □

The following continuous-time extension of Theorem 2.2.14 can be derived by reasoning as in discrete-time. The only slight difference is that for a continuous-time process X, L^1-convergence of X_t as $t \to \infty$ need not imply that X is UI.

Theorem 2.24. *Let M be a right-continuous supermartingale that is bounded in L^1.*

1. *If M is uniformly integrable, then $M_t \to M_\infty$ a.s. and in L^1, and*

$$\mathbb{E}(M_\infty \mid \mathcal{F}_t) \le M_t \quad a.s.$$

with equality if M is a martingale.

2. *If M is a martingale and $M_t \to M_\infty$ in L^1 as $t \to \infty$, then M is uniformly integrable.*

Proof. See Exercise 2.25. □

Inequalities

Doob's submartingale inequality and L^p-inequality are very easily extended to the setting of general right-continuous martingales.

Theorem 2.25 (Doob's submartingale inequality)**.** *Let M be a right-continuous submartingale. Then for all $\lambda > 0$ and $t \geq 0$*

$$\mathbb{P}\Big(\sup_{s\leq t} M_s \geq \lambda\Big) \leq \lambda^{-1}\,\mathbb{E}|M_t|.$$

Proof. Let T be a countable, dense subset of $[0,t]$ and choose an increasing sequence of finite subsets $T_n \subseteq T$, with $0, t \in T_n$ for every n and $T_n \uparrow T$ as $n \to \infty$. By right-continuity of M we have that

$$\sup_n \max_{s\in T_n} M_s = \sup_{s\in T} M_s = \sup_{s\in[0,t]} M_s.$$

This implies that

$$\Big\{\max_{s\in T_n} M_s > c\Big\} \uparrow \Big\{\sup_{s\in T} M_s > c\Big\},$$

and so, by monotone convergence of sets,

$$\mathbb{P}\Big(\max_{s\in T_n} M_s > c\Big) \uparrow \mathbb{P}\Big(\sup_{s\in T} M_s > c\Big).$$

By the discrete-time version of the submartingale inequality, for each $m > 0$ sufficiently large,

$$\mathbb{P}\Big(\sup_{s\in[0,t]} M_s > \lambda - \tfrac{1}{m}\Big) = \mathbb{P}\Big(\sup_{s\in T} M_s > \lambda - \tfrac{1}{m}\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\max_{s\in T_n} M_s > \lambda - \tfrac{1}{m}\Big) \leq \tfrac{1}{\lambda - 1/m}\,\mathbb{E}|M_t|.$$

Letting $m \to \infty$ gives the result. □

By exactly the same reasoning, we can generalise the L^p-inequality to continuous time.

Theorem 2.26 (Doob's L^p-inequality)**.** *Let M be a right-continuous martingale or a right-continuous, nonnegative submartingale. Then for all $p > 1$ and $t \geq 0$*

$$\mathbb{E}\Big(\sup_{s\leq t} |M_s|^p\Big) \leq \left(\tfrac{p}{p-1}\right)^p \mathbb{E}|M_t|^p.$$

Some processes become martingales when they are *stopped*, that is, when we replace the process $t \mapsto X_t$ by $t \mapsto X_{t\wedge\tau}$, where τ is a stopping time. When there exists a sequence $(\tau_k)_{k\geq 0}$ of stopping times such that

$$\lim_{k\to+\infty} \tau_k = +\infty \quad \text{a.s.},$$

and the processes $t \mapsto X_{t\wedge\tau_k}$ are martingales, we say that X is a **local martingale**. Local martingales are a useful tool for certain proofs.

Exercise 2.3. *Consider a positive martingale $(X_n)_{n\in\mathbb{N}}$ and for $a > 0$ define*

$$\tau_a := \inf\{k \in \mathbb{N} : X_k \geq a\}$$

the first passage time of X_n above the level a.

1. *Is the time τ_a a stopping time?*
2. *Fix an integer n. Justify the equalities*
$$\mathbb{E}\Big(X_{\min(\tau_a,n)}\Big) = \mathbb{E}(X_0) = \mathbb{E}(X_n)$$
and deduce the equality
$$\mathbb{E}\Big(X_{\tau_a}\mathbf{1}_{\{\tau_a\le n\}}\Big) = \mathbb{E}\Big(X_n\mathbf{1}_{\{\tau_a\le n\}}\Big)\,.$$
3. *From the previous equality, derive the bound (called* Doob's maximal inequality*)*
$$\mathbb{P}\Big(\max_{k=1,\dots,n} X_k \ge a\Big) \le \frac{1}{a}\,\mathbb{E}\Big(X_n\mathbf{1}_{\{\max_{k=1,\dots,n} X_k\ge a\}}\Big)\,.$$
4. *Show that this inequality also holds for a positive submartingale.*

Solution.

1. *Using the definition, the time τ_a is indeed a stopping time.*
2. *Since (X_n) is a martingale, we have $\mathbb{E}(X_n) = \mathbb{E}(X_0)$. Similarly, $(X_{n\wedge\tau_a})$ is a martingale by the Optional Stopping Theorem, hence*
$$\mathbb{E}(X_{n\wedge\tau_a}) = \mathbb{E}(X_0).$$
It follows that $\mathbb{E}(X_{n\wedge\tau_a}) = \mathbb{E}(X_n)$. Decomposing according to $\{\tau_a\le n\}$ or $\{\tau_a>n\}$, we obtain
$$\mathbb{E}\Big(X_{\tau_a}\mathbf{1}_{\{\tau_a\le n\}}\Big) + \mathbb{E}\Big(X_n\mathbf{1}_{\{\tau_a>n\}}\Big) = \mathbb{E}\Big(X_n\mathbf{1}_{\{\tau_a\le n\}}\Big) + \mathbb{E}\Big(X_n\mathbf{1}_{\{\tau_a>n\}}\Big)\,,$$
which gives the desired relation.
3. *By definition of τ_a, we have $X_{\tau_a}\ge a$, hence*
$$a\,\mathbb{P}(\tau_a\le n) \le \mathbb{E}\Big(X_{\tau_a}\mathbf{1}_{\{\tau_a\le n\}}\Big) = \mathbb{E}\Big(X_n\mathbf{1}_{\{\tau_a\le n\}}\Big)\,.$$
To conclude, note that
$$\{\tau_a\le n\} = \Big\{\max_{k=1,\dots,n} X_k\ge a\Big\}\,.$$
Indeed, $\tau_a\le n$ means that the process (X_k) exceeds the level a before n, which is equivalent to $\max_{k=1,\dots,n} X_k\ge a$.
4. *When (X_n) is a submartingale, we proceed differently. Observe first that for every $k\le n$,*
$$X_k \le \mathbb{E}(X_n\mid\mathcal{F}_k).$$
Using this inequality and the properties of conditional expectation, we obtain
$$\mathbb{E}\Big(X_{\tau_a}\mathbf{1}_{\{\tau_a\le n\}}\Big) = \sum_{k=0}^{n}\mathbb{E}\Big(X_k\mathbf{1}_{\{\tau_a=k\}}\Big) \le \sum_{k=0}^{n}\mathbb{E}\Big(\mathbb{E}(X_n\mid\mathcal{F}_k)\,\mathbf{1}_{\{\tau_a=k\}}\Big)$$
$$\le \sum_{k=0}^{n}\mathbb{E}\Big(\mathbb{E}(X_n\mathbf{1}_{\{\tau_a=k\}}\mid\mathcal{F}_k)\Big) \le \sum_{k=0}^{n}\mathbb{E}\Big(X_n\mathbf{1}_{\{\tau_a=k\}}\Big) = \mathbb{E}\Big(X_n\mathbf{1}_{\{\tau_a\le n\}}\Big)\,.$$
To conclude, we argue as in the previous question and obtain the same Doob's inequality.

Exercise 2.4 (A counterexample)**.** *Find a process $(M_n)_{n\geq 0}$ with $\mathbb{E}[|M_n|] < +\infty$ for all n such that*

$$\mathbb{E}[M_{n+1} \mid M_n] = M_n \quad \text{for all } n,$$

but where M is not a martingale.
Solution *Consider a simple random walk starting from 0 with independent steps ± 1, but at the first return to 0, the walk is forced to take the same step as its very first one.*

For $n \geq 1$, we then have:

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \begin{cases} M_n, & \text{if } M_n \neq 0, \\ -1, & \text{if } M_n = 0 \text{ and } M_1 = -1, \\ 1, & \text{if } M_n = 0 \text{ and } M_1 = 1. \end{cases}$$

In particular, $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \neq M_n$ when $M_n = 0$, so M is not a martingale.

Exercise 2.5. *Let $X = (X_1, X_2, X_3)$ be a Gaussian vector with covariance matrix*

$$C = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & 1 \\ -2 & 1 & 5 \end{pmatrix}.$$

1. *Does the random vector X have a density?*
2. *Compute $\mathbb{E}[X_1|X_2]$ and $\mathbb{E}[X_1|X_2, X_3]$.*
3. *Compute $\mathbb{E}\Big[e^{i\xi X_1} \,\Big|\, X_2\Big]$ and $\mathbb{E}\Big[e^{i\xi X_1} \,\Big|\, X_2, X_3\Big]$.*

Solution

1. *Let C_1, C_2, C_3 denote the columns of C. We observe that*

$$2C_1 - C_2 + C_3 = 0,$$

hence C is not invertible: the Gaussian vector X does not have a density. Moreover,

$$2X_1 - X_2 + X_3 = 0 \quad a.s.$$

2. *Write $X_1 = aX_2 + (X_1 - aX_2)$ with $a \in \mathbb{R}$ such that $(X_1 - aX_2)$ is independent of X_2, that is,*

$$\mathrm{Cov}(X_1 - aX_2, X_2) = 0 \quad \Rightarrow \quad 2 - 5a = 0.$$

Therefore $a = \dfrac{2}{5}$ and

$$\mathbb{E}[X_1|X_2] = \frac{2}{5}X_2.$$

Since $2X_1 - X_2 + X_3 = 0$ a.s.,

$$X_1 = \frac{X_2 - X_3}{2} \quad \text{and} \quad \mathbb{E}[X_1|X_2, X_3] = \frac{X_2 - X_3}{2}.$$

3. *With* $a = \frac{2}{5}$, *the variance of* $X_1 - aX_2$ *is*

$$\mathrm{Var}(X_1 - aX_2) = 2 - 4a + 5a^2 = \frac{6}{5}.$$

Thus

$$\begin{aligned}\mathbb{E}\left[e^{i\xi X_1} \,\middle|\, X_2\right] &= e^{i\xi \frac{2}{5}X_2}\mathbb{E}\left[e^{i\xi(X_1 - \frac{2}{5}X_2)}\right] \\ &= e^{i\xi\frac{2}{5}X_2}e^{-\frac{1}{2}\frac{6}{5}\xi^2} \\ &= e^{i\xi\frac{2}{5}X_2}e^{-\frac{3}{5}\xi^2}.\end{aligned}$$

Moreover,

$$\mathbb{E}\left[e^{i\xi X_1} \,\middle|\, X_2, X_3\right] = e^{i\xi\frac{X_2 - X_3}{2}}.$$

2.5 Exercises

Exercise 2.6. *(a) Show that if* X *is* $N(\mu, \sigma^2)$, *and* Y *is* $N(\nu, \tau^2)$, *where* X *and* Y *are independent, then* $X + Y$ *is* $N(\mu + \nu, \sigma^2 + \tau^2)$.

*(b) * Show that if* X *and* Y *are independent* $N(0, 1)$, *then* $Z = \frac{X}{Y}$ *has the Cauchy density*

$$f(z) = \frac{1}{\pi}\frac{1}{1 + z^2}, \quad -\infty < z < \infty.$$

*[**Hint:** Consider the map* $W = Y$, $Z = \frac{X}{Y}$.*] Deduce that* Z^{-1} *also has the Cauchy density.*

(c) If X *and* Y *are independent continuous random variables, show that* $U = XY$ *and* $V = \frac{X}{Y}$ *have respective densities*

$$f_U(u) = \int f_X(v) f_Y\left(\frac{u}{v}\right)\left|\frac{1}{v}\right| dv.$$

$$f_V(v) = \int f_X(uv) f_Y(u)|u| \, du$$

Exercise 2.7. *Suppose we roll two dice, a red and a green one, and let* X *be the value on the red die and* Y *the value on the green die. Let* $Z = XY$.

1. *Let* $W = \mathbb{E}(Z \mid X)$. *What are the possible values for* W*? Give the distribution of* W.

2. *Do the same exercise for* $U = \mathbb{E}(X \mid Z)$.

3. *Do the same exercise for* $V = \mathbb{E}(Y \mid X, Z)$.

Exercise 2.8. *Suppose we roll two dice, a red and a green one, and let* X *be the value on the red die and* Y *the value on the green die. Let* $Z = X/Y$.

1. *Find* $\mathbb{E}[(X + 2Y)^2 \mid X]$.

2. *Find* $\mathbb{E}[(X + 2Y)^2 \mid X, Z]$.

3. *Find* $\mathbb{E}[X + 2Y \mid Z]$.

Exercise 2.9. *Suppose* $X_1, X_2, \ldots$ *are independent random variables with*

$$\mathbb{P}\{X_j = 2\} = 1 - \mathbb{P}\{X_j = -1\} = \tfrac{1}{3}.$$

Let $S_n = X_1 + \cdots + X_n$ *and let* $\mathcal{F}_n$ *denote the information in* $X_1, \ldots, X_n$.

1. *Find* $\mathbb{E}[S_n]$, $\mathbb{E}[S_n^2]$, $\mathbb{E}[S_n^3]$.
2. *If* $m < n$, *find* $\mathbb{E}[S_n \mid \mathcal{F}_m]$, $\mathbb{E}[S_n^2 \mid \mathcal{F}_m]$, $\mathbb{E}[S_n^3 \mid \mathcal{F}_m]$.
3. *If* $m < n$, *find* $\mathbb{E}[X_m \mid S_n]$.

Exercise 2.10. *Let* $X_1, X_2, \ldots$ *be independent, identically distributed random variables with*

$$\mathbb{P}\{X_j = 1\} = q, \qquad \mathbb{P}\{X_j = -1\} = 1 - q.$$

Let $S_0 = 0$ *and for* $n \geq 1$, $S_n = X_1 + X_2 + \cdots + X_n$. *Define* $Y_n = e^{S_n}$.

1. *For which value of* q *is* Y_n *a martingale?*
2. *For the remaining parts of this exercise assume* q *takes the value from part (1). Explain why* Y_n *satisfies the conditions of the martingale convergence theorem.*
3. *Let* $Y_\infty = \lim_{n\to\infty} Y_n$. *Explain why* $Y_\infty = 0$.
   Hint: *there are at least two ways to show this. One is to consider* $\log Y_n$ *and use the law of large numbers. Another is to note that with probability one* Y_{n+1}/Y_n *does not converge.*
4. *Use the optional sampling theorem to determine the probability that* Y_n *ever attains a value greater than* 100.
5. *Does there exist a* $C < \infty$ *such that* $\mathbb{E}[Y_n^2] \leq C$ *for all* n*?*

Exercise 2.11. *(Basic example). Let* X *be an integrable random variable. Show that*

$$\Big(\mathbb{E}(X \mid \mathcal{F}_t),\ t \geq 0\Big)$$

is a martingale.

Exercise 2.12. *(Square-integrable martingale). Let* $(M_t, t \geq 0)$ *be an* $\mathcal{F}_t$*-martingale that is square-integrable (that is,* M_t^2 *has finite expectation for every* t*). Show that:*

1. $\mathbb{E}((M_t - M_s)^2 \mid \mathcal{F}_s) = \mathbb{E}(M_t^2 \mid \mathcal{F}_s) - M_s^2$, *for* $t > s$.
2. $\mathbb{E}((M_t - M_s)^2) = \mathbb{E}(M_t^2) - \mathbb{E}(M_s^2)$, *for* $t > s$.
3. *The function* φ *defined by* $\varphi(t) = \mathbb{E}(M_t^2)$ *is nondecreasing.*

Exercise 2.13. *(Examples of martingales).*

1. *Let* X *be a process with independent increments (PII). Show that* X *is a martingale and that, if* X *is square-integrable, then*

$$X_t^2 - \mathbb{E}(X_t^2)$$

   is a martingale.

Exercise 2.14. *Let M be a positive continuous uniformly integrable martingale and*

$$\tau = \inf\{t : M_t = 0\}.$$

Show that M is identically zero for $t > \tau$.

Exercise 2.15. *a) Show that if X_n and Y_n are submartingales, then* $\sup(X_n, Y_n)$ *is a submartingale.*

b) If X_n is a submartingale, then $\mathbb{E}[X_n] \geq \mathbb{E}[X_{n-1}]$.

c) If X_n is a martingale, then $\mathbb{E}[X_n] = \mathbb{E}[X_{n-1}]$.

Chapter 3

Discrete-Time Markov Chains

3.1 Markov chains

In this section, we introduce discrete-time Markov chains [5, 6, 20, 23, 26, 28]. Their importance stems from two main reasons:

- There are numerous physical, biological, economic, and social phenomena that can be modeled using Markov chains.
- A well-developed theoretical framework exists, allowing for effective analysis and computation.

3.1.1 Homogeneous Markov chains

In this subsection we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family of discrete random variables

$$X_n : \Omega \to I, \quad n \in \mathbb{Z}_+,$$

where I is a countable set known as the *state space*, and every $i \in I$ is called a *state*.

Recall that every random variable $X : \Omega \to I$ has an associated distribution ν, i.e., its law, given by

$$\nu_i = \mathbb{P}(X = i).$$

A stochastic process is a collection of random variables $\{X_t;\, t \in T\}$, where each $X_t : \Omega \to \mathbb{R}$ represents the evolution of a system of random values over time.

We focus on the *discrete case*, where

$$T = \mathbb{Z}_+ = \{0, 1, 2, \dots\},$$

so that our process is $\{X_t,\, t \geq 0\}$. In particular, we consider *Markov chains*, which are memoryless processes: the future behavior depends only on the present state.

Definition 3.1 (Stochastic Matrix)**.** *A stochastic matrix, also called a transition (probability) matrix,*

$$\Pi = (p_{i,j};\ i, j \in I),$$

is a matrix which satisfies the following conditions:

1. $p_{i,j} \in [0, 1], \quad \forall i, j \in I;$
2. $\forall i \in I, \quad \sum_{j \in I} p_{i,j} = 1.$

The element $p_{i,j}$ is called the transition probability. That is to say, it is the probability that at time k the process is in state j, given that at time $k-1$ it was in state i.

Then, using the concepts defined so far, we are going to study two important issues for the Markov theory.

Definition 3.2 (Markov Chain)**.** *Given a stochastic process $\{X_n;\ n \geq 0\}$ that takes values in the state space I, we have:*

1. $\mathbb{P}(X_0 = i) = \nu_i, \quad \forall i \in I$, *which is called the initial distribution.*
2. $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n), \quad \forall n \geq 0,\ i_0, \ldots, i_n, i_{n+1} \in I$.

Therefore, the stochastic process $\{X_n;\ n \geq 0\}$ is called a Markov chain with initial distribution $\nu = \{\nu_i;\ i \in I\}$ and transition matrix $\Pi = (p_{i,j};\ i, j \in I)$.

The last equality is known as the *Markov property*, which states that the probability of a future event depends only on the present state, not on the entire past history. This implies that Markov chains are *memoryless processes*.

Definition 3.3 (Homogeneous Markov Chain)**.** *A* homogeneous Markov chain *is a Markov chain for which, for all $n \geq 0$ and all $i, j \in I$, the transition probability*

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

is independent of n.

We often denote a homogeneous Markov chain with initial distribution ν and transition matrix Π by HMC(ν, Π).

Theorem 3.1. *A stochastic process $\{X_n;\ n \geq 0\}$ with values in I is a* HMC(ν, Π) *if and only if, for all $n \geq 0$ and all $i_0, \ldots, i_{n-1}, i \in I$, we have*

$$\mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i) = \nu_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i}. \tag{3.1}$$

Proof. We start with the "only if" direction. Assume that $\{X_n\}$ is a homogeneous Markov chain (ν, Π). Applying the chain rule of probability and the Markov property, we obtain

$$\mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i) = \mathbb{P}(X_0 = i_0)\, \mathbb{P}(X_1 = i_1 \mid X_0 = i_0) \cdots \mathbb{P}(X_n = i \mid X_{n-1} = i_{n-1}).$$

Therefore,

$$\mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i) = \nu_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i}.$$

Now we prove the "if" direction. Suppose that (3.1) holds. We need to check that $\{X_n\}$ is a Markov chain and that it is homogeneous.

Initial distribution: For any $i_0 \in I$, we compute

$$\mathbb{P}(X_0 = i_0) = \sum_{j \in I} \nu_{i_0} p_{i_0,j} = \nu_{i_0} \sum_{j \in I} p_{i_0,j} = \nu_{i_0}.$$

Markov property: For any $i_0, \ldots, i_{n-1}, i, j \in I$, we have

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, \ldots, X_n = i) = \frac{\mathbb{P}(X_0 = i_0, \ldots, X_n = i, X_{n+1} = j)}{\mathbb{P}(X_0 = i_0, \ldots, X_n = i)}.$$

Using (3.1), this becomes

$$\frac{\nu_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i}\, p_{i,j}}{\nu_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i}} = p_{i,j}.$$

On the other hand,

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

Expanding with (3.1), we obtain

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \frac{\sum_{i_0,\dots,i_{n-1} \in I} \nu_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i} p_{i,j}}{\sum_{i_0,\dots,i_{n-1} \in I} \nu_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i}}.$$

The factor cancels out, giving

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

Thus,

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, \dots, X_n = i) = p_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i),$$

so the Markov property holds, and the transition probabilities do not depend on n. Therefore $\{X_n\}$ is a homogeneous Markov chain HMC(ν, Π). □

Corollary 3.1. *Let $\{Y_n;\, n \geq 1\}$ be a family of independent and identically distributed random variables with values in I. If*

$$X_n = \sum_{i=1}^{n} Y_i, \quad n \geq 1,$$

then the process $\{X_n;\, n \geq 0\}$ is a homogeneous Markov chain.

3.1.2 Examples

We start by presenting a few examples, and then we discuss the key property that characterizes Markov chains.

Example 3.1. *(Gambler's ruin.) Consider a gambling game in which, on each turn, you win \$1 with probability $p = 0.4$ or lose \$1 with probability $1 - p = 0.6$. Suppose you adopt the rule of quitting if your fortune reaches \$N, while reaching \$0 also ends the game. Let X_n denote your fortune after n plays.*

The process $(X_n)_{n\geq 0}$ satisfies the Markov property, *meaning that given the current state X_n, any information about the past is irrelevant for predicting the next state X_{n+1}. For example, if at time n your fortune is $X_n = i$ with $0 < i < N$, then for any possible history $i_0, i_1, \dots, i_{n-1}$,*

$$P(X_{n+1} = i + 1 \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = 0.4,$$

because increasing your wealth by one unit requires winning the next bet.

Formally, a discrete-time process X_n is a Markov chain *with transition matrix $p(i, j)$ if, for all $j, i, i_0, \dots, i_{n-1}$,*

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p(i, j). \tag{3.2}$$

Equation (3.2) expresses the memoryless property: the conditional distribution of the next state depends only on the current state. Here, we focus on the temporally homogeneous case where $p(i, j)$ is independent of n.

$$p(i, j) = P(X_{n+1} = j \mid X_n = i)$$

does not depend on the time n.

Intuitively, the transition probability gives the rules of the game. It is the basic information needed to describe a Markov chain. In the case of the gambler's ruin chain, the transition probability has:

$$p(i, i+1) = 0.4, \quad p(i, i-1) = 0.6, \quad \text{if } 0 < i < N$$

$$p(0, 0) = 1, \quad p(N, N) = 1$$

When $N = 5$, the matrix is:

	0	1	2	3	4	5
0	1.0	0	0	0	0	0
1	0.6	0	0.4	0	0	0
2	0	0.6	0	0.4	0	0
3	0	0	0.6	0	0.4	0
4	0	0	0	0.6	0	0.4
5	0	0	0	0	0	1.0

Example 3.2. *(Weather chain.) Let X_n be the weather on day n in Ithaca, NY, which we assume is either: $1 =$ rainy, or $2 =$ sunny. Even though the weather is not exactly a Markov chain, we can propose a Markov chain model for the weather by writing down a transition probability:*

	1	2
1	0.6	0.4
2	0.2	0.8

*The table says, for example, the probability a rainy day (state 1) is followed by a sunny day (state 2) is $p(1, 2) = 0.4$. A typical question of interest is: **Q.** What is the long-run fraction of days that are sunny?*

Example 3.3. *(Social mobility.) Let X_n be a family's social class in the n-th generation, which we assume is either $1 =$ lower, $2 =$ middle, or $3 =$ upper. In our simple version of sociology, changes of status are a Markov chain with the following transition probability:*

	1	2	3
1	0.7	0.2	0.1
2	0.3	0.5	0.2
3	0.2	0.4	0.4

***Q.** Do the fractions of people in the three classes approach a limit?*

3.2 Continuous-time Markov process

In this section, we present the definition of a continuous-time Markov process. Let $\{X(t) : t \geq 0\}$ be a stochastic process with state space $\mathcal{X}$. We say that $\{X(t)\}$ is a **continuous-time Markov process** if it satisfies the following property:

$$P\Big(X(t+s)=j \,\Big|\, X(s)=i,\, X(u)=x(u),\, 0\le u<s\Big) = P\Big(X(t+s)=j \,\Big|\, X(s)=i\Big),$$

for all $i, j \in \mathcal{X}$, all $s, t \geq 0$, and all possible histories $\{x(u), 0 \leq u < s\}$.

In words, the Markov property states that, given the present state, the future evolution of the process is independent of its past history.

If, in addition, the process satisfies

$$P(X(t+s)=j \mid X(s)=i) = P(X(t)=j \mid X(0)=i), \quad \forall\, i, j \in \mathcal{X},\ s, t \geq 0,$$

then the process is said to have **stationary transition probabilities**, or equivalently, it is a **time-homogeneous Markov process**.

3.2.1 Continuous-time examples

Example 3.4 (Wiener Process / Brownian Motion)**.** *This is a continuous-time Markov process with a continuous state space, where the state of the system evolves continuously, similar to the random movement of particles in a fluid.*

Example 3.5 (Poisson Process)**.** *This process models the number of events that occur in a continuous time interval. The state represents the count of events, and each transition increases the count by one. A typical application is modeling the number of customers arriving at a store over time.*

Example 3.6 (Birth–Death Process)**.** *A special type of continuous-time Markov process where the system can only transition to an adjacent state, either increasing (*birth*) or decreasing (*death*) the state variable.*

Example 3.7 (System Reliability)**.** *Consider a simple machine that can be in either a* working *state or a* failed *state. The time until failure is exponentially distributed, and the time required for repair (to return to the working state) is also exponentially distributed. The transitions between these two states satisfy the Markov property: the future state depends only on the current state, not on the past history of failures or repairs.*

3.3 Multistep transition probabilities

The *transition probability*

$$p(i, j) = P(X_{n+1} = j \mid X_n = i)$$

gives the probability of moving from state i to state j in one step.

Our objective is to compute the probability of transitioning from i to j in $m > 1$ steps:

$$p^m(i, j) = P(X_{n+m} = j \mid X_n = i).$$

As suggested by the notation, p^m corresponds to the m-th power of the transition matrix.

As an example, consider the *social mobility chain* to illustrate this concept: Let X_n be a family's social class in the nth generation, which we assume is either 1 = lower, 2 =

middle, or 3 = upper. In our simple version of sociology, changes of status are a Markov chain with the following transition probability

	1	2	3
1	0.7	0.2	0.1
2	0.3	0.5	0.2
3	0.2	0.4	0.4

and consider the following concrete question:
Q1. Your parents were middle class (state 2). What is the probability that you are in the upper class (state 3) but your children are lower class (state 1)?

Solution. Intuitively, the Markov property implies that starting from state 2, the probability of jumping to state 3 and then to state 1 is given by

$$p(2,3)p(3,1)$$

To get this conclusion from the definitions, we note that using the definition of conditional probability,

$$P(X_2 = 1, X_1 = 3 \mid X_0 = 2) = \frac{P(X_2 = 1, X_1 = 3, X_0 = 2)}{P(X_0 = 2)}$$

$$= \frac{P(X_2 = 1, X_1 = 3, X_0 = 2)}{P(X_1 = 3, X_0 = 2)} \cdot \frac{P(X_1 = 3, X_0 = 2)}{P(X_0 = 2)} = P(X_2 = 1 \mid X_1 = 3, X_0 = 2) \cdot P(X_1 = 3 \mid X_0 = 2)$$

By the Markov property, the last expression is

$$P(X_2 = 1 \mid X_1 = 3) \cdot P(X_1 = 3 \mid X_0 = 2) = p(2,3)p(3,1)$$

Moving on to the real question:
Q2. What is the probability your children are lower class (1) given your parents were middle class (2)?

Solution. To do this we simply have to consider the three possible states for your class and use the solution of the previous problem:

$$P(X_2 = 1 \mid X_0 = 2) = \sum_{k=1}^{3} P(X_2 = 1, X_1 = k \mid X_0 = 2) = \sum_{k=1}^{3} p(2,k)p(k,1)$$

$$= (0.3)(0.7) + (0.5)(0.3) + (0.2)(0.2) = 0.21 + 0.15 + 0.04 = 0.21$$

There is nothing special here about the states 2 and 1. By the same reasoning,

$$P(X_2 = j \mid X_0 = i) = \sum_{k=1}^{3} p(i,k)p(k,j)$$

The right-hand side of the last equation gives the (i,j)-th entry of the matrix p multiplied by itself.

To explain this, we note that to compute $p^2(2,1)$, we multiplied the entries of the second row by those in the first column:

$$\begin{pmatrix} \cdots & & \\ 0.3 & 0.5 & 0.2 \\ \cdots & & \end{pmatrix} \begin{pmatrix} 0.7 & \cdots \\ 0.3 & \cdots \\ 0.2 & \cdots \end{pmatrix} = \begin{pmatrix} \cdots & \\ 0.40 & \cdots \\ \cdots & \end{pmatrix}$$

If we wanted $p^2(1,3)$, we would multiply the first row by the third column:

$$\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ \cdots & & \\ \cdots & & \end{pmatrix} \begin{pmatrix} \cdots & 0.1 \\ \cdots & 0.2 \\ \cdots & 0.4 \end{pmatrix} = \begin{pmatrix} \cdots & 0.15 \\ \cdots & \\ \cdots & \end{pmatrix}$$

When all of the computations are done, we have

$$\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{pmatrix} \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.57 & 0.28 & 0.15 \\ 0.40 & 0.39 & 0.21 \\ 0.34 & 0.40 & 0.26 \end{pmatrix}$$

If we use this procedure to compute A^{20}, we get a matrix with three rows that agree in the first six decimal places with:

$$\begin{pmatrix} 0.468085 & 0.340425 & 0.191489 \end{pmatrix}$$

Later, we will see that as $n \to \infty$, p_n converges to a matrix with all three rows equal to $\left(\frac{22}{47}, \frac{16}{47}, \frac{9}{47}\right)$.

To explain our interest in p^m, we will now prove:

Theorem 3.2. *The m-step transition probability $P(X_{n+m} = j \mid X_n = i)$ is the m-th power of the transition matrix p.*

The key ingredient in proving this is the Chapman–Kolmogorov equation:

$$p^{m+n}(i,j) = \sum_k p^m(i,k)p^n(k,j)$$

Once this is proved, since taking $n = 1$, we see that

$$p^{m+1}(i,j) = \sum_k p^m(i,k)p(k,j)$$

That is, the $m+1$-step transition probability is the m-step transition probability times p. Why the Chapman–Kolmogorov equation is true? To go from i to j in $m+n$ steps, we have to go from i to some state k in m steps and then from k to j in n steps. The Markov property implies that the two parts of our journey are independent and it is represented in the following representation

0	m	$m+n$
.	.	.
i	.	.
.	.	j
.	.	.

Proof of the Chapman–Kolmogorov equation According to the state at time m, we have

$$P(X_{m+n} = j \mid X_0 = i) = \sum_k P(X_{m+n} = j, X_m = k \mid X_0 = i)$$

Using the definition of conditional probability, we get

$$\begin{aligned} P(X_{m+n} = j, X_m = k \mid X_0 = i) &= \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_0 = i)} \qquad (3.3) \\ &= \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_m = k, X_0 = i)} \cdot \frac{P(X_m = k, X_0 = i)}{P(X_0 = i)} \\ &= P(X_{m+n} = j \mid X_m = k, X_0 = i) \cdot P(X_m = k \mid X_0 = i) \end{aligned}$$

By the Markov property, the last expression is

$$P(X_{m+n} = j \mid X_m = k) \cdot P(X_m = k \mid X_0 = i) = p^m(i,k)p^n(k,j)$$

and we have proved the results.

Example 3.8. *(Gambler's ruin.) Suppose for simplicity that $N = 4$ in Example 3.1, so that the transition probability is*

$$\begin{pmatrix} & 0 & 1 & 2 & 3 & 4 \\ 0 & 1.0 & 0 & 0 & 0 & 0 \\ 1 & 0.6 & 0 & 0.4 & 0 & 0 \\ 2 & 0 & 0.6 & 0 & 0.4 & 0 \\ 3 & 0 & 0 & 0.6 & 0 & 0.4 \\ 4 & 0 & 0 & 0 & 0 & 1.0 \end{pmatrix}$$

To compute p_2 one row at a time, we note:

$$p^2(0,0) = 1 \quad \textit{and} \quad p^2(4,4) = 1, \quad \textit{since these are absorbing states.}$$

$$p^2(1,3) = (0.4)^2 = 0.16, \quad \textit{since the chain has to go up twice.}$$

$$p^2(1,1) = (0.4)(0.6) = 0.24, \quad \textit{the chain must go from 1 to 2 to 1.}$$

$$p^2(1,0) = 0.6, \quad \textit{to be at 0 at time 2, the first jump must be to 0.}$$

We obtain

$$p^2 = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0.6 & 0.24 & 0 & 0.16 & 0 \\ 0.36 & 0 & 0.48 & 0 & 0.16 \\ 0 & 0.36 & 0 & 0.24 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$p^{20} = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0.87655 & 0.00032 & 0 & 0.00022 & 0.12291 \\ 0.69186 & 0 & 0.00065 & 0 & 0.30749 \\ 0.41842 & 0.00049 & 0 & 0.00032 & 0.58437 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

0 and 4 are absorbing states. Here we see that the probability of avoiding absorption for 20 steps is 0.00054 from state 3, 0.00065 from state 2, and 0.00081 from state 1. Later, we will see that

$$\lim_{n\to\infty} p^n = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ \frac{57}{65} & 0 & 0 & 0 & \frac{8}{65} \\ \frac{45}{65} & 0 & 0 & 0 & \frac{20}{65} \\ \frac{27}{65} & 0 & 0 & 0 & \frac{38}{65} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

3.4 Classification of states

In this section, we study how to classify the states of a Markov chain. This classification is based on the *communication* relationships between states: we examine whether it is possible to move from one state to another and then return to the initial state, or whether a given state is absorbing (i.e., the process cannot leave it).

Formally, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of discrete random variables

$$X_n : \Omega \to I, \quad n \geq 0,$$

where I is a countable set. Assume that $\{X_n; n \geq 0\}$ forms a homogeneous Markov chain, denoted HMC(ν, Π), with initial distribution ν and transition matrix

$$\Pi = (p_{i,j})_{i,j \in I}.$$

3.4.1 Basic concepts

We begin by introducing two fundamental concepts that allow us to partition the states of a Markov chain.

Definition 3.4 (Accessibility). *Given states $i, j \in I$, we say that j is* accessible *from i, denoted $i \to j$, if the chain can reach state j from state i in a finite number of steps with positive probability. Formally, there exists some $n \geq 0$ such that*

$$p_{i,j}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i) > 0.$$

Definition 3.5 (Communication). *States i and j* communicate *if they are mutually accessible; that is, i is accessible from j and j is accessible from i. We denote this relation by $i \leftrightarrow j$.*

The communication relation is an equivalence relation. Hence, it satisfies reflexivity, symmetry, and transitivity, as we now show:

1. $p_{i,i}^{(0)} = \mathbb{P}(X_0 = i \mid X_0 = i) = 1 > 0$, so $i \leftrightarrow i$ in 0 steps.

2. If $i \leftrightarrow j$, there exist $n, m \geq 0$ such that $p_{i,j}^{(n)} > 0$ and $p_{j,i}^{(m)} > 0$. Then $j \leftrightarrow i$.

3. If $i \leftrightarrow j$, there exist $n_1, n_2 \geq 0$ such that $p_{i,j}^{(n_1)} > 0$ and $p_{j,i}^{(n_2)} > 0$. Also, if $j \leftrightarrow \ell$, there exist $m_1, m_2 \geq 0$ such that $p_{j,\ell}^{(m_1)} > 0$ and $p_{\ell,j}^{(m_2)} > 0$. Then $i \leftrightarrow \ell$ because, using the Chapman–Kolmogorov equation, we have

$$p_{i,\ell}^{(n_1+m_1)} = \sum_{k \in I} p_{i,k}^{(n_1)} \, p_{k,\ell}^{(m_1)} \geq p_{i,j}^{(n_1)} \, p_{j,\ell}^{(m_1)} > 0,$$

$$p_{\ell,i}^{(n_2+m_2)} = \sum_{k \in I} p_{\ell,k}^{(m_2)} \, p_{k,i}^{(n_2)} \geq p_{\ell,j}^{(m_2)} \, p_{j,i}^{(n_2)} > 0.$$

Thus, the communication relation $\leftrightarrow$ is an *equivalence relation*, which partitions the state space I into equivalence classes called *communication classes*. In particular, any two states that communicate belong to the same class. This notion allows us to define several important properties of the states.

Definition 3.6 (Irreducible Chain). *A Markov chain is called irreducible if there exists a unique equivalence class; in other words, when every state communicates with every other state.*

Definition 3.7 (Closed Class)**.** *A subset $C \subseteq I$ is called a closed class if $i \in C$ and $i \to j$ imply $j \in C$. That is, from a state in C we can never reach a state in $I \setminus C$ (hence, the chain cannot leave C).*

To continue, once given the last definition, we say that a class C is *closed* if the elements of the stochastic matrix satisfy the following property

$$\sum_{j \in C} p_{i,j} = 1, \quad \forall i \in C.$$

An alternative way to study whether a class C is closed is by using the n-step transition probability, which must take the following value

$$p_{i,j}^{(n)} = 0, \quad \forall i \in C, \; j \in I \setminus C, \; n \geq 1.$$

3.4.2 Classification

We now introduce key concepts for classifying states in a Markov chain, focusing on whether a process can return to its initial state.

Definition 3.8 (Recurrent and Transient States)**.** *For any state $i \in I$, we define*

$$f_i = \mathbb{P}_i(X_n = i \textit{ for infinitely many } n) = \mathbb{P}_i(\{X_n = i \textit{ infinitely often}\}).$$

We say that state i is:

- *recurrent if $f_i = 1$,*
- *transient if $f_i = 0$.*

In other words, a state is recurrent if the chain will return to it in the future; otherwise, it is transient.

Definition 3.9 (First Passage Time)**.** *The first passage time to state $i \in I$ is defined as*

$$T_i = \inf\{n \geq 1 : X_n = i\},$$

that is, the first time at which the chain visits state i.

Definition 3.10 (Hitting Probability)**.** *The probability that a chain which starts in state i passes through j, denoted $\rho_{i,j}$, is defined as*

$$\rho_{i,j} = \mathbb{P}_i(T_j < \infty).$$

Equivalently,

$$\rho_{i,j} = \sum_{k=1}^{\infty} \mathbb{P}_i(T_j = k).$$

In particular, $\rho_{i,i}$ denotes the probability that the chain, starting from i, eventually returns to i.

Now, we introduce another useful concept. Let $\mathbf{1}_{\{X=k\}}$ denote the indicator function of the event $\{X = k\}$.

Definition 3.11 (Number of Visits)**.** *The number of visits of the chain to state j is defined as*

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

Given $X_0 = i$, we note that the events $\{N(j) \geq 1\}$ and $\{T_j < \infty\}$ are equivalent. Thus, we can express

$$\rho_{i,j} = \mathbb{P}_i(T_j < \infty) = \mathbb{P}_i(N(j) \geq 1).$$

Using this fact, we prove the following recursive property:

$$\mathbb{P}_i(N(j) \geq k) = \rho_{i,j}\, \rho_{j,j}^{k-1}, \quad \forall k \geq 2.$$

For illustration, consider the probability that a Markov chain starting at state i first reaches state j at time k, with the subsequent return to j occurring additional n steps.

$$\mathbb{P}(X_{n+k} = j,\, X_{n+k-1} \neq j, \ldots, X_{k+1} \neq j, X_k = j, X_{k-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i).$$

This equal to

$$\mathbb{P}_j(T_j = n)\, \mathbb{P}_i(T_j = k).$$

Hence, the probability of visiting j at least twice is

$$\mathbb{P}_i(N(j) \geq 2) = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \mathbb{P}_i(T_j = k)\, \mathbb{P}_j(T_j = n) = \left(\sum_{k=1}^{\infty} \mathbb{P}_i(T_j = k) \right) \left(\sum_{n=1}^{\infty} \mathbb{P}_j(T_j = n) \right).$$

That is,

$$\mathbb{P}_i(N(j) \geq 2) = \rho_{i,j}\, \rho_{j,j}.$$

Proceeding by induction on the number of visits, one can show that for all $k \geq 2$,

$$\mathbb{P}_i(N(j) \geq k) = \rho_{i,j}\, \rho_{j,j}^{k-1}.$$

We calculate the probability that a chain which starts at i visits the state j for the first time at time l, and after this, the chain will visit again the state j $k-1$ times:

$$\begin{aligned} &\mathbb{P}\Big(X_{n+k+\cdots+m+l} = j, \ldots, X_{k+\cdots+m+l} = j, \ldots, X_{m+l} = j, \ldots, X_l = j, \ldots, X_1 \neq j \,\Big|\, X_0 = i\Big) \\ &= \mathbb{P}_j(T_j = n) \cdot \mathbb{P}_j(T_j = m) \cdots \mathbb{P}_j(T_j = m) \cdot \mathbb{P}_i(T_j = l). \end{aligned}$$

As we have done previously, once computed the last equality, we have the following probability:

$$\begin{aligned} \mathbb{P}_i(N(j) \geq k) &= \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \cdots \sum_{n=1}^{\infty} \mathbb{P}_i(T_j = l)\, \mathbb{P}_j(T_j = m) \cdots \mathbb{P}_j(T_j = n) \\ &= \sum_{l=1}^{\infty} \mathbb{P}_i(T_j = l) \sum_{m=1}^{\infty} \mathbb{P}_j(T_j = m) \cdots \sum_{n=1}^{\infty} \mathbb{P}_j(T_j = n) \\ &= \rho_{i,j}\, \rho_{j,j}^{k-1} \cdots \rho_{j,j} = \rho_{i,j}\, \rho_{j,j}^{k-1}. \end{aligned}$$

This result is particularly important, as it will assist in proving several theorems concerning the classification of states that we will encounter later. Considering the concepts previously defined and given one state $i \in I$, if the probability satisfies the following equality $\rho_{i,i} = 1$, then

$$\mathbb{P}_i(N(i) = 1) = \lim_{k\to+\infty} \mathbb{P}_i(N(i) \geq k) = \lim_{k\to+\infty} \rho_{i,i}\, \rho_{i,i}^{k-1} = \lim_{k\to+\infty} 1^k = 1,$$

and hence, the state i is **recurrent**, because we can visit this state infinitely many times.

Moreover, the probability $\rho_{i,i}$ may also be less than 1, i.e. $\rho_{i,i} < 1$, and in this case we have that

$$\mathbb{P}_i(N(i) = 1) = \lim_{k\to+\infty} \mathbb{P}_i(N(i) \geq k) = \lim_{k\to+\infty} \rho_{i,i}\, \rho_{i,i}^{k-1} = 0,$$

and now the state i is **transient**, because the number of times we visit this state is finite. After examining the previous equalities, we now focus on defining the number of times a Markov chain visits a given state. In this context, we will introduce the corresponding calculations. of the expected value. Therefore, given that

$$\mathbb{E}_i\Big(\mathbf{1}_{\{X_n=j\}}\Big) = \mathbb{P}_i(X_n = j) = p_{i,j}^{(n)},$$

the expected number of visits to state j, for a chain that begins in i, is the following:

$$\mathbb{E}_i(N(j)) = \sum_{n=1}^{\infty} p_{i,j}^{(n)}.$$

We now turn to the relationship between the classification of states in a Markov chain and the expected number of visits to each state.

Theorem 3.3 (Expected visits and recurrence)**.** *Given a state $j \in I$, which can be transient or recurrent, we have*

$$\mathbb{E}_j(N(j)) = \begin{cases} \frac{\rho_{j,j}}{1-\rho_{j,j}}, & \textit{if } j \textit{ is transient,} \\ \infty, & \textit{if } j \textit{ is recurrent.} \end{cases}$$

We now focus on key properties of recurrent and transient states, essential for classifying all state classes.

Proposition 3.1. *Given one state i, if i is recurrent and $i \to j$, then $j \to i$.*

Proof. We begin by assuming that the property $j \to i$ does not hold, so that $j \not\to i$. This means that once the chain reaches state j, it can never return to the state i in which it was initially located. Consequently, after leaving state i the chain cannot return to it, which implies that i cannot be recurrent. Hence i would be transient, a contradiction. □

By the last proposition, we can check that given both states i, j, the state i is recurrent if once we have given up it we are not able to come back, this means that, $i \to j$ but $j \not\to i$.

Proposition 3.2. *Consider two states i, j. If i is recurrent and $i \to j$, then j is also recurrent.*

Proof. To begin with, we know that if i is recurrent and $i \to j$, then also we have that $j \to i$. Therefore, there exist integers $M, m \geq 1$ such that $p_{j,i}^{(M)} > 0$ and $p_{i,j}^{(m)} > 0$. Then, using the Chapman–Kolmogorov equation, we obtain the following outcome:

$$p_{j,j}^{(M+n+m)} \;\geq\; p_{j,i}^{(M)}\, p_{i,i}^{(n)}\, p_{i,j}^{(m)},$$

which means that to go from state j to the state j in $M+n+m$ steps, we can go from j to i in M steps, then from i to i in n steps and finally from i to j in m steps.

With this, we can compute the following expected value:

$$E_j(N(j)) \;=\; \sum_{k=1}^{\infty} p_{j,j}^{(k)} \;\geq\; \sum_{k=M+n+m}^{\infty} p_{j,j}^{(k)} \;=\; \sum_{n=1}^{\infty} p_{j,j}^{(M+n+m)} \;\geq\; p_{j,i}^{(M)}\, p_{i,j}^{(m)} \sum_{n=1}^{\infty} p_{i,i}^{(n)}.$$

Thus,

$$E_j(N(j)) \;\geq\; p_{j,i}^{(M)}\, p_{i,j}^{(m)}\, E_i(N(i)) \;=\; \infty,$$

because i is a recurrent state. Hence, j is also recurrent. □

To carry on, as an immediate consequence of both properties we have already studied, we have the following results.

Corollary 3.2. *Considering a communication class C, then all its elements are recurrent or transient.*

Corollary 3.3. *Given a class C, which is finite, irreducible and also closed, then all its states are recurrent.*

Proof. First, consider a class C that is finite and closed. This implies that there exists at least one recurrent state within C. Moreover, if C is irreducible, meaning that all states in C communicate with each other, then by the previously established results, every state in C must be recurrent. □

Now, we introduce an example in which we analyze all the concepts we have just studied.

Example 3.9. *Consider the Markov chain with state space*

$$I = \{1, 2, 3, 4, 5\}$$

and transition matrix

$$\Pi = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}.$$

In this case there are two communication classes which are

$$C_1 = \{2, 4\}, \quad C_2 = \{1, 3, 5\}.$$

To continue, we analyze the communication relation that exists between the different states in order to determine whether the previous classes are recurrent or transient. For this, we use the diagram associated to the stochastic matrix, which is omitted here for simplicity.

We first consider the states of class C_1. Since state 4 is transient (it leads to state 1 without the possibility of returning), all states in C_1 are transient as well.

Next, for class C_2, we observe that it is closed, finite, and irreducible, which implies that all its states are recurrent.

Having established the notions of recurrent and transient states through these examples, we are now able to introduce further classifications of the states of a Markov chain.

Definition 3.12. *A state i of a Markov chain is called **absorbing** if $p_{i,i} = 1$. Thus, an absorbing state is necessarily recurrent.*

With the last definition, we obtain that if a Markov chain is in an absorbing state i, it will not leave it since $p_{i,j} = 0,\ \forall j \neq i$. Hence, the closed class is only made up by state i, i.e.

$$C = \{i\}.$$

Definition 3.13 (Essential and Inessential States)**.** *An essential state is one that allows exit and return, in other words, given an essential state i for all $j \in I$ such that $i \to j$ it is also true that $j \to i$.*

Given the prior definition, if the state allows to leave it but not to return, then it is called inessential*, hence state i is inessential if for $j \in I$ there exists $n \geq 1$ such that $p_{i,j}^{(n)} > 0$ but $p_{j,i}^{(m)} = 0,\ \forall m \geq 1$.*

3.4.3 Periodicity and cyclic classes

We now assume the state space is partitioned into disjoint sets, with the chain moving from one set to another and eventually returning to the initial set. These sets are called *cyclic subclasses*. Before analyzing them, we first study the period of essential states, since it determines the number of cyclic subclasses associated with each state.

Definition 3.14 (Period of a State)**.** *Given essential state $i \in I$, the period of the state i denoted $d(i)$ is*

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

so we say that the state i is periodic *if $d(i) > 1$. Moreover i is said to be* aperiodic *if $d(i) = 1$.*

Hence, we have seen that a state $i \in I$ is periodic if the greatest common divisor of the number of steps to return to the starting point is greater than one.

The following proposition studies the fact that if all states are in the same class, then they have the same period.

Proposition 3.3. *Consider two states $i, j \in I$ which satisfy the following property $i \leftrightarrow j$, then the period of the states are the same, $d(i) = d(j)$.*

Proof. Given $i, j \in I$ a pair of distinct states. Then $i \leftrightarrow j$ and there exists $n, m \geq 1$ such that $p_{i,j}^{(n)} > 0$ and $p_{j,i}^{(m)} > 0$, thereby we have

$$p_{i,i}^{(n+m)} \geq p_{i,j}^{(n)}\, p_{j,i}^{(m)} > 0,$$

and hence $n + m$ must be divisible by $d(i)$.

Let $k \geq 1$ such that $p_{j,j}^{(k)} > 0$, now we have

$$p_{i,i}^{(n+k+m)} \geq p_{i,j}^{(n)}\, p_{j,j}^{(k)}\, p_{j,i}^{(m)} > 0,$$

in this case $d(i) \mid (n + k + m)$, thus k is divisible by $d(i)$. This is true for every k such that $p_{j,j}^{(k)} > 0$, so $d(i) \mid d(j)$ by the definition of period. Reversing the roles of i and j we have $d(j) \mid d(i)$, so $d(i) = d(j)$. □

Consequently, the period of a class can be defined as the period of its states.

Now, we consider the congruence equivalence relation, so given $m, n \in \mathbb{Z}$ we have $m \equiv n \pmod d$ which means that $m - n$ is divisible by d.

For the following results, we suppose that C forms an essential states class and we fix a reference state $i \in C$ with period d.

To continue, given $j \in C$ and considering $r \geq 1$ such that $p_{j,i}^{(r)} > 0$, if the probabilities satisfy $p_{i,j}^{(m)} > 0$ and $p_{i,j}^{(n)} > 0$, then we have that

$$p_{i,i}^{(m+r)} \geq p_{i,j}^{(m)} \, p_{j,i}^{(r)} > 0 \quad \text{and} \quad p_{i,i}^{(n+r)} \geq p_{i,j}^{(n)} \, p_{j,i}^{(r)} > 0.$$

Suppose that the period of the state i is d, so it divides $m+r$ and also $n+r$. Consequently, $m-n$ is divisible by d and hence we can rewrite this fact as $m \equiv n \pmod d$. Finally, we define s_j as the remainder when n is divided by d for any n with $p_{i,j}^{(n)} > 0$.

Corollary 3.4. *For any $j \in C$ corresponds an integer $s_j \in \{0, 1, \ldots, d-1\}$ such that $p_{i,j}^{(n)} > 0$, which implies that $n \equiv s_j \pmod d$.*

To carry on, we are going to present the cyclic classes. In order to study this concept, we will need the outcomes obtained right now, where we have been studying the congruence relation. So, considering $h \in \{0, 1, \ldots, d-1\}$, we define

$$C_h = \{j \in C \; ; \; p_{i,j}^{(n)} > 0 \text{ for } s_j \equiv h \pmod d\} = \{j \in C \; ; \; p_{i,j}^{(n)} > 0 \text{ for } n \equiv h \pmod d\},$$

the last equality is due to the congruence for a fixed modulus being an equivalence relation. Moreover, $C_0 = C_d$. Once given this, we can express

$$C = \bigcup_{h=0}^{d-1} C_h,$$

so the sets $C_0, \ldots, C_{d-1}$, which are disjoint, are called the **cyclic subclasses** of I.

Now, we study a result that will be helpful when classifying the states that compose a Markov chain into cyclical subclasses. This result takes into account all the non-null elements of the transition matrix.

Proposition 3.4. *Given $j \in C_k$ and $p_{j,l} > 0$, by convention $C_d = C_0$, it follows that if $k < d-1$, then $l \in C_{k+1}$ and if $k = d-1$, then $l \in C_0$.*

Proof. Let $n \geq 1$ such that $p_{i,j}^{(n)} > 0$. We have that

$$p_{i,l}^{(n+1)} \geq p_{i,j}^{(n)} \, p_{j,l} > 0.$$

As $j \in C_k$, we get that $n \equiv k \pmod d$ and, as the congruence relation is preserved by sums, we obtain $n + 1 \equiv k + 1 \pmod d$, which implies that $l \in C_{k+1}$. □

Example 3.10. *Consider the Markov chain with state space $I = \{1, 2, 3, 4, 5, 6, 7\}$ and stochastic matrix*

$$\Pi = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}.$$

In this case, we have that the Markov chain is irreducible, due to the existence of a unique equivalence class. Now, we study the cyclic subclasses that the chain has.

We note that all states are essential, thus, the period of this class is the period of any state. Consider for example the state $i = 1$, and then $d(1) = 3$. Therefore, there are three cyclic subclasses and these are the following ones

$$C_0 = \{1, 2\}, \quad C_1 = \{4, 7\}, \quad C_2 = \{3, 5, 6\}.$$

Example 3.11. *(An analysis of the random walk on $\mathbb{Z}$)*
In this example, we analyze all the concepts that we have studied previously. As we explained in the first chapter, this example focuses on the real line of integers. We start at state i, and at each time step we can move one step forward or one backward with probabilities p and $q = 1 - p$, respectively. Using the previous explanation, it is easy to check that all the states communicate, thereby there is only one class which is irreducible. To continue, we know that all the states of a given class should be recurrent or transient. In our case, we will focus on checking if the initial state, which is the state 0 (i.e., $X_0 = 0$ to simplify), is recurrent or transient. Then we will deduce that the rest of the states belong to the same type. Before carrying on, we need to mention that the stochastic process is composed of random variables ζ_i that follow a Bernoulli distribution. Then the stochastic process

$$X_n = \sum_{i=1}^{n} \zeta_i$$

follows a Binomial distribution, so we have

$$\mathbb{P}(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Now, to classify the state 0, so it is only necessary to consider the following sum

$$\sum_{m=1}^{\infty} p_{00}^{(m)},$$

which is the expected value of visits to the state 0, for a chain that begins in 0.

Firstly, if we consider m as an odd number we have $p_{00}^{(2n+1)} = 0$, so we cannot come back to the initial state 0 with an odd number of movements, due to the requirement that the number of steps to the right should equal the number of steps to the left.

Then, if the number of movements is even we have that

$$p_{00}^{(2n)} = \mathbb{P}(X_{2n} = 0) = \binom{2n}{n} p^n (1-p)^n.$$

Therefore we get

$$\sum_{m=1}^{\infty} p_{00}^{(m)} \;=\; \sum_{n=1}^{\infty} p_{00}^{(2n)} \;=\; \sum_{n=1}^{\infty} \binom{2n}{n} p^n (1-p)^n.$$

In this case, Stirling's formula, if n is large, provides a good approximation to $n!$, we have

$$n! \;\approx\; n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}, \quad \text{as } n \to \infty.$$

Hence, we can approximate the probability

$$p_{0,0}^{(2n)} \approx \frac{1}{\sqrt{\pi n}} \Big(4p(1-p)\Big)^n.$$

To continue, we have to consider two cases.

Case 1: *When* $p = \frac{1}{2}$*, then*

$$p_{0,0}^{(2n)} \approx \frac{1}{\sqrt{\pi n}}.$$

Thus, the sum becomes

$$\sum_{m=1}^{\infty} p_{0,0}^{(m)} \approx \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = +\infty.$$

In this case, the state 0 *is* ***recurrent****, as the last sum diverges, and hence, all states are recurrent.*

Case 2: *When* $p \neq \frac{1}{2}$*, then*

$$p_{0,0}^{(2n)} \approx \frac{r^n}{\sqrt{\pi n}}, \qquad \text{where } 0 < r = 4p(1-p) < 1.$$

Thus, the sum becomes

$$\sum_{m=1}^{\infty} p_{0,0}^{(m)} \approx \sum_{n=1}^{\infty} \frac{r^n}{\sqrt{\pi n}} \leq \sum_{n=1}^{\infty} r^n < +\infty.$$

In this case, the state 0 *is* ***transient****, as the last sum is finite, and hence, all states are transient.*

As we said before, the chain is irreducible, so there exists a unique class which has a period of 2*. Thus, we have two cyclic subclasses:*

$$C_0 = \{j \in \mathbb{Z} \,;\, p_{0,j}^{(n)} > 0 \ \text{for } n \equiv 0 \pmod 2\} = \{0, \pm 2, \pm 4, \dots\},$$

$$C_1 = \{j \in \mathbb{Z} \,;\, p_{0,j}^{(n)} > 0 \ \text{for } n \equiv 1 \pmod 2\} = \{\pm 1, \pm 3, \pm 5, \dots\}.$$

3.5 Distribution and measure

This section examines the limiting behavior of Markov chains as n approaches infinity, focusing on invariant probability distributions. These describe the long-run fraction of time the chain spends in each state and reveal further useful properties in the study of Markov chains.

In this section, we should take into account a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family of discrete random variables $X_n : \Omega \to I$ where $\{X_n, n \geq 0\}$ is a HMC(ν, Π) and I is a countable set called state space.

3.5.1 Stationary or invariant distribution

This subsection explores the link between the long-term behavior of a Markov chain and its invariant distribution (or measure). A measure $\mu = (\mu_i,\, i \in I)$ with non-negative entries is invariant if it satisfies

$$\mu \Pi = \mu.$$

Definition 3.15 (Invariant distribution)**.** *Given* $\gamma = (\gamma_i, i \in I)$ *a row vector which is a probability distribution such that satisfies*

$$0 \leq \gamma_i \leq 1, \quad \forall i \in I, \qquad \text{and} \qquad \sum_{i \in I} \gamma_i = 1,$$

and considering $\Pi = (p_{i,j}, i, j \in I)$ *a stochastic matrix, then if*

$$\sum_{i \in I} \gamma_i p_{i,j} = \gamma_j, \qquad \forall j \in I, \tag{3.4}$$

we say that γ *is an* ***invariant distribution*** *or* ***stationary distribution*** *for the stochastic matrix* Π.

For a Markov chain $\{X_n, n \geq 0\}$ with transition matrix Π and invariant distribution γ, we can rewrite in matrix form the condition 3.4 as follows

$$\gamma \Pi = \gamma,$$

where γ is a row vector as we said above.

Given γ an invariant distribution, we will study that the distribution (or law) of the HMC(γ, Π), $\{X_n, n \geq 0\}$, is independent of n, for all $n \geq 0$. Therefore, we will see that all random variables X_n have the same law.

For $n = 1$ we have the equality 3.4 which we have analyzed previously. Now we study the case in which $n = 2$, for this we will use the Chapman-Kolmogorov equation so we have

$$\sum_{i \in I} \gamma_i p_{i,j}^{(2)} = \sum_{i \in I} \gamma_i \sum_{k \in I} p_{i,k} p_{k,j} = \sum_{k \in I} \left(\sum_{i \in I} \gamma_i p_{i,k} \right) p_{k,j} = \sum_{k \in I} \gamma_k p_{k,j} = \gamma_j.$$

and hence we get

$$\gamma_j^{(2)} = \mathbb{P}(X_2 = j) = \sum_{i \in I} \gamma_i p_{i,j}^{(2)} = \gamma_j$$

which is the same law as the previous case, where $n = 1$.

After that, we use mathematical induction over the number of steps needed in order for the chain to reach the state j. Assuming the result is true for $n - 1$, we will see that it also holds for n. To continue, we want to compute the distribution of the random variable X_n. For this, first of all, we study whether equality (6.1) holds but in this case n steps are needed to go from state i to state j, so we have

$$\sum_{i \in I} \gamma_i p_{i,j}^{(n)} = \sum_{i \in I} \gamma_i \sum_{i_1, \ldots, i_{n-1} \in I} p_{i,i_1} \cdots p_{i_{n-1},j} = \sum_{i \in I} \gamma_i \sum_{i_{n-1} \in I} p_{i,i_{n-1}}^{(n-1)} p_{i_{n-1},j}$$

$$= \sum_{i_{n-1} \in I} \left(\sum_{i \in I} \gamma_i p_{i,i_{n-1}}^{(n-1)} \right) p_{i_{n-1},j} = \sum_{i_{n-1} \in I} \gamma_{i_{n-1}} p_{i_{n-1},j} = \gamma_j.$$

In this case we get

$$\gamma_j^{(n)} = \mathbb{P}(X_n = j) = \sum_{i \in I} \gamma_i p_{i,j}^{(n)} = \gamma_j,$$

and this is the law of the random variable X_n.

If the distribution of the initial state X_0 is γ, then

$$\sum_{i \in I} \gamma_i p_{i,j}^{(n)} = \gamma_j$$

implies that, for all n,

$$\mathbb{P}(X_n = j) = \gamma_j.$$

Thus, all random variables share the same distribution, which is independent of n. After defining invariant distributions and their properties, we study their existence for any stochastic matrix.

Proposition 3.5 (Existence of invariant distribution)**.** *Given a homogeneous Markov chain $\{X_n, n \geq 0\}$ with finite state space I, then every stochastic matrix Π has an invariant distribution.*

Now, to study the previous concepts, we will use two practical examples that differ on the number of invariant distributions for the stochastic matrix.

Example 3.12. *Consider the Markov chain with state space $I = \{1, 2\}$ and transition matrix*

$$\Pi = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{5} & \frac{4}{5} \end{pmatrix}.$$

In this example, we want to study that this chain has an invariant distribution γ. To find this distribution, we have to prove the equality $\gamma\Pi = \gamma$, this is

$$\begin{pmatrix} \gamma_1 & \gamma_2 \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{5} & \frac{4}{5} \end{pmatrix} = \begin{pmatrix} \gamma_1 & \gamma_2 \end{pmatrix}.$$

With this we obtain the following equations:

$$\gamma_1 = \tfrac{1}{4}\gamma_1 + \tfrac{1}{5}\gamma_2, \qquad \gamma_2 = \tfrac{3}{4}\gamma_1 + \tfrac{4}{5}\gamma_2.$$

Moreover, we know the vector γ is a probability distribution, so it also has to satisfy the equality $\gamma_1 + \gamma_2 = 1$.

Once resolved the system of linear equations, we find that

$$\gamma = \left(\tfrac{19}{24}, \tfrac{5}{24}\right),$$

all the components are positive and satisfy the three equations above, so they represent a unique invariant distribution for the chain.

Example 3.13 (Random walk with absorbing barriers)**.** *In this case, the Markov chain with state space*

$$I = \{0, 1, \ldots, M\}$$

is given by the stochastic matrix

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ 0 & 0 & q & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}.$$

Now we want to compute the invariant distribution for the stochastic matrix Π. For this, we have to analyze the following equality

$$\gamma\Pi = \gamma \qquad \textit{where } \gamma = (\gamma_0, \gamma_1, \ldots, \gamma_M),$$

so we have the following system

$$\begin{cases} \gamma_0 = \gamma_0 + \gamma_1 q, \\ \gamma_1 = \gamma_2 q, \\ \gamma_j = \gamma_{j-1} p + \gamma_{j+1} q, \quad j = 2, \ldots, M-2, \\ \gamma_{M-1} = \gamma_{M-2} p, \\ \gamma_M = \gamma_{M-1} p + \gamma_M. \end{cases}$$

Moreover, we know that the vector γ is a probability distribution, so it also has to satisfy

$$\gamma_0 + \cdots + \gamma_M = 1, \qquad \gamma_i \geq 0 \quad \forall i \in I.$$

Once solved, the system of linear equations yields

$$\gamma_1 = \cdots = \gamma_{M-1} = 0, \qquad \gamma_0 + \gamma_M = 1.$$

Hence the family of invariant distributions is

$$\gamma = (1-\alpha, 0, \ldots, 0, \alpha), \qquad \alpha \in [0,1].$$

The previous examples show that an invariant distribution is not necessarily unique, although at least one always exists. The next theorem establishes the relation between invariant distributions and n-step transition probabilities, and demonstrates that an invariant distribution serves as an equilibrium distribution.

Theorem 3.4. *Given the state space I which is finite, we suppose that for some $i \in I$ it is satisfied that*

$$p_{i,j}^{(n)} \longrightarrow \gamma_j \quad \text{as } n \to \infty \quad \text{for all } j \in I,$$

then $\gamma = (\gamma_j;\, j \in I)$ is an invariant distribution.

Proof. Firstly, we know that $0 \leq \gamma_j \leq 1$ for all $j \in I$, this also holds for the probability $p_{i,j}^{(n)}$, so that

$$0 \leq p_{i,j}^{(n)} \leq 1 \quad \text{for all } n \geq 1 \text{ and } i,j \in I.$$

Now, we analyze the vector γ and see that it is a probability distribution. Using the commutation between the summation and the limit, due to the finiteness of the state space, we have

$$\sum_{j\in I} \gamma_j = \sum_{j \in I} \lim_{n\to+\infty} p_{i,j}^{(n)} = \lim_{n\to+\infty} \sum_{j\in I} p_{i,j}^{(n)} = \lim_{n\to+\infty} 1 = 1.$$

Finally, using the Chapman-Kolmogorov equation, we analyze if γ is an invariant distribution. We get

$$\gamma_j = \lim_{n\to+\infty} p_{i,j}^{(n)} = \lim_{n\to+\infty} p_{i,j}^{(n+1)} = \lim_{n\to+\infty} \sum_{k\in I} p_{i,k}^{(n)}\, p_{k,j} = \sum_{k\in I} \left(\lim_{n\to+\infty} p_{i,k}^{(n)}\right) p_{k,j} = \sum_{k\in I} \gamma_k p_{k,j}.$$

Thus, the vector γ is indeed an invariant distribution, since the equality holds. □

Example 3.14. *Consider the Markov chain of Example 3.12 Our goal is to compute the transition probabilities $p_{1,1}^{(n)}$ and $p_{2,2}^{(n)}$.*

First we have to compute the eigenvalues of the transition matrix Π. For this, we calculate the characteristic equation:

$$\det(\Pi - \lambda I_d) = 0 \quad \Longrightarrow \quad (1/4-\lambda)(4/5-\lambda) - \tfrac{20}{3} = 0.$$

Solving this equation we obtain the eigenvalues $\lambda = 1, \frac{1}{20}$. Now, we can rewrite the transition matrix Π as a diagonal matrix and hence there exists an invertible matrix A such that

$$\Pi = A \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} A^{-1} \quad \Longrightarrow \quad \Pi^{(n)} = A \begin{pmatrix} 1^n & 0 \\ 0 & \left(\frac{1}{2}\right)^n \end{pmatrix} A^{-1}.$$

Now, we want to compute the transition probability $p_{1,1}^{(n)}$. Using the n-th power of the stochastic matrix we obtain

$$p_{1,1}^{(n)} = \alpha \cdot 1^n + \beta \left(\tfrac{1}{2}\right)^n = \alpha + \beta \left(\tfrac{1}{2}\right)^n .$$

To continue, we calculate the values of the constants α and β, so we have to solve the following system of linear equations

$$\begin{cases} 1 = \alpha + \beta, \\ \frac{1}{4} = \alpha + \beta \cdot \frac{1}{20}, \end{cases}$$

so, the values of the constants are the following

$$\alpha = \frac{19}{4}, \quad \beta = \frac{15}{19}.$$

Now, we use the last theorem applied on the probability $p_{1,1}^{(n)}$ and we have

$$\lim_{n\to+\infty} p_{1,1}^{(n)} = \lim_{n\to+\infty} \left(\frac{19}{4} + \frac{15}{19} \left(\frac{1}{20}\right)^n \right) = \frac{19}{4} = \gamma_1,$$

and with this we obtain that the distribution γ_1 is invariant. If we had computed the probability $p_{2,1}^{(n)}$, the result would have been the same.

Now, we repeat the same idea on the probability $p_{2,2}^{(n)}$, then we get

$$p_{2,2}^{(n)} = \mu \cdot 1^n + \lambda \left(\frac{1}{20}\right)^n = \mu + \lambda \left(\frac{1}{20}\right)^n ,$$

and now we want to compute the constants μ and λ, so we have to solve the following system of linear equations

$$\begin{cases} 1 = \mu + \lambda, \\ \frac{4}{5} = \mu + \lambda \cdot \frac{1}{20}. \end{cases}$$

In this case, we obtain the values

$$\mu = \frac{15}{19}, \quad \lambda = \frac{19}{4},$$

and hence the limit is the following

$$\lim_{n\to+\infty} p_{2,2}^{(n)} = \lim_{n\to+\infty} \left(\frac{15}{19} + \frac{19}{4} \left(\frac{1}{20}\right)^n \right) = \frac{15}{19} = \gamma_2.$$

With this we obtain that the distribution γ_2 is invariant. If we had computed the probability $p_{1,2}^{(n)}$ the result would have been the same.

Therefore, we get that

$$\gamma = (\gamma_1, \gamma_2) = \left(\frac{19}{4}, \frac{15}{19} \right),$$

and it is an invariant distribution. We observe that the invariant distribution may not exist, be unique or more than one distribution may exist. We are going to introduce some results that allow us to set conditions in order to guarantee the existence and uniqueness of the invariant measure.

Definition 3.16 (Measure and invariant measure)**.** *A measure is any row vector* $\mu = (\mu_i,\ i \in I)$ *with* $\mu_i \geq 0$ *for all* $i \in I$*. Moreover, we say a measure* μ *is invariant, if for any transition matrix* $\Pi = (p_{i,j},\ i,j \in I)$ *satisfies*

$$\sum_{i \in I} \mu_i p_{i,j} = \mu_j, \quad \forall j \in I.$$

This equality in matrix form is the following:

$$\mu \Pi = \mu.$$

Now, in the following result, we analyze the existence of the invariant measure.

Theorem 3.5 (Existence of the invariant measure)**.** *Given the stochastic matrix* $\Pi = (p_{i,j};\ i,j \in I)$ *of the Markov chain* $\{X_n,\ n \geq 0\}$*, which is irreducible and where all states are recurrent. Consider*

$$\lambda_i^r = \mathbb{E}_r \left[\sum_{n=0}^{T_r - 1} \mathbf{1}_{\{X_n = i\}} \right],$$

then

1. $\lambda_i^r \in (0,1)$ *for all* $i \in I$*.*
2. $\lambda^r = (\lambda_i^r,\ i \in I)$ *satisfies* $\lambda^r \Pi = \lambda^r$*,*

where $T_r = \inf\{n \geq 1\ ;\ X_n = r\}$ *is the first time the chain visits the state* r*.*

Proof. Given $n \geq 1$, we have that the event $\{T_r \geq n\}$ depends only on $X_0, \ldots, X_{n-1}$ because none of the random variables X_m, for all $m \geq 1$, are in state r. Using the Markov property at time $n-1$ we get

$$\mathbb{P}_r(X_1 = l, \ldots, X_{n-1} = i, X_n = j \text{ and } T_r \geq n) = \mathbb{P}_r(X_{n-1} = i \text{ and } T_r \geq n)\, p_{i,j}.$$

We also know that the states of the Markov chain are recurrent, so

$$\mathbb{P}_r(T_r < \infty) = 1,$$

which is the same as

$$\mathbb{P}(X_0 = X_{T_r} = r) = 1.$$

Now, we have

$$\lambda_j^r = \mathbb{E}_r \left[\sum_{n=1}^{T_r} \mathbf{1}_{\{X_n = j\}} \right] = \sum_{n=1}^{\infty} \mathbb{P}_r(X_n = j \text{ and } T_r \geq n).$$

To continue, note that before visiting state j at time n, the chain has been in some state $i \in I$ at time $n-1$. Hence

$$\lambda_j^r = \sum_{i \in I} \sum_{n=1}^{\infty} \mathbb{P}_r(X_{n-1} = i, X_n = j, T_r \geq n).$$

By the Markov property, this is

$$\lambda_j^r = \sum_{i \in I} p_{i,j} \sum_{n=1}^{\infty} \mathbb{P}_r(X_{n-1} = i, T_r \geq n).$$

Now we rewrite the last sum as

$$\sum_{n=1}^{\infty} \mathbb{P}_r(X_{n-1} = i, T_r \geq n) = \mathbb{E}_r\left[\sum_{\ell=0}^{T_r-1} \mathbf{1}_{\{X_\ell = i\}}\right] = \lambda_i^r.$$

Therefore,

$$\lambda_j^r = \sum_{i \in I} \lambda_i^r p_{i,j}.$$

This shows that $\lambda^r \Pi = \lambda^r$, which proves part (ii). To prove part (i), recall that all states of the Markov chain communicate among them. Then, for each state $i \in I$, there exist $n, m \geq 0$ such that $p_{i,r}^{(m)} > 0$ and $p_{r,i}^{(n)} > 0$.

From the definition, we have

$$\lambda_i^r = \sum_{k \in I} \lambda_k^r p_{k,i}^{(n)} \geq \lambda_r^r p_{r,i}^{(n)} = p_{r,i}^{(n)} > 0.$$

In addition, since

$$1 = \lambda_r^r = \sum_{k \in I} \lambda_k^r p_{k,r}^{(m)} \geq \lambda_i^r p_{i,r}^{(m)},$$

we deduce that

$$\lambda_i^r \leq \frac{1}{p_{i,r}^{(m)}} < 1.$$

Hence, the vector λ^r satisfies $\lambda_i^r \in (0, 1)$ for all $i \in I$, which shows (i). □

First, remember that a state i is *recurrent* if it satisfies the equality

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1.$$

Now, we studied that this is equivalent to the following condition

$$\rho_{i,i} = \mathbb{P}_i(T_i < \infty) = 1.$$

To continue, we introduce some new concepts.

3.6 Positive recurrence

Definition 3.17 (Positive and Null Recurrence). *A state $i \in I$ is said to be* positive recurrent *if, starting at the state i, the expected time till the chain returns to the state i is finite. This is equivalent to*

$$m_i = \mathbb{E}_i(T_i) < \infty.$$

In the case in which the state $i \in I$ is recurrent and does not satisfy the previous condition, in other words if

$$m_i = \mathbb{E}_i(T_i) = \infty,$$

the state i is called null recurrent.

In the last definition, the variable m_i is called the mean recurrence time *at state i, which is the expected return time to this state.*

Corollary 3.5. *Positive recurrent states and null recurrent states are both recurrent.*

Proof. For a positive recurrent state, we have $m_i < \infty$ and this means that T_i cannot be ∞ with strictly positive probability. Hence, state i is recurrent. On the other hand, for null recurrent states, it is given by definition that these states are recurrent. □

Now, we analyze that for a Markov chain with stochastic matrix Π, to say that the chain has positive recurrent states is equivalent to say that the stochastic matrix Π has an invariant distribution. In our case, the study focuses on irreducible Markov chains.

Theorem 3.6. *Given an irreducible Markov chain with stochastic matrix $\Pi = (p_{i,j};\, i, j \in I)$, the following properties are equivalent:*

1. *some state i is positive recurrent,*
2. *all states are positive recurrent,*
3. *the stochastic matrix Π has an invariant distribution γ.*

In particular, when (iii) holds we have

$$m_i = \frac{1}{\gamma_i}, \quad \forall i \in I.$$

Example 3.15 (Symmetric random walk)**.** *Consider the example of the random walk on $\mathbb{Z}$ for the case in which $p = \frac{1}{2} = q$. We know that it is an irreducible Markov chain. We know that there is an invariant measure μ and any invariant measure is a scalar multiple of μ.*

Now, we analyze whether the symmetric random walk is null recurrent or positive recurrent. We have

$$\sum_{i\in I} \mu_i = \infty,$$

so there can be no invariant distribution and, all states of the walk are null recurrent.

Finally, we introduce an important result which is based on irreducible Markov chains with finite state space I. In this result, a relationship with positive recurrent states is established.

Proposition 3.6. *Given an irreducible homogeneous Markov chain with finite state space I, then it is positive recurrent.*

Proof. First, we show recurrence. For this, assume that the Markov chain is transient. So, for all $i, j \in I$, we get

$$\sum_{n=1}^{\infty} p_{i,j}^{(n)} = \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = j) = \sum_{n=1}^{\infty} \mathbb{E}_i(\mathbf{1}_{\{X_n = j\}}) = \mathbb{E}_i(N(j)) < \infty,$$

which contradicts irreducibility. Hence the chain is recurrent.

Since the state space is finite, every recurrent state is necessarily positive recurrent, and the proof is complete. □

and hence, as the state space is finite we have

$$\sum_{j\in I} \sum_{n=1}^{\infty} p_{i,j}^{(n)} < \infty,$$

but the previous sum is equal to

$$\sum_{n=1}^{\infty}\sum_{j\in I} p_{i,j}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty,$$

which is a contradiction. Therefore, this implies that the Markov chain is recurrent.

Now, by previous theorems, there exists an invariant measure λ^r and, as the state space is finite,

$$\sum_{i\in I} \lambda_i^r = \mathbb{E}_r(T_r) = m_r < \infty.$$

Therefore, the chain is positive recurrent.

3.7 Finite state space

In this section, we study the uniqueness of the invariant distribution considering the case in which the state space I is finite, so $I = \{1,\dots,N\}$. For this, we refer to the non-degenerate probability distribution of I, which we denote by $\pi = (\pi_i;\, i \in I)$. This distribution satisfies $0 < \pi_i \leq 1$ and also $\sum_{i\in I} \pi_i = 1$.

Now, we give a definition which is necessary to study the ergodic theorem with finite state space.

Definition 3.18 (Ergodicity)**.** *A homogeneous Markov chain, with stochastic matrix* $\Pi = (p_{i,j};\, i,j \in I)$*, is* ergodic *if, for all* $i \in I$*, it satisfies the following property:*

$$\pi_j = \lim_{n\to+\infty} p_{i,j}^{(n)}, \quad \forall j \in I,$$

where π_j *is a non-degenerate probability distribution.*

In the previous definition, we observe that the limit is independent of the state $i \in I$. To continue, the ergodicity helps us to study the problem of the uniqueness of the invariant distribution. In the following result we analyze this fact.

Theorem 3.7. *Given a homogeneous Markov chain* $\{X_n;\, n \geq 0\}$*, with finite state space* I *and stochastic matrix* Π*. Suppose that there exists* $k \geq 1$ *such that*

$$\min_{i,j\in I} p_{i,j}^{(k)} > 0.$$

Then, for all $i \in I$*, there exists* $\pi = (\pi_j;\, j \in I)$ *which satisfies*

$$\pi_j = \lim_{n\to+\infty} p_{i,j}^{(n)}, \quad \forall j \in I,$$

where π *is a non-degenerate probability distribution. Furthermore, for all* $j \in I$*, it satisfies*

$$\sum_{l\in I} \pi_l\, p_{l,j} = \pi_j,$$

and therefore π *is the unique invariant distribution for* Π*.*

To continue, once studied the ergodic theorem for finite state space, we will analyze that the left implication is also true.

Corollary 3.6. *Given a non-degenerate probability distribution* $\pi = (\pi_j;\ j \in I)$ *over space* I *which satisfies*

$$\lim_{n\to+\infty} p_{i,j}^{(n)} = \pi_j,$$

then there exists an integer $k \geq 1$ *for which*

$$\min_{i,j\in I} p_{i,j}^{(k)} > 0.$$

Definition 3.19 (Regular Markov Chain)**.** *A Markov chain with finite state space is said to be* regular *if there exists a power of the transition matrix* Π *with only positive entries.*

Example 3.16. *Consider the following stochastic matrix*

$$\Pi = \begin{pmatrix} 1-b & a \\ b & 1-a \end{pmatrix},$$

where $a, b > 0$ *and* $a + b > 0$.

Now, we want to compute the m-step transition probability of every state. For this, it is necessary to compute the m-th power of the stochastic matrix, $\Pi^{(m)}$*. Firstly, we have to calculate the eigenvalues of* Π*. For this, we compute the characteristic equation:*

$$\det(\Pi - \lambda I) = 0 \implies \lambda^2 - \lambda + (a+b)(\lambda - 1) + 1 = 0.$$

Thus, solving the equation we obtain the following eigenvalues:

$$\lambda_1 = 1, \qquad \lambda_2 = 1 - (a+b).$$

The corresponding eigenvectors are $(1, 1)$ *and* $(a, -b)$*, respectively. With this, we can define the following matrices:*

$$Q = \begin{pmatrix} 1 & 1 \\ a & -b \end{pmatrix}, \qquad Q^{-1} = \frac{1}{a+b}\begin{pmatrix} b & 1 \\ a & -1 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 1-a-b \end{pmatrix}.$$

Hence, we can rewrite the transition matrix as

$$\Pi = QDQ^{-1}, \qquad \text{which implies that} \qquad \Pi^{(m)} = QD^{(m)}Q^{-1}.$$

Therefore, we obtain

$$\Pi^{(m)} = \frac{1}{a+b}\begin{pmatrix} 1 & 1 \\ a & -b \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & (1-a-b)^m \end{pmatrix}\begin{pmatrix} b & 1 \\ a & -1 \end{pmatrix}.$$

Thus, each entry of $\Pi^{(m)}$ *represents the m-step transition probability of going from one state to another.*

To continue, we introduce a practical example in which we analyze the concept which we have just defined.

Example 3.17. *Consider the Markov chain with state space* $I = \{1, 2, 3\}$ *and transition matrix*

$$\Pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ 1 & 0 & 0 \end{pmatrix}.$$

We will study if the Markov chain is regular or not. For this, we are going to compute the successive powers of the stochastic matrix in order to check if there exists a matrix where all of its entries are non-null. Firstly, we calculate Π^2*, and we get*

$$\Pi^2 = \begin{pmatrix} \frac{7}{20} & \frac{7}{20} & \frac{3}{10} \\ \frac{37}{50} & \frac{7}{50} & \frac{3}{25} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

Since this matrix has a null entry, we proceed to calculate Π^3*, which is the following:*

$$\Pi^3 = \begin{pmatrix} \frac{109}{200} & \frac{49}{200} & \frac{21}{100} \\ \frac{259}{500} & \frac{199}{500} & \frac{21}{250} \\ \frac{7}{20} & \frac{7}{20} & \frac{3}{10} \end{pmatrix}.$$

Now, all entries of the matrix are positive. Therefore, this implies that the Markov chain is regular.

To continue, using the concept which we have just introduced, the *regularity*, we will study the uniqueness of the invariant distribution of Theorem 7.2.2.

Proposition 3.7. *Consider the stochastic matrix*

$$\Pi = (p_{i,j},\ i,j \in I)$$

which is regular. Then the invariant distribution

$$\pi = (\pi_j,\ j \in I)$$

is unique.

Example 3.18. *(The Tower Guard)*

A guard can stand at the four corners of a tower in the following way: after staying 5 minutes in a corner, he tosses a coin; if it lands heads (with probability p*) he moves to the left, otherwise (tails, with probability* $q = 1 - p$*) he moves to the right. The process is repeated indefinitely. Study the Markov chain associated with this process.*

We number the corners $1, 2, 3, 4$ *in circular order (for example clockwise). If* $p = \frac{1}{2}$ *then* $q = \frac{1}{2}$*, and the transition matrix* Π *(rows corresponding to current states, columns to next states) is written as*

$$\Pi = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix},$$

because from one corner the guard moves with probability $\frac{1}{2}$ *to each of the two neighboring corners.*

Note that, for any $n \geq 1$*,*

$$\Pi^{(n)} = \begin{cases} \Pi, & \text{if } n \text{ is odd,} \\ \Pi^2, & \text{if } n \text{ is even,} \end{cases}$$

and explicitly

$$\Pi^2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

1. ***Case: initial position chosen at random.***
   If the initial position X_0 is chosen uniformly from $\{1, 2, 3, 4\}$, then the initial distribution is
   $$\mu_0 = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$
   Since each row of Π and Π^2 is just a permutation of the components (and the uniform distribution is invariant under permutations), for all $n \geq 0$:
   $$\mu_n = \mu_0 \Pi^{(n)} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$
   In other words, the distribution remains uniform for all n.

2. ***Case: starting from a fixed corner.***
   Suppose the guard starts at corner 1, so that $\mu_0 = (1, 0, 0, 0)$. Then, for any integer $n \geq 1$:
   $$\mu_n = \mu_0 \Pi^{(n)} = \begin{cases} (0, \frac{1}{2}, 0, \frac{1}{2}), & \text{if } n \text{ is odd,} \\ (\frac{1}{2}, 0, \frac{1}{2}, 0), & \text{if } n \text{ is even.} \end{cases}$$
   That is, the guard alternates at each step between being with probability $\frac{1}{2}$ on the corners of opposite parity and on the corners of the same parity as the starting point.

Remark. The uniform distribution $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ is an invariant distribution of the chain; the chain is periodic (period 2), which explains the oscillation between two distributions when starting from a fixed corner. Secondly, if n is an odd number, then the law of X_n is given by

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix}.$$

Example 3.19. *Consider the Markov chain with state space*

$$I = \{1, 2, 3, 4, 5\}$$

and transition matrix

$$\Pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

*The class C_1 is **recurrent** because once we reach class C_1 we can no longer leave it; moreover, it is finite and irreducible.*

*The state of class C_2 is **recurrent**, in particular it is an **absorbing state**, because if we start from state 4, we can move to state 3, and if we start from state 3, we always remain there.*

*The state of class C_3 is **transient**, because if we start from state 4, we can move to state 3, and once we reach 3 we can no longer return to 4.*

*The state of class C_4 is **transient**, because if we start from state 5, we can move to state 1, and once we reach 1 we can no longer return to 5.*

Example 3.20. *Consider the Markov chain with state space*

$$I = \{1, 2, 3, 4, 5\}$$

and transition matrix

$$\Pi = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

In this case, there exists only one communication class:

$$C_1 = \{1, 2, 3, 4, 5\}.$$

*Next, we analyze the communication relation that exists between the different states. In this case we have a single class C_1, which contains all the states of the Markov chain. Moreover, by using the previous diagram, it is easy to see that all states communicate, so the chain is **irreducible**. In particular, class C_1 is **recurrent** because we can visit each state infinitely many times.*

Example 3.21. *Consider the Markov chain with state space*

$$I = \{1, 2, 3, 4\}$$

and transition matrix

$$\Pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

*In this case, we have a single class C_1, which contains all the states of the Markov chain. Moreover, by using the previous diagram, it is easy to see that all states communicate; therefore, the chain is **irreducible**. In particular, class C_1 is **recurrent** because we can visit each state infinitely many times.*

3.8 Exercises

Exercise 3.1. *For $p, q \in [0,1]$, let X be the two-state chain $(1,2)$, with transition matrix*

$$P = \begin{pmatrix} 1-q & p \\ q & 1-p \end{pmatrix}.$$

1. *For which values of p, q is the chain irreducible? aperiodic?*
2. *For each p, q, find the set D of all invariant distributions of X.*
3. *Compute P^n, $n \in \mathbb{N}$.*
4. *When X is irreducible, compute*

$$d_1(n) := \tfrac{1}{2}\Big(\big|\mathbb{P}_1(X_n = 1) - \pi(1)\big| + \big|\mathbb{P}_1(X_n = 2) - \pi(2)\big| \Big),$$

and

$$d_2(n) := \tfrac{1}{2}\Big(\big|\mathbb{P}_2(X_n = 1) - \pi(1)\big| + \big|\mathbb{P}_2(X_n = 2) - \pi(2)\big| \Big).$$

5. *Draw the shape of the graphs $n \mapsto d_i(n)$, $i = 1, 2$, for:*

$$(p,q) = (0.5, 0.5), \quad (p,q) = (0.4, 0.1), \quad (p,q) = (0.9, 0.95), \quad (p,q) = (1,1).$$

Solutions.

1. *The chain is irreducible iff $p > 0$ and $q > 0$. It is aperiodic if $(p,q) \neq (1,1)$.*
2. *If $p = q = 0$, then*

$$D = \Big\{ \alpha\delta_1 + (1-\alpha)\delta_2 \ : \ \alpha \in [0,1] \Big\}.$$

Otherwise,

$$D = \left\{ \frac{q}{p+q}\delta_1 + \frac{p}{p+q}\delta_2 \right\}.$$

3. *If $p = q = 0$, then $P^n = I_2$ for any $n \in \mathbb{N}$. Otherwise, we use diagonalization:*

$$P = ADA^{-1}, \qquad A = \begin{pmatrix} 1 & 1 \\ -q & p \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix}, \quad A^{-1} = \frac{1}{p+q}\begin{pmatrix} -q & p \\ 1 & 1 \end{pmatrix}.$$

Hence

$$P^n = AD^nA^{-1} = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix} + (1-p-q)^n \cdot \frac{1}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}.$$

4. *For $n \in \mathbb{N}$,*

$$d_1(n) = \tfrac{1}{2}\Big(|P^n(1,1) - \pi(1)| + |P^n(1,2) - \pi(2)| \Big) = \frac{p}{p+q}\,|1-p-q|^n.$$

By symmetry,

$$d_2(n) = \frac{q}{p+q}\,|1-p-q|^n.$$

5. *The graphs $n \mapsto d_i(n)$ show exponential decay in n when the chain is irreducible, except in the degenerate case $(p, q) = (1, 1)$ where the chain alternates between states.*

Exercise 3.2. *Let X be an E-valued Markov chain whose transition kernel P is assumed to be symmetric.*

1. *Assume that E is finite. Show that the uniform distribution on E is invariant.*

2. *Assume again that E is finite. Under what condition can it be said that the uniform distribution is the unique invariant distribution? What happens if this condition is not fulfilled?*

3. *Can you find an example with E infinite (countable), P irreducible and symmetric, in which the chain X does not possess any invariant distribution?*

4. *Can you find an example with E infinite (countable), P irreducible and symmetric, in which the chain X possesses a unique invariant distribution?*

Solution.

1. *Let π be the uniform distribution on E. If $|E| = n$, then $\pi(x) = \frac{1}{n}$ for all $x \in E$. Since P is symmetric,*

$$\sum_{x\in E} \pi(x)P(x, y) = \frac{1}{n}\sum_{x\in E} P(x, y) = \frac{1}{n}\sum_{x\in E} P(y, x) = \frac{1}{n} = \pi(y).$$

Thus π is invariant.

2. *If X is irreducible, then its stationary distribution is unique, so it must be π. If the chain is not irreducible, then E splits into several closed recurrent classes. On each finite class, the uniform distribution is stationary, and any convex combination of such uniform distributions is also stationary. Hence there are infinitely many invariant distributions in this case, and π is only one possible example.*

3. *Example: the simple random walk (SRW) on $\mathbb{Z}^d$ with $d \geq 1$. The transition matrix is symmetric and the chain is irreducible. However, there is no stationary distribution since $\mathbb{Z}^d$ is infinite and no probability measure can remain invariant under the uniform mass distribution.*

4. *Such an example cannot exist. Indeed, suppose X is irreducible, symmetric, and possesses an invariant distribution π. Then any invariant measure must be a multiple of π. But the infinite counting measure μ, assigning mass 1 to each state $x \in E$, is invariant (by the same argument as in part (1)). Since μ cannot be normalized into a probability distribution on an infinite E, it cannot be a multiple of π. This contradiction shows that no such example exists.*

Exercise 3.3. *Suppose that $\xi_0, \xi_1, \xi_2, \ldots$ are independent random variables with common probability function*

$$f(k) = \mathbb{P}(\xi_0 = k), \quad k \in \mathbb{Z}.$$

Let $S = \{1, \ldots, N\}$. Let X_0 be another random variable, independent of the sequence (ξ_n), taking values in S, and let $f : S \times \mathbb{Z} \to S$ be a given function. Define new random variables $X_1, X_2, \ldots$ by

$$X_{n+1} = f(X_n, \xi_n), \quad n = 0, 1, 2, \ldots$$

1. *Show that* (X_n) *forms a Markov chain.*
2. *Find its transition probabilities.*

Solution.

1. Fix a time $n \geq 1$. Suppose that we know $X_n = x$. The goal is to show that

   $$\text{Past} = (X_0, \ldots, X_{n-1}) \quad \text{is independent of} \quad \text{Future} = (X_{n+1}, X_{n+2}, \ldots).$$

   The variables in the Past are functions of

   $$X_0, \xi_1, \ldots, \xi_{n-2}.$$

   The variables in the future are functions of

   $$x, \xi_n, \xi_{n+1}, \ldots$$

   But $X_0, \xi_1, \ldots, \xi_{n-2}$ are independent of $\xi_n, \xi_{n+1}, \ldots$. Therefore, the past and the future are independent, which proves the Markov property.

2. We compute the transition probabilities:

   $$\begin{aligned} \mathbb{P}(X_{n+1} = y \mid X_n = x) &= \mathbb{P}(f(X_n, \xi_n) = y \mid X_n = x) \\ &= \mathbb{P}(f(x, \xi_n) = y \mid X_n = x) \\ &= \mathbb{P}(f(x, \xi_n) = y) \\ &= \mathbb{P}(f(x, \xi_0) = y) \\ &= \mathbb{P}(\xi_0 \in A_{x,y}), \end{aligned}$$

   where

   $$A_{x,y} := \{\xi \in \mathbb{Z} : f(x, \xi) = y\}.$$

Exercise 3.4. *Discuss the topological properties of the graphs of the following Markov chains:*

$$(a) \quad P = \begin{pmatrix} 0 & 0.5 \\ 0.5 & 0 \end{pmatrix}, \qquad (b) \quad P = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix},$$

$$(c) \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 0 & 2/3 \\ 4/5 & 1/3 & 0 \end{pmatrix}, \qquad (d) \quad P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$(e) \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

Solution. *Draw the transition diagram for each case.*

1. *(a)Irreducible?* ***YES****, because there is a path from every state to any other state. Aperiodic?* ***YES****, because the times n for which* $p_{1,1}^{(n)} > 0$ *are* $1, 2, 3, 4, 5, \ldots$ *and their gcd is* 1.

2. *(b) Irreducible?* ***YES****, because there is a path from every state to any other state. Aperiodic?* ***YES****, for the same reason: the set of return times has gcd equal to* 1.

3. *(c) Irreducible?* ***NO****, because starting from state* 2 *it remains at* 2 *forever. However, it can be checked that all states have period* 1*, since* $p_{i,i} > 0$ *for all* $i = 1, 2, 3$*.*

4. *(d) Irreducible?* ***YES****, because there is a path from every state to any other state. Aperiodic?* ***NO****, because the times* n *for which* $p_{1,1}^{(n)} > 0$ *are* $2, 4, 6, \ldots$ *and their gcd is* 2*.*

5. *(e) Irreducible?* ***YES****, because there is a path from every state to any other state. Aperiodic?* ***YES****, because the set of return times to each state contains all integers* $n \geq 1$*, hence the gcd is* 1*.*

Exercise 3.5. *Consider a Markov chain with two states 1,2. Suppose that* $p_{1,2} = a, p_{2,1} = b$*. For which values of* a *and* b *do we obtain an absorbing Markov chain?*
Solution *One of them (or both) should be zero. Because, if they are both positive, the chain will keep moving between 1 and 2 forever.*

Exercise. Consider a Markov chain with states

$$S = \{0, 1, \ldots, N\}$$

and transition probabilities

$$p_{i,i+1} = p, \quad p_{i,i-1} = q, \quad 1 \leq i \leq N-1,$$

where $p + q = 1$, $0 < p < 1$; and assume

$$p_{0,1} = 1, \qquad p_{N,N-1} = 1.$$

1. Draw the graph (transition diagram).
2. Is the Markov chain irreducible?
3. Is it aperiodic?
4. What is the period of the chain?
5. Find the stationary distribution.

Solution.

2. Yes, it is possible to go from any state to any other state (the chain is irreducible).
3. Yes, because $p_{0,0} > 0$ (hence the chain is aperiodic).
4. The period of the chain is 1.
5. We write the balance equations by equating fluxes:

$$\pi(i)q = \pi(i-1)p, \qquad 1 \leq i \leq N.$$

As long as $1 \leq i \leq N$, we have

$$\pi(i) = \frac{p}{q}\pi(i-1) = \left(\frac{p}{q}\right)^2 \pi(i-2) = \cdots = \left(\frac{p}{q}\right)^i \pi(0), \qquad 0 \leq i \leq N.$$

Since

$$\pi(0) + \pi(1) + \cdots + \pi(N-1) + \pi(N) = 1,$$

we obtain

$$\pi(0)\left(1 + \tfrac{p}{q} + \left(\tfrac{p}{q}\right)^2 + \cdots + \left(\tfrac{p}{q}\right)^N\right) = 1.$$

This gives

$$\pi(0) = \left(1 + \tfrac{p}{q} + \left(\tfrac{p}{q}\right)^2 + \cdots + \left(\tfrac{p}{q}\right)^N\right)^{-1}.$$

For $p \neq q$, the sum is geometric, hence

$$\pi(0) = \frac{1 - \frac{p}{q}}{1 - \left(\frac{p}{q}\right)^{N+1}}.$$

As long as $p \neq q$. Hence, if $p \neq q$,

$$\pi(i) = \frac{\left(\frac{p}{q}\right)^N - 1}{\frac{p}{q} - 1} \left(\tfrac{p}{q}\right)^i, \quad 0 \leq i \leq N.$$

If $p = q = \frac{1}{2}$, then

$$\pi(0) = (1 + 1 + \cdots + 1)^{-1} = \frac{1}{N+1},$$

and so

$$\pi(i) = \frac{1}{N+1}, \qquad 0 \leq i \leq N.$$

Thus, in this case, $\pi(i)$ is the uniform distribution on the set of states.

Chapter 4

Poisson Processes

The Poisson point process is a stochastic process that assigns a probability distribution to configurations of points on $\mathbb{R}^+$ [14,19,22,27]. These points may represent, for example, the times at which a bus arrives at a stop, the arrival times of telephone calls at a central office, and so on. In some cases, such as when buses follow a regular schedule without traffic disturbances, the arrival times are fairly evenly spaced. In other situations, for instance when roadworks disrupt the traffic, the arrival times become much more irregular, and one may observe both long waiting times and buses arriving almost immediately one after another.For a Poisson point process the number of points in a given set has a Poisson distribution. Moreover, the numbers of points in disjoint sets are stochastically independent. A Poisson process exists on a general s-finite measure space. Its distribution is characterised by a specific exponential form of the Laplace functional.

4.1 Exponential distribution

The gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt, \quad \alpha > 0.$$

Theorem 4.1. *The gamma function satisfies the following properties:*

1. *For each $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.*
2. *For each integer $n \geq 1$, $\Gamma(n) = (n - 1)!$.*
3. *$\Gamma(1/2) = \sqrt{\pi}$.*

Proof. For each $\alpha > 1$, by integration by parts,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt = \int_0^\infty t^{\alpha-1} d(-e^{-t}) = \left[-t^{\alpha-1}e^{-t}\right]_0^\infty + \int_0^\infty (\alpha-1)t^{\alpha-2}e^{-t}\, dt = (\alpha-1)\Gamma(\alpha-1).$$

Next, we prove by induction that for each integer $n \geq 1$, $\Gamma(n) = (n-1)!$. The case $n = 1$ holds because

$$\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1.$$

Assume $\Gamma(n) = (n-1)!$. Then,

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)! = n!.$$

Finally, by the change of variables $t = \frac{x^2}{2}$, $dt = x\,dx$,

$$\Gamma(1/2) = \int_0^\infty t^{-1/2} e^{-t}\,dt = \int_0^\infty \left(\frac{x^2}{2}\right)^{-1/2} e^{-x^2/2} x\,dx$$

$$= \sqrt{2}\int_0^\infty e^{-x^2/2}\,dx = \frac{\sqrt{2}}{2}\int_{-\infty}^\infty e^{-x^2/2}\,dx = \sqrt{2}\cdot\sqrt{\frac{\pi}{2}} = \sqrt{\pi}.$$

□

In particular, we have:

$$\int_0^\infty x^n e^{-x}\,dx = \Gamma(n+1) = n! \tag{4.1}$$

By the change of variable $y = \frac{x}{\theta}$, we get:

$$\int_0^\infty x^{\alpha-1} e^{-x/\theta}\,dx = \int_0^\infty y^{\alpha-1}\theta^\alpha e^{-y}\,dy = \theta^\alpha\Gamma(\alpha) \tag{4.2}$$

Definition 4.1. *A random variable X follows an* ***exponential distribution*** *with parameter $\lambda > 0$ if its density is given by:*

$$f(x) = \begin{cases} 0 & \text{if } x < 0, \\ \lambda e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

We write this as $X \sim$ Exponential(λ).

The above function f is indeed a probability density function because it is nonnegative and

$$\int_{-\infty}^\infty f(t)\,dt = \int_0^\infty \lambda e^{-\lambda t}\,dt = \left[-e^{-\lambda t}\right]_0^\infty = 1.$$

Theorem 4.2. *Let $X \sim$ Exponential(λ), then:*

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}, \quad \mathbb{E}[X^k] = \frac{k!}{\lambda^k}$$

Proof. We have

$$\mathbb{E}[X^k] = \int_0^\infty x^k\cdot\lambda e^{-\lambda x}\,dx = \lambda\int_0^\infty x^k e^{-\lambda x}\,dx = \lambda\cdot\frac{\Gamma(k+1)}{\lambda^{k+1}} = \frac{k!}{\lambda^k}.$$

□

In particular,

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \mathbb{E}[X^2] = \frac{2}{\lambda^2}, \quad \mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{\lambda^2}.$$

The cumulative distribution function (CDF) of an exponential distribution with parameter $\lambda > 0$ is given by:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t)\,dt = 1 - e^{-\lambda x}, \quad x \geq 0.$$

The exponential distribution satisfies, for all $s, t \geq 0$,

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s).$$

This property is called the *memoryless property* of the exponential distribution.

Let X be a random variable. The cumulative distribution function of X is $F(x) = \mathbb{P}(X \leq x)$, and the *survival function* is defined by:

$$S(x) = \mathbb{P}(X > x) = 1 - F(x).$$

If X has a continuous density f, the **hazard rate function** (or failure rate) is defined by:

$$r(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\ln(S(t)).$$

Then:

$$\mathbb{P}(X \in (t, t+dt) \mid X > t) \approx \frac{\mathbb{P}(X \in (t, t+dt))}{\mathbb{P}(X > t)} = \frac{f(t)\,dt}{S(t)} = r(t)\,dt.$$

If X models a lifetime, then $r(t)$ is the conditional mortality rate of individuals surviving up to time t. Since $S(t)$ is decreasing, we have $r(t) \geq 0$.

Assuming X is a positive random variable, then:

$$1 - F(t) = e^{-\int_0^t r(s)\,ds}.$$

If X follows an exponential distribution with parameter λ, then $r(t) = 1/\lambda$ for all $t \geq 0$. Conversely, if the hazard rate is constant, then the random variable follows an exponential distribution.

Definition 4.2. *A random variable X follows a **Gamma distribution** with parameters $\alpha > 0$ and $\theta > 0$ if its density is given by:*

$$f(x) = \begin{cases} 0 & \text{if } x < 0, \\ \dfrac{x^{\alpha-1}e^{-x/\theta}}{\theta^\alpha \Gamma(\alpha)} & \text{if } x \geq 0. \end{cases}$$

We denote this by $X \sim Gamma(\alpha, \theta)$.

The above function f defines a bona fide probability density because,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^\infty \frac{x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} e^{-x/\theta}\,dx = 1.$$

A gamma distribution with parameter $\alpha = 1$ is an exponential distribution.

Theorem 4.3. *If X has a gamma distribution with parameters α and θ, then:*

$$\mathbb{E}[X] = \alpha\theta, \quad Var(X) = \alpha\theta^2, \quad \mathbb{E}[X^k] = \frac{\Gamma(\alpha+k)\theta^k}{\Gamma(\alpha)}.$$

Proof. We have

$$\mathbb{E}[X^k] = \int_0^\infty x^k \frac{x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} e^{-x/\theta}\, dx = \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty x^{k+\alpha-1} e^{-x/\theta}\, dx.$$

By the change of variable $y = \frac{x}{\theta}$, we get:

$$= \frac{1}{\theta^\alpha \Gamma(\alpha)} \cdot \theta^{k+\alpha} \Gamma(k+\alpha) = \frac{\Gamma(k+\alpha)\theta^k}{\Gamma(\alpha)}.$$

In particular,

$$\mathbb{E}[X] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \theta = \alpha\theta, \quad \mathbb{E}[X^2] = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \theta^2 = (\alpha+1)\alpha\theta^2,$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \alpha\theta^2.$$

□

Example 4.1. *The lifetime (in hours) of an electronic part is a random variable with probability density:*

$$f(x) = cx^9 e^{-x}, \quad x > 0.$$

Find c ***and the expected lifetime of the part.***

Solution 4.1. *To normalize the density:*

$$1 = \int_0^\infty cx^9 e^{-x}\, dx = c\Gamma(10) = c \cdot 9!.$$

So $c = \frac{1}{9!}$.

The expected value of X *is:*

$$\mathbb{E}[X] = \int_0^\infty \frac{1}{9!} x^{10} e^{-x}\, dx = \frac{1}{9!}\Gamma(11) = \frac{10!}{9!} = 10.$$

Exercise 4.1. *The time* T *required to repair a machine is an exponentially distributed random variable with mean* $\frac{1}{2}$.

1. *What is the probability that a repair time exceeds* $\frac{1}{2}$ *hours?*
2. *What is the probability that a repair time takes at least* 12.5 *hours given that its duration exceeds 12 hours?*

Solution Since $T \sim \text{Exponential}(\lambda)$ with mean $\frac{1}{2}$, we have $\lambda = 2$. Thus, the probability density function is:

$$f(t) = 2e^{-2t}, \quad t \geq 0.$$

1. We compute:

$$\mathbb{P}(T \geq \tfrac{1}{2}) = \int_{1/2}^\infty 2e^{-2t}\, dt = \left[-e^{-2t}\right]_{1/2}^\infty = e^{-1}.$$

2. We use the memoryless property of the exponential distribution:

$$\mathbb{P}(T \geq 12.5 \mid T \geq 12) = \frac{\mathbb{P}(T \geq 25/2)}{\mathbb{P}(T \geq 12)} = \frac{e^{-2\cdot 12.5}}{e^{-2\cdot 12}} = \frac{e^{-25}}{e^{-24}} = e^{-1}.$$

Exercise 4.2. *The service times of customers going through a checkout counter in a retail store are independent random variables with an exponential distribution with mean 1.5 minutes. Approximate the probability that 100 customers can be serviced in less than 3 hours of total service time.*

Solution Let $X_1, \ldots, X_{100}$ be the service times of the 100 customers. For each $1 \leq i \leq 100$, we have:

$$\mathbb{E}[X_i] = 1.5, \quad \mathrm{Var}(X_i) = (1.5)^2 = 2.25.$$

Then:

$$\mathbb{E}\left[\sum_{j=1}^{100} X_j\right] = 100 \times 1.5 = 150, \quad \mathrm{Var}\left(\sum_{j=1}^{100} X_j\right) = 100 \times 2.25 = 225.$$

We want to approximate:

$$\mathbb{P}\left(\sum_{j=1}^{100} X_j \leq 180\right),$$

since 3 hours = 180 minutes. By the Central Limit Theorem:

$$\sum_{j=1}^{100} X_j \approx \mathcal{N}(150, 225),$$

so we standardize:

$$\mathbb{P}\left(\sum_{j=1}^{100} X_j \leq 180\right) \approx \mathbb{P}\left(Z \leq \frac{180 - 150}{\sqrt{225}}\right) = \mathbb{P}(Z \leq 2),$$

where $Z \sim \mathcal{N}(0, 1)$. Using the standard normal table:

$$\mathbb{P}(Z \leq 2) = 0.9772.$$

Exercise 4.3. *Find the density of a gamma random variable with mean 8 and variance 16.*

Solution Recall that if $X \sim \mathrm{Gamma}(\alpha, \theta)$, then:

$$\mathbb{E}[X] = \alpha\theta, \quad \mathrm{Var}(X) = \alpha\theta^2.$$

We are given:

$$\mathbb{E}[X] = 8, \quad \mathrm{Var}(X) = 16.$$

Solving the system:

$$\begin{cases} \alpha\theta = 8, \\ \alpha\theta^2 = 16, \end{cases} \quad \Rightarrow \quad \theta = 2, \quad \alpha = 4.$$

Thus, the density is:

$$f(x) = \begin{cases} \dfrac{x^3 e^{-x/2}}{2^4 \Gamma(4)} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 4.4. *Let $X \sim$ Gamma(α, θ) and $Y \sim$ Gamma(β, θ). Suppose X and Y are independent. Then:*

$$X + Y \sim \mathit{Gamma}(\alpha + \beta, \theta).$$

Proof. The moment generating functions of X and Y are:

$$M_X(t) = (1 - \theta t)^{-\alpha}, \quad M_Y(t) = (1 - \theta t)^{-\beta}.$$

Since X and Y are independent, the moment generating function of $X + Y$ is:

$$M_{X+Y}(t) = M_X(t) M_Y(t) = (1 - \theta t)^{-\alpha}(1 - \theta t)^{-\beta} = (1 - \theta t)^{-(\alpha+\beta)}.$$

This is the moment generating function of a Gamma distribution with parameters $\alpha + \beta$ and θ. ☐

Theorem 4.5. *Let $X_1, \dots, X_n$ be independent and identically distributed random variables with exponential distribution with parameter λ. Then:*

$$\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda).$$

Proof. Each $X_i \sim \text{Gamma}(1, \lambda)$. By Theorem 5.4, the sum $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda)$. ☐

Exercise 4.4. *Suppose that you arrive at a single-teller bank and find 5 other customers: one being served and four waiting in line. You join the end of the line. If the service times are all exponential with rate 2 minutes, what are the expected value and variance of the amount of time you will spend in the bank?*

Solution The total time you spend in the bank is the sum of 5 independent exponential random variables with rate 2, since you wait for the 5 customers ahead of you.

Each $X_i \sim \text{Exponential}(\lambda = 2)$, so:

$$\mathbb{E}[X_i] = \frac{1}{2}, \quad \text{Var}(X_i) = \frac{1}{4}.$$

Thus:

$$\mathbb{E}[T] = 5 \cdot \frac{1}{2} = 2.5 \text{ minutes}, \quad \text{Var}(T) = 5 \cdot \frac{1}{4} = 1.25.$$

Reliability application:

Suppose a system has n parts, and the system functions only if all n parts work. Let X_i be the lifetime of the i-th part, with $X_1, \dots, X_n$ independent and $X_i \sim \text{Exponential}(\theta_i)$.

Let $Y = \min(X_1, \dots, X_n)$. Then:

$$\mathbb{P}(Y > t) = \mathbb{P}(\min(X_1, \dots, X_n) > t) = \prod_{i=1}^{n} \mathbb{P}(X_i > t) = \prod_{i=1}^{n} e^{-t/\theta_i} = e^{-t \sum_{i=1}^{n} \frac{1}{\theta_i}}.$$

So, $Y \sim \text{Exponential}\left(\lambda = \sum_{i=1}^{n} \frac{1}{\theta_i}\right)$.

Now consider $Z = \max(X_1, \dots, X_n)$. Then:

$$\mathbb{P}(Z \le t) = \mathbb{P}(\max(X_1, \dots, X_n) \le t) = \prod_{i=1}^{n} \mathbb{P}(X_i \le t) = \prod_{i=1}^{n} \left(1 - e^{-t/\theta_i}\right).$$

Since $\max(x, y) + \min(x, y) = x + y$,

$$\mathbb{E}[\max(X_1, X_2)] = \theta_1 + \theta_2 - \frac{1}{1/\theta_1 + 1/\theta_2}.$$

Given $x_1, \dots, x_n$, we have that

$$\max(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i - \sum_{1 \le i_1 < i_2 \le n} \min(x_{i_1}, x_{i_2})$$
$$+ \sum_{1 \le i_1 < i_2 < i_3 \le n} \min(x_{i_1}, x_{i_2}, x_{i_3}) - \cdots + (-1)^{n+1} \min(x_1, \dots, x_n).$$

Theorem 4.6. *1. (i) Let X and Y be two independent random variables such that $X \sim Exp(\theta_1)$ and $Y \sim Exp(\theta_2)$. Then:*

$$\mathbb{P}(X < Y) = \frac{1/\theta_1}{1/\theta_1 + 1/\theta_2}.$$

2. Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \sim Exp(\theta_i)$. Then:

$$\mathbb{P}\left(X_i = \min_{1 \le j \le n} X_j\right) = \frac{1/\theta_i}{\sum_{j=1}^{n} 1/\theta_j}.$$

Proof. The joint density of X and Y is:

$$f_{X,Y}(x,y) = \frac{e^{-\frac{x}{\theta_1}} e^{-\frac{y}{\theta_2}}}{\theta_1 \theta_2}, \quad x, y > 0.$$

So,

$$\mathbb{P}(X < Y) = \int_0^\infty \int_x^\infty \frac{e^{-\frac{x}{\theta_1}} e^{-\frac{y}{\theta_2}}}{\theta_1 \theta_2}\, dy\, dx.$$

Simplifying the integrals:

$$\mathbb{P}(X < Y) = \int_0^\infty \frac{e^{-\frac{x}{\theta_1}}}{\theta_1} \left(\int_x^\infty \frac{e^{-\frac{y}{\theta_2}}}{\theta_2}\, dy\right) dx = \frac{1/\theta_1}{1/\theta_1 + 1/\theta_2}.$$

For the second part, we have:

$$\mathbb{P}\left(X_i = \min_{1 \le j \le n} X_j\right) = \mathbb{P}\left(X_i < \min_{1 \le j \le n, j \ne i} X_j\right) = \frac{1/\theta_i}{\sum_{j=1}^{n} 1/\theta_j},$$

4.2 Poisson process

Poisson distributions are used in the modeling of random phenomena where the future is independent of the past (machine failures, accidents, telephone calls to a switchboard, queues, mortality, healing times of minor injuries, inventory levels, number of shooting stars in the summer sky, and so on).

Definition 4.3. *A random variable X follows a Poisson distribution $\mathcal{P}(\lambda)$ with parameter $\lambda > 0$ if its probability distribution is given by*

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}.$$

The expectation of X is $\mathbb{E}(X) = \lambda$ and its variance is $\text{Var}(X) = \lambda$. Consequently, its standard deviation is $\sigma(X) = \sqrt{\lambda}$. When n takes large values and p is small, the binomial distribution $B(n,p)$ can be approximated by the Poisson distribution $\mathcal{P}(np)$ (preserving the mean). The usual conditions for this approximation are

$$n \ge 30, \qquad p \le 0.1, \qquad np < 15.$$

Exercise 4.5. *A medical service receives on average 4 calls during an 8-hour period. Let X be the number of calls received by this service in an 8-hour period.*

1. *Which probability distribution can be applied here?*
2. *Use this distribution to compute:*

$$P(X > 3), \quad P(X < 4), \quad P(X > 0).$$

Solution:

1. *The random variable X follows a Poisson distribution with parameter $\lambda = 4$, since we are dealing with the number of events occurring during a fixed time interval.*

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad X \sim \mathcal{P}(\lambda).$$

2. *We calculate:*

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - \Big[P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\Big] \\ &= 1 - \left[e^{-4} + 4e^{-4} + 8e^{-4} + \frac{32}{6}e^{-4}\right] \\ &= 1 - \frac{71}{6}e^{-4} \\ &\approx 1 - 0.43347 = 0.56653 \quad (\textit{or } 56.653\%). \end{aligned}$$

Similarly,

$$\begin{aligned} P(X < 4) = P(X \leq 3) &= e^{-4} + 4e^{-4} + 8e^{-4} + \frac{32}{6}e^{-4} \\ &\approx 0.43347 \quad (\textit{or } 43.347\%). \end{aligned}$$

Finally,

$$P(X > 0) = 1 - P(X = 0) = 1 - e^{-4} \approx 1 - 0.0183 = 0.9817 \quad (\textit{or } 98.17\%).$$

Exercise 4.6. ***Poisson law approximated by a normal distribution***

Suppose that the number of asbestos particles in one square meter of dust on a surface follows a Poisson distribution with a mean of 1000. *If one square meter of dust is analyzed, what is the probability that* 950 *particles or fewer are found?*

Solution:

Let X be the random variable representing the (integer) number of asbestos particles in one square meter. Then $X \sim \mathcal{P}(1000)$.

The exact probability can be written using the Poisson distribution:

$$P(X \leq 950) = \sum_{k=0}^{950} e^{-1000}\frac{1000^k}{k!}.$$

This cumulative probability can only be computed numerically and is approximately

$$P(X \leq 950) \approx 0.05783629553\ldots$$

Because this exact computation is difficult, we approximate it using the normal distribution. For large λ, the Poisson distribution can be approximated by a normal distribution:

$$X \approx Y \sim \mathcal{N}(\mu = 1000,\ \sigma^2 = 1000).$$

Applying the continuity correction,

$$\begin{aligned} P(X \le 950) &\approx P\Big(Y \le 950.5\Big) \\ &= P\left(Z \le \frac{950.5 - 1000}{\sqrt{1000}}\right) \\ &= P(Z \le -1.57) \\ &= 1 - \Phi(1.57) \\ &\approx 1 - 0.9418 \\ &= 0.0582. \end{aligned}$$

Practical interpretation: *Poisson probabilities that are difficult to compute exactly can be well approximated using probabilities based on the normal distribution, which are easier to calculate.*

Definition 4.4. *A stochastic process $\{N(t) : t \ge 0\}$ is said to be a counting process if $N(t)$ represents the total number of "events" that have occurred up to time t.*

A counting process $N(t)$ must satisfy:

- *(i) $N(t) \ge 0$.*
- *(ii) $N(t)$ is integer-valued.*
- *(iii) If $s < t$, then $N(s) \le N(t)$.*
- *(iv) For $s < t$, $N(t) - N(s)$ is the number of events that have occurred in the interval $(s, t]$.*

A counting process is said to possess independent increments if for each $0 \le t_1 < t_2 < \cdots < t_m$, the random variables $N(t_1), N(t_2) - N(t_1), N(t_3) - N(t_2), \ldots, N(t_m) - N(t_{m-1})$ are independent.

A counting process is said to have stationary increments if for each $0 \le t_1 \le t_2$, $N(t_2) - N(t_1)$ and $N(t_2 - t_1)$ have the same distribution.

4.2.1 Construction via waiting times

The second construction of the Poisson point process is based on the distribution of the position differences

$$Z_n = X_n - X_{n-1}.$$

These increments also uniquely characterize the process, through the relation

$$X_n(\omega) = \sum_{j=1}^{n} Z_j(\omega).$$

The remarkable result is that the random variables Z_j are i.i.d. and follow a very specific law, namely an exponential distribution with parameter λ.

Theorem 4.7. *For all n, the random variables $Z_1, \ldots, Z_n$ are independent and identically distributed with common distribution* $\mathrm{Exp}(\lambda)$.

Proof. Fix the time instants

$$t_0 = 0 < s_1 < t_1 < s_2 < t_2 < \cdots < s_n < t_n.$$

We then compute

$$\begin{aligned}
&\mathbb{P}\Big(X_1 \in (s_1, t_1],\, X_2 \in (s_2, t_2],\, \ldots,\, X_n \in (s_n, t_n]\Big) \\
&= \mathbb{P}\Big(N(0, s_1] = 0,\, N(s_1, t_1] = 1,\, N(t_1, s_2] = 0, \ldots,\, N(t_{n-1}, s_n] = 0,\, N(s_n, t_n] > 1\Big) \\
&= \prod_{k=1}^{n} \mathbb{P}\Big(N(t_{k-1}, s_k] = 0\Big) \prod_{k=1}^{n-1} \mathbb{P}\Big(N(s_k, t_k] = 1\Big)\, \mathbb{P}\Big(N(s_n, t_n] > 1\Big).
\end{aligned}$$

Using the distribution of Poisson increments, this becomes

$$\begin{aligned}
&= \prod_{k=1}^{n} e^{-\lambda(s_k - t_{k-1})} \prod_{k=1}^{n-1} \Big(\lambda(t_k - s_k) e^{-\lambda(t_k - s_k)}\Big)\Big(1 - e^{-\lambda(t_n - s_n)}\Big) \\
&= \lambda^{n-1} \prod_{k=1}^{n-1} (t_k - s_k)\Big(e^{-\lambda s_n} - e^{-\lambda t_n}\Big).
\end{aligned}$$

This can be expressed as the multiple integral

$$\int_{s_1}^{t_1} \int_{s_2}^{t_2} \cdots \int_{s_n}^{t_n} \lambda^n e^{-\lambda x_n}\, dx_n\, dx_{n-1} \cdots dx_2\, dx_1.$$

The joint density of $(X_1, \ldots, X_n)$ is given by

$$f(x_1, \ldots, x_n) = \begin{cases} \lambda^n e^{-\lambda x_n}, & \text{if } 0 < x_1 < \cdots < x_n, \\ 0, & \text{otherwise.} \end{cases}$$

We can then compute the distribution function of the interarrival times Z_k:

$$\begin{aligned}
&\mathbb{P}(Z_1 \le z_1, \ldots, Z_n \le z_n) \\
&= \mathbb{P}(X_1 \le z_1,\, X_2 - X_1 \le z_2, \ldots, X_n - X_{n-1} \le z_n) \\
&= \mathbb{P}(X_1 \le z_1,\, X_2 \le z_2 + X_1, \ldots, X_n \le z_n + X_{n-1}) \\
&= \int_0^{z_1} \int_{x_1}^{z_2 + x_1} \cdots \int_{x_{n-1}}^{z_n + x_{n-1}} f(x_1, \ldots, x_n)\, dx_n \cdots dx_1.
\end{aligned}$$

The joint density of $(Z_1, \ldots, Z_n)$ is then obtained by differentiation:

$$\begin{aligned}
\frac{\partial^n}{\partial z_1 \cdots \partial z_n} \mathbb{P}(Z_1 \le z_1, \ldots, Z_n \le z_n) &= f(z_1,\, z_1 + z_2,\, \ldots,\, z_1 + \cdots + z_n) \\
&= \lambda^n e^{-\lambda(z_1 + \cdots + z_n)}.
\end{aligned}$$

This density is exactly the joint density of n independent exponential random variables with parameter λ.

This result provides a method to construct the Poisson process: each X_n is obtained from X_{n-1} by adding an exponential random variable, independent of the previous ones. It follows that X_n has a Gamma distribution with parameters (λ, n). Moreover, the Markov property can be written as

$$\mathbb{P}(Z_n > t + \varepsilon \mid Z_n > t) = e^{-\lambda \varepsilon}, \quad \forall \varepsilon > 0,$$

and hence

$$\lim_{\varepsilon\to 0}\frac{\mathbb{P}(Z_n \le t+\varepsilon \mid Z_n > t)}{\varepsilon} = \lim_{\varepsilon\to 0}\frac{1-e^{-\lambda\varepsilon}}{\varepsilon} = \lambda.$$

Thus, the parameter λ represents the rate of occurrence of a new point, which is independent of the past. This result can also be obtained directly using the counting function.

□

Exercise 4.7 (Queueing System)**.** *Consider a queue at a service counter. The number of customers in the queue follows a Poisson distribution with parameter λ. The probability that a customer can obtain the desired service is equal to p. What is the distribution of the random variable that counts the number of satisfied requests? We assume that, conditional on there being n customers, the number of satisfied requests follows a binomial distribution with parameter p.*

Solution. *The number of customers in the queue is assumed to follow a Poisson distribution with parameter λ. Let N denote the random variable representing the number of customers in the queue. Thus,*

$$P(N=n) = \frac{\lambda^n}{n!}e^{-\lambda}, \quad n \in \mathbb{N}.$$

Let Y be the random variable representing the number of satisfied requests. We compute the distribution of Y, assuming that the conditional law of Y given $N=n$ is binomial. Hence, for $n \in \mathbb{N}$ and $k \le n$,

$$P(Y=k \mid N=n) = \binom{n}{k}p^k(1-p)^{n-k}.$$

The event "k requests are satisfied" corresponds to

$$\{Y=k\} = \bigcup_{n\ge k}\{Y=k,\, N=n\}.$$

Since the events $\{Y=k, N=n\}$ are pairwise disjoint, we can write

$$P(Y=k) = \sum_{n=k}^{\infty} P(Y=k, N=n) = \sum_{n=k}^{\infty} P(Y=k \mid N=n)\, P(N=n).$$

Substituting the distributions, we obtain

$$P(Y=k) = \sum_{n=k}^{\infty}\binom{n}{k}p^k(1-p)^{n-k}\,\frac{\lambda^n}{n!}e^{-\lambda}.$$

Rewriting the binomial coefficient as $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and making the change of index $m = n-k$, we get

$$P(Y=k) = \frac{(\lambda p)^k}{k!}e^{-\lambda}\sum_{m=0}^{\infty}\frac{\big(\lambda(1-p)\big)^m}{m!}.$$

Since

$$\sum_{m=0}^{\infty}\frac{\big(\lambda(1-p)\big)^m}{m!} = e^{\lambda(1-p)},$$

it follows that

$$P(Y=k) = \frac{(\lambda p)^k}{k!} e^{-\lambda}\, e^{\lambda(1-p)} = \frac{(\lambda p)^k}{k!} e^{-\lambda p}.$$

Thus, the random variable Y follows a Poisson distribution with parameter λp:

$$Y \sim \mathcal{P}(\lambda p).$$

Proof: It is very easy to get it.

Theorem 4.8. *Let $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$. Suppose that X and Y are independent. Then,*

$$X + Y \sim Poisson(\lambda_1 + \lambda_2).$$

Proof: The moment generating function of X is $M_X(t) = e^{\lambda_1(e^t-1)}$. The moment generating function of Y is $M_Y(t) = e^{\lambda_2(e^t-1)}$. Since X and Y are independent random variables, the moment generating function of $X+Y$ is

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}.$$

This is the moment generating function of a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Definition 4.5. *The stochastic process $\{N(t) : t \geq 0\}$ is said to be a Poisson process with rate function $\lambda > 0$, if:*

- *(i) $N(0) = 0$.*
- *(ii) The process has independent increments.*
- *(iii) $P(N(h) \geq 2) = o(h)$.*
- *(iv) $P(N(h) = 1) = \lambda h + o(h)$.*

Definition 4.6. *) The stochastic process $\{N(t) : t \geq 0\}$ is said to be a Poisson process with rate function $\lambda > 0$, if:*

- *(i) $N(0) = 0$.*
- *(ii) The process has independent increments.*
- *(iii) For each $0 \leq s, t$, $N(s+t) - N(s)$ has a Poisson distribution with mean λt.*

The two definitions of Poisson processes are equivalent.

In the previous definition, $N(t)$ is the number of occurrences until time t. The rate of occurrences per unit of time is a constant. The average number of occurrences in the time interval $(s, s+t]$ is λt.

It follows that for each $0 \leq t_1 < t_2 < \cdots < t_m$, and each $0 \leq k_1 \leq k_2 \leq \cdots \leq k_m$,

$$P(N(t_1) = k_1, N(t_2) = k_2, \ldots, N(t_m) = k_m)$$

$$= P(N(t_1) = k_1)P(N(t_2) - N(t_1) = k_2 - k_1) \cdots P(N(t_m) - N(t_{m-1}) = k_m - k_{m-1}).$$

$$= e^{-\lambda t_1}\frac{(\lambda t_1)^{k_1}}{k_1!} e^{-\lambda(t_2-t_1)}\frac{(\lambda(t_2-t_1))^{k_2-k_1}}{(k_2-k_1)!} \cdots e^{-\lambda(t_m-t_{m-1})}\frac{(\lambda(t_m-t_{m-1}))^{k_m-k_{m-1}}}{(k_m-k_{m-1})!}.$$

It is easy to see that for each $0 \leq s \leq t$,

$$E[N(s)] = \lambda s, \quad \mathrm{Var}(N(s)) = \lambda s, \quad \mathrm{Cov}(N(s), N(t)) = \lambda s.$$

Exercise 4.8. *Let* $\{N(t) : t \geq 0\}$ *be a Poisson process with rate* $\lambda = 2$*. Compute:*

- *(i)* $P(N(5) = 4)$.
- *(ii)* $P(N(5) = 4, N(6) = 9)$.
- *(iii)* $E[2N(3) - 4N(5)]$.
- *(iv)* $Var(2N(3) - 4N(5))$.

Solution: Let $\{N(t) : t \geq 0\}$ be a Poisson process with rate $\lambda = 2$. Compute:

1. We know that $N(5) \sim \text{Pois}(\lambda \cdot 5) = \text{Pois}(10)$, hence
$$P\Big(N(5) = 4\Big) = e^{-10}\frac{10^4}{4!}.$$
2. By independent increments,
$$P\Big(N(5) = 4,\, N(6) = 9\Big) = P\Big(N(5) = 4\Big)\, P\Big(N(6) - N(5) = 5\Big).$$
Since $N(6) - N(5) \sim \text{Pois}(\lambda \cdot 1) = \text{Pois}(2)$,
$$P\Big(N(5) = 4,\, N(6) = 9\Big) = e^{-10}\frac{10^4}{4!}\, e^{-2}\frac{2^5}{5!} = e^{-12}\frac{10^4 2^5}{4!\,5!}.$$
3. Using $E[N(t)] = \lambda t$,
$$\mathbb{E}[2N(3) - 4N(5)] = 2E[N(3)] - 4E[N(5)] = 2(2 \cdot 3) - 4(2 \cdot 5) = 12 - 40 = -28.$$
4. Using $\text{Var}(aX + bY) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) + 2ab\,\text{Cov}(X, Y)$ and $\text{Cov}(N(s), N(t)) = \lambda s$ for $s \leq t$,
$$\begin{aligned}\text{Var}[2N(3) - 4N(5)] &= 4\ \text{Var}(N(3)) + 16\ \text{Var}(N(5)) + 2(2)(-4)\,\text{Cov}(N(3), N(5)) \\ &= 4(6) + 16(10) - 16(6) = 24 + 160 - 96 = 88.\end{aligned}$$

Exercise 4.9. *Let* $\{N(t) : t \geq 0\}$ *be a Poisson process with rate* $\lambda = 2$*. Compute:*

- *(i)* $P(N(5) = 4, N(6) = 9, N(10) = 15)$.
- *(ii)* $P(N(5) - N(2) = 3)$.
- *(iii)* $P(N(5) - N(2) = 3, N(7) - N(6) = 4)$.
- *(iv)* $P(N(2) + N(5) = 4)$.
- *(v)* $E[N(5) - 2N(6) + 3N(10)]$.
- *(vi)* $Var(N(5) - 2N(6) + 3N(10))$.
- *(vii)* $Cov(N(5) - 2N(6), 3N(10))$.

Solution. Recall that for a Poisson process of rate λ,
$$N(t) - N(s) \sim \text{Poisson}(\lambda(t - s))$$
and disjoint increments are independent.

(i)

$$\begin{aligned} P &= P(N(5) = 4)\, P(N(6) - N(5) = 5)\, P(N(10) - N(6) = 6) \\ &= e^{-10}\frac{10^4}{4!}\; e^{-2}\frac{2^5}{5!}\; e^{-8}\frac{8^6}{6!} \\ &= e^{-20}\frac{10^4}{4!}\frac{2^5}{5!}\frac{8^6}{6!}. \end{aligned}$$

(ii)

$$P = P\Big(N(5) - N(2) = 3\Big) = e^{-6}\frac{6^3}{3!}.$$

(iii)

$$\begin{aligned} P &= P\Big(N(5) - N(2) = 3\Big)\, P\Big(N(7) - N(6) = 4\Big) \\ &= e^{-6}\frac{6^3}{3!}\; e^{-2}\frac{2^4}{4!} = e^{-8}\frac{6^3}{3!}\frac{2^4}{4!}. \end{aligned}$$

(iv) Let $k = N(2)$. We need $k + N(5) = 4$ with $N(5) = k + \Big(N(5) - N(2)\Big)$. Hence $k + \Big(k + \Delta\Big) = 4$ with $\Delta \geq 0$, giving $\Delta = 4 - 2k$, $k = 0, 1, 2$. Thus

$$\begin{aligned} P &= \sum_{k=0}^{2} P(N(2) = k)\, P\Big(N(5) - N(2) = 4 - 2k\Big) \\ &= e^{-10}\sum_{k=0}^{2}\frac{4^k}{k!}\frac{6^{4-2k}}{(4-2k)!}. \end{aligned}$$

(v)

$$\begin{aligned} E[N(5) - 2N(6) + 3N(10)] &= 2\cdot 5 - 2(2\cdot 6) + 3(2\cdot 10) \\ &= 10 - 24 + 60 = 46. \end{aligned}$$

(vi) Using $\mathrm{Var}(N(t)) = 2t$, $\mathrm{Cov}(N(s), N(t)) = 2\min(s,t)$, and writing $X = N(5) - 2N(6) + 3N(10)$:

$$\begin{aligned} \mathrm{Var}(X) = {} & 1^2\,\mathrm{Var}(N(5)) + (-2)^2\,\mathrm{Var}(N(6)) + 3^2\,\mathrm{Var}(N(10)) \\ & + 2(1)(-2)\,\mathrm{Cov}(N(5), N(6)) + 2(1)(3)\,\mathrm{Cov}(N(5), N(10)) + 2(-2)(3)\,\mathrm{Cov}(N(6), N(10)) \\ = {} & 10 + 48 + 180 - 40 + 60 - 144 = 114. \end{aligned}$$

(vii)

$$\begin{aligned} \mathrm{Cov}(N(5) - 2N(6), 3N(10)) &= 3\Big[\mathrm{Cov}(N(5), N(10)) - 2\,\mathrm{Cov}(N(6), N(10))\Big] \\ &= 3\Big[10 - 2\cdot 12\Big] = -42. \end{aligned}$$

Theorem 4.9. *Let $\{N(t) : t \geq 0\}$ be a Poisson process with rate λ. Let $0 < s, t$ and $k \geq j \geq 0$. Then,*

$$P(N(s+t) = k | N(s) = j) = P(N(t) = k - j),$$

i.e., the distribution of $N(s+t)$ given $N(s) = j$ is $j +$ Poisson(λt). So,

$$E[N(s+t)|N(s) = j] = j + \lambda t \quad \text{and} \quad Var(N(s+t)|N(s) = j) = \lambda t.$$

Proof. Since $N(s)$ and $N(s+t)-N(s)$ are independent,

$$P(N(s+t)=k|N(s)=j)=P(N(s+t)-N(t)=k-j|N(s)=j)=P(N(s+t)-N(t)=k-j).$$

Theorem 4.10. *(Markov property of the Poisson process) Let $\{N(t):t\geq 0\}$ be a Poisson process with rate λ. Let $0\leq t_1<t_2<\cdots<t_m<s$ and let $k_1\leq k_2\leq\cdots\leq k_m\leq j$. Then,*

$$P(N(s)=j\mid N(t_1)=k_1,\dots,N(t_m)=k_m)=P(N(s)=j\mid N(t_m)=k_m)$$

Proof. Since $N(t_1),N(t_2)-N(t_1),\dots,N(t_m)-N(t_{m-1}),N(s)-N(t_m)$ are independent,

$$P(N(s)=j\mid N(t_1)=k_1,\dots,N(t_m)=k_m)=P(N(s)-N(t_m)=j-k_m)$$

$$=P(N(t_1)=k_1,N(t_2)-N(t_1)=k_2-k_1,\dots,N(t_m)-N(t_{m-1})=k_m-k_{m-1},N(s)-N(t_m)=j-k_m)$$

$$=P(N(t_1)=k_1)P(N(t_2)-N(t_1)=k_2-k_1)$$

$$\cdots P(N(t_m)-N(t_{m-1})=k_m-k_{m-1})P(N(s)-N(t_m)=j-k_m)$$

$$=P(N(t_1)=k_1)P(N(t_2)-N(t_1)=k_2-k_1)\cdots P(N(t_m)-N(t_{m-1})=k_m-k_{m-1})$$

$$=P(N(s)-N(t_m)=j-k_m)=P(N(s)=j\mid N(t_m)=k_m).$$

Theorem 4.11. *Let $\{N(t):t\geq 0\}$ be a Poisson process with rate λ. Let $s,t\geq 0$. Then,*

$$P(N(t)=k\mid N(s+t)=n)=\binom{n}{k}\left(\frac{t}{s+t}\right)^k\left(\frac{s}{s+t}\right)^{n-k},$$

i.e., the distribution of $N(t)$ given $N(s+t)=n$ is binomial with parameters n and $p=\frac{t}{s+t}$. So,

$$E[N(t)\mid N(s+t)=n]=n\frac{t}{s+t}$$

and

$$Var(N(t)\mid N(s+t)=n)=n\frac{t}{s+t}\left(\frac{s}{s+t}\right).$$

The previous theorem can be extended as follows: Given $0\leq t_1<t_2<\cdots<t_m$, the conditional distribution of $(N(t_1),N(t_2)-N(t_1),\dots,N(t_m)-N(t_{m-1}))$ given $N(t_m)=n$ is a multinomial distribution with parameter

$$\left(\frac{t_1}{t_m},\frac{t_2-t_1}{t_m},\dots,\frac{t_m-t_{m-1}}{t_m}\right).$$

Given $N(t_m)=n$, we know that events happen in the interval $[0,t_m]$, and each of these events happens independently. The probability that one of these events happens in a particular subinterval is the fraction of the total length of this interval.

4.2.2 Interarrival times

For $n \geq 1$, let S_n be the arrival time of the n-th event, i.e.

$$S_n = \inf\{t \geq 0 : N(t) = n\}.$$

Let $T_n = S_n - S_{n-1}$ be the elapsed time between the $(n-1)$-th and the n-th event.

Theorem 4.12. *$T_n, n = 1, 2, \dots$ are independent identically distributed exponential random variables having mean $\frac{1}{\lambda}$.*

Proof. We have that

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}.$$

So, T_1 has an exponential distribution with mean $\frac{1}{\lambda}$. By the Markov property,

$$P(T_2 > t \mid T_1 = s) = P(\text{no occurrences in } (s, s+t] \mid T_1 = s) = P(\text{no occurrences in } (s, s+t]) = e^{-\lambda t}.$$

So, T_1 and T_2 are independent, and T_2 has an exponential distribution with mean $\frac{1}{\lambda}$. By induction, the claim follows.

The previous theorem says that if the rate of events is λ events per unit of time, then the expected waiting time between events is $\frac{1}{\lambda}$. A useful relation between $N(t)$ and S_n is

$$\{S_n \leq t\} = \{\text{the } n\text{-th occurrence happens before time } t\}$$

$$= \{\text{there are at least } n \text{ occurrences in the interval } [0, t]\} = \{N(t) \geq n\}.$$

Theorem 4.13. *S_n has a gamma distribution with parameters $\alpha = n$ and $\theta = \frac{1}{\lambda}$.*

Proof. $T_1, \dots, T_n$ are i.i.d. random variables with an exponential distribution with mean $\frac{1}{\lambda}$. So, S_n has a gamma distribution with parameters $\alpha = n$ and $\theta = \frac{1}{\lambda}$. Q.E.D.

By the previous theorem,

$$E[S_n] = \frac{n}{\lambda}, \quad \text{Var}(S_n) = \frac{n}{\lambda^2},$$

and S_n has the density function

$$f_{S_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}, \quad t \geq 0.$$

Theorem 4.14. *Given that $N(t) = n$, the n arrival times $S_1, \dots, S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $(0, t)$.*

Exercise 4.10. *Let $N(t)$ be a Poisson process with rate $\lambda = 3$.*

(i) What is the probability of 4 occurrences of the random event in the interval $(2.5, 4]$?

$$P(N(4) - N(2.5) = 4) = e^{-4.5} \frac{(4.5)^4}{4!}.$$

(ii) What is the probability that $T_2 > 5$ given that $N(4) = 1$?

$$P(T_2 > 5 \mid N(4) = 1) = P(T_2 > 5 \mid T_1 \leq 4) = P(T_2 > 5) = e^{-15}.$$

(iii) What is the distribution of T_n?

$$T_n \textit{ has an exponential distribution with mean } \frac{1}{3}.$$

(iv) What is the expected time of the occurrence of the 5-th event?

$$E[S_5] = 5 \cdot \frac{1}{3}.$$

Exercise 4.11. *In a telephone call centre, calls arrive according to a Poisson process with rate $\lambda = 10$ calls per hour.*

1. *If an operator takes a break of 10 minutes starting at 10:30, how many calls will she miss on average during her break?*
2. *What is the probability that she misses at most 2 calls?*
3. *Knowing that 4 calls arrive between 10:00 and 11:00, what is the probability that she missed no calls during her break? That she missed exactly one call?*
4. *Knowing that there will be 2 calls between 10:30 and 11:00, what is the probability that they both arrive between 10:30 and 10:45?*

Solution.
Calls follow a Poisson process with rate $\lambda = 10$ calls/hour.

1. The break length is 10 minutes $= \frac{10}{60} = \frac{1}{6}$ hour. For a Poisson process the number of arrivals in an interval of length t is Poisson(λt). Thus the mean number missed is
$$\mathbb{E}[N] = \lambda t = 10 \times \frac{1}{6} = \frac{10}{6} = \frac{5}{3} \approx 1.6667.$$
2. Let $N \sim$ Poisson(μ) be the number missed during the break with $\mu = \frac{5}{3}$. The probability of at most 2 calls is
$$\mathbb{P}(N \le 2) = e^{-\mu} \sum_{k=0}^{2} \frac{\mu^k}{k!} = e^{-5/3} \left(1 + \frac{5/3}{1!} + \frac{(5/3)^2}{2!}\right).$$
Numerically,
$$\mathbb{P}(N \le 2) \approx e^{-1.6667}\Big(1 + 1.6667 + 1.3889\Big) \approx 0.809.$$
3. Conditional on 4 arrivals in the hour $[10:00, 11:00]$, the unordered arrival times are i.i.d. uniform on the hour. Equivalently the number falling in any subinterval of length t (here $t = \frac{1}{6}$ hour) is Binomial$(4, p)$ with $p = t/1 = 1/6$.
   - Probability she missed no calls (i.e. 0 of the 4 fall in the break):
   $$\mathbb{P}(0) = \binom{4}{0} p^0 (1-p)^4 = \left(\frac{5}{6}\right)^4 \approx 0.4823.$$

- Probability she missed exactly one call:

$$\mathbb{P}(1) = \binom{4}{1} p(1-p)^3 = 4 \cdot \tfrac{1}{6} \cdot \left(\tfrac{5}{6}\right)^3 \approx 4 \times 0.1667 \times 0.5787 \approx 0.3858.$$

4. Given there are 2 calls in [10:30, 11:00] (a 30-minute interval), their conditional joint distribution is that the two arrival times are i.i.d. uniform on that half-hour. The probability that both occur in the first 15 minutes [10:30, 10:45] (length 15 min $= \frac{1}{2}$ of the 30-min interval) is

$$\left(\frac{15}{30}\right)^2 = \left(\tfrac{1}{2}\right)^2 = \tfrac{1}{4} = 0.25.$$

Equivalently, view the count in the subinterval [10:30, 10:45] as Binomial$(2, p)$ with $p = \frac{1}{2}$; the probability both fall there is $p^2 = (1/2)^2$.

Exercise 4.12. *Let $N(t)$ be a Poisson process with rate λ. Let S_n denote the time of the n-th event. Find:*

(i) $E[S_4]$*?*

$$E[S_4] = \frac{4}{\lambda}.$$

(ii) $E[S_4 \mid N(1) = 2]$*?*

$$E[S_4 \mid N(1) = 2] = 1 + E[S_2] = 1 + \frac{2}{\lambda}.$$

(iii) $E[N(4) - N(2) \mid N(1) = 3]$*?*

$$E[N(4) - N(2) \mid N(1) = 3] = E[N(4) - N(2)] = 2\lambda.$$

Exercise 4.13. *For a Poisson process, the expected waiting time between events is 0.10 years.*

(i) What is the probability that 10 or fewer events occur during a 2-year time span?

$$P(N(2) \leq 10) = \sum_{j=0}^{10} e^{-20} \frac{20^j}{j!}.$$

(ii) What is the probability that the waiting time between 2 consecutive events is at least 0.2 years?

$$P(T_2 \geq 0.2) = e^{-2}.$$

(iii) If $N(2) = 20$, what is the probability that exactly 10 events occur during $(0, 1]$?

$$P(N(1) = 10 \mid N(2) = 20) = \binom{20}{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^{10}.$$

4.2.3 Superposition and decomposition of a Poisson process

Theorem 4.15. *Let $X \sim$ Poisson(λ_1) and $Y \sim$ Poisson(λ_2). Suppose that X and Y are independent. Then, $X+Y \sim$ Poisson$(\lambda_1+\lambda_2)$.*

Theorem 4.16. *If $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are two independent Poisson processes with respective rates λ_1 and λ_2, then $\{N_1(t)+N_2(t) : t \geq 0\}$ is a Poisson process with rate $\lambda_1+\lambda_2$.*

Proof. Let $N(t) = N_1(t) + N_2(t)$. Given $0 \leq t_1 < t_2 < \cdots < t_m$,

$$N_1(t_1), N_1(t_2)-N_1(t_1), \ldots, N_1(t_m)-N_1(t_{m-1}), N_2(t_1), N_2(t_2)-N_2(t_1), \ldots, N_2(t_m)-N_2(t_{m-1})$$

are independent random variables. So,

$$N(t_1), N(t_2)-N(t_1), \ldots, N(t_m)-N(t_{m-1})$$

are independent random variables. Also,

$$N_1(t_j)-N_1(t_{j-1}) \sim \text{Poisson}(\lambda_1(t_j-t_{j-1})) \quad \text{and} \quad N_2(t_j)-N_2(t_{j-1}) \sim \text{Poisson}(\lambda_2(t_j-t_{j-1})).$$

Hence,

$$N(t_j)-N(t_{j-1}) \sim \text{Poisson}((\lambda_1+\lambda_2)(t_j-t_{j-1})).$$

Thus, $\{N(t) : t \geq 0\}$ is a Poisson process with rate $\lambda_1+\lambda_2$.

Theorem 4.17. *Let $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ be two independent Poisson processes with respective rates λ_1 and λ_2. Let $\lambda = \lambda_1+\lambda_2$. Then, the conditional distribution of $N_1(t)$ given $N(t) = n$ is binomial with parameters n and $p = \frac{\lambda_1}{\lambda}$.*

Proof. We have that

$$P(N_1(t)=k \mid N(t)=n) = \frac{P(N_1(t)=k, N(t)=n)}{P(N(t)=n)} = \frac{P(N_1(t)=k, N_2(t)=n-k)}{P(N(t)=n)}$$

$$= \frac{e^{-\lambda_1 t}\frac{(\lambda_1 t)^k}{k!} \cdot e^{-\lambda_2 t}\frac{(\lambda_2 t)^{n-k}}{(n-k)!}}{e^{-\lambda t}\frac{(\lambda t)^n}{n!}} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Theorem 4.18. *Let λ be an s-finite measure on X. Then there exists a Poisson process on X with intensity measure λ.*

Proof. [19] □

4.2.4 Decomposition of a Poisson process

Let given a Poisson process $\{N(t) : t \geq 0\}$ with rate λ. Suppose that each time an event occurs it is classified as either a type I or a type II event. Suppose further that each event is classified as type I with probability p or as type II with probability $1-p$. Let $N_1(t)$ denote the number of type I events occurring in $[0,t]$. Let $N_2(t)$ denote the number of type II events occurring in $[0,t]$. Note that $N(t) = N_1(t) + N_2(t)$.

Theorem 4.19. *$\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are independent Poisson processes with respective rates $\lambda_1 = \lambda p$ and $\lambda_2 = \lambda(1-p)$.*

Exercise 4.14. *Arrivals of customers into a store follow a Poisson process with rate* $\lambda = 20$ *arrivals per hour. Suppose that the probability that a customer buys something is* $p = 0.30$.

(a) Find the expected number of sales made during an eight-hour business day.

(b) Find the probability that 10 or more sales are made in one hour.

(c) Find the expected time of the first sale of the day, if the store opens at 8 a.m.

Solution: *Let* $N_1(t)$ *be the number of arrivals who buy something, and* $N_2(t)$ *be the number of arrivals who do not buy anything. Then* N_1 *and* N_2 *are two independent Poisson processes.*

The rate for N_1 *is:*

$$\lambda_1 = \lambda p = 20 \times 0.3 = 6.$$

The rate for N_2 *is:*

$$\lambda_2 = \lambda(1-p) = 20 \times 0.7 = 14.$$

(a) The expected number of sales during an 8-hour day is:

$$\mathbb{E}[X_1(8)] = 8 \times \lambda_1 = 8 \times 6 = 48.$$

(b) The probability that 10 or more sales are made in one hour is:

$$\mathbb{P}(N_1(1) \geq 10) = 1 - \sum_{j=0}^{9} \mathbb{P}(N_1(1) = j) = 1 - \sum_{j=0}^{9} \frac{e^{-6} \cdot 6^j}{j!}.$$

(c) The expected time until the first sale is:

$$\mathbb{E}[T_{1,1}] = \frac{1}{\lambda_1} = \frac{1}{6} \; hours = 10 \; minutes.$$

Since the store opens at 8:00 a.m., the expected time of the first sale is 8:10 a.m.

Exercise 4.15. *Lucky Tom finds coins on his way to work at a Poisson rate of 0.5 coins per minute. The denominations are randomly distributed:*

(i) 60% of the coins are worth 1;

(ii) 20% of the coins are worth 5; and

(iii) 20% of the coins are worth 10.

Calculate the conditional expected value of the coins Tom found during his one-hour walk today, given that among the coins he found exactly ten were worth 5 each.

Choices:

(A) 108 (B) 115 (C) 128 (D) 165 (E) 180

Solution:

Let $N_1(t)$ be the number of coins of value 1 found until time t, $N_2(t)$ the number of coins of value 5, and $N_3(t)$ the number of coins of value 10. Then N_1, N_2, N_3 are independent Poisson processes.

Their respective rates are:

$$\lambda_1 = 0.5 \times 0.6 = 0.3, \quad \lambda_2 = 0.5 \times 0.2 = 0.1, \quad \lambda_3 = 0.5 \times 0.2 = 0.1.$$

We want to compute:

$$\mathbb{E}\left[N_1(60) + 5N_2(60) + 10N_3(60) \mid N_2(60) = 10\right].$$

Since N_1 and N_3 are independent of N_2, we have:

$$\mathbb{E}\left[N_1(60)\right] = 60 \times 0.3 = 18, \quad \mathbb{E}\left[N_3(60)\right] = 60 \times 0.1 = 6.$$

Hence, the expected total value is:

$$\mathbb{E}\left[\textit{Total Value} \mid N_2(60) = 10\right] = 18 + 5 \times 10 + 10 \times 6 = 128.$$

Answer: (C) 128

Exercise 4.16. *Customers arrive at a service center in accordance with a Poisson process with mean 10 per hour. There are two servers, and the service time for each server is exponentially distributed with a mean of 10 minutes. If both servers are busy, customers wait and are served by the first available server. Customers always wait for service, regardless of how many other customers are in line.*

Calculate the percent of the time, on average, when there are no customers being served.

Choices:

(A) 0.8% (B) 3.9% (C) 7.2% (D) 9.1% (E) 12.7%

Solution:

This is a classic M/M/2 queue. The arrival rate is $\lambda = 10$ per hour. The service rate for each server is:

$$\mu = \frac{1}{10/60} = 6 \textit{ per hour.}$$

Let $\rho = \frac{\lambda}{2\mu} = \frac{10}{12} = \frac{5}{6}$. The probability that no customer is in the system (i.e., both servers are idle) is:

$$P_0 = \frac{1}{\sum_{n=0}^{1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^2}{2!} \cdot \frac{1}{1-\rho}}.$$

Compute:

$$\frac{\lambda}{\mu} = \frac{10}{6} = \frac{5}{3}.$$

Then:

$$P_0 = \left(1 + \frac{5}{3} + \frac{(5/3)^2}{2(1-5/6)}\right)^{-1} = \left(1 + \frac{5}{3} + \frac{25/9}{2 \cdot (1/6)}\right)^{-1} = \left(1 + \frac{5}{3} + \frac{25/9}{1/3}\right)^{-1}$$

$$= \left(1 + \frac{5}{3} + \frac{25}{3}\right)^{-1} = \left(\frac{3+5+25}{3}\right)^{-1} = \left(\frac{33}{3}\right)^{-1} = \frac{1}{11} \approx 0.0909.$$

So, the percent of the time when no customers are being served is about 9.1%.

Answer: (D) 9.1%

Exercise 4.17. *Traffic flow of cars along a road is modelled by a Poisson process with intensity $\lambda = 2$ cars per minute. Because of roadworks, traffic is stopped alternately in each direction. Assume that when stopped each vehicle occupies on average 8 metres.*

1. *What is the distribution of the arrival time X_n of the n-th car?*
2. *Using the central limit theorem, give a Gaussian approximation for the law of X_n.*
3. *For how long can traffic be stopped if we want the queue thus formed to exceed 250 m with probability only 0.2? (You may use the normal quantile $z_{0.2} \approx -0.85$ or $z_{0.8} \approx 0.8416$.)*

Solution.

1. ***Law of X_n.*** *In a Poisson process with rate λ, the arrival time X_n of the n-th event has an Erlang (Gamma) distribution with shape n and rate λ. Its density is*

$$f_{X_n}(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}, \qquad x > 0,$$

with mean and variance

$$\mathbb{E}[X_n] = \frac{n}{\lambda}, \qquad \mathrm{Var}(X_n) = \frac{n}{\lambda^2}.$$

2. ***CLT (Gaussian) approximation.*** *By the Central Limit Theorem (or using the fact that the Gamma(n, λ) distribution is approximately normal for large n),*

$$X_n \approx \mathcal{N}\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right).$$

Equivalently, the standardized variable

$$\frac{X_n - n/\lambda}{\sqrt{n}/\lambda} \overset{approx}{\sim} \mathcal{N}(0,1).$$

3. ***Stopping time to keep queue length ≤ 250 m with probability 0.8.*** *Let $N(t)$ be the number of arrivals during a stop of length t minutes. Then $N(t) \sim \mathrm{Poisson}(\mu)$ with $\mu = \lambda t = 2t$.*

The queue length (in metres) when traffic is stopped for t minutes is approximately

$$L(t) \approx 8 \cdot N(t) \quad \textit{metres.}$$

We want the probability that the queue exceeds 250 m *to be 0.2, i.e.*

$$\mathbb{P}\Big(L(t) > 250\Big) = 0.2 \quad \Longrightarrow \quad \mathbb{P}\Big(L(t) \leq 250\Big) = 0.8.$$

The condition $L(t) \leq 250$ is equivalent to $N(t) \leq \lfloor 250/8 \rfloor = 31$ (since $250/8 = 31.25$). So we seek t such that

$$\mathbb{P}\Big(N(t) \leq 31\Big) \approx 0.8.$$

Use the normal approximation to the Poisson: $N(t) \approx \mathcal{N}(\mu, \mu)$ with $\mu = 2t$. Applying a continuity correction, write

$$\mathbb{P}(N(t) \le 31) \approx \mathbb{P}\left(\frac{31.5 - \mu}{\sqrt{\mu}} \le z\right),$$

and set the right-hand side equal to 0.8, so the critical z-value is $z_{0.8} \approx 0.8416$. Thus

$$\frac{31.5 - \mu}{\sqrt{\mu}} = 0.8416.$$

Let $y = \sqrt{\mu}$. Then $y^2 + 0.8416\, y - 31.5 = 0$, so

$$y = \frac{-0.8416 + \sqrt{0.8416^2 + 4 \cdot 31.5}}{2} \approx 5.20743,$$

hence $\mu = y^2 \approx 27.1173$ and

$$t = \frac{\mu}{\lambda} = \frac{27.1173}{2} \approx 13.56 \text{ minutes.}$$

Therefore, stopping traffic for about $\boxed{13.6 \text{ minutes}}$ yields a probability ≈ 0.2 that the queue exceeds 250 m.

Remark. *If one omits the continuity correction and solves $(31 - \mu)/\sqrt{\mu} = 0.8416$, one finds $\mu \approx 26.655$ and $t \approx 13.33$ minutes. The continuity-corrected value 13.56 minutes is slightly more conservative.*

Exercise 4.18. *Taxicabs leave a hotel with a group of passengers at a Poisson rate $\lambda = 10$ per hour. The number of people in each group taking a cab is independent and follows the distribution:*

Number of People	***Probability***
1	*0.60*
2	*0.30*
3	*0.10*

Using the normal approximation, calculate the probability that at least 1050 people leave the hotel in a cab during a 72-hour period.

Choices:

(A) 0.60 (B) 0.65 (C) 0.70 (D) 0.75 (E) 0.80

Solution:

Let $N_1(t)$, $N_2(t)$, and $N_3(t)$ be the number of taxicabs leaving with 1, 2, and 3 persons respectively, by time t. These are independent Poisson processes with rates:

$$\lambda_1 = 10 \times 0.60 = 6, \quad \lambda_2 = 10 \times 0.30 = 3, \quad \lambda_3 = 10 \times 0.10 = 1.$$

Let Y be the total number of people who leave the hotel in 72 hours:

$$Y = N_1(72) + 2N_2(72) + 3N_3(72).$$

Mean:

$$\mathbb{E}[Y] = 72 \cdot 6 + 2 \cdot 72 \cdot 3 + 3 \cdot 72 \cdot 1 = 432 + 432 + 216 = 1080.$$

Variance:

$$Var(Y) = Var(N_1(72)) + 4\,Var(N_2(72)) + 9\,Var(N_3(72)) = 432 + 864 + 648 = 1944.$$

Using the normal approximation:

$$P(Y \geq 1050) \approx P\left(Z \geq \frac{1050 - 1080}{\sqrt{1944}}\right) = P(Z \geq -0.68) = P(Z \leq 0.68) \approx 0.75.$$

Answer: (D) 0.75

Exercise 4.19. *The number of accidents follows a Poisson distribution with mean 12. Each accident generates 1, 2, or 3 claimants with probabilities $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$, respectively.*
Calculate the variance in the total number of claimants.
Choices:

(A) 20 (B) 25 (C) 30 (D) 35 (E) 40

Solution:
Let X_1, X_2, X_3 be the number of accidents resulting in 1, 2, and 3 claimants respectively. These are independent Poisson random variables with means:

$$\lambda_1 = 12 \cdot \frac{1}{2} = 6, \quad \lambda_2 = 12 \cdot \frac{1}{3} = 4, \quad \lambda_3 = 12 \cdot \frac{1}{6} = 2.$$

The total number of claimants is:

$$T = X_1 + 2X_2 + 3X_3.$$

Variance:

$$Var(T) = Var(X_1) + 4\,Var(X_2) + 9\,Var(X_3) = 6 + 4 \cdot 4 + 9 \cdot 2 = 6 + 16 + 18 = 40.$$

Answer: (E) 40

Exercise 4.20. *Workers' compensation claims are reported according to a Poisson process with mean 100 per month. The number of claims reported and the claim amounts are independently distributed. 2% of the claims exceed \$30,000. Calculate the number of complete months of data that must be gathered to have at least a 90% chance of observing at least 3 claims each exceeding \$30,000.*
Choices:

(A) 1 (B) 2 (C) 3 (D) 4 (E) 5

Solution:
Let $N_1(t)$ be the number of claims exceeding \$30,000 received up to time t months. Then:

$$\lambda_1 = 100 \times 0.02 = 2 \text{ (per month)}.$$

We compute:

$$P(N_1(2) \geq 3) = 1 - e^{-4}\left(1 + 4 + \frac{4^2}{2!}\right) \approx 0.761,$$

$$P(N_1(3) \geq 3) = 1 - e^{-6}\left(1 + 6 + \frac{6^2}{2!}\right) \approx 0.93.$$

Thus, 3 months are needed to ensure at least a 90% chance.
Answer: (C) 3

Exercise 4.21. *You are given:*

- *A loss occurrence in excess of $1 billion may be caused by a hurricane, an earthquake, or a fire.*
- *Hurricanes, earthquakes, and fires occur independently.*
- *The number of hurricanes causing such losses follows a Poisson process with expected inter-arrival time of 2.0 years.*
- *The number of earthquakes follows a Poisson process with expected inter-arrival time of 5.0 years.*
- *The number of fires follows a Poisson process with expected inter-arrival time of 10.0 years.*

Determine the expected time between all types of loss occurrences exceeding $1 billion.
Solution:
Let:

$$\lambda_1 = \frac{1}{2}, \quad \lambda_2 = \frac{1}{5}, \quad \lambda_3 = \frac{1}{10}.$$

These represent the rates of hurricanes, earthquakes, and fires respectively. The total rate of all loss events is:

$$\lambda = \lambda_1 + \lambda_2 + \lambda_3 = \frac{1}{2} + \frac{1}{5} + \frac{1}{10} = \frac{4}{5}.$$

Hence, the expected time between loss occurrences is:

$$\mathbb{E}[T] = \frac{1}{\lambda} = \frac{5}{4} = 1.25 \text{ years.}$$

Answer: 1.25 years

Exercise 4.22. *An insurance company has two insurance portfolios. Claims in Portfolio P occur in accordance with a Poisson process with mean 3 per year. Claims in Portfolio Q occur in accordance with a Poisson process with mean 5 per year. The two processes are independent.*

Calculate the probability that 3 claims occur in Portfolio P before 3 claims occur in Portfolio Q.

Choices: *(A) 0.28 (B) 0.33 (C) 0.38 (D) 0.43 (E) 0.48*
Solution:

Let $N_1(t)$ and $N_2(t)$ be the number of claims in portfolios P and Q by time t, respectively. Then $N_1 \sim Poisson(3t)$, $N_2 \sim Poisson(5t)$, and they are independent.

Let $N(t) = N_1(t) + N_2(t)$. Then $N \sim Poisson(8t)$. Each claim is from portfolio P with probability:

$$p = \frac{3}{3+5} = \frac{3}{8}.$$

The probability that 3 claims occur in Portfolio P before 3 claims occur in Portfolio Q is the probability that among the first 5 claims, at least 3 are from P:

$$P(Bin(5, 3/8) \geq 3) = \sum_{k=3}^{5} \binom{5}{k} \left(\frac{3}{8}\right)^k \left(\frac{5}{8}\right)^{5-k}.$$

$$= \binom{5}{3} \left(\frac{3}{8}\right)^3 \left(\frac{5}{8}\right)^2 + \binom{5}{4} \left(\frac{3}{8}\right)^4 \left(\frac{5}{8}\right) + \binom{5}{5} \left(\frac{3}{8}\right)^5.$$

$$= 10 \cdot \frac{27}{512} \cdot \frac{25}{64} + 5 \cdot \frac{81}{4096} \cdot \frac{5}{8} + \frac{243}{32768} \approx 0.2752.$$

Answer: (A) 0.28

Exercise 4.23. *Subway trains arrive at a station at a Poisson rate of 20 per hour.*

- *25% are express (travel time: 16 minutes)*
- *75% are local (travel time: 28 minutes)*

You take the first train that arrives. Your co-worker waits for the first express.

Choices: *(A) You arrive 6 minutes earlier than your co-worker*
(B) You arrive 4.5 minutes earlier than your co-worker
(C) Same arrival time
(D) You arrive 4.5 minutes later
(E) You arrive 6 minutes later

Solution:

Your Expected Time:

$$Wait = \frac{1}{20}\ hours = 3\ minutes$$

$$Expected\ travel\ time = 0.25 \cdot 16 + 0.75 \cdot 28 = 25$$

$$Total\ expected\ time = 3 + 25 = 28\ minutes$$

Co-worker's Expected Time:

$$Express\ train\ rate = 0.25 \cdot 20 = 5\ per\ hour \Rightarrow Expected\ wait = \frac{1}{5}\ hours = 12\ minutes$$

$$Travel\ time = 16 \Rightarrow Total\ time = 12 + 16 = 28\ minutes$$

Conclusion: *Both have the same expected arrival time.*

Answer: (C) Your expected arrival times are the same

4.2.5 Nonhomogeneous Poisson Process

Definition 4.7. *The counting process* $\{N(t) : t \geq 0\}$ *is said to be a* ***nonhomogeneous Poisson process*** *with intensity function* $\lambda(t) \geq 0$, $t \geq 0$, *if:*

(i) $N(0) = 0$,

(ii) $\{N(t) : t \geq 0\}$ *has independent increments,*

(iii) $\mathbb{P}(N(t+h) - N(t) \geq 2) = o(h)$,

(iv) $\mathbb{P}(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$.

Definition 4.8. *The counting process* $\{N(t) : t \geq 0\}$ *is said to be a* ***nonhomogeneous Poisson process*** *with intensity function* $\lambda(t)$, $t \geq 0$, *if:*

(i) $N(0) = 0$,

(ii) For each $t > 0$, $N(t)$ *has a Poisson distribution with mean*

$$m(t) = \int_0^t \lambda(s)\, ds,$$

(iii) For each $0 \leq t_1 < t_2 < \cdots < t_m$, *the random variables* $N(t_1)$, $N(t_2) - N(t_1)$, ..., $N(t_m) - N(t_{m-1})$ *are independent.*

The function $m(t)$ *is called the* ***mean value function*** *of the nonhomogeneous Poisson process.*

It follows from the previous definition that for each $0 \leq t_1 < t_2 < \cdots < t_m$ and for all integers $k_1, \ldots, k_m \geq 0$,

$$\begin{aligned}\mathbb{P}(N(t_1) = k_1, N(t_2) = k_2, \ldots, N(t_m) = k_m) &= \mathbb{P}(N(t_1) = k_1, N(t_2) - N(t_1) = k_2 - k_1, \ldots, \\ &\quad N(t_m) - N(t_{m-1}) = k_m - k_{m-1}) \\ &= \frac{e^{-m(t_1)}(m(t_1))^{k_1}}{k_1!} \cdot \frac{e^{-(m(t_2)-m(t_1))}(m(t_2) - m(t_1))^{k_2-k_1}}{(k_2-k_1)!} \\ &\quad \cdots \cdot \frac{e^{-(m(t_m)-m(t_{m-1}))}(m(t_m) - m(t_{m-1}))^{k_m-k_{m-1}}}{(k_m-k_{m-1})!}.\end{aligned}$$

If $\lambda(t) = \lambda$ for each $t \geq 0$, then the process is a **homogeneous Poisson process**. Let S_n be the time of the n–th occurrence. Then,

$$\mathbb{P}(S_n > t) = \mathbb{P}(N_t \leq n-1) = \sum_{j=0}^{n-1} \frac{e^{-m(t)}(m(t))^j}{j!}.$$

Exercise 4.24. *For a nonhomogeneous Poisson process, the intensity function is given by*

$$\lambda(t) = \begin{cases} 10 & \text{if } t \in (0, 1/2], \\ 2 & \text{if } t \in (1/2, 1], \\ 10 & \text{if } t \in (1, 3/2], \\ \vdots & \end{cases}$$

(i) How many occurrences are expected in the time period $(0, 1]$*? During* $(0, 3/2]$*?*

(ii) If $S_{10} = 0.45$ *is given, calculate the probability that* $S_{11} > 0.75$.

Answer:

(i) The expected number in the time period $(0, 1]$ *is*

$$m(1) = \int_0^1 \lambda(t)\,dt = \int_0^{1/2} 10\,dt + \int_{1/2}^1 2\,dt = 5 + 1 = 6.$$

The expected number in the time period $(0, 3/2]$ *is*

$$m(3/2) = \int_0^{3/2} \lambda(t)\,dt = \int_0^{1/2} 10\,dt + \int_{1/2}^1 2\,dt + \int_1^{3/2} 10\,dt = 5 + 1 + 5 = 11.$$

(ii)

$$\begin{aligned}
\mathbb{P}(S_{11} > 0.75 \mid S_{10} = 0.45) &= \mathbb{P}(N(0.75) = 10 \mid S_{10} = 0.45) \\
&= \mathbb{P}(N(0.75) - N(0.45) = 0 \mid S_{10} = 0.45) \\
&= \mathbb{P}(N(0.75) - N(0.45) = 0) \\
&= e^{-1},
\end{aligned}$$

because

$$m(0.75) - m(0.45) = \int_{0.45}^{0.5} 10\,dt + \int_{0.5}^{0.75} 2\,dt = 0.5 + 0.5 = 1.$$

4.3 Compound Poisson process

Definition 4.9. *A stochastic process* $\{X(t) : t \geq 0\}$ *is said to be a* ***compound Poisson process*** *if it can be represented as*

$$X(t) = \sum_{i=1}^{N(t)} Y_i, \quad t \geq 0,$$

where $\{N(t) : t \geq 0\}$ *is a Poisson process and* $\{Y_i\}$ *is a sequence of i.i.d. random variables independent of* $\{N(t)\}$*.*

Using the law of total expectation, we have:

$$\begin{aligned}
\mathbb{E}[X(t) \,|\, N(t) = n] &= \mathbb{E}\left[\sum_{i=1}^n Y_i \,\middle|\, N(t) = n\right] = n\mathbb{E}[Y_1], \\
\mathbb{E}[X(t)] &= \mathbb{E}[\mathbb{E}[X(t) \,|\, N(t)]] = \mathbb{E}[N(t)] \cdot \mathbb{E}[Y_1] = \lambda t\mathbb{E}[Y_1].
\end{aligned}$$

Similarly, for the variance:

$$\begin{aligned}
\mathrm{Var}(X(t) \,|\, N(t) = n) &= \mathrm{Var}\left(\sum_{i=1}^n Y_i\right) = n\mathrm{Var}(Y_1), \\
\mathrm{Var}(X(t)) &= \mathbb{E}[\mathrm{Var}(X(t) \,|\, N(t))] + \mathrm{Var}(\mathbb{E}[X(t) \,|\, N(t)]) \\
&= \mathbb{E}[N(t)] \cdot \mathrm{Var}(Y_1) + \mathrm{Var}(N(t)) \cdot (\mathbb{E}[Y_1])^2 \\
&= \lambda t\mathrm{Var}(Y_1) + \lambda t(\mathbb{E}[Y_1])^2 = \lambda t\mathbb{E}[Y_1^2].
\end{aligned}$$

Hence,

$$\mathbb{E}[X(t)] = \lambda t\mathbb{E}[Y_1], \quad \mathrm{Var}(X(t)) = \lambda t\mathbb{E}[Y_1^2].$$

Exercise 4.25. *The claims department of an insurance company receives envelopes with claims at a Poisson rate of $\lambda = 50$ envelopes per week. For any period of time, the number of envelopes and the number of claims per envelope are independent. The number of claims per envelope follows the distribution:*

Number of Claims	*Probability*
1	*0.20*
2	*0.25*
3	*0.40*
4	*0.15*

Using the normal approximation, calculate the 90th percentile of the number of claims received in 13 weeks.

Solution

Let $N(t)$ be the number of envelopes received until time t. Then $N(t)$ is a Poisson process with rate $\lambda = 50$. Let $\{X_j\}$ be the number of claims in each envelope. Then the total number of claims received in 13 weeks is

$$Y = \sum_{j=1}^{N(13)} X_j.$$

Compute the moments:

$$\mathbb{E}[X] = (1)(0.2) + (2)(0.25) + (3)(0.4) + (4)(0.15) = 2.5,$$
$$\mathbb{E}[X^2] = (1)^2(0.2) + (2)^2(0.25) + (3)^2(0.4) + (4)^2(0.15) = 7.2.$$

Then:

$$\mathbb{E}[Y] = \mathbb{E}[N(13)] \cdot \mathbb{E}[X] = 50 \cdot 13 \cdot 2.5 = 1625,$$
$$\mathrm{Var}(Y) = \mathbb{E}[N(13)] \cdot \mathbb{E}[X^2] = 50 \cdot 13 \cdot 7.2 = 4680.$$

Let q be the 90th percentile. Then,

$$0.90 \approx \mathbb{P}(Y \leq q) \approx \mathbb{P}\left(Z \leq \frac{q - 1625}{\sqrt{4680}}\right) = \mathbb{P}(Z \leq 1.282).$$

So,

$$q = 1625 + \sqrt{4680} \cdot 1.282 \approx 1625 + 124.03 = 1749.03.$$

Answer: ***(D) 1750***

Exercise 4.26. *A company provides insurance to a concert hall for losses due to power failure. You are given:*

(i) The number of power failures in a year has a Poisson distribution with mean 1.

(ii) The distribution of ground-up losses due to a single power failure is:

Loss (x)	*Probability*
10	*0.3*
20	*0.3*
50	*0.4*

(iii) The number of power failures and the amounts of losses are independent.

(iv) There is an annual deductible of 30.

Calculate the expected amount of claims paid by the insurer in one year.
Solution: *Let N be the number of power failures. Let $\{X_j\}$ be the ground-up losses. Define the total loss:*

$$S = \sum_{j=1}^{N} X_j.$$

We are to compute the expected amount paid by the insurer, which is:

$$\mathbb{E}[(S-30)^+] = \mathbb{E}[S] - \mathbb{E}[S \wedge 30].$$

First, compute the expected total loss:

$$\mathbb{E}[X] = 10 \cdot 0.3 + 20 \cdot 0.3 + 50 \cdot 0.4 = 29,$$
$$\mathbb{E}[S] = \mathbb{E}[N] \cdot \mathbb{E}[X] = 1 \cdot 29 = 29.$$

Compute specific probabilities:

$$\begin{aligned}
\mathbb{P}(S=0) &= \mathbb{P}(N=0) = e^{-1} \approx 0.3679, \\
\mathbb{P}(S=10) &= \mathbb{P}(N=1, X_1=10) = e^{-1} \cdot 0.3 = 0.1104, \\
\mathbb{P}(S=20) &= \mathbb{P}(N=1, X_1=20) + \mathbb{P}(N=2, X_1+X_2=20) \\
&= e^{-1} \cdot 0.3 + \frac{e^{-1}}{2!} \cdot (0.3)^2 = 0.1104 + 0.0270 = 0.1374.
\end{aligned}$$

Then,

$$\begin{aligned}
\mathbb{E}[S \wedge 30] &= 10 \cdot \mathbb{P}(S=10) + 20 \cdot \mathbb{P}(S=20) + 30 \cdot \mathbb{P}(S \geq 30) \\
&= 10 \cdot 0.1104 + 20 \cdot 0.1374 + 30 \cdot (1 - 0.3679 - 0.1104 - 0.1374) \\
&= 1.104 + 2.748 + 30 \cdot 0.3843 = 1.104 + 2.748 + 11.529 = 16.481.
\end{aligned}$$

Finally,

$$\mathbb{E}[(S-30)^+] = \mathbb{E}[S] - \mathbb{E}[S \wedge 30] = 29 - 16.481 = 12.519.$$

Answer: (D) 12 *section*Problem 5.20 (Problem #14, Sample Test)*
You are given:

- *An aggregate loss distribution has a compound Poisson distribution with expected number of claims equal to 1.25.*
- *Individual claim amounts can take only the values 1, 2, or 3, with equal probability.*

Determine the probability that aggregate losses exceed 3.78.
Solution: *Let N be the number of claims. Let $\{X_j\}$ be the individual claim amounts. Let*

$$S = \sum_{j=1}^{N} X_j.$$

We compute the following probabilities:

$$\begin{aligned}
\mathbb{P}(S=0) &= \mathbb{P}(N=0) = e^{-1.25} \approx 0.2865,\\
\mathbb{P}(S=1) &= \mathbb{P}(N=1, X_1=1) = e^{-1.25}\cdot\frac{(1.25)^1}{1!}\cdot\frac{1}{3} \approx 0.1194,\\
\mathbb{P}(S=2) &= \mathbb{P}(N=1, X_1=2) + \mathbb{P}(N=2, X_1+X_2=2)\\
&= e^{-1.25}\cdot\frac{(1.25)^1}{1!}\cdot\frac{1}{3} + e^{-1.25}\cdot\frac{(1.25)^2}{2!}\cdot\left(\frac{1}{3}\right)^2 \approx 0.1194 + 0.0249 = 0.1442,\\
\mathbb{P}(S=3) &= \mathbb{P}(N=1, X_1=3) + \mathbb{P}(N=2, X_1+X_2=3) + \mathbb{P}(N=3, X_1+X_2+X_3=3)\\
&= e^{-1.25}\cdot\frac{(1.25)^1}{1!}\cdot\frac{1}{3} + e^{-1.25}\cdot\frac{(1.25)^2}{2!}\cdot\left(\frac{1}{3}\right)^2 + e^{-1.25}\cdot\frac{(1.25)^3}{3!}\cdot\left(\frac{1}{3}\right)^3\\
&\approx 0.1194 + 0.0488 + 0.0035 = 0.1717.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{P}(S>3) &= 1 - \mathbb{P}(S=0) - \mathbb{P}(S=1) - \mathbb{P}(S=2) - \mathbb{P}(S=3)\\
&\approx 1 - 0.2865 - 0.1194 - 0.1442 - 0.1717 = 0.2782.
\end{aligned}$$

Exercise 4.27. *At a taxi station, cars of brand A and brand B arrive according to two independent Poisson processes with respective intensities* 10 *and* 15 *per hour.*

1. *Let T be the (random) minute of arrival of the first taxi. Find the distribution of T and the probability that the first arriving taxi is of brand A.*
2. *If the first arriving taxi is of brand A, find the distribution of the additional waiting time (after the arrival of this taxi) before the arrival of the first taxi of brand B.*
3. *Show that the arrival times of taxis (regardless of brand) form a Poisson process and give its intensity.*

Solution.

1. *Let T_1 and T_2 be the arrival times of the first taxis of brands A and B respectively. Then T_1 and T_2 are independent exponential random variables with parameters $\lambda_1 = 10$ and $\lambda_2 = 15$ (per hour). By definition,*

$$T = \min(T_1, T_2).$$

   From the previous exercise,

$$T \sim \mathrm{Exp}(\lambda_1+\lambda_2) = \mathrm{Exp}(25).$$

   Moreover,

$$\mathbb{P}(\textit{first taxi is brand A}) = \frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{10}{25} = \frac{2}{5}.$$

2. *Conditional on the first taxi being of brand A, we seek the distribution of*

$$T_2 - T_1 \,\Big|\, T_2 \geq T_1.$$

   Using the memoryless property of the exponential law (or Bayes' formula),

$$\mathbb{P}(T_2 \geq x + T_1 \mid T_2 \geq T_1) = \frac{\mathbb{P}(T_2 \geq x + T_1)}{\mathbb{P}(T_2 \geq T_1)} = e^{-\lambda_2 x}.$$

   Therefore,

$$T_2 - T_1 \,\Big|\, T_2 \geq T_1 \sim \mathrm{Exp}(\lambda_2) = \mathrm{Exp}(15).$$

3. *Let* N_t^1 *and* N_t^2 *be the numbers of taxis of brands A and B arriving up to time t (in hours). These are independent Poisson processes with intensities* $\lambda_1 = 10$ *and* $\lambda_2 = 15$ *respectively. Set*

$$N_t = N_t^1 + N_t^2.$$

For $k \geq 0$,

$$\begin{aligned}
\mathbb{P}(N_t = k) &= \sum_{l=0}^{k} \mathbb{P}(N_t^1 = l,\ N_t^2 = k - l) \\
&= \sum_{l=0}^{k} \mathbb{P}(N_t^1 = l)\, \mathbb{P}(N_t^2 = k - l) \\
&= e^{-(\lambda_1+\lambda_2)t} \sum_{l=0}^{k} \frac{(\lambda_1 t)^l}{l!} \frac{(\lambda_2 t)^{k-l}}{(k-l)!} \\
&= e^{-(\lambda_1+\lambda_2)t} \frac{(\lambda_1 + \lambda_2)^k t^k}{k!}.
\end{aligned}$$

Thus $N_t \sim \mathrm{Poisson}((\lambda_1 + \lambda_2)t) = \mathrm{Poisson}(25t)$.

To verify that (N_t) *is indeed a Poisson process, we need to check that it has independent increments and right-continuous sample paths. Right-continuity follows immediately from the definition of* N_t *as the sum of two Poisson processes.*

For independence of increments, let I and J be two disjoint time intervals and $k, l \geq 0$. *Using the independence of increments of* N^1 *and* N^2,

$$\mathbb{P}(N(I) = k,\ N(J) = l) = \sum_{i=0}^{k} \sum_{j=0}^{l} \mathbb{P}(N^1(I) = i, N^2(I) = k-i) \mathbb{P}(N^1(J) = j, N^2(J) = l-j)$$

$$= \mathbb{P}(N(I) = k)\, \mathbb{P}(N(J) = l).$$

Hence $N = (N_t)_{t \geq 0}$ *is a Poisson process with intensity*

$$\lambda = \lambda_1 + \lambda_2 = 25.$$

Chapter 5

Brownian Motion

5.1 Introduction to Brownian Motion

Brownian motion is one of the most fundamental and intriguing stochastic processes [1,3,4,13,16]. Its origin traces back to 1827, when the botanist Robert Brown observed under a microscope that tiny pollen particles suspended in water exhibited a perpetual, irregular motion.

This seemingly chaotic movement is now understood as the result of countless collisions between the particle and surrounding water molecules. Each collision produces only a minuscule displacement, but when aggregated over an immense number of impacts, the total displacement appears random. This phenomenon is intimately connected to the *central limit theorem*, since the particle's position can be viewed as the cumulative effect of many small, independent random contributions.

A similar line of reasoning arises in *financial mathematics*. Asset prices fluctuate due to the continuous influence of innumerable, unpredictable market events. Although each event typically has only a negligible effect on the price, their accumulation leads to significant randomness. Thus, the modeling of asset price dynamics parallels the modeling of Brownian motion, which explains why this process lies at the heart of modern stochastic finance. In this chapter, we formally introduce a stochastic process

$$\{B(t) : t \geq 0\},$$

called *Brownian motion* (or the *Wiener process*), which provides the canonical mathematical framework for the phenomena described above.

5.2 A simple model for Brownian motion

We now construct a mathematical model for the motion of a small pollen grain suspended in a fluid. For simplicity, we first consider the motion in one dimension (along the real line). The extension to three dimensions is straightforward, since the coordinate processes can be treated independently.

The particle moves randomly along the line as a result of collisions with surrounding fluid molecules. Each collision causes a small displacement, either to the left or to the right. Without loss of generality, we assume that at time $t = 0$ the particle is located at the origin, i.e. at position 0. Let N denote the (large) number of collisions per unit time between the particle and the surrounding molecules. Suppose further that each collision

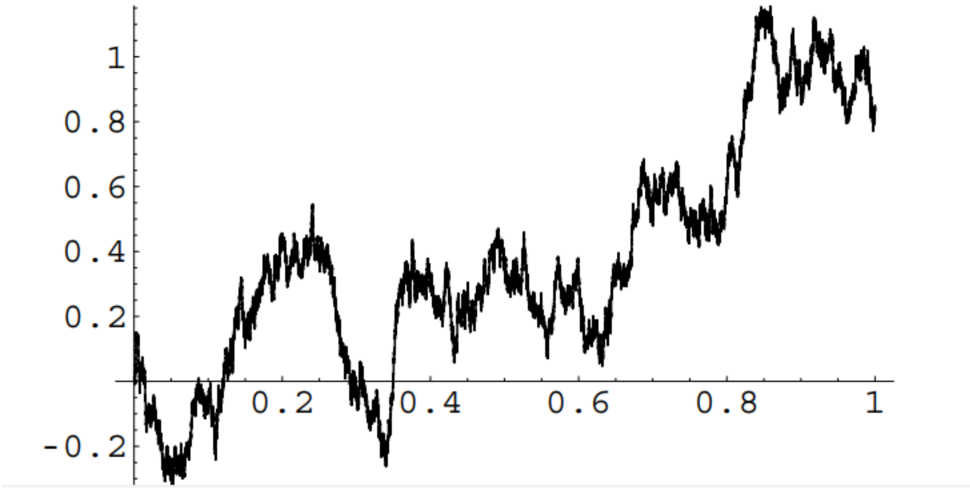

Figure 5.1: A sample path of the Brownian motion.

shifts the particle by a fixed, very small distance, and that the displacement is equally likely to be to the left or to the right, each with probability 1/2. A typical sample path of Brownian motion as a function of time is illustrated in Figure 5.1 It is important to emphasize that inertia of the particle is neglected in this model. That is, each collision instantaneously changes the particle's position, and we do not attempt to describe its velocity. This assumption is justified in the case of a highly viscous fluid, where inertial effects are negligible compared to the random displacements.

5.2.1 From random walks to Brownian motion

A natural model for the motion of a particle in d dimensions can be described in terms of a random walk. We let the position of the particle after n steps be

$$S_n = Y_1 + Y_2 + \cdots + Y_n,$$

where $(Y_i)_{i\geq 1}$ are i.i.d. random vectors uniformly distributed on $\{-1,1\}^d$. Equivalently, at each step the d coordinates are updated by independent fair ± 1 displacements.

We are interested in the large-scale behavior of the particle, as observed from far away. Since $n^{-1}S_n \to 0$ almost surely, the central limit theorem suggests that the proper spatial scaling is of order $\sqrt{n}$. This leads us to introduce the rescaled process

$$S_n^*(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor},$$

where $\lfloor x \rfloor$ denotes the integer part of x.

Now consider a partition $0 = t_0 < t_1 < \cdots < t_p$. The increments

$$S_n^*(t_i) - S_n^*(t_{i-1}), \qquad i = 1, 2, \dots, p,$$

are independent, and by the central limit theorem they converge in distribution to centered Gaussian vectors with covariance matrix $(t_i - t_{i-1})I_d$. It is clear that this property still holds (up to a multiplicative factor) if the distribution of the steps is changed, provided they remain centered with finite variance. This observation naturally motivates the definition of a continuous limiting process, which will serve as the mathematical model for Brownian motion.

5.3 Definition of the Brownian motion

Motivated by the preceding discussion, we now give the formal definition.

Definition 5.1 (Brownian Motion)**.** *A stochastic process* $B = \{B(t) : t \geq 0\}$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called* Brownian motion *or* Wiener process *if:*

1. $B(0) = 0$,
2. *B has independent increments, that is, for all* $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$*, the random variables*
$$B(t_1),\ B(t_2) - B(t_1),\ \ldots,\ B(t_n) - B(t_{n-1})$$
are independent,
3. *B has normal increments, that is, for all* $t \geq 0$ *and* $h > 0$*,*
$$B(t + h) - B(t) \sim \mathcal{N}(0, h),$$
4. *B has continuous sample paths, that is, for all* $\omega \in \Omega$*, the function* $t \mapsto B(t, \omega)$ *is continuous in* t*.*

First of all, one has to ask whether a process satisfying these four requirements actually exists. This question is non-trivial and will be addressed positively in Section 4.5. Here, we sketch an idea of a possible approach to proving existence.

The first three properties in the definition of Brownian motion concern only the finite-dimensional distributions of the process B. By Kolmogorov's existence theorem, one can show that a process with finite-dimensional distributions satisfying conditions (1), (2), and (3) indeed exists.

To apply Kolmogorov's theorem, it is necessary to verify that the family of finite-dimensional distributions defined by conditions (1)–(3) is consistent; that is, these conditions do not contradict one another. This verification essentially reduces to the following argument. Suppose that for some $0 \leq t_1 \leq t_2 \leq t_3$, the increments satisfy

$$B(t_2) - B(t_1) \sim \mathcal{N}(0, t_2 - t_1) \quad \text{and} \quad B(t_3) - B(t_2) \sim \mathcal{N}(0, t_3 - t_2),$$

and are independent. Then, by the convolution property of the normal distribution, we obtain

$$B(t_3) - B(t_1) = \big(B(t_3) - B(t_2)\big) + \big(B(t_2) - B(t_1)\big) \sim \mathcal{N}\big(0, (t_3 - t_2) + (t_2 - t_1)\big) = \mathcal{N}(0, t_3 - t_1).$$

Since this is consistent with condition (3), there is no apparent contradiction among conditions (1), (2), and (3). Therefore, Kolmogorov's existence theorem can be applied to construct a process satisfying these conditions.

However, Kolmogorov's theorem alone does not ensure that the resulting process also satisfies condition (4). An additional refinement of the construction is required in order to incorporate this property. For this reason, we adopt a different approach to establish the existence of a process satisfying all four conditions.

The following example illustrates that condition (4) is indispensable and cannot be omitted from the definition of Brownian motion.

Example 5.1 (Necessity of Continuity)**.** *Assume that we have a process* $\{B(t) : t \geq 0\}$ *satisfying conditions (1), (2), (3), and (4). We now show how, by modifying* B*, one can construct a process* $\{\widetilde{B}(t) : t \geq 0\}$ *that still satisfies properties (1), (2), and (3), but fails to satisfy property (4). This demonstrates that condition (4) does not follow from the first three conditions alone.*

Let $U \sim \text{Uniform}[0, 1]$ *be a random variable, independent of the process* B*. Define*

$$\widetilde{B}(t) = \begin{cases} B(t), & \text{if } t \neq U, \\ 0, & \text{if } t = U. \end{cases}$$

Step 1: Finite-dimensional distributions. *We claim that* $\widetilde{B}$ *has the same finite-dimensional distributions as* B*. Fix times* $0 \leq t_1 < t_2 < \cdots < t_n$*. The vectors*

$$\Big(B(t_1), \ldots, B(t_n)\Big) \quad \text{and} \quad \Big(\widetilde{B}(t_1), \ldots, \widetilde{B}(t_n)\Big)$$

are equal unless $U \in \{t_1, \ldots, t_n\}$*. But since* U *is continuously distributed, the probability of this event is zero. Hence, with probability one, the two random vectors coincide, and therefore they have the same distribution. This shows that* $\widetilde{B}$ *satisfies properties (1), (2), and (3).*

Step 2: Path continuity. *We now examine whether* $\widetilde{B}$ *satisfies condition (4). Consider the point* $t = U$*. By the continuity of* B*, we have*

$$\lim_{\substack{t \to U \\ t \neq U}} \widetilde{B}(t) = \lim_{\substack{t \to U \\ t \neq U}} B(t) = B(U).$$

However, by definition,

$$\widetilde{B}(U) = 0.$$

Thus, $\widetilde{B}$ *has a discontinuity at* U *unless* $B(U) = 0$*.*

Step 3: Probability of discontinuity. *Since* $B(U) \sim \mathcal{N}(0, U)$ *for each fixed* U*, we know that*

$$\mathbb{P}[B(U) = 0] = 0.$$

Therefore,

$$\mathbb{P}\Big[\widetilde{B} \text{ has a discontinuity at } U\Big] = 1.$$

Hence, with probability one, the process $\widetilde{B}$ *fails to have continuous sample paths.*

This example shows that property (4) (continuity of sample paths) is not implied by conditions (1), (2), and (3). It must therefore be explicitly included in the definition of Brownian motion.

5.3.1 Multivariate gaussian distributions and gaussian processes

It follows directly from the definition of Brownian motion that its one-dimensional distributions are Gaussian, namely

$$B(t) \sim \mathcal{N}(0, t).$$

A natural question is: *what do the finite-dimensional (multivariate) distributions of Brownian motion look like?* It turns out that these distributions belong to the class of *multivariate Gaussian distributions.* The purpose of this subsection is to introduce these distributions.

Recall first the one-dimensional case. A random variable X is said to have a Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ (written $X \sim \mathcal{N}(\mu, \sigma^2)$) if its density is

$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right), \qquad t \in \mathbb{R}.$$

It is convenient to extend this definition to the degenerate case $\sigma^2 = 0$ by declaring that

$$X \sim \mathcal{N}(\mu, 0) \quad \Longleftrightarrow \quad X = \mu \text{ almost surely.}$$

The characteristic function of a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$\varphi_X(u) = \mathbb{E}[e^{iuX}] = \exp\left(i\mu u - \tfrac{1}{2}\sigma^2 u^2\right), \qquad u \in \mathbb{R}.$$

A Gaussian random variable with $\mu = 0$ and $\sigma^2 = 1$ is called a *standard Gaussian.* Its density is given explicitly by

$$f_X(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{t^2}{2}\right), \qquad t \in \mathbb{R}.$$

We now extend this notion from random variables to random vectors. We begin with the definition of a *standard Gaussian random vector.*

Definition 5.2 (Standard Gaussian Vector)**.** *Fix dimension $d \in \mathbb{N}$. A random vector $X = (X_1, \ldots, X_d)^T$ is called d-*dimensional standard Gaussian *if:*

1. *$X_1, \ldots, X_d \sim \mathcal{N}(0, 1)$ are standard Gaussian random variables, and*
2. *$X_1, \ldots, X_d$ are independent random variables.*

By independence, the joint density of a d-dimensional standard Gaussian vector X is given by

$$f_{X_1,\ldots,X_d}(t_1, \ldots, t_d) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(t_1^2 + \cdots + t_d^2)\right) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\langle t, t\rangle\right),$$

where $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. The expectation vector of X is equal to zero (because all components X_i have zero mean by definition). The covariance matrix of X is the $d \times d$ identity matrix (because the variance of any component X_i is 1 and different components are independent and hence uncorrelated):

$$\mathbb{E}[X] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \operatorname{Cov}(X) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

The next lemma states that the standard Gaussian distribution remains unchanged under rotations of the space around the origin.

Lemma 5.1. *If X is a d-dimensional standard Gaussian random vector and A an orthogonal $d \times d$ matrix, then the random vector AX is also standard Gaussian.*

Proof. Recall that the orthogonality of the matrix A means that $AA^T = A^T A = I_d$. It follows that $\det A = \pm 1$, and in particular, A is invertible. By the transformation formula, the density of the random vector AX is

$$f_{AX}(t) = f_X(A^{-1}t)\left|\det(A^{-1})\right| = f_X(A^{-1}t) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\langle A^{-1}t, A^{-1}t\rangle\right).$$

Now observe that

$$\langle A^{-1}t, A^{-1}t\rangle = \langle (A^{-1})^T A^{-1}t, t\rangle = \langle (AA^T)^{-1}t, t\rangle = \langle t, t\rangle.$$

Hence,

$$f_{AX}(t) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\langle t, t\rangle\right) = f_X(t),$$

which shows that AX has the same distribution as X. □

The next lemma will be used in the construction of the Brownian motion in Section 4.5.

Lemma 5.2. *Let X_1 and X_2 be independent Gaussian random variables with mean* 0 *and variance σ^2. Then the random variables*

$$Y_1 = \frac{X_1 + \sqrt{2}\,X_2}{\sqrt{2}}, \qquad Y_2 = \frac{X_1 - \sqrt{2}\,X_2}{\sqrt{2}}$$

are also independent Gaussian random variables with mean 0 *and variance σ^2.*

Proof. Since X_1 and X_2 are independent $\mathcal{N}(0, \sigma^2)$ random variables, the vector

$$\begin{pmatrix} X_1/\sigma \\ X_2/\sigma \end{pmatrix}$$

is a 2-dimensional standard Gaussian vector.

Now consider the linear transformation

$$\begin{pmatrix} Y_1/\sigma \\ Y_2/\sigma \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}}_{=:Q} \begin{pmatrix} X_1/\sigma \\ X_2/\sigma \end{pmatrix}.$$

The matrix Q is orthogonal, i.e. $QQ^\top = I$. It is a standard fact that if Z is a standard Gaussian vector in $\mathbb{R}^d$ and Q is an orthogonal matrix, then QZ is also a standard Gaussian vector in $\mathbb{R}^d$.

Thus, $(Y_1/\sigma, Y_2/\sigma)^\top$ is a 2-dimensional standard Gaussian vector, which implies that Y_1/σ and Y_2/σ are independent $\mathcal{N}(0, 1)$ variables. Rescaling back by σ, we conclude that

$$Y_1,\, Y_2 \overset{\text{indep.}}{\sim} \mathcal{N}(0, \sigma^2).$$

□

We are now ready to introduce the general (non-standard) multivariate Gaussian distribution. The guiding principle is the following: a random vector is said to be Gaussian if it can be expressed as an affine transformation of a standard Gaussian random vector.

Definition 5.3. *A random vector $Y = (Y_1, \dots, Y_d)^T$ is called* d-dimensional Gaussian *if there exists some $m \in \mathbb{N}$, an m-dimensional standard Gaussian vector $X = (X_1, \dots, X_m)^T$, a $d \times m$ matrix A, and a vector $\mu \in \mathbb{R}^d$ such that*

$$Y \overset{d}{=} AX + \mu.$$

Exercise 5.1. *Show that the expectation and the covariance matrix of Y are given by*

$$E(Y) = \mu, \quad \operatorname{Cov}(Y) = AA^T.$$

Remark 5.1. *We usually denote the covariance matrix by $\Sigma := \operatorname{Cov}(Y) = AA^T$ (not by Σ^2), and write $Y \sim \mathcal{N}_d(\mu, \Sigma)$. Note that the parameter μ takes values in $\mathbb{R}^d$, whereas the covariance matrix Σ can be any symmetric, positive semidefinite matrix.*

Any affine transformation of a Gaussian vector is again a Gaussian vector:

Lemma 5.3. *If $Y \sim \mathcal{N}_d(\mu, \Sigma)$ is a d-dimensional Gaussian vector, A_0 is a $d_0 \times d$ matrix, and $\mu_0 \in \mathbb{R}^{d_0}$, then*

$$A_0 Y + \mu_0 \sim \mathcal{N}_{d_0}(A_0 \mu + \mu_0, A_0 \Sigma A_0^T).$$

Proof. By definition, we can represent Y in the form $Y = AX + \mu$, where $AA^T = \Sigma$ and X is an m-dimensional standard Gaussian vector. The d_0-dimensional random vector

$$A_0 Y + \mu_0 = A_0(AX + \mu) + \mu_0 = (A_0 A)X + (A_0 \mu + \mu_0)$$

is an affine transformation of X, and hence, it is multivariate Gaussian. The parameters of $A_0 Y + \mu_0$ are given by

$$E(A_0 Y + \mu_0) = A_0 \mu + \mu_0, \quad \operatorname{Cov}(A_0 Y + \mu_0) = (A_0 A)(A_0 A)^T = A_0 A A^T A_0^T = A_0 \Sigma A_0^T.$$

□

Remark 5.2. *In particular, any component Y_i of a Gaussian random vector $(Y_1, \dots, Y_d)^T$ is a Gaussian random variable. The converse is not true: If $Y_1, \dots, Y_d$ are Gaussian random variables, then it is in general not true that $(Y_1, \dots, Y_d)^T$ is a Gaussian random vector. However, if we additionally require that $Y_1, \dots, Y_d$ are independent, the statement becomes true.*

Lemma 5.4. *Let $Y_1, \dots, Y_d$ be independent Gaussian random variables. Then, the vector $(Y_1, \dots, Y_d)^T$ is a Gaussian random vector.*

Proof. Let $Y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Then, we can write $Y_i = \sigma_i X_i + \mu_i$, where X_i are independent standard normal random variables. Therefore, the random vector $(Y_1, \dots, Y_d)^T$ is an affine transformation of the standard Gaussian vector $(X_1, \dots, X_d)^T$, and hence is itself a d-dimensional Gaussian vector. □

Lemma 5.5. *The characteristic function of a d-dimensional Gaussian random vector $Y \sim \mathcal{N}_d(\mu, \Sigma)$ is given by*

$$\varphi_Y(t) := E\left(e^{i\langle t, Y\rangle}\right) = \exp\left(i\langle \mu, t\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\right), \quad t \in \mathbb{R}^d.$$

Proof. Fix $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. The mapping $y \mapsto \langle t, y\rangle$ is a linear map from $\mathbb{R}^d$ to $\mathbb{R}$ with associated matrix $(t_1, \ldots, t_d)$. By previous Lemma, the random variable $Z := \langle t, Y\rangle$ is Gaussian with expectation $\langle \mu, t\rangle$ and variance $\langle t, \Sigma t\rangle$. We have

$$\varphi_Y(t) = E\left(e^{i\langle t,Y\rangle}\right) = E\left(e^{iZ}\right) = \varphi_Z(1) = \exp\left(i\langle \mu, t\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\right),$$

where the last step uses the known formula for the characteristic function of a Gaussian random variable. □

Exercise 5.2. *Let $X_1, X_2, \ldots$ be a sequence of d-dimensional Gaussian vectors whose expectations μ_n converge to μ and covariance matrices Σ_n converge to Σ. Show that X_n converges in distribution to $\mathcal{N}_d(\mu, \Sigma)$.*

What is the density of a multivariate Gaussian distribution $\mathcal{N}_d(\mu, \Sigma)$? First of all, this density does not always exist, as the following example shows.

Example 5.2. *Let us construct a two-dimensional Gaussian random vector which has no density. Let X be a standard normal random variable. The two-dimensional vector $(X, X)^\top$ is Gaussian because it can be represented as a linear transformation AX, where*

$$A : x \mapsto \begin{pmatrix} x \\ x \end{pmatrix}.$$

However, the random vector $(X, X)^\top$ has no density with respect to the two-dimensional Lebesgue measure because it takes values in the line $\{(x, x) : x \in \mathbb{R}\}$, which has Lebesgue measure zero. The covariance matrix of $(X, X)^\top$ is

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

which is degenerate (its determinant is 0).

The next lemma gives a formula for the density of a multivariate Gaussian distribution when Σ is non-degenerate.

Lemma 5.6. *The density of a d-dimensional Gaussian random vector $Y \sim \mathcal{N}_d(\mu, \Sigma)$, where Σ is a non-degenerate matrix, is given by*

$$f_Y(t) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}\langle t - \mu, \Sigma^{-1}(t - \mu)\rangle\right).$$

If the matrix Σ is degenerate, then Y has no density with respect to the Lebesgue measure on $\mathbb{R}^d$.

Proof. Since Σ is positive semidefinite, we can write $\Sigma = \Sigma^{1/2} \cdot \Sigma^{1/2}$ for some symmetric matrix $\Sigma^{1/2}$. Then we have the representation

$$Y = \Sigma^{1/2} X + \mu,$$

where X is a standard normal vector in $\mathbb{R}^d$. □

Let X be a standard Gaussian vector on $\mathbb{R}^d$. Consider the transformation

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto \Sigma^{1/2}x + \mu.$$

Then, $T(X) \overset{d}{=} Y$.

1. If Σ is degenerate, then the image of T is a subspace of $\mathbb{R}^d$ of dimension strictly less than d. Therefore, the image of T has Lebesgue measure 0. Hence, Y takes values in a subset of $\mathbb{R}^d$ of Lebesgue measure zero, and so Y has no density.

2. If we assume that $\det \Sigma \neq 0$, then the inverse transformation is

$$T^{-1}(y) = \Sigma^{-1/2}(y - \mu).$$

The density of X is

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\langle x, x\rangle\right), \quad x \in \mathbb{R}^d.$$

We compute the density of Y using the change-of-variable formula:

$$f_Y(y) = f_X(T^{-1}(y)) \left|\det T^{-1}\right| = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}\left\langle \Sigma^{-1/2}(y-\mu), \Sigma^{-1/2}(y-\mu)\right\rangle\right), y \in \mathbb{R}^d.$$

Using the symmetry of the matrix $\Sigma^{1/2}$, we obtain

$$f_Y(y) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}\langle y-\mu, \Sigma^{-1}(y-\mu)\rangle\right), \quad y \in \mathbb{R}^d,$$

which is the desired formula.

For general random vectors, independence of components always implies that they are uncorrelated, but the converse need not hold. A distinctive feature of the multivariate Gaussian distribution is that, in this case, uncorrelated components are also independent.

Theorem 5.1. *Let $Y = (Y_1, \dots, Y_d)^\top$ be a random vector with a multivariate Gaussian distribution. Then the following properties are equivalent:*

1. *The random variables $Y_1, \dots, Y_d$ are independent.*

2. $\mathrm{Cov}(Y_i, Y_j) = 0$ *for all $i \neq j$.*

Proof. It is known that (1) implies (2) even without the multivariate Gaussian assumption. We prove that (2) implies (1). Assume that $\mathrm{Cov}(Y_i, Y_j) = 0$ for all $i \neq j$. Then the components Y_k are Gaussian, say $Y_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$. Since the variables are uncorrelated, the covariance matrix of Y is diagonal, and the expectation vector is

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix}.$$

The characteristic function of Y is

$$\varphi_{Y_1,\dots,Y_d}(t_1, \dots, t_d) = \exp\left(i\langle \mu, t\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\right) = \exp\left(i\sum_{k=1}^d \mu_k t_k - \frac{1}{2}\sum_{k=1}^d \sigma_k^2 t_k^2\right) = \prod_{k=1}^d \exp\left(i\mu_k t_k - \frac{1}{2}\sigma_k^2 t_k^2\right)$$

$$= \prod_{k=1}^{d} \varphi_{Y_k}(t_k),$$

which is the product of the characteristic functions of the components Y_k. Hence, the Y_k are independent. □

Recall that two random vectors $X = (X_1, \dots, X_n)^\top$ and $Y = (Y_1, \dots, Y_m)^\top$ defined on a common probability space are called *independent* if for every Borel sets $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$, we have

$$P(X \in A,\ Y \in B) = P(X \in A) \cdot P(Y \in B).$$

Exercise 5.3. *Let $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ be a Gaussian random vector. Show that the random vectors $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_m)$ are independent if and only if*

$$\operatorname{Cov}(X_i, Y_j) = 0 \quad \textit{for all } i = 1, \dots, n \textit{ and } j = 1, \dots, m.$$

Solution

Write the joint vector as

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \quad \text{with} \quad X = (X_1, \dots, X_n)^\top, \qquad Y = (Y_1, \dots, Y_m)^\top.$$

Since Z is Gaussian, it is completely determined by its mean vector $\mu = \mathbb{E}[Z]$ and its covariance matrix

$$\Sigma = \operatorname{Cov}(Z) = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix},$$

where $\Sigma_{XY} = \operatorname{Cov}(X, Y)$ and $\Sigma_{YX} = \Sigma_{XY}^\top$.

(Only if.) If X and Y are independent then for every i, j, $\operatorname{Cov}(X_i, Y_j) = 0$. This is a standard property: independence implies zero covariance (when covariances exist).

(If.) Conversely, assume $\operatorname{Cov}(X_i, Y_j) = 0$ for all i, j, i.e. $\Sigma_{XY} = 0$ (the zero matrix). Because Z is multivariate Gaussian, its joint probability density (when a density exists) is

$$f_Z(z) = \frac{1}{(2\pi)^{(n+m)/2} \det(\Sigma)^{1/2}} \exp\Big(-\tfrac{1}{2}(z-\mu)^\top \Sigma^{-1} (z-\mu) \Big), \qquad z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

When $\Sigma_{XY} = 0$ the covariance matrix Σ is block-diagonal:

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{pmatrix}.$$

A block-diagonal covariance matrix has determinant $\det(\Sigma) = \det(\Sigma_{XX}) \det(\Sigma_{YY})$ and inverse

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{XX}^{-1} & 0 \\ 0 & \Sigma_{YY}^{-1} \end{pmatrix}.$$

Thus the quadratic form splits:

$$(z-\mu)^\top \Sigma^{-1} (z-\mu) = (x-\mu_X)^\top \Sigma_{XX}^{-1} (x-\mu_X) + (y-\mu_Y)^\top \Sigma_{YY}^{-1} (y-\mu_Y),$$

and the joint density factorizes as

$$f_Z(x, y) = f_X(x)\, f_Y(y),$$

where f_X and f_Y are the marginal Gaussian densities of X and Y (respectively). Factorization of the joint density into a product of the marginals implies that X and Y are independent.

Equivalently (and often more conceptually), one may argue using characteristic functions. The characteristic function of the Gaussian vector Z is

$$\varphi_Z(t,s) = \exp\left(it^\top \mu_X + is^\top \mu_Y - \tfrac{1}{2}\begin{pmatrix} t \\ s \end{pmatrix}^\top \Sigma \begin{pmatrix} t \\ s \end{pmatrix}\right).$$

If $\Sigma_{XY} = 0$ the exponent splits into a sum depending only on t and only on s, so $\varphi_Z(t,s) = \varphi_X(t)\varphi_Y(s)$; the factorization of the characteristic function implies independence of X and Y. Therefore for Gaussian vectors the condition $\mathrm{Cov}(X_i, Y_j) = 0$ for all i, j is equivalent to independence of X and Y.

Brownian motion as a gaussian process

A stochastic process is called Gaussian if its finite-dimensional distributions are multivariate Gaussian. More precisely:

Definition 5.4. *A stochastic process $\{X(t) : t \in T\}$ is called* Gaussian *if for every $n \in \mathbb{N}$ and every $t_1, \ldots, t_n \in T$, the random vector $(X(t_1), \ldots, X(t_n))^\top$ is an n-dimensional Gaussian vector.*

Example 5.3. *Let us show that the Brownian motion is a Gaussian process. Take some $0 \le t_1 \le t_2 \le \cdots \le t_n$. We show that the vector $(B(t_1), \ldots, B(t_n))$ is Gaussian. Consider the random variables*

$$\Delta_i = B(t_i) - B(t_{i-1}),$$

where we define $t_0 := 0$ and $B(0) = 0$. By the definition of Brownian motion, the increments Δ_i are independent and each has a Gaussian distribution. So, it follows that the vector $(\Delta_1, \ldots, \Delta_n)$ is n-dimensional Gaussian. We can represent $(B(t_1), \ldots, B(t_n))$ as a linear transformation of $(\Delta_1, \ldots, \Delta_n)$:

$$B(t_i) = \Delta_1 + \cdots + \Delta_i.$$

Then, the vector $(B(t_1), \ldots, B(t_n))$ is also n-dimensional Gaussian.

Remark 5.3. *The finite-dimensional distributions of a Gaussian process are uniquely determined by the* expectation function

$$\mu(t) = E(X(t))$$

and the covariance function

$$\Gamma(t_1, t_2) = \mathrm{Cov}(X(t_1), X(t_2)).$$

Example 5.4. *If B is a Brownian motion, then*

$$E(B(t)) = 0, \quad \Gamma(t_1, t_2) = \min(t_1, t_2).$$

Conversely, we have the following characterization of Brownian motion.

Theorem 5.2. *A stochastic process $\{B(t) : t \ge 0\}$ is a Brownian motion if and only if:*

1. *B is a Gaussian process;*
2. *$E[B(t)] = 0$ for all $t \geq 0$;*
3. *$\mathrm{Cov}(B(t_1), B(t_2)) = \min(t_1, t_2)$ for all $t_1, t_2 \geq 0$;*
4. *B has continuous sample paths.*

Proof. It is straightforward to verify that these four properties are equivalent to the usual defining properties of Brownian motion. □

The following result is known as the **weak Markov property** of Brownian motion.

Theorem 5.3. *Let $\{B(t) : t \geq 0\}$ be a Brownian motion, and fix $u \geq 0$. Then:*

1. *The process $B_u(s) := B(u+s) - B(u)$, for $s \geq 0$, is itself a Brownian motion.*
2. *The processes $\{B(t) : 0 \leq t \leq u\}$ and $\{B_u(s) : s \geq 0\}$ are independent.*

Proof. The process B_u is Gaussian. Indeed, for any finite collection $s_1, \dots, s_n \geq 0$, the random vector

$$(B_u(s_1), \dots, B_u(s_n))$$

is a linear transformation of the Gaussian vector $(B(u+s_1), \dots, B(u+s_n), B(u))$, hence Gaussian. Since B has continuous sample paths, the same holds for B_u.

Next, we compute expectation and covariance. Clearly,

$$E[B_u(s)] = E[B(u+s) - B(u)] = 0.$$

For the covariance, we have

$$\begin{aligned}\mathrm{Cov}(B_u(s_1), B_u(s_2)) &= \mathrm{Cov}(B(u+s_1) - B(u),\, B(u+s_2) - B(u)) \\ &= \min(u+s_1, u+s_2) - u \\ &= \min(s_1, s_2).\end{aligned}$$

Thus B_u is a Brownian motion.

Finally, to prove independence of $\{B(t) : 0 \leq t \leq u\}$ and $\{B_u(s) : s \geq 0\}$, we recall that for Gaussian processes, independence is equivalent to vanishing covariance. By direct computation, every $B(t)$ with $t \leq u$ is uncorrelated with every $B_u(s)$, which establishes independence. We know that Two stochastic processes $\{X(t) : t \in T\}$ and $\{Y(s) : s \in S\}$ defined on the same probability space are called *independent* if for all $t_1, \dots, t_n \in T$ and $s_1, \dots, s_m \in S$, the vector $(X(t_1), \dots, X(t_n))$ is independent of $(Y(s_1), \dots, Y(s_m))$. To show that $\{B(t) : 0 \leq t \leq u\}$ and $\{B_u(s) : s \geq 0\}$ are independent, it suffices to show that there is no correlation between them. Take some $0 \leq t \leq u$ and $s \geq 0$. Then:

$$\mathrm{Cov}(B(t), B_u(s)) = \mathrm{Cov}(B(t), B(u+s) - B(u)) = \mathrm{Cov}(B(t), B(u+s)) - \mathrm{Cov}(B(t), B(u)) = t - t = 0.$$

This proves the independence. □

The next theorem states the self-similarity property of the Brownian motion.

Theorem 5.4. *Let $\{B(t) : t \geq 0\}$ be a Brownian motion and let $a > 0$. Then the process*

$$\left\{ \frac{1}{\sqrt{a}} B(at) : t \geq 0 \right\}$$

is again a Brownian motion.

Proof. Exercise. □

5.4 Lévy's Construction of the Brownian motion

Theorem 5.5. *[] The Brownian motion exists. Concretely: it is possible to construct a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a stochastic process* $\{B(t) : t \geq 0\}$ *on this probability space such that:*

1. $B(0) = 0$.
2. *B has independent increments.*
3. $B(t+h) - B(t) \sim \mathcal{N}(0, h)$ *for all* $t, h \geq 0$.
4. *For every* $\omega \in \Omega$*, the function* $t \mapsto B(t, \omega)$ *is continuous in* t.

5.4.1 Brownian motion as the limit of a symmetric random walk

Here, we introduce a construction of Brownian motion from a symmetric random walk. Divide the half-line $[0, \infty)$ into tiny subintervals of length δ, as shown in Figure 5.4.1.

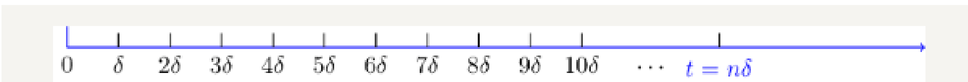

Figure 5.2: Dividing the half-line $[0, \infty)$ into tiny subintervals of length δ.

Each subinterval corresponds to a time slot of length δ. Thus, the intervals are

$$(0, \delta], \quad (\delta, 2\delta], \quad (2\delta, 3\delta], \quad \ldots$$

More generally, the k-th interval is $((k-1)\delta, k\delta]$.

We assume that in each time slot, we toss a fair coin. We define the random variables X_i as follows:

$$X_i = \begin{cases} \sqrt{\delta}, & \text{with probability } \frac{1}{2}, \\ -\sqrt{\delta}, & \text{with probability } \frac{1}{2}. \end{cases}$$

Moreover, the X_i's are independent. Note that

$$\mathbb{E}[X_i] = 0, \qquad \operatorname{Var}(X_i) = \delta.$$

Now, we define the process $W(t)$ as follows. We let $W(0) = 0$. At time $t = n\delta$, the value of $W(t)$ is given by

$$W(t) = W(n\delta) = \sum_{i=1}^{n} X_i.$$

Since $W(t)$ is the sum of n i.i.d. random variables, we know how to find $\mathbb{E}[W(t)]$ and $\operatorname{Var}(W(t))$. In particular,

$$\mathbb{E}[W(t)] = \sum_{i=1}^{n} \mathbb{E}[X_i] = 0, \qquad \operatorname{Var}(W(t)) = \sum_{i=1}^{n} \operatorname{Var}(X_i) = n\delta = t.$$

For any $t \in (0, \infty)$, as $n \to \infty$, $\delta \to 0$. By the Central Limit Theorem, $W(t)$ will become a normal random variable:

$$W(t) \sim \mathcal{N}(0, t).$$

Since the coin tosses are independent, we conclude that $W(t)$ has independent increments. That is, for all

$$0 \le t_1 < t_2 < t_3 < \cdots < t_n,$$

the random variables

$$W(t_2) - W(t_1), \quad W(t_3) - W(t_2), \quad \ldots, \quad W(t_n) - W(t_{n-1})$$

are independent.

Stationary increments.

We say that a random process $X(t)$ has *stationary increments* if, for all $t_2 > t_1 \ge 0$ and all $r > 0$, the two random variables

$$X(t_2) - X(t_1) \quad \text{and} \quad X(t_2 + r) - X(t_1 + r)$$

have the same distribution. In other words, the distribution of the difference depends only on the length of the interval $(t_1, t_2]$, and not on the exact location of the interval on the real line.

We now claim that the random process $W(t)$, defined above, has stationary increments. To see this, we argue as follows. For $0 \le t_1 < t_2$, if $t_1 = n_1\delta$ and $t_2 = n_2\delta$, we obtain

$$W(t_1) = W(n_1\delta) = \sum_{i=1}^{n_1} X_i, \qquad W(t_2) = W(n_2\delta) = \sum_{i=1}^{n_2} X_i.$$

Then, we can write

$$W(t_2) - W(t_1) = \sum_{i=n_1+1}^{n_2} X_i.$$

Therefore, we conclude

$$\mathbb{E}[W(t_2) - W(t_1)] = \sum_{i=n_1+1}^{n_2} \mathbb{E}[X_i] = 0,$$

$$\operatorname{Var}(W(t_2) - W(t_1)) = \sum_{i=n_1+1}^{n_2} \operatorname{Var}(X_i) = (n_2 - n_1)\delta = t_2 - t_1.$$

Thus, for any $0 \le t_1 < t_2$, the distribution of $W(t_2) - W(t_1)$ only depends on the length of the interval $[t_1, t_2]$, i.e., how many coin tosses are in that interval. In particular,

$$W(t_2) - W(t_1) \implies \mathcal{N}(0, t_2 - t_1),$$

as $\delta \to 0$. Hence, $W(t)$ has stationary increments. The above construction can be made more rigorous. The random process $W(t)$ is called the *standard Brownian motion* or the *standard Wiener process*. Brownian motion has continuous sample paths, i.e., $W(t)$ is a continuous function of t. However, it can be shown that it is nowhere differentiable (see Figure 5.2), this property can be proven later.

Theorem 5.6 (Kolmogorov's continuity theorem)**.** *Suppose that the process $\{X_t\}_{t\ge 0}$ satisfies the following condition: for all $T > 0$, there exist positive constants α, β, D such that*

$$\mathbb{E}\Big[\|X_t - X_s\|^{\alpha}\Big] \le D \cdot \|t - s\|^{1+\beta}, \qquad 0 \le s, t \le T. \tag{5.1}$$

Then there exists a continuous version of X.

Proposition 5.1 (Continuous version of Brownian motion)**.** *Brownian motion satisfies Kolmogorov's condition* (5.1) *with* $\alpha = 4$*,* $D = n(n+2)$ *and* $\beta = 1$*. Therefore, the Brownian motion has a continuous version.*

Proof. Recall that if $Z \sim \mathcal{N}(0, \sigma^2)$, then for $k \in \mathbb{N}$,

$$\mathbb{E}[Z^k] = \mu_k = \begin{cases} 0, & \text{if } k \text{ is odd,} \\ (k-1)!! \cdot \sigma^k, & \text{if } k \text{ is even.} \end{cases}$$

For Brownian motion, we have

$$B_t^{(i)} \sim \mathcal{N}(0, t) \quad \forall i, \qquad B_t^{(i)} - B_s^{(i)} \sim \mathcal{N}\Big(0,\, t-s\Big) \quad \forall i.$$

Hence,

$$\mathbb{E}\Big[\|B_t - B_s\|^4\Big] = \sum_{i=1}^n \mathbb{E}\Big[(B_t^{(i)} - B_s^{(i)})^4\Big] + \sum_{\substack{i,j=1 \\ j \neq i}}^n \mathbb{E}\Big[(B_t^{(i)} - B_s^{(i)})^2 (B_t^{(j)} - B_s^{(j)})^2\Big].$$

For the first term, using the Gaussian moment formula,

$$\mathbb{E}\Big[(B_t^{(i)} - B_s^{(i)})^4\Big] = \frac{4!}{2!\,2^2}(t-s)^2 = 3(t-s)^2.$$

Thus,

$$\sum_{i=1}^n \mathbb{E}\Big[(B_t^{(i)} - B_s^{(i)})^4\Big] = n \cdot 3(t-s)^2.$$

For the cross terms,

$$\mathbb{E}\Big[(B_t^{(i)} - B_s^{(i)})^2 (B_t^{(j)} - B_s^{(j)})^2\Big] = (t-s)^2, \qquad i \neq j.$$

Therefore,

$$\sum_{\substack{i,j=1 \\ j \neq i}}^n \mathbb{E}\Big[(B_t^{(i)} - B_s^{(i)})^2 (B_t^{(j)} - B_s^{(j)})^2\Big] = n(n-1)(t-s)^2.$$

Putting everything together,

$$\mathbb{E}\Big[\|B_t - B_s\|^4\Big] = n \cdot 3(t-s)^2 + n(n-1)(t-s)^2 = n(n+2)(t-s)^2.$$

This shows that Brownian motion satisfies condition (5.1) with $\alpha = 4$, $\beta = 1$, and $D = n(n+2)$. Hence, Brownian motion admits a continuous version. □

We conclude that Brownian motion admits a continuous version, and thus our construction indeed satisfies the definition provided above. Another important regularity property, however, which Brownian motion fails to possess, is bounded variation. Before addressing this, we first recall the definition of a function of bounded variation.

Definition 5.5. *A right-continuous function* $f : (0, t) \to \mathbf{R}$ *is a function of bounded variation if*

$$V_f^{(1)}(t) := \sup_{k \in \mathbb{N}} \sup_{0 = t_0 \leq t_1 \leq \cdots \leq t_k = t} \sum_{j=1}^k |f(t_j) - f(t_{j-1})| < \infty,$$

where the supremum is taken over all partitions of the interval $(0, t)$*. If the supremum is infinite,* f *is said to be of unbounded variation.*

Remark 5.4. *It is not hard to show that f is of bounded variation if and only if it can be written as the difference of two increasing functions.*

Proposition 5.2 (Brownian Motion is of unbounded variation)**.** *The Brownian Motion is almost surely of unbounded variation (for all $t > 0$).*

Proof. Without loss of generality, let $t = 1$ (for $t \neq 1$ use the scaling properties of the Brownian Motion). Let $(X_t)_{t\geq 0}$ be a Brownian Motion, then define

$$Z_n = \sum_{k=1}^{2^n} \left| X_{k2^{-n}} - X_{(k-1)2^{-n}} \right| = \sqrt{2^n} \frac{1}{2^n} \sum_{k=1}^{2^n} \left| \sqrt{2^n} \Big(X_{k2^{-n}} - X_{(k-1)2^{-n}} \Big) \right|.$$

By the law of large numbers, as $n \to \infty$,

$$\frac{1}{2^n} \sum_{k=1}^{2^n} \left| \sqrt{2^n} \Big(X_{k2^{-n}} - X_{(k-1)2^{-n}} \Big) \right| \xrightarrow{\mathbb{P}} \mathbb{E}[|X_1|].$$

Hence $Z_n \xrightarrow{\mathbb{P}} \infty$. Because of the triangle inequality, the random variables Z_n are monotonically increasing; thus $Z_n \to \infty$ almost surely. □

Proposition 5.3 (Brownian Motion is of finite quadratic variation)**.** *Let $\pi^m : 0 = t_0^m < t_1^m < \cdots < t_m^m = t$ ($m \in \mathbb{N}$) be a partition of $[0, t]$ whose mesh size converges to zero. Then*

$$\lim_{m\to\infty} \sum_{k=1}^{m} \Big(X_{t_k^m} - X_{t_{k-1}^m} \Big)^2 = t, \quad \textit{in probability.}$$

Proof. Define

$$Z_m = \sum_{k=1}^{m} \Big(X_{t_k^m} - X_{t_{k-1}^m} \Big)^2.$$

We compute

$$\mathbb{E}[Z_m] = \sum_{k=1}^{m} \mathbb{E}\Big[\Big(X_{t_k^m} - X_{t_{k-1}^m} \Big)^2 \Big] = \sum_{k=1}^{m} (t_k^m - t_{k-1}^m) = t.$$

Next,

$$\mathrm{Var}(Z_m) = \sum_{k=1}^{m} \mathrm{Var}\Big(\Big(X_{t_k^m} - X_{t_{k-1}^m} \Big)^2 \Big).$$

Since $X_{t_k^m} - X_{t_{k-1}^m} \sim \mathcal{N}(0, t_k^m - t_{k-1}^m)$,

$$\mathrm{Var}\Big(\Big(X_{t_k^m} - X_{t_{k-1}^m} \Big)^2 \Big) = (t_k^m - t_{k-1}^m)^2 \, \mathrm{Var}(X_1^2),$$

where X_1 is a standard gaussian random variable. Thus

$$\mathrm{Var}(Z_m) \leq \left(\sup_{k=1,\ldots,m} (t_k^m - t_{k-1}^m) \right) \sum_{k=1}^{m} (t_k^m - t_{k-1}^m) \, \mathrm{Var}(X_1^2) = \mathrm{Var}(X_1^2)\, t \cdot \sup_{k=1,\ldots,m} (t_k^m - t_{k-1}^m).$$

As $m \to \infty$, $\sup_k (t_k^m - t_{k-1}^m) \to 0$. Hence $\mathrm{Var}(Z_m) \to 0$, and by Chebyshev's inequality we obtain $Z_m \to t$ in probability. □

5.4.2 Non-differentiability of Brownian motion paths

Theorem 5.7. *(Paley, Wiener, Zygmund) Let $\{B(t) : t \geq 0\}$ be a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, with probability 1, the function $t \mapsto B(t)$ is nowhere differentiable.*

Concretely: There is a measurable set $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$ and for all $t_0 \geq 0$, the function $t \mapsto B(t; \omega)$ has no derivative at t_0.

Remark 5.5. *We will prove even more. For a function $f : \mathbb{R} \to \mathbb{R}$, define*

$$D^+ f(t) := \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \quad \textit{(upper right derivative)},$$

$$D^- f(t) := \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \quad \textit{(lower right derivative)}.$$

If $D^+ f(t) = D^- f(t) \in \mathbb{R}$, then we say that f is differentiable from the right. In a similar way, one can define the upper and lower left derivatives. Consider the set

$$A := \Big\{\omega \in \Omega : \exists t_0 \in (0,1) \text{ such that } -\infty < D^- B(t_0; \omega) \leq D^+ B(t_0; \omega) < \infty \Big\}.$$

We would like to show that $P(A) = 0$, i.e., for almost every sample path of the Brownian motion and every $t_0 \geq 0$, we have $D^+ B(t_0) = +\infty$, or $D^- B(t_0) = -\infty$, or both. However, it is not immediately clear whether the set A is measurable. Therefore, we will prove a slightly weaker statement: There is a measurable set A_0 with $P(A_0) = 0$ such that $A \subset A_0$.

Proof. We have $A \subset \bigcup_{M \in \mathbb{N}} A_M$, where

$$A_M := \left\{\omega \in \Omega : \exists t_0 \in (0,1) \text{ such that } \sup_{h \in (0,1)} \frac{B(t_0 + h; \omega) - B(t_0; \omega)}{h} \leq M \right\}.$$

Fix some $M \in \mathbf{N}$. We show that $P(A_M) = 0$. Take some $n \in \mathbf{N}$, $n \geq 3$. Any $t_0 \in (0,1)$ must lie in some interval $(k2^{-n-1}, k2^{-n})$, for $k = 1, \ldots, 2^n$. If the event A_M occurs and $t_0 \in (k2^{-n-1}, k2^{-n})$, then the following three events occur:

1. $F_{n,k}^{(1)} : |B(k2^{-n+1}) - B(k2^{-n})| \leq |B(k2^{-n+1}) - B(t_0)| + |B(t_0) - B(k2^{-n})| \leq \frac{3}{2^n} M,$
2. $F_{n,k}^{(2)} : |B(k2^{-n+2}) - B(k2^{-n+1})| \leq \frac{5}{2^n} M,$
3. $F_{n,k}^{(3)} : |B(k2^{-n+3}) - B(k2^{-n+2})| \leq \frac{7}{2^n} M.$

Consider the event $F_{n,k} := F_{n,k}^{(1)} \cap F_{n,k}^{(2)} \cap F_{n,k}^{(3)}$. Then, for every $n \geq 3$, we have

$$A_M \subset \bigcup_{k=1}^{2^n} F_{n,k}.$$

We estimate the probabilities. For instance, for $F_{n,k}^{(3)}$, we compute

$$P(F_{n,k}^{(3)}) = P\left(\left|B\left(k2^{-n+3}\right) - B\left(k2^{-n+2}\right)\right| \leq \frac{7}{2^n} M\right) = \mathbb{P}\left(|N| \leq \frac{7M}{\sqrt{2^n}}\right),$$

where $N \sim \mathcal{N}(0,1)$. Denoting by $f_N(t)$ the standard Gaussian density (which is less than $1/\sqrt{2\pi} < 1/2$), we have

$$P(F^{(3)}_{n,k}) = \int_{-\frac{7M}{\sqrt{2^n}}}^{\frac{7M}{\sqrt{2^n}}} f_N(t)dt \leq \frac{14M}{\sqrt{2^n}}.$$

Similarly, we obtain

$$P(F^{(1)}_{n,k}) \leq \frac{7M}{\sqrt{2^n}}, \quad P(F^{(2)}_{n,k}) \leq \frac{27M}{\sqrt{2^n}}.$$

□

Since the events $F^{(1)}_{n,k}$, $F^{(2)}_{n,k}$, and $F^{(3)}_{n,k}$ are independent (by the independence of increments of the Brownian motion), we have

$$P(F_{n,k}) = P(F^{(1)}_{n,k}) \cdot P(F^{(2)}_{n,k}) \cdot P(F^{(3)}_{n,k}) \leq \left(\frac{7M}{\sqrt{2^n}}\right)^3 = \frac{(7M)^3}{2^{3n/2}}.$$

It follows that

$$P(A_M) \leq P(\left(\bigcup_{k=1}^{2^n} F_{n,k}\right) \leq \sum_{k=1}^{2^n} P(F_{n,k}) \leq 2^n \cdot \frac{(7M)^3}{2^{3n/2}} = \frac{(7M)^3}{2^{n/2}}.$$

Since this bound tends to zero as $n \to \infty$, we conclude that $P(A_M) = 0$. Hence, the set

$$A_0 := \bigcup_{M \in \mathbb{N}} A_M$$

satisfies $P(A_0) = 0$. We can now take $\Omega_0 := \Omega \setminus A_0$.

5.5 The Markov property and Blumenthal's law

For the study of the Markov property, it is useful to extend the discussion to higher-dimensional Brownian motion. This can be defined naturally by requiring that each component is a one-dimensional Brownian motion and that the components are mutually independent.

Definition 5.6. *If $B_1, \dots, B_d$ are independent linear Brownian motions started in $x_1, \dots, x_d$, then the stochastic process $\{B(t) : t \geq 0\}$ given by*

$$B(t) = \Big(B_1(t), \dots, B_d(t)\Big)^T$$

is called a d-dimensional Brownian motion started in $\Big(x_1, \dots, x_d\Big)^T$. The d-dimensional Brownian motion started at the origin is also called standard Brownian motion. *Two-dimensional Brownian motion is also called* planar Brownian motion.

Let $\mathbb{P}_x$ denote the probability measure under which the d-dimensional process $\{B(t) : t \geq 0\}$ is a Brownian motion starting at $x \in \mathbb{R}^d$, and let $\mathbb{E}_x$ denote the corresponding expectation. Suppose that $\{X(t) : t \geq 0\}$ is a stochastic process. Intuitively, the *Markov property* states that if we know the trajectory of the process on the interval $(0, s)$, then for predicting its future evolution $\{X(t) : t \geq s\}$ it is sufficient to know only the present state $X(s)$. In other words, the process "forgets" its past once the current position is known.

Definition 5.7 (Time-homogeneous Markov process)**.** *A process* $\{X(t) : t \geq 0\}$ *is called a* (time-homogeneous) Markov process *if, at any fixed time* $s \geq 0$*, it starts afresh from* $X(s)$*. More precisely, assuming the process can be started at any initial point* $X(0) = x \in \mathbb{R}^d$*, the time-shifted process* $\{X(s+t) : t \geq 0\}$ *has the same distribution as the process started at* $X(s) \in \mathbb{R}^d$*.*

Remark 5.6. *We shall give a formal definition of Markov processes later in this chapter, but begin by presenting a direct formulation of these ideas in the special case of Brownian motion.*

Definition 5.8 (Independence of processes)**.** *Two stochastic processes* $\{X(t) : t \geq 0\}$ *and* $\{Y(t) : t \geq 0\}$ *are said to be* independent *if, for any finite sets of times* $t_1, \ldots, t_n \geq 0$ *and* $s_1, \ldots, s_m \geq 0$*, the random vectors*

$$\Big(X(t_1), \ldots, X(t_n)\Big) \quad \textit{and} \quad \Big(Y(s_1), \ldots, Y(s_m)\Big)$$

are independent.

Example 5.5. *For a one-dimensional Brownian motion* $\{B(t) : t \geq 0\}$*, we have*

$$E\Big(\, B(t) \,\Big|\, \mathcal{F}^+(s)\Big) = E\Big(\, B(t) - B(s) \,\Big|\, \mathcal{F}^+(s)\Big) + B(s) = E\Big(\, B(t) - B(s)\Big) + B(s) = B(s),$$

hence, Brownian motion is a martingale.

Proposition 5.4. *(Doob's maximal inequality) Suppose* $\{X(t) : t \geq 0\}$ *is a continuous submartingale and* $p > 1$*. Then, for any* $t \geq 0$*,*

$$E\Big(\sup_{0 \leq s \leq t} |X(s)|^p\Big) \;\leq\; \Big(\frac{p}{p-1}\Big)^p \,\mathbb{E}\Big(\,|X(t)|^p\Big).$$

Lemma 5.7. *Suppose* $\{B(t) : t \geq 0\}$ *is a linear Brownian motion. Then the process*

$$\Big\{\, B(t)^2 - t : t \geq 0\Big\}$$

is a martingale.

Proof. The process is adapted to the natural filtration of Brownian motion, and

$$E\Big(\, B(t)^2 - t \,\Big|\, \mathcal{F}^+(s)\Big) = E\Big(\, (B(t) - B(s))^2 \,\Big|\, \mathcal{F}^+(s)\Big) + 2\,E\Big(\, B(t)B(s) \,\Big|\, \mathcal{F}^+(s)\Big) - B(s)^2 - t.$$

Using independence and properties of Brownian increments:

$$E\Big(\, (B(t) - B(s))^2 \,\Big|\, \mathcal{F}^+(s)\Big) = t - s, \quad E\Big(B(t) \,\Big|\, \mathcal{F}^+(s)\Big) = B(s),$$

and so

$$E\Big(B(t)B(s) \,\Big|\, \mathcal{F}^+(s)\Big) = B(s)\,E\Big(B(t) \,\Big|\, \mathcal{F}^+(s)\Big) = B(s)^2.$$

Therefore,

$$E\Big(\, B(t)^2 - t \,\Big|\, \mathcal{F}^+(s)\Big) = (t - s) \;+\; 2\,B(s)^2 \;-\; B(s)^2 \;-\; t = B(s)^2 - s$$

□

5.6 Exercises

Exercise 5.4. *Verify that with respect to the canonical filtration:*

- *the Brownian motion,*
- *the process* $\tilde{N} = (\tilde{N}_t)_{t\geq 0}$ *defined by* $\tilde{N}_t = N_t - \lambda t$*, where* $N = (N_t)_{t\geq 0}$ *is a Poisson process with intensity* λ*,*

are martingales with respect to their canonical filtration.

Exercise 5.5. *Assume* $(B_t)_{t\geq 0}$ *is a Brownian motion. Then each of the following processes is also a Brownian motion:*

1. $(-B_t)_{t\geq 0}$ *(symmetry by reflection);*
2. $(B_{t+s} - B_s)_{t\geq 0}$*, for fixed* $s \geq 0$ *(translation of the origin);*
3. $\left(cB_{t/c^2}\right)_{t\geq 0}$*, for fixed* $c > 0$ *(scaling transform);*
4. $\left(tB_{1/t}\right)_{t\geq 0}$ *with the convention that it is zero at* $t = 0$ *(time reciprocal);*
5. *For fixed* $u > 0$*,* $(B_u - B_{u-t})_{0\leq t\leq u}$ *(time reversal).*

Solution
We prove only (4). Define $X_t = tB_{1/t}$*, for* $t > 0$*.*

1. *Since* $B_t \sim \mathcal{N}(0, t)$*, by the scaling property of the normal distribution, we have*
$$X_t \sim \mathcal{N}\Big(0,\, t^2 \cdot \tfrac{1}{t}\Big) = \mathcal{N}(0, t).$$
2. *Moreover, if* $s \leq t$*, then*
$$\mathbb{E}[B_s B_t] = \mathbb{E}\Big[B_s(B_t - B_s + B_s)\Big] = \mathbb{E}[B_s^2] = s.$$
Thus, $\mathbb{E}[B_s B_t] = s \wedge t$*. Similarly,*
$$\mathbb{E}[X_s X_t] = st\, \mathbb{E}[B_{1/s} B_{1/t}] = s \wedge t.$$
Therefore, for $0 < t_1 < t_2$*,*
$$\mathbb{E}[(X_{t_2} - X_{t_1})^2] = t_2 - t_1,$$
so $X_{t_2} - X_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$*.*
3. *On the other hand, for* $0 < t_1 < t_2 < t_3 < t_4$*, by independence of Brownian increments,*
$$\mathbb{E}[(X_{t_2} - X_{t_1})(X_{t_4} - X_{t_3})] = 0.$$
Thus the process has independent increments. Since Gaussian random vectors are determined by their covariance, $(X_t)_{t>0}$ *is a Gaussian process with independent increments.*

4. *Finally, each trajectory of (X_t) is continuous on $(0,\infty)$. At $t=0$, by a standard property of Brownian motion, one can choose a separable dense subset $D \subset \mathbb{R}_+$ with 0 as an accumulation point such that*

$$\lim_{\substack{t\in D\\ t\downarrow 0}} X_t = 0, \quad a.s.$$

Thus, by defining $X_0 := 0$, we extend (X_t) continuously at $t=0$.

Hence, $(X_t)_{t\geq 0}$ is a Brownian motion.

Exercise 5.6. 1. *Recall the transition density (Gaussian kernel) of one-dimensional Brownian motion:*

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} \exp\Big(-\frac{(x-y)^2}{2t}\Big), \qquad t>0,\ x,y\in\mathbb{R}.$$

Show that

$$\frac{\partial}{\partial t} p(t,x,y) = \frac{1}{2}\frac{\partial^2}{\partial x^2} p(t,x,y).$$

Solution
Write $p(t,x,y) = (2\pi t)^{-1/2}\exp\Big(-\frac{(x-y)^2}{2t}\Big)$. Differentiating, we obtain

$$\partial_t p = (2\pi t)^{-1/2} e^{-(x-y)^2/(2t)} \left(-\frac{1}{2t} + \frac{(x-y)^2}{2t^2}\right).$$

On the other hand,

$$\partial_{xx} p = (2\pi t)^{-1/2} e^{-(x-y)^2/(2t)} \left(\frac{(x-y)^2}{t^2} - \frac{1}{t}\right).$$

Thus $\partial_t p = \frac{1}{2}\partial_{xx} p$, as required. Show that, for the dyadic partition of order n on $[0,t]$,

$$\sum_{k=1}^{2^n} \Big(B_{k2^{-n}t} - B_{(k-1)2^{-n}t}\Big)^2 \xrightarrow[n\to\infty]{} t \qquad \text{in } L^2 \text{ and almost surely.}$$

Deduce that Brownian motion has no bounded variation on any finite interval.
Solution
Let

$$Q_n := \sum_{k=1}^{2^n} \Big(B_{k2^{-n}t} - B_{(k-1)2^{-n}t}\Big)^2.$$

The increments are independent and

$$\mathbb{E}\Big[(B_{k2^{-n}t} - B_{(k-1)2^{-n}t})^2\Big] = 2^{-n}t.$$

Hence $\mathbb{E}[Q_n] = t$ for all n.

For the variance, since increments are independent:

$$\mathrm{Var}(Q_n) = \sum_{k=1}^{2^n} \mathrm{Var}\left((\Delta_k^n)^2\right),$$

where $\Delta_k^n \sim \mathcal{N}(0, 2^{-n}t)$. *If* $Z \sim \mathcal{N}(0, \sigma^2)$ *then* $\mathrm{Var}(Z^2) = 2\sigma^4$. *Thus*

$$\mathrm{Var}(Q_n) = 2^n \cdot 2(2^{-n}t)^2 = 2t^2 2^{-n} \to 0.$$

So $Q_n \to t$ *in* L^2, *and also almost surely along dyadics. Therefore the quadratic variation of* B *on* $[0,t]$ *equals* t.

If B *had bounded variation, then the sum of squared increments would vanish as the mesh goes to zero, contradicting the above. Hence* B *has infinite variation on every interval.*

2. *Show that, for any* $t > 0$ *and* $r > 0$,

$$\mathbb{P}\Big(\sup_{0\le s\le t} |B_s| > r\Big) \le \frac{t}{r^2}.$$

Solution
Let $\tau = \inf\{s \le t : |B_s| > r\}$. *Then*

$$r^2 \mathbf{1}_{\{\tau\le t\}} \le |B_{\tau\wedge t}|^2.$$

Taking expectation and using the optional stopping theorem,

$$r^2\mathbb{P}(\tau \le t) \le \mathbb{E}[|B_{\tau\wedge t}|^2] = \mathbb{E}[\tau\wedge t] \le t,$$

which yields the desired inequality.

3. *Prove*

$$\mathbb{E}\Big(\sup_{0\le s\le t} |B_s|^2\Big) \le 4\,\mathbb{E}[B_t^2] = 4t.$$

Solution
This is the classical Doob L^2 *inequality: if* $(M_s)_{0\le s\le t}$ *is an* L^2 *martingale, then*

$$\mathbb{E}\Big(\sup_{0\le s\le t} |M_s|^2\Big) \le 4\,\mathbb{E}[|M_t|^2].$$

Apply it to $M_s = B_s$.

4. *Let* $W_t = \int_0^t B_s\,ds$

   (a) Show $\mathbb{E}[W_t] = 0$ *and* $\mathbb{E}[W_t^2] = t^3/3$.

   (b) Find the conditional distribution of W_t *given* $B_t = x$.

   (c) Prove that $W_t - tB_t$ *is a martingale.*

Solution
(a) Since $\mathbb{E}[B_s] = 0$, *we have* $\mathbb{E}[W_t] = 0$. *For the variance,*

$$\mathbb{E}[W_t^2] = \int_0^t\int_0^t \mathbb{E}[B_sB_u]\,ds\,du = \int_0^t\int_0^t \min(s,u)\,ds\,du = \tfrac{t^3}{3}.$$

(b) For $s \le t$, $\mathbb{E}[B_s \mid B_t = x] = \frac{s}{t}x$. *Hence*

$$\mathbb{E}[W_t \mid B_t = x] = \int_0^t \frac{s}{t}x\,ds = \tfrac{t}{2}x.$$

Also $\mathrm{Cov}(W_t, B_t) = \int_0^t s\,ds = \frac{t^2}{2}$ *and* $\mathrm{Var}(W_t) = t^3/3$, $\mathrm{Var}(B_t) = t$. *Thus*

$$\mathrm{Var}(W_t \mid B_t) = \frac{t^3}{3} - \frac{(t^2/2)^2}{t} = \tfrac{t^3}{12}.$$

Therefore

$$W_t \mid (B_t = x) \sim \mathcal{N}\!\left(\tfrac{t}{2}x,\ \tfrac{t^3}{12}\right).$$

(c) Define $M_t = W_t - tB_t$. *Then heuristically,*

$$dW_t = B_t\,dt, \qquad d(tB_t) = B_t\,dt + t\,dB_t,$$

so

$$dM_t = -t\,dB_t.$$

This has no drift term, hence M_t *is a martingale.*

Bibliography

[1] Arnold, L. (1974). *Stochastic Differential Equations: Theory and Applications*. Wiley-Interscience, John Wiley & Sons, New York. Translated from the German.

[2] Ash, R. B. (1970). *Basic Probability Theory*. Wiley, New York.

[3] Asmussen, S., and Glynn, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Volume 57 of Stochastic Modelling and Applied Probability. Springer, New York.

[4] Berestycki, N., Lubetzky, E., Peres, Y., and Sly, A. (2018). Random walks on the random graph. *The Annals of Probability*, 46(1), 456–490.

[5] Brémaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, New York.

[6] Ching, W., Huang, X., Ng, M. K., and Siu, T. K. (2013). *Markov Chains: Models, Algorithms and Applications*. 2nd edition. Springer, Boston, MA.

[7] Durrett, R. (2012). *Essentials of Stochastic Processes*. 2nd edition. Springer.

[8] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*. Vol. I, 3rd edition. John Wiley & Sons, New York.

[9] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. Vol. II, 2nd edition. John Wiley & Sons, New York.

[10] Freedman, D. (1983). *Markov Chains*. Springer, New York. (Reprint of the 1971 Holden-Day edition.)

[11] Friedman, A. (1975). *Stochastic Differential Equations and Applications*. Vol. 1. Academic Press, New York. Probability and Mathematical Statistics, Vol. 28.

[12] Friedman, A. (1976). *Stochastic Differential Equations and Applications*. Vol. 2. Academic Press, New York. Probability and Mathematical Statistics, Vol. 28.

[13] Gard, T. C. (1988). *Introduction to Stochastic Differential Equations*. Volume 114 of Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker Inc., New York.

[14] Gonzàlez Serrano, I. (2024). The Poisson process and some of its applications. Grau de mathematiques, Final project, Universitat de Barcelona.

[15] Kaimanovich, V. A., and Vershik, A. M. (1983). Random walks on discrete groups: boundary and entropy. *The Annals of Probability*, 11(3), 457–490.

[16] Karlin, S., and Taylor, H. M. (1975). *A Second Course in Stochastic Processes*. Academic Press, New York.

[17] Kemeny, J. G., and Snell, J. L. (1960). *Finite Markov Chains*. Van Nostrand, Princeton, N.J.

[18] Kutoyants, Y. A. (2023). *Introduction to the Statistics of Poisson Processes and Applications*. Springer.

[19] Last, G., and Penrose, M. (2017). *Lectures on the Poisson Process*. Version from 21 August 2017, IMS Textbook, Cambridge University Press.

[20] Levin, D. A., and Peres, Y. (2017). *Markov Chains and Mixing Times*. Vol. 107. American Mathematical Society.

[21] Levin, D. A., Peres, Y., and Wilmer, E. L. (2008). *Markov Chains and Mixing Times*. American Mathematical Society. Available at `http://pages.uoregon.edu/dlevin/MARKOV/`

[22] Lockhart, R. (2011). Poisson Processes. Simon Fraser University, STAT 870 — Summer 2011.

[23] Norris, J. R. (1997). *Markov Chains*. Cambridge University Press, Cambridge.

[24] Ross, S. (1996). *Stochastic Processes*. 2nd edition. Wiley, New York.

[25] Sanz, M. (1999). *Probabilitats*. Edicions Universitat de Barcelona.

[26] Schmidt, V. (2006). *Markov Chains and Monte Carlo Simulation*. `http://www.mathematik.uni-ulm.de/stochastik/lehre/ss06/markov/skriptengl/node10.html`. Ulm University, lecture notes (May 2016).

[27] Schneider, R., and Weil, W. (2008). *Stochastic and Integral Geometry*. Springer, Berlin.

[28] Stroock, D. W. (2014). *An Introduction to Markov Process*. 2nd edition. Springer, Berlin.

[29] Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.