STOCHASTIC PROCESSES AND APPLICATIONS

G.A. Pavliotis
Department of Mathematics
Imperial College London

February 2, 2014

Contents

Chapter 1

Introduction

In this chapter we introduce some of the concepts and techniques that we will study in this book. In Section 1.1 we present a brief historical overview on the development of the theory of stochastic processes in the twentieth century. In Section 1.2 we introduce the one-dimensional random walk an we use this example in order to introduce several concepts such Brownian motion, the Markov property. Some comments on the role of probabilistic modeling in the physical sciences are offered in Section 1.3. Discussion and bibliographical comments are presented in Section 1.4. Exercises are included in Section 1.5.

1.1 Historical Overview

The theory of stochastic processes, at least in terms of its application to physics, started with Einstein's work on the theory of Brownian motion: *Concerning the motion, as required by the molecular-kinetic theory of heat, of particles suspended in liquids at rest* (1905) and in a series of additional papers that were published in the period $1905 - 1906$. In these fundamental works, Einstein presented an explanation of Brown's observation (1827) that when suspended in water, small pollen grains are found to be in a very animated and irregular state of motion. In developing his theory Einstein introduced several concepts that still play a fundamental role in the study of stochastic processes and that we will study in this book. Using modern terminology, Einstein introduced a Markov chain model for the motion of the particle (molecule, pollen grain...). Furthermore, he introduced the idea that it makes more sense to talk about the probability of finding the particle at position x at time t, rather than about individual trajectories.

In his work many of the main aspects of the modern theory of stochastic processes can be found:

- The assumption of Markovianity (no memory) expressed through the Chapman-Kolmogorov equation.
- The Fokker–Planck equation (in this case, the diffusion equation).
- The derivation of the Fokker-Planck equation from the master (Chapman-Kolmogorov) equation through a Kramers-Moyal expansion.
- The calculation of a transport coefficient (the diffusion equation) using macroscopic (kinetic theory-based) considerations:
$$D = \frac{k_B T}{6\pi\eta a}.$$
- k_B is Boltzmann's constant, T is the temperature, η is the viscosity of the fluid and a is the diameter of the particle.

Einstein's theory is based on an equation for the probability distribution function, the *Fokker–Planck equation*. Langevin (1908) developed a theory based on a *stochastic differential equation*. The equation of motion for a Brownian particle is

$$m\frac{d^2x}{dt^2} = -6\pi\eta a\frac{dx}{dt} + \xi,$$

where ξ is a random force. It can be shown that there is complete agreement between Einstein's theory and Langevin's theory. The theory of Brownian motion was developed independently by Smoluchowski, who also performed several experiments.

The approaches of Langevin and Einstein represent the two main approaches in the modelling of physical systems using the theory of stochastic processes and, in particular, diffusion processes: either study individual trajectories of Brownian particles. Their evolution is governed by a stochastic differential equation:

$$\frac{dX}{dt} = F(X) + \Sigma(X)\xi(t),$$

where $\xi(t)$ is a random force or study the probability $\rho(x,t)$ of finding a particle at position x at time t. This probability distribution satisfies the Fokker–Planck equation:

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(F(x)\rho) + \frac{1}{2}D^2 : (A(x)\rho),$$

where $A(x) = \Sigma(x)\Sigma(x)^T$. The theory of stochastic processes was developed during the 20th century by several mathematicians and physicists including Smoluchowksi, Planck, Kramers, Chandrasekhar, Wiener, Kolmogorov, Itô, Doob.

1.2 The One-Dimensional Random Walk

We let time be discrete, i.e. $t = 0, 1, \dots$. Consider the following stochastic process S_n: S_0 = 0; at each time step it moves to ± 1 with equal probability $\frac{1}{2}$.

In other words, at each time step we flip a fair coin. If the outcome is heads, we move one unit to the right. If the outcome is tails, we move one unit to the left.

Alternatively, we can think of the random walk as a sum of independent random variables:

$$S_n = \sum_{j=1}^{n} X_j,$$

where $X_j \in \{-1, 1\}$ with $\mathbb{P}(X_j = \pm 1) = \frac{1}{2}$.

We can simulate the random walk on a computer:

- We need a (pseudo)random number generator to generate n independent random variables which are uniformly distributed in the interval [0,1].
- If the value of the random variable is $\geqslant \frac{1}{2}$ then the particle moves to the left, otherwise it moves to the right.
- We then take the sum of all these random moves.
- The sequence $\{S_n\}_{n=1}^{N}$ indexed by the discrete time $T = \{1, 2, \dots N\}$ is the path of the random walk. We use a linear interpolation (i.e. connect the points $\{n, S_n\}$ by straight lines) to generate a continuous path.

Every path of the random walk is different: it depends on the outcome of a sequence of independent random experiments. We can compute statistics by generating a large number of paths and computing averages. For example, $\mathbb{E}(S_n) = 0$, $\mathbb{E}(S_n^2) = n$. The paths of the random walk (without the linear interpolation) are not continuous: the random walk has a jump of size 1 at each time step. This is an example of a discrete time, discrete space stochastic processes. The random walk is a time-homogeneous Markov process. If we take a large number of steps, the random walk starts looking like a continuous time process with continuous paths.

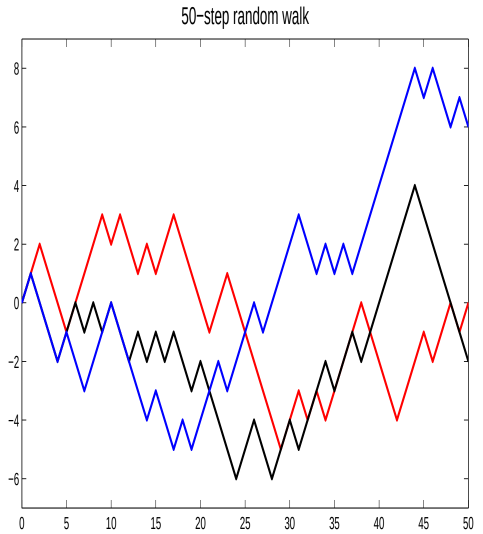

Figure 1.1: Three paths of the random walk of length $N = 50$.

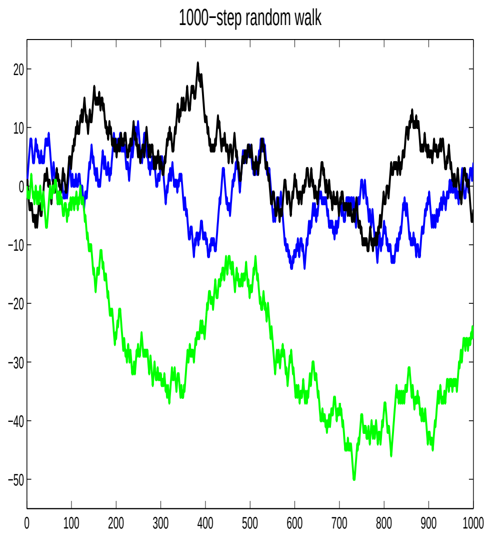

Figure 1.2: Three paths of the random walk of length $N = 1000$.

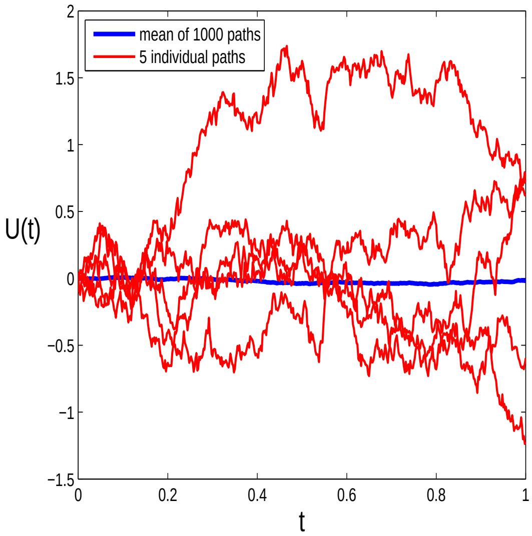

Figure 1.3: Sample Brownian paths.

We can quantify this observation by introducing an appropriate rescaled process and by taking an appropriate limit. Consider the sequence of continuous time stochastic processes

$$Z^n_t := \frac{1}{\sqrt{n}} S_{nt}.$$

In the limit as $n \to \infty$, the sequence $\{Z^n_t\}$ converges (in some appropriate sense, that will be made precise in later chapters) to a Brownian motion with diffusion coefficient $D = \frac{\Delta x^2}{2\Delta t} = \frac{1}{2}$. Brownian motion $W(t)$ is a continuous time stochastic processes with continuous paths that starts at 0 ($W(0) = 0$) and has independent, normally. distributed Gaussian increments. We can simulate the Brownian motion on a computer using a random number generator that generates normally distributed, independent random variables. We can write an equation for the evolution of the paths of a Brownian motion X_t with diffusion coefficient D starting at x:

$$dX_t = \sqrt{2D} dW_t, \quad X_0 = x.$$

This is the simplest example of a stochastic differential equation. The probability of finding X_t at y at time t, given that it was at x at time $t = 0$, the transition probability density $\rho(y,t)$ satisfies the PDE

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial y^2}, \quad \rho(y,0) = \delta(y-x).$$

This is the simplest example of the Fokker–Planck equation. The connection between Brownian motion and the diffusion equation was made by Einstein in 1905.

1.3 Why Randomness

Why introduce randomness in the description of physical systems?

- To describe outcomes of a repeated set of experiments. Think of tossing a coin repeatedly or of throwing a dice.
- To describe a deterministic system for which we have incomplete information: we have imprecise knowledge of initial and boundary conditions or of model parameters.
  - ODEs with random initial conditions are equivalent to stochastic processes that can be described using stochastic differential equations.
- To describe systems for which we are not confident about the validity of our mathematical model.
- To describe a dynamical system exhibiting very complicated behavior (chaotic dynamical systems). Determinism versus predictability.
- To describe a high dimensional deterministic system using a simpler, low dimensional stochastic system. Think of the physical model for Brownian motion (a heavy particle colliding with many small particles).
- To describe a system that is inherently random. Think of quantum mechanics.

Stochastic modeling is currently used in many different areas ranging from biology to climate modeling to economics.

1.4 Discussion and Bibliography

The fundamental papers of Einstein on the theory of Brownian motion have been reprinted by Dover [7]. The readers of this book are strongly encouraged to study these papers. Other fundamental papers from the early period of the development of the theory of stochastic processes include the papers by Langevin, Ornstein and Uhlenbeck [25], Doob [5], Kramers [13] and Chandrashekhar's famous review article [3]. Many of these early papers on the theory of stochastic processes have been reprinted in [6]. Many of the early papers on the theory of Brownian motion are available from `http://www.physik.uni-augsburg.de/theo1/hanggi/History/BM-History.html`. Very useful historical comments can be founds in the books by Nelson [19] and Mazo [18].

The figures in this chapter were generated using matlab programs from `http://www-math.bgsu.edu/z/sde/matlab/index.html`.

1.5 Exercises

1. Read the papers by Einstein, Ornstein-Uhlenbeck, Doob etc.

2. Write a computer program for generating the random walk in one and two dimensions. Study numerically the Brownian limit and compute the statistics of the random walk.

Chapter 2

Elements of Probability Theory

In this chapter we put together some basic definitions and results from probability theory that will be used later on. In Section 2.1 we give some basic definitions from the theory of probability. In Section 2.2 we present some properties of random variables. In Section 2.3 we introduce the concept of conditional expectation and in Section 2.4 we define the characteristic function, one of the most useful tools in the study of (sums of) random variables. Some explicit calculations for the multivariate Gaussian distribution are presented in Section 2.5. Different types of convergence and the basic limit theorems of the theory of probability are discussed in Section 2.6. Discussion and bibliographical comments are presented in Section 2.7. Exercises are included in Section 2.8.

2.1 Basic Definitions from Probability Theory

In Chapter 1 we defined a stochastic process as a dynamical system whose law of evolution is probabilistic. In order to study stochastic processes we need to be able to describe the outcome of a random experiment and to calculate functions of this outcome. First we need to describe the set of all possible experiments.

Definition 2.1. *The set of all possible outcomes of an experiment is called the* sample space *and is denoted by* Ω*.*

Example 2.2.

- *The possible outcomes of the experiment of tossing a coin are* H *and* T*. The sample space is* $\Omega = \big\{H, T\big\}$*.*
- *The possible outcomes of the experiment of throwing a die are* $1, 2, 3, 4, 5$ *and* 6*. The sample space is* $\Omega = \big\{1, 2, 3, 4, 5, 6\big\}$*.*

We define *events* to be subsets of the sample space. Of course, we would like the unions, intersections and complements of events to also be events. When the sample space Ω is uncountable, then technical difficulties arise. In particular, not all subsets of the sample space need to be events. A definition of the collection of subsets of events which is appropriate for finite additive probability is the following.

Definition 2.3. *A collection $\mathcal{F}$ of Ω is called a field on Ω if*

i. $\emptyset \in \mathcal{F}$;

ii. if $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$;

iii. If $A,\, B \in \mathcal{F}$ *then* $A \cup B \in \mathcal{F}$.

From the definition of a field we immediately deduce that $\mathcal{F}$ is closed under finite unions and finite intersections:

$$A_1, \dots A_n \in \mathcal{F} \quad \Rightarrow \quad \cup_{i=1}^n A_i \in \mathcal{F}, \quad \cap_{i=1}^n A_i \in \mathcal{F}.$$

When Ω is infinite dimensional then the above definition is not appropriate since we need to consider countable unions of events.

Definition 2.4 ($\sigma-$algebra)**.** *A collection $\mathcal{F}$ of Ω is called a σ-field or σ-algebra on Ω if*

i. $\emptyset \in \mathcal{F}$;

ii. if $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$;

iii. If $A_1,\, A_2, \dots \in \mathcal{F}$ *then* $\cup_{i=1}^\infty A_i \in \mathcal{F}$.

A σ-algebra is closed under the operation of taking countable intersections.

Example 2.5.

- $\mathcal{F} = \big\{\emptyset,\, \Omega\big\}$.
- $\mathcal{F} = \big\{\emptyset,\, A,\, A^c,\, \Omega\big\}$ *where A is a subset of Ω.*
- *The power set of Ω, denoted by $\{0,1\}^\Omega$ which contains all subsets of Ω.*

Let $\mathcal{F}$ be a collection of subsets of Ω. It can be extended to a $\sigma-$algebra (take for example the power set of Ω). Consider all the $\sigma-$algebras that contain $\mathcal{F}$ and take their intersection, denoted by $\sigma(\mathcal{F})$, i.e. $A \subset \Omega$ if and only if it is in every $\sigma-$algebra containing $\mathcal{F}$. $\sigma(\mathcal{F})$ is a $\sigma-$algebra (see Exercise 1). It is the smallest algebra containing $\mathcal{F}$ and it is called the $\sigma-$algebra generated by $\mathcal{F}$.

Example 2.6. *Let $\Omega = \mathbb{R}^n$. The σ-algebra generated by the open subsets of $\mathbb{R}^n$ (or, equivalently, by the open balls of $\mathbb{R}^n$) is called the Borel σ-algebra of $\mathbb{R}^n$ and is denoted by $\mathcal{B}(\mathbb{R}^n)$.*

Let X be a closed subset of $\mathbb{R}^n$. Similarly, we can define the Borel σ-algebra of X, denoted by $\mathcal{B}(X)$.

A sub-σ–algebra is a collection of subsets of a σ–algebra which satisfies the axioms of a σ–algebra.

The $\sigma-$field $\mathcal{F}$ of a sample space Ω contains all possible outcomes of the experiment that we want to study. Intuitively, the $\sigma-$field contains all the information about the random experiment that is available to us.

Now we want to assign probabilities to the possible outcomes of an experiment.

Definition 2.7 (Probability measure)**.** *A probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$ satisfying*

i. $\mathbb{P}(\emptyset) = 0, \ \mathbb{P}(\Omega) = 1$*;*

ii. For $A_1, A_2, \ldots$ with $A_i \cap A_j = \emptyset, \ i \neq j$ then

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Definition 2.8. *The triple $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ comprising a set Ω, a σ-algebra $\mathcal{F}$ of subsets of Ω and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a called a probability space.*

Example 2.9. *A biased coin is tossed once:* $\Omega = \{H, T\}, \ \mathcal{F} = \{\emptyset, H, T, \Omega\} = \{0, 1\}, \ \mathbb{P} : \mathcal{F} \mapsto [0, 1]$ *such that* $\mathbb{P}(\emptyset) = 0, \ \mathbb{P}(H) = p \in [0, 1], \ \mathbb{P}(T) = 1 - p, \ \mathbb{P}(\Omega) = 1.$

Example 2.10. *Take $\Omega = [0, 1], \ \mathcal{F} = \mathcal{B}([0, 1]), \ \mathbb{P} = Leb([0, 1])$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.*

2.1.1 Conditional Probability

One of the most important concepts in probability is that of the dependence between events.

Definition 2.11. *A family $\{A_i \, : \, i \in I\}$ of events is called independent if*

$$\mathbb{P}\big(\cap_{j \in J} A_j\big) = \Pi_{j \in J} \mathbb{P}(A_j)$$

for all finite subsets J of I.

When two events A, B are dependent it is important to know the probability that the event A will occur, given that B has already happened. We define this to be *conditional probability*, denoted by $\mathbb{P}(A|B)$. We know from elementary probability that

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

A very useful result is that of the *law of total probability*.

Definition 2.12. *A family of events $\{B_i \,:\, i \in I\}$ is called a partition of Ω if*

$$B_i \cap B_j = \emptyset, \;\; i \neq j \quad \textit{and} \quad \cup_{i \in I} B_i = \Omega.$$

Proposition 2.13. *Law of total probability. For any event A and any partition $\{B_i \,:\, i \in I\}$ we have*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

The proof of this result is left as an exercise. In many cases the calculation of the probability of an event is simplified by choosing an appropriate partition of Ω and using the law of total probability.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and fix $B \in \mathcal{F}$. Then $\mathbb{P}(\cdot|B)$ defines a probability measure on $\mathcal{F}$. Indeed, we have that

$$\mathbb{P}(\emptyset|B) = 0, \quad \mathbb{P}(\Omega|B) = 1$$

and (since $A_i \cap A_j = \emptyset$ implies that $(A_i \cap B) \cap (A_j \cap B) = \emptyset$)

$$P(\cup_{j=1}^{\infty} A_i|B) = \sum_{j=1}^{\infty} \mathbb{P}(A_i|B),$$

for a countable family of pairwise disjoint sets $\{A_j\}_{j=1}^{+\infty}$. Consequently, $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|B))$ is a probability space for every $B \in cF$.

2.2 Random Variables

We are usually interested in the consequences of the outcome of an experiment, rather than the experiment itself. The function of the outcome of an experiment is a *random variable*, that is, a map from Ω to $\mathbb{R}$.

Definition 2.14. *A sample space* Ω *equipped with a* $\sigma-$*field of subsets* $\mathcal{F}$ *is called a measurable space.*

Definition 2.15. *Let* $(\Omega, \mathcal{F})$ *and* $(E, \mathcal{G})$ *be two measurable spaces. A function* $X : \Omega \to E$ *such that the event*

$$\{\omega \in \Omega : X(\omega) \in A\} =: \{X \in A\} \tag{2.1}$$

belongs to $\mathcal{F}$ *for arbitrary* $A \in \mathcal{G}$ *is called a measurable function or random variable.*

When E is $\mathbb{R}$ equipped with its Borel σ-algebra, then (2.1) can by replaced with

$$\{X \leqslant x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

Let X be a random variable (measurable function) from $(\Omega, \mathcal{F}, \mu)$ to $(E, \mathcal{G})$. If E is a metric space then we may define *expectation* with respect to the measure μ by

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

More generally, let $f : E \mapsto \mathbb{R}$ be $\mathcal{G}$–measurable. Then,

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

Let U be a topological space. We will use the notation $\mathcal{B}(U)$ to denote the Borel σ–algebra of U: the smallest σ–algebra containing all open sets of U. Every random variable from a probability space $(\Omega, \mathcal{F}, \mu)$ to a measurable space $(E, \mathcal{B}(E))$ induces a probability measure on E:

$$\mu_X(B) = \mathbb{P}X^{-1}(B) = \mu(\omega \in \Omega; X(\omega) \in B), \quad B \in \mathcal{B}(E). \tag{2.2}$$

The measure μ_X is called the *distribution* (or sometimes the *law*) of X.

Example 2.16. *Let* $\mathcal{I}$ *denote a subset of the positive integers. A vector* $\rho_0 = \{\rho_{0,i},\ i \in \mathcal{I}\}$ *is a distribution on* $\mathcal{I}$ *if it has nonnegative entries and its total mass equals* 1*:* $\sum_{i \in \mathcal{I}} \rho_{0,i} = 1$.

Consider the case where $E = \mathbb{R}$ equipped with the Borel $\sigma-$algebra. In this case a random variable is defined to be a function $X : \Omega \to \mathbb{R}$ such that

$$\{\omega \in \Omega : X(\omega) \leqslant x\} \subset \mathcal{F} \quad \forall x \in \mathbb{R}.$$

We can now define the probability distribution function of X, $F_X : \mathbb{R} \to [0,1]$ as

$$F_X(x) = \mathbb{P}\big(\{\omega \in \Omega \big| X(\omega) \leqslant x\big)\big\} =: \mathbb{P}(X \leqslant x). \tag{2.3}$$

In this case, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X)$ becomes a probability space.

The distribution function $F_X(x)$ of a random variable has the properties that $\lim_{x\to -\infty} F_X(x) = 0$, $\lim_{x\to +\infty} F(x) = 1$ and is right continuous.

Definition 2.17. *A random variable X with values on $\mathbb{R}$ is called discrete if it takes values in some countable subset $\{x_0,\, x_1,\, x_2, \dots\}$ of $\mathbb{R}$. i.e.: $\mathbb{P}(X = x) \neq x$ only for $x = x_0,\, x_1, \dots$.*

With a random variable we can associate the probability mass function $p_k = \mathbb{P}(X = x_k)$. We will consider nonnegative integer valued discrete random variables. In this case $p_k = \mathbb{P}(X = k),\ k = 0, 1, 2, \dots$.

Example 2.18. *The Poisson random variable is the nonnegative integer valued random variable with probability mass function*

$$p_k = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots,$$

where $\lambda > 0$.

Example 2.19. *The binomial random variable is the nonnegative integer valued random variable with probability mass function*

$$p_k = \mathbb{P}(X = k) = \frac{N!}{n!(N-n)!} p^n q^{N-n} \quad k = 0, 1, 2, \dots N,$$

where $p \in (0,1),\ q = 1 - p$.

Definition 2.20. *A random variable X with values on $\mathbb{R}$ is called continuous if $\mathbb{P}(X = x) = 0\, \forall x \in \mathbb{R}$.*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable with distribution F_X. This is a probability measure on $\mathcal{B}(\mathbb{R})$. We will assume that it is absolutely continuous with respect to the Lebesgue measure with density ρ_X: $F_X(dx) = \rho(x)\, dx$. We will call the density $\rho(x)$ the probability density function (PDF) of the random variable X.

Example 2.21. *i. The exponential random variable has PDF*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x < 0, \end{cases}$$

with $\lambda > 0$*.*

ii. The uniform random variable has PDF

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & x \notin (a, b), \end{cases}$$

with $a < b$*.*

Definition 2.22. *Two random variables* X *and* Y *are independent if the events* $\{\omega \in \Omega \,|\, X(\omega) \leqslant x\}$ *and* $\{\omega \in \Omega \,|\, Y(\omega) \leqslant y\}$ *are independent for all* x, $y \in \mathbb{R}$*.*

Let X, Y be two continuous random variables. We can view them as a random vector, i.e. a random variable from Ω to $\mathbb{R}^2$. We can then define the joint distribution function

$$F(x, y) = \mathbb{P}(X \leqslant x, \, Y \leqslant y).$$

The mixed derivative of the distribution function $f_{X,Y}(x, y) := \frac{\partial^2 F}{\partial x \partial y}(x, y)$, if it exists, is called the joint PDF of the random vector $\{X, \, Y\}$:

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

If the random variables X and Y are independent, then

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

and

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

The joint distribution function has the properties

$$\begin{aligned} F_{X,Y}(x, y) &= F_{Y,X}(y, x), \\ F_{X,Y}(+\infty, y) &= F_Y(y), \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) \, dx. \end{aligned}$$

We can extend the above definition to random vectors of arbitrary finite dimensions. Let X be a random variable from $(\Omega, \mathcal{F}, \mu)$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The (joint) distribution function $F_X \mathbb{R}^d \to [0, 1]$ is defined as

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

Let X be a random variable in $\mathbb{R}^d$ with distribution function $f(x_N)$ where $x_N = \{x_1, \dots x_N\}$. We define the marginal or reduced distribution function $f^{N-1}(x_{N-1})$ by

$$f^{N-1}(x_{N-1}) = \int_{\mathbb{R}} f^N(x_N)\, dx_N.$$

We can define other reduced distribution functions:

$$f^{N-2}(x_{N-2}) = \int_{\mathbb{R}} f^{N-1}(x_{N-1})\, dx_{N-1} = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x_N)\, dx_{N-1} dx_N.$$

2.2.1 Expectation of Random Variables

We can use the distribution of a random variable to compute expectations and probabilities:

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\, dF_X(x) \tag{2.4}$$

and

$$\mathbb{P}[X \in G] = \int_G dF_X(x), \quad G \in \mathcal{B}(E). \tag{2.5}$$

The above formulas apply to both discrete and continuous random variables, provided that we define the integrals in (2.4) and (2.5) appropriately.

When $E = \mathbb{R}^d$ and a PDF exists, $dF_X(x) = f_X(x)\, dx$, we have

$$F_X(x) := \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_X(x)\, dx..$$

When $E = \mathbb{R}^d$ then by $L^p(\Omega; \mathbb{R}^d)$, or sometimes $L^p(\Omega; \mu)$ or even simply $L^p(\mu)$, we mean the Banach space of measurable functions on Ω with norm

$$\|X\|_{L^p} = \Big(\mathbb{E}|X|^p\Big)^{1/p}.$$

Let X be a nonnegative integer valued random variable with probability mass function p_k. We can compute the expectation of an arbitrary function of X using the formula

$$\mathbb{E}(f(X)) = \sum_{k=0}^{\infty} f(k) p_k.$$

Let X, Y be random variables we want to know whether they are correlated and, if they are, to calculate how correlated they are. We define the covariance of the two random variables as

$$\mathrm{cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\big] = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y.$$

The correlation coefficient is

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(X)}} \tag{2.6}$$

The Cauchy-Schwarz inequality yields that $\rho(X,Y) \in [-1,1]$. We will say that two random variables X and Y are uncorrelated provided that $\rho(X,Y) = 0$. It is not true in general that two uncorrelated random variables are independent. This is true, however, for Gaussian random variables (see Exercise 5).

Example 2.23.

- *Consider the random variable $X : \Omega \mapsto \mathbb{R}$ with pdf*

$$\gamma_{\sigma,b}(x) := (2\pi\sigma)^{-\frac{1}{2}} \exp\left(-\frac{(x-b)^2}{2\sigma}\right).$$

Such an X is termed a Gaussian or normal random variable. The mean is

$$\mathbb{E}X = \int_{\mathbb{R}} x\gamma_{\sigma,b}(x)\,dx = b$$

and the variance is

$$\mathbb{E}(X-b)^2 = \int_{\mathbb{R}} (x-b)^2\gamma_{\sigma,b}(x)\,dx = \sigma.$$

- *Let $b \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ be symmetric and positive definite. The random variable $X : \Omega \mapsto \mathbb{R}^d$ with pdf*

$$\gamma_{\Sigma,b}(x) := \Big((2\pi)^d det\Sigma\Big)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\langle\Sigma^{-1}(x-b),(x-b)\rangle\right)$$

is termed a multivariate Gaussian or normal random variable. The mean is

$$\mathbb{E}(X) = b \tag{2.7}$$

and the covariance matrix is

$$\mathbb{E}\Big((X-b)\otimes(X-b)\Big) = \Sigma. \tag{2.8}$$

Since the mean and variance specify completely a Gaussian random variable on $\mathbb{R}$, the Gaussian is commonly denoted by $\mathcal{N}(m,\sigma)$. The standard normal random variable is $\mathcal{N}(0,1)$. Similarly, since the mean and covariance matrix completely specify a Gaussian random variable on $\mathbb{R}^d$, the Gaussian is commonly denoted by $\mathcal{N}(m,\Sigma)$.

Some analytical calculations for Gaussian random variables will be presented in Section 2.5.

2.3 Conditional Expecation

Assume that $X \in L^1(\Omega, \mathcal{F}, \mu)$ and let $\mathcal{G}$ be a sub–σ–algebra of $\mathcal{F}$. The conditional expectation of X with respect to $\mathcal{G}$ is defined to be the function (random variable) $\mathbb{E}[X|\mathcal{G}] : \Omega \mapsto E$ which is $\mathcal{G}$–measurable and satisfies

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

We can define $\mathbb{E}[f(X)|\mathcal{G}]$ and the conditional probability $\mathbb{P}[X \in F|\mathcal{G}] = \mathbb{E}[I_F(X)|\mathcal{G}]$, where I_F is the indicator function of F, in a similar manner.

We list some of the most important properties of conditional expectation.

Theorem 2.24. *[Properties of Conditional Expectation]. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and let $\mathcal{G}$ be a sub–σ–algebra of $\mathcal{F}$.*

(a) If X is $\mathcal{G}-$measurable and integrable then $\mathbb{E}(X|\mathcal{G}) = X$.

(b) (Linearity) If $X_1,\ X_2$ are integrable and $c_1,\ c_2$ constants, then

$$\mathbb{E}(c_1 X_1 + c_2 X_2|\mathcal{G}) = c_1 \mathbb{E}(X_1|\mathcal{G}) + c_2 \mathbb{E}(X_2|\mathcal{G}).$$

(c) (Order) If $X_1,\ X_2$ are integrable and $X_1 \leqslant X_2$ a.s., then $\mathbb{E}(X_1|\mathcal{G}) \leqslant \mathbb{E}(X_2|\mathcal{G})$ a.s.

(d) If Y and XY are integrable, and X is $\mathcal{G}-$measurable then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.

(e) (Successive smoothing) If $\mathcal{D}$ is a sub–σ–algebra of $\mathcal{F}$, $\mathcal{D} \subset \mathcal{G}$ and X is integrable, then $\mathbb{E}(X|\mathcal{D}) = \mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{D}] = \mathbb{E}[\mathbb{E}(X|\mathcal{D})|\mathcal{G}]$.

(f) (Convergence) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables such that, for all n, $|X_n| \leqslant Z$ where Z is integrable. If $X_n \to X$ a.s., then $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ a.s. and in L^1.

Proof. See Exercise 10. □

2.4 The Characteristic Function

Many of the properties of (sums of) random variables can be studied using the Fourier transform of the distribution function. Let $F(\lambda)$ be the distribution function of a (discrete or continuous) random variable X. The characteristic function of X is defined to be the Fourier transform of the distribution function

$$\phi(t) = \int_{\mathbb{R}} e^{it\lambda}\, dF(\lambda) = \mathbb{E}(e^{itX}). \tag{2.9}$$

For a continuous random variable for which the distribution function F has a density, $dF(\lambda) = p(\lambda) d\lambda$, (2.9) gives

$$\phi(t) = \int_{\mathbb{R}} e^{it\lambda} p(\lambda)\, d\lambda.$$

For a discrete random variable for which $\mathbb{P}(X = \lambda_k) = \alpha_k$, (2.9) gives

$$\phi(t) = \sum_{k=0}^{\infty} e^{it\lambda_k} a_k.$$

From the properties of the Fourier transform we conclude that the characteristic function determines uniquely the distribution function of the random variable, in the sense that there is a one-to-one correspondance between $F(\lambda)$ and $\phi(t)$. Furthermore, in the exercises at the end of the chapter the reader is asked to prove the following two results.

Lemma 2.25. *Let $\{X_1, X_2, \dots X_n\}$ be independent random variables with characteristic functions $\phi_j(t),\ j = 1, \dots n$ and let $Y = \sum_{j=1}^{n} X_j$ with characteristic function $\phi_Y(t)$. Then*

$$\phi_Y(t) = \Pi_{j=1}^{n} \phi_j(t).$$

Lemma 2.26. *Let X be a random variable with characteristic function $\phi(t)$ and assume that it has finite moments. Then*

$$E(X^k) = \frac{1}{i^k} \phi^{(k)}(0).$$

2.5 Gaussian Random Variables

In this section we present some useful calculations for Gaussian random variables. In particular, we calculate the normalization constant, the mean and variance and the characteristic function of multidimensional Gaussian random variables.

Theorem 2.27. *Let $\mathbf{b} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ a symmetric and positive definite matrix. Let $\mathbf{X}$ be the multivariate Gaussian random variable with probability density function*

$$\gamma_{\Sigma,b}(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{1}{2}\langle \Sigma^{-1}(\mathbf{x}-\mathbf{b}), \mathbf{x}-\mathbf{b}\rangle\right).$$

Then

i. The normalization constant is

$$Z = (2\pi)^{d/2}\sqrt{det(\Sigma)}.$$

ii. The mean vector and covariance matrix of $\mathbf{X}$ are given by

$$\mathbb{E}\mathbf{X} = \mathbf{b}$$

and

$$\mathbb{E}((\mathbf{X}-\mathbb{E}\mathbf{X})\otimes(\mathbf{X}-\mathbb{E}\mathbf{X})) = \Sigma.$$

iii. The characteristic function of $\mathbf{X}$ is

$$\phi(\mathbf{t}) = e^{i\langle \mathbf{b},\mathbf{t}\rangle - \frac{1}{2}\langle \mathbf{t},\Sigma\mathbf{t}\rangle}.$$

Proof. i. From the spectral theorem for symmetric positive definite matrices we have that there exists a diagonal matrix Λ with positive entries and an orthogonal matrix B such that

$$\Sigma^{-1} = B^T\Lambda^{-1}B.$$

Let $\mathbf{z} = \mathbf{x}-\mathbf{b}$ and $\mathbf{y} = B\mathbf{z}$. We have

$$\begin{aligned} \langle \Sigma^{-1}\mathbf{z},\mathbf{z}\rangle &= \langle B^T\Lambda^{-1}B\mathbf{z},\mathbf{z}\rangle \\ &= \langle \Lambda^{-1}B\mathbf{z}, B\mathbf{z}\rangle = \langle \Lambda^{-1}\mathbf{y},\mathbf{y}\rangle \\ &= \sum_{i=1}^d \lambda_i^{-1}y_i^2. \end{aligned}$$

Furthermore, we have that $\det(\Sigma^{-1}) = \Pi_{i=1}^d \lambda_i^{-1}$, that $\det(\Sigma) = \Pi_{i=1}^d \lambda_i$ and that the Jacobian of an orthogonal transformation is $J = \det(B) = 1$.

Hence,

$$
\begin{aligned}
\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle \Sigma^{-1}(\mathbf{x}-\mathbf{b}), \mathbf{x}-\mathbf{b}\rangle\right) d\mathbf{x} &= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle \Sigma^{-1}\mathbf{z}, \mathbf{z}\rangle\right) d\mathbf{z} \\
&= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\sum_{i=1}^d \lambda_i^{-1} y_i^2\right) |J|\, d\mathbf{y} \\
&= \prod_{i=1}^d \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\lambda_i^{-1} y_i^2\right) dy_i \\
&= (2\pi)^{d/2} \Pi_{i=1}^n \lambda_i^{1/2} = (2\pi)^{d/2}\sqrt{\det(\Sigma)},
\end{aligned}
$$

from which we get that

$$
Z = (2\pi)^{d/2}\sqrt{\det(\Sigma)}.
$$

In the above calculation we have used the elementary calculus identity

$$
\int_{\mathbb{R}} e^{-\alpha \frac{x^2}{2}}\, dx = \sqrt{\frac{2\pi}{\alpha}}.
$$

ii. From the above calculation we have that

$$
\begin{aligned}
\gamma_{\Sigma,b}(\mathbf{x})\, d\mathbf{x} &= \gamma_{\Sigma,b}(B^T\mathbf{y}+\mathbf{b})\, d\mathbf{y} \\
&= \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \prod_{i=1}^d \exp\left(-\frac{1}{2}\lambda_i y_i^2\right) dy_i.
\end{aligned}
$$

Consequently

$$
\begin{aligned}
\mathbb{E}\mathbf{X} &= \int_{\mathbb{R}^d} \mathbf{x}\gamma_{\Sigma,b}(\mathbf{x})\, d\mathbf{x} \\
&= \int_{\mathbb{R}^d} (B^T\mathbf{y}+\mathbf{b})\gamma_{\Sigma,b}(B^T\mathbf{y}+\mathbf{b})\, d\mathbf{y} \\
&= \mathbf{b}\int_{\mathbb{R}^d} \gamma_{\Sigma,b}(B^T\mathbf{y}+\mathbf{b})\, d\mathbf{y} = \mathbf{b}.
\end{aligned}
$$

We note that, since $\Sigma^{-1} = B^T\Lambda^{-1}B$, we have that $\Sigma = B^T\Lambda B$. Furthermore, $\mathbf{z} = B^T \mathbf{y}$. We calculate

$$
\begin{aligned}
\mathbb{E}((X_i - b_i)(X_j - b_j)) = \int_{\mathbb{R}^d} z_i z_j \gamma_{\Sigma,b}(\mathbf{z} + \mathbf{b})\, d\mathbf{z} & \\
= \quad & \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \int_{\mathbb{R}^d} \sum_k B_{ki} y_k \sum_m B_{mi} y_m \exp\left(-\frac{1}{2}\sum_\ell \lambda_\ell^{-1} y_\ell^2\right) d\mathbf{y} \\
= \quad & \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \sum_{k,m} B_{ki} B_{mj} \int_{\mathbb{R}^d} y_k y_m \exp\left(-\frac{1}{2}\sum_\ell \lambda_\ell^{-1} y_\ell^2\right) d\mathbf{y} \\
= \quad & \sum_{k,m} B_{ki} B_{mj} \lambda_k \delta_{km} \\
= \quad & \Sigma_{ij}.
\end{aligned}
$$

iii. Let $\mathbf{y}$ be a multivariate Gaussian random variable with mean $\mathbf{0}$ and covariance I. Let also $C = B\sqrt{\Lambda}$. We have that $\Sigma = CC^T = C^T C$. We have that

$$\mathbf{X} = C\mathbf{Y} + \mathbf{b}.$$

To see this, we first note that $\mathbf{X}$ is Gaussian since it is given through a linear transformation of a Gaussian random variable. Furthermore,

$$\mathbb{E}\mathbf{X} = \mathbf{b} \quad \text{and} \quad \mathbb{E}((X_i - b_i)(X_j - b_j)) = \Sigma_{ij}.$$

Now we have:

$$
\begin{aligned}
\phi(\mathbf{t}) &= \mathbb{E}e^{i\langle \mathbf{X}, \mathbf{t}\rangle} = e^{i\langle \mathbf{b}, \mathbf{t}\rangle}\mathbb{E}e^{i\langle C\mathbf{Y}, \mathbf{t}\rangle} \\
&= e^{i\langle \mathbf{b}, \mathbf{t}\rangle}\mathbb{E}e^{i\langle \mathbf{Y}, C^T\mathbf{t}\rangle} \\
&= e^{i\langle \mathbf{b}, \mathbf{t}\rangle}\mathbb{E}e^{i\sum_j(\sum_k C_{jk} t_k) y_j} \\
&= e^{i\langle \mathbf{b}, \mathbf{t}\rangle} e^{-\frac{1}{2}\sum_j \left|\sum_k C_{jk} t_k\right|^2} \\
&= e^{i\langle \mathbf{b}, \mathbf{t}\rangle} e^{-\frac{1}{2}\langle C\mathbf{t}, C\mathbf{t}\rangle} \\
&= e^{i\langle \mathbf{b}, \mathbf{t}\rangle} e^{-\frac{1}{2}\langle \mathbf{t}, C^T C\mathbf{t}\rangle} \\
&= e^{i\langle \mathbf{b}, \mathbf{t}\rangle} e^{-\frac{1}{2}\langle \mathbf{t}, \Sigma\mathbf{t}\rangle}.
\end{aligned}
$$

Consequently,

$$\phi(\mathbf{t}) \quad = \quad e^{i\langle \mathbf{b}, \mathbf{t}\rangle - \frac{1}{2}\langle \mathbf{t}, \Sigma\mathbf{t}\rangle}.$$

$\square$

2.6 Types of Convergence and Limit Theorems

One of the most important aspects of the theory of random variables is the study of limit theorems for sums of random variables. The most well known limit theorems in probability theory are the law of large numbers and the central limit theorem. There are various different types of convergence for sequences or random variables. We list the most important types of convergence below.

Definition 2.28. *Let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of random variables. We will say that*

(a) Z_n converges to Z with probability one if

$$\mathbb{P}\big(\lim_{n\to+\infty} Z_n = Z\big) = 1.$$

(b) Z_n converges to Z in probability if for every $\varepsilon > 0$

$$\lim_{n\to+\infty} \mathbb{P}\big(|Z_n - Z| > \varepsilon\big) = 0.$$

(c) Z_n converges to Z in L^p if

$$\lim_{n\to+\infty} \mathbb{E}\big[\big|Z_n - Z\big|^p\big] = 0.$$

(d) Let $F_n(\lambda), n = 1, \cdots + \infty$, $F(\lambda)$ be the distribution functions of $Z_n\, n = 1, \cdots + \infty$ and Z, respectively. Then Z_n converges to Z in distribution if

$$\lim_{n\to+\infty} F_n(\lambda) = F(\lambda)$$

for all $\lambda \in \mathbb{R}$ at which F is continuous.

Recall that the distribution function F_X of a random variable from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}$ induces a probability measure on $\mathbb{R}$ and that $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X)$ is a probability space. We can show that the convergence in distribution is equivalent to the weak convergence of the probability measures induced by the distribution functions.

Definition 2.29. *Let (E, d) be a metric space, $\mathcal{B}(E)$ the $\sigma-$algebra of its Borel sets, P_n a sequence of probability measures on $(E, \mathcal{B}(E))$ and let $C_b(E)$ denote the space of bounded continuous functions on E. We will say that the sequence of P_n converges weakly to the probability measure P if, for each $f \in C_b(E)$,*

$$\lim_{n\to+\infty} \int_E f(x)\, dP_n(x) = \int_E f(x)\, dP(x).$$

Theorem 2.30. *Let $F_n(\lambda), n = 1, \cdots + \infty,\ F(\lambda)$ be the distribution functions of $Z_n\, n = 1, \cdots + \infty$ and Z, respectively. Then Z_n converges to Z in distribution if and only if, for all $g \in C_b(\mathbb{R})$*

$$\lim_{n \to +\infty} \int_X g(x)\, dF_n(x) = \int_X g(x)\, dF(x). \tag{2.10}$$

Notice that (2.10) is equivalent to

$$\lim_{n \to +\infty} \mathbb{E}_n g(X_n) = \mathbb{E} g(X),$$

where E_n and E denote the expectations with respect to F_n and F, respectively.

When the sequence of random variables whose convergence we are interested in takes values in $\mathbb{R}^d$ or, more generally, a metric space space (E, d) then we can use weak convergence of the sequence of probability measures induced by the sequence of random variables to define convergence in distribution.

Definition 2.31. *A sequence of real valued random variables X_n defined on a probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ and taking values on a metric space (E, d) is said to converge in distribution if the indued measures $F_n(B) = P_n(X_n \in B)$ for $B \in \mathcal{B}(E)$ converge weakly to a probability measure P.*

Let $\{X_n\}_{n=1}^{\infty}$ be iid random variables with $\mathbb{E} X_n = V$. Then, the *strong law of large numbers* states that average of the sum of the iid converges to V with probability one:

$$\mathbb{P}\Big(\lim_{N \to +\infty} \frac{1}{N} \sum_{n=1}^{N} X_n = V \Big) = 1.$$

The strong law of large numbers provides us with information about the behavior of a sum of random variables (or, a large number or repetitions of the same experiment) on average. We can also study fluctuations around the average behavior. Indeed, let $\mathbb{E}(X_n - V)^2 = \sigma^2$. Define the centered iid random variables $Y_n = X_n - V$. Then, the sequence of random variables $\frac{1}{\sigma\sqrt{N}} \sum_{n=1}^{N} Y_n$ converges in distribution to a $\mathcal{N}(0, 1)$ random variable:

$$\lim_{n \to +\infty} \mathbb{P}\left(\frac{1}{\sigma\sqrt{N}} \sum_{n=1}^{N} Y_n \leqslant a \right) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx.$$

This is the *central limit theorem*.

2.7 Discussion and Bibliography

The material of this chapter is very standard and can be found in many books on probability theory. Well known textbooks on probability theory are [2, 8, 9, 16, 17, 12, 23].

The connection between conditional expectation and orthogonal projections is discussed in [4].

The reduced distribution functions defined in Section 2.2 are used extensively in statistical mechanics. A different normalization is usually used in physics textbooks. See for instance [1, Sec. 4.2].

The calculations presented in Section 2.5 are essentially an exercise in linear algebra. See [15, Sec. 10.2].

Random variables and probability measures can also be defined in infinite dimensions. More information can be found in [20, Ch. 2].

The study of limit theorems is one of the cornerstones of probability theory and of the theory of stochastic processes. A comprehensive study of limit theorems can be found in [10].

2.8 Exercises

1. Show that the intersection of a family of σ-algebras is a σ-algebra.

2. Prove the law of total probability, Proposition 2.13.

3. Calculate the mean, variance and characteristic function of the following probability density functions.

   (a) The exponential distribution with density

   $$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x < 0, \end{cases}$$

   with $\lambda > 0$.

   (b) The uniform distribution with density

   $$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & x \notin (a, b), \end{cases}$$

   with $a < b$.

(c) The Gamma distribution with density

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(\alpha)}(\lambda x)^{\alpha-1}e^{-\lambda x} & x > 0, \\ 0 & x < 0, \end{cases}$$

with $\lambda > 0,\ \alpha > 0$ and $\Gamma(\alpha)$ is the Gamma function

$$\Gamma(\alpha) = \int_0^\infty \xi^{\alpha-1}e^{-\xi}\,d\xi, \quad \alpha > 0.$$

4. Le X and Y be independent random variables with distribution functions F_X and F_Y. Show that the distribution function of the sum $Z = X + Y$ is the convolution of F_X and F_Y:

$$F_Z(x) = \int F_X(x-y)\,dF_Y(y).$$

5. Let X and Y be Gaussian random variables. Show that they are uncorrelated if and only if they are independent.

6. (a) Let X be a continuous random variable with characteristic function $\phi(t)$. Show that

$$\mathbb{E}X^k = \frac{1}{i^k}\phi^{(k)}(0),$$

where $\phi^{(k)}(t)$ denotes the k-th derivative of ϕ evaluated at t.

(b) Let X be a nonnegative random variable with distribution function $F(x)$. Show that

$$\mathbb{E}(X) = \int_0^{+\infty}(1 - F(x))\,dx.$$

(c) Let X be a continuous random variable with probability density function $f(x)$ and characteristic function $\phi(t)$. Find the probability density and characteristic function of the random variable $Y = aX + b$ with $a,\ b \in \mathbb{R}$.

(d) Let X be a random variable with uniform distribution on $[0, 2\pi]$. Find the probability density of the random variable $Y = \sin(X)$.

7. Let X be a discrete random variable taking vales on the set of nonnegative integers with probability mass function $p_k = \mathbb{P}(X = k)$ with $p_k \geqslant 0$, $\sum_{k=0}^{+\infty} p_k = 1$. The *generating function* is defined as

$$g(s) = \mathbb{E}(s^X) = \sum_{k=0}^{+\infty} p_k s^k.$$

(a) Show that

$$\mathbb{E}X = g'(1) \quad \text{and} \quad \mathbb{E}X^2 = g''(1) + g'(1),$$

where the prime denotes differentiation.

(b) Calculate the generating function of the Poisson random variable with

$$p_k = \mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad \text{and} \quad \lambda > 0.$$

(c) Prove that the generating function of a sum of independent nonnegative integer valued random variables is the product of their generating functions.

8. Write a computer program for studying the law of large numbers and the central limit theorem. Investigate numerically the rate of convergence of these two theorems.

9. Study the properties of Gaussian measures on separable Hilbert spaces from [20, Ch. 2].

10. Prove Theorem 2.24.

Chapter 3

Basics of the Theory of Stochastic Processes

In this chapter we present some basic results form the theory of stochastic processes and we investigate the properties of some of the standard stochastic processes in continuous time. In Section 3.1 we give the definition of a stochastic process. In Section 3.2 we present some properties of stationary stochastic processes. In Section 3.3 we introduce Brownian motion and study some of its properties. Various examples of stochastic processes in continuous time are presented in Section 3.4. The Karhunen-Loeve expansion, one of the most useful tools for representing stochastic processes and random fields, is presented in Section 3.5. Further discussion and bibliographical comments are presented in Section 3.6. Section 3.7 contains exercises.

3.1 Definition of a Stochastic Process

Stochastic processes describe dynamical systems whose evolution law is of probabilistic nature. The precise definition is given below.

Definition 3.1 (stochastic process)**.** *Let T be an ordered set, $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $(E, \mathcal{G})$ a measurable space. A stochastic process is a collection of random variables $X = \{X_t; t \in T\}$ where, for each fixed $t \in T$, X_t is a random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(E, \mathcal{G})$. Ω is called the sample space. and E is the state space of the stochastic process X_t.*

The set T can be either discrete, for example the set of positive integers $\mathbb{Z}_+$, or continuous, $T = [0, +\infty)$. The state space E will usually be $\mathbb{R}^d$ equipped with the σ–algebra of Borel sets.

A stochastic process X may be viewed as a function of both $t \in T$ and $\omega \in \Omega$. We will sometimes write $X(t), X(t, \omega)$ or $X_t(\omega)$ instead of X_t. For a fixed sample point $\omega \in \Omega$, the function $X_t(\omega) : T \mapsto E$ is called a (realization, trajectory) of the process X.

Definition 3.2 (finite dimensional distributions)**.** *The finite dimensional distributions (fdd) of a stochastic process are the 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 $t_i \in T$, $i \in \{1, \ldots, k\}$:*

$$F(\mathbf{x}) = \mathbb{P}(X(t_i) \leqslant x_i,\ i = 1, \ldots, k)$$

with $\mathbf{x} = (x_1, \ldots, x_k)$.

From experiments or numerical simulations we can only obtain information about the finite dimensional distributions of a process. A natural question arises: are the finite dimensional distributions of a stochastic process sufficient to determine a stochastic process uniquely? This is true for processes with continuous paths [1]. This is the class of stochastic processes that we will study in these notes.

Definition 3.3. *We will say that two processes X_t and Y_t are equivalent if they have same finite dimensional distributions.*

Definition 3.4. *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 $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is a $\mathcal{N}(\mu_k, K_k)$ random variable for some vector μ_k and a symmetric nonnegative definite matrix K_k for all $k = 1, 2, \ldots$ and for all $t_1, t_2, \ldots, t_k$.*

From the above definition we conclude that the finite dimensional distributions of a Gaussian continuous time stochastic process are Gaussian with probability distribution function

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

where $\mathbf{x} = (x_1, x_2, \ldots x_k)$.

[1] In fact, what we need is the stochastic process to be *separable*. See the discussion in Section 3.6

It is straightforward to extend the above definition to arbitrary dimensions. A Gaussian process $x(t)$ is characterized by its mean

$$m(t) := \mathbb{E}x(t)$$

and the covariance (or autocorrelation) matrix

$$C(t,s) = \mathbb{E}\Big(\big(x(t) - m(t)\big) \otimes \big(x(s) - m(s)\big)\Big).$$

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

3.2 Stationary Processes

3.2.1 Strictly Stationary Processes

In many stochastic processes that appear in applications their statistics remain invariant under time translations. Such stochastic processes are called *stationary*. It is possible to develop a quite general theory for stochastic processes that enjoy this symmetry property.

Definition 3.5. *A stochastic process is called (strictly) stationary if all finite dimensional distributions are invariant under time translation: for any integer k and times $t_i \in T$, 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 any s such that $s + t_i \in T$ for all $i \in \{1, \dots, k\}$. In other words,*

$$\mathbb{P}(X_{t_1+t} \in A_1, X_{t_2+t} \in A_2 \dots X_{t_k+t} \in A_k) = \mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2 \dots X_{t_k} \in A_k), \quad \forall t \in T.$$

Example 3.6. *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 (see Exercise 1). Assume furthermore that $\mathbb{E}Y_0 = \mu < +\infty$. Then, by the strong law of large numbers, we have that*

$$\frac{1}{N}\sum_{j=0}^{N-1} X_j = \frac{1}{N}\sum_{j=0}^{N-1} Y_j \to \mathbb{E}Y_0 = \mu,$$

almost surely. In fact, the Birkhoff ergodic theorem states that, for any function f such that $\mathbb{E}f(Y_0) < +\infty$, we have that

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{j=0}^{N-1} f(X_j) = \mathbb{E}f(Y_0), \tag{3.1}$$

almost surely. The sequence of iid random variables is an example of an ergodic strictly stationary processes.

Ergodic strictly stationary processes satisfy (3.1) Hence, we can calculate the statistics of a sequence stochastic process X_n using a single sample path, provided that it is long enough ($N \gg 1$).

Example 3.7. *Let Z be a random variable and define the stochastic process $X_n = Z,\ n = 0, 1, 2, \dots$. Then X_n is a strictly stationary process (see Exercise 2). We can calculate the long time average of this stochastic process:*

$$\frac{1}{N}\sum_{j=0}^{N-1} X_j = \frac{1}{N}\sum_{j=0}^{N-1} Z = Z,$$

which is independent of N and does not converge to the mean of the stochastic processes $\mathbb{E}X_n = \mathbb{E}Z$ (assuming that it is finite), or any other deterministic number. This is an example of a non-ergodic processes.

3.2.2 Second Order Stationary Processes

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

$$\mathbb{E}(X_{t+s}) = \mathbb{E}X_t, \quad s \in T \tag{3.2}$$

from which we conclude that $\mathbb{E}X_t$ is constant. and

$$\mathbb{E}((X_{t_1+s} - \mu)(X_{t_2+s} - \mu)) = \mathbb{E}((X_{t_1} - \mu)(X_{t_2} - \mu)), \quad s \in T \tag{3.3}$$

from which we conclude that the covariance or autocorrelation or correlation function $C(t, s) = \mathbb{E}((X_t - \mu)(X_s - \mu))$ depends on the difference between the two times, t and s, i.e. $C(t, s) = C(t - s)$. This motivates the following definition.

Definition 3.8. *A stochastic process $X_t \in L^2$ is called second-order stationary or wide-sense stationary or weakly stationary if the first moment $\mathbb{E}X_t$ is a 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 the expectation of the process X_t. Without loss of generality, we can set $\mu = 0$, since if $\mathbb{E}X_t = \mu$ then the process $Y_t = X_t - \mu$ is mean zero. A mean zero process with be called a centered process. The function $C(t)$ is the *covariance* (sometimes also called autocovariance) or the autocorrelation function of the X_t. Notice that $C(t) = \mathbb{E}(X_t X_0)$, whereas $C(0) = \mathbb{E}(X_t^2)$, which is finite, by assumption. Since we have assumed that X_t is a real valued process, we have that $C(t) = C(-t)$, $t \in \mathbb{R}$.

Remark 3.9. *Let X_t be a strictly stationary stochastic process with finite second moment (i.e. $X_t \in L^2$). The definition of strict stationarity implies that $\mathbb{E}X_t = \mu$, a constant, and $\mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t-s)$. Hence, a strictly stationary process with finite second moment is also stationary in the wide sense. The converse is not true.*

Example 3.10. *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 3.6 we know that this is a strictly stationary process, irrespective of whether Y_0 is such that $\mathbb{E}Y_0^2 < +\infty$. 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 we have no correlation between the values of the stochastic process at different times n and k.*

Example 3.11. *Let Z be a single random variable and consider the stochastic process $X_n = Z$, $n = 0, 1, 2, \dots$. From Example 3.7 we know that this is a strictly stationary process irrespective of whether $\mathbb{E}|Z|^2 < +\infty$ or not. Assume now that $\mathbb{E}Z = 0$, $EZ^2 = \sigma^2$. 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.*

We will see in Section 3.2.3 that for second order stationary processes, ergodicity is related to fast decay of correlations. In the first of the examples above, there was no correlation between our stochastic processes at different times and the stochastic process is ergodic. On the contrary, in our second example there is very strong correlation between the stochastic process at different times and this process is not ergodic.

Remark 3.12. *The first two moments of a Gaussian process are sufficient for a complete characterization of the process. Consequently, a Gaussian stochastic process is strictly stationary if and only if it is weakly stationary.*

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

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

Lemma 3.13. *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) set $\mathbb{E}X_t = 0$. We calculate:

$$\begin{aligned}
|C(t+h) - C(t)|^2 &= |\mathbb{E}(X_{t+h}X_0) - \mathbb{E}(X_t X_0)|^2 = \mathbb{E}|((X_{t+h} - X_t)X_0)|^2 \\
&\leqslant \mathbb{E}(X_0)^2 \mathbb{E}(X_{t+h} - X_t)^2 \\
&= C(0)(\mathbb{E}X_{t+h}^2 + \mathbb{E}X_t^2 - 2\mathbb{E}X_t X_{t+h}) \\
&= 2C(0)(C(0) - C(h)) \to 0,
\end{aligned}$$

as $h \to 0$. Thus, continuity of $C(\cdot)$ at 0 implies continuity for all t.

Assume now that $C(t)$ is continuous. From the above calculation we have

$$\mathbb{E}|X_{t+h} - X_t|^2 = 2(C(0) - C(h)), \tag{3.4}$$

which converges to 0 as $h \to 0$. Conversely, assume that X_t is L^2-continuous. Then, from the above equation we get $\lim_{h \to 0} C(h) = C(0)$. □

Notice that form (3.4) we immediately conclude that $C(0) > C(h),\ h \in \mathbb{R}$.

The Fourier transform of the covariance function of a second order stationary process always exists. This enables us to study second order stationary processes using tools from Fourier analysis. To make the link between second order stationary processes and Fourier analysis we will use Bochner's theorem, which applies to all nonnegative functions.

Definition 3.14. *A function $f(x) : \mathbb{R} \mapsto \mathbb{R}$ is called nonnegative definite if*

$$\sum_{i,j=1}^{n} f(t_i - t_j) c_i \bar{c}_j \geqslant 0 \tag{3.5}$$

for all $n \in \mathbb{N},\ t_1, \dots t_n \in \mathbb{R},\ c_1, \dots c_n \in \mathbb{C}$.

Lemma 3.15. *The covariance function of 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.

$$\begin{aligned}\sum_{i,j=1}^n C(t_i - t_j) c_i \bar{c}_j &= \sum_{i,j=1}^n \mathbb{E} X_{t_i} X_{t_j} c_i \bar{c}_j \\ &= \mathbb{E}\left(\sum_{i=1}^n X_{t_i} c_i \sum_{j=1}^n X_{t_j} \bar{c}_j\right) = \mathbb{E}\left(X_t^c \bar{X}_t^c\right) \\ &= \mathbb{E}|X_t^c|^2 \geqslant 0.\end{aligned}$$

□

Theorem 3.16. *[Bochner] Let $C(t)$ be a continuous positive definite function. Then there exists a unique nonnegative measure ρ on $\mathbb{R}$ such that $\rho(\mathbb{R}) = C(0)$ and*

$$C(t) = \int_{\mathbb{R}} e^{i\omega t} \, \rho(d\omega) \quad \forall t \in \mathbb{R}. \tag{3.6}$$

Definition 3.17. *Let X_t be a second order stationary process with autocorrelation function $C(t)$ whose Fourier transform is the measure $\rho(d\omega)$. The measure $\rho(d\omega)$ is called the* spectral measure *of the process X_t.*

In the following we will assume that the spectral measure is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ with density $S(\omega)$, i.e. $\rho(d\omega) = S(\omega) d\omega$. The Fourier transform $S(\omega)$ of the covariance function is called the spectral density of the process:

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\omega} C(t) \, dt. \tag{3.7}$$

From (3.6) it follows that that the autocorrelation function of a mean zero, second order stationary process is given by the inverse Fourier transform of the spectral density:

$$C(t) = \int_{-\infty}^{\infty} e^{it\omega} S(\omega) \, d\omega. \tag{3.8}$$

There are various cases where the experimentally measured quantity is the spectral density (or power spectrum) of a stationary stochastic process. Conversely, from a time series of observations of a stationary processes we can calculate the autocorrelation function and, using (3.8) the spectral density.

The autocorrelation function of a second order stationary process enables us to associate a time scale to X_t, the *correlation time* τ_{cor}:

$$\tau_{cor} = \frac{1}{C(0)} \int_0^\infty C(\tau)\, d\tau = \int_0^\infty \mathbb{E}(X_\tau X_0)/\mathbb{E}(X_0^2)\, d\tau.$$

The slower the decay of the correlation function, the larger the correlation time is. Notice that when the correlations do not decay sufficiently fast so that $C(t)$ is integrable, then the correlation time will be infinite.

Example 3.18. *Consider a mean zero, second order stationary process with correlation function*

$$R(t) = R(0)e^{-\alpha|t|} \tag{3.9}$$

where $\alpha > 0$. We will write $R(0) = \frac{D}{\alpha}$ where $D > 0$. The spectral density of this process is:

$$\begin{aligned} S(\omega) &= \frac{1}{2\pi}\frac{D}{\alpha}\int_{-\infty}^{+\infty} e^{-i\omega t}e^{-\alpha|t|}\, dt \\ &= \frac{1}{2\pi}\frac{D}{\alpha}\left(\int_{-\infty}^{0} e^{-i\omega t}e^{\alpha t}\, dt + \int_{0}^{+\infty} e^{-i\omega t}e^{-\alpha t}\, dt\right) \\ &= \frac{1}{2\pi}\frac{D}{\alpha}\left(\frac{1}{-i\omega+\alpha} + \frac{1}{i\omega+\alpha}\right) \\ &= \frac{D}{\pi}\frac{1}{\omega^2+\alpha^2}. \end{aligned}$$

This function is called the Cauchy or the Lorentz distribution. The correlation time is (we have that $R(0) = D/\alpha$)

$$\tau_{cor} = \int_0^\infty e^{-\alpha t}\, dt = \alpha^{-1}.$$

A Gaussian process with an exponential correlation function is of particular importance in the theory and applications of stochastic processes.

Definition 3.19. *A real-valued Gaussian stationary process defined on $\mathbb{R}$ with correlation function given by* (3.9) *is called the (stationary) Ornstein-Uhlenbeck process.*

The Ornstein Uhlenbeck process is used as a model for the velocity of a Brownian particle. It is of interest to calculate the statistics of the position of the Brownian particle, i.e. of the integral

$$X(t) = \int_0^t Y(s)\, ds, \tag{3.10}$$

where $Y(t)$ denotes the stationary OU process.

Lemma 3.20. *Let $Y(t)$ denote the stationary OU process with covariance function* (3.9) *and set $\alpha = D = 1$. Then the position process* (3.10) *is a mean zero Gaussian process with covariance function*

$$\mathbb{E}(X(t)X(s)) = 2\min(t,s) + e^{-\min(t,s)} + e^{-max(t,s)} - e^{-|t-s|} - 1. \tag{3.11}$$

Proof. See Exercise 8. ☐

3.2.3 Ergodic Properties of Second-Order Stationary Processes

Second order stationary processes have nice ergodic properties, provided that the correlation between values of the process at different times decays sufficiently fast. In this case, it is possible to show that we can calculate expectations by calculating time averages. An example of such a result is the following.

Theorem 3.21. *Let $\{X_t\}_{t \geqslant 0}$ be a second order stationary process on a probability space Ω, $\mathcal{F}$, $\mathbb{P}$ with mean μ and covariance $R(t)$, and assume that $R(t) \in L^1(0, +\infty)$. Then*

$$\lim_{T \to +\infty} \mathbb{E} \left| \frac{1}{T} \int_0^T X(s)\, ds - \mu \right|^2 = 0. \tag{3.12}$$

For the proof of this result we will first need an elementary lemma.

Lemma 3.22. *Let $R(t)$ be an integrable symmetric function. Then*

$$\int_0^T \int_0^T R(t-s)\, dtds = 2 \int_0^T (T-s) R(s)\, ds. \tag{3.13}$$

Proof. We make the change of variables $u = t - s$, $v = t + s$. The domain of integration in the t, s variables is $[0,T] \times [0,T]$. In the u, v variables it becomes $[-T,T] \times [0, 2(T - |u|)]$. The Jacobian of the transformation is

$$J = \frac{\partial(t,s)}{\partial(u,v)} = \frac{1}{2}.$$

The integral becomes

$$
\begin{aligned}
\int_0^T \int_0^T R(t-s)\,dtds &= \int_{-T}^{T} \int_0^{2(T-|u|)} R(u) J\,dvdu \\
&= \int_{-T}^{T} (T-|u|) R(u)\,du \\
&= 2\int_0^T (T-u)R(u)\,du,
\end{aligned}
$$

where the symmetry of the function $R(u)$ was used in the last step. $\square$

Proof of Theorem 3.21. We use Lemma (3.22) to calculate:

$$
\begin{aligned}
\mathbb{E}\left|\frac{1}{T}\int_0^T X_s\,ds - \mu\right|^2 &= \frac{1}{T^2}\mathbb{E}\left|\int_0^T (X_s-\mu)\,ds\right|^2 \\
&= \frac{1}{T^2}\mathbb{E}\int_0^T\int_0^T (X(t)-\mu)(X(s)-\mu)\,dtds \\
&= \frac{1}{T^2}\int_0^T\int_0^T R(t-s)\,dtds \\
&= \frac{2}{T^2}\int_0^T (T-u)R(u)\,du \\
&\leqslant \frac{2}{T}\int_0^{+\infty}\left|\left(1-\frac{u}{T}\right)R(u)\right|\,du \leqslant \frac{2}{T}\int_0^{+\infty} R(u)\,du \to 0,
\end{aligned}
$$

using the dominated convergence theorem and the assumption $R(\cdot) \in L^1$. $\square$

Assume that $\mu = 0$ and define

$$
D = \int_0^{+\infty} R(t)\,dt, \tag{3.14}
$$

which, from our assumption on $R(t)$, is a finite quantity. [2] The above calculation suggests that, for $T \gg 1$, we have that

$$
\mathbb{E}\left(\int_0^t X(t)\,dt\right)^2 \approx 2DT.
$$

This implies that, at sufficiently long times, the mean square displacement of the integral of the ergodic second order stationary process X_t scales linearly in time, with proportionality coefficient $2D$.

[2] Notice however that we do not know whether it is nonzero. This requires a separate argument.

Assume that X_t is the velocity of a (Brownian) particle. In this case, the integral of X_t

$$Z_t = \int_0^t X_s \, ds,$$

represents the particle position. From our calculation above we conclude that

$$\mathbb{E}Z_t^2 = 2Dt.$$

where

$$D = \int_0^\infty R(t) \, dt = \int_0^\infty \mathbb{E}(X_t X_0) \, dt \tag{3.15}$$

is the *diffusion coefficient*. Thus, one expects that at sufficiently long times and under appropriate assumptions on the correlation function, the time integral of a stationary process will approximate a Brownian motion with diffusion coefficient D. The diffusion coefficient is an example of a transport coefficient and (3.15) is an example of the Green-Kubo formula: a transport coefficient can be calculated in terms of the time integral of an appropriate autocorrelation function. In the case of the diffusion coefficient we need to calculate the integral of the velocity autocorrelation function.

Example 3.23. *Consider the stochastic processes with an exponential correlation function from Example 3.18, and assume that this stochastic process describes the velocity of a Brownian particle. Since $R(t) \in L^1(0, +\infty)$ Theorem 3.21 applies. Furthermore, the diffusion coefficient of the Brownian particle is given by*

$$\int_0^{+\infty} R(t) \, dt = R(0)\tau_c^{-1} = \frac{D}{\alpha^2}.$$

3.3 Brownian Motion

The most important continuous time stochastic process is Brownian motion. Brownian motion is a mean zero, continuous (i.e. it has continuous sample paths: for a.e $\omega \in \Omega$ the function X_t is a continuous function of time) process with independent Gaussian increments. A process X_t has independent increments if for every sequence $t_0 < t_1 < \dots t_n$ the random variables

$$X_{t_1} - X_{t_0}, \, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$$

are independent. If, furthermore, for any $t_1,\ t_2,\ s \in T$ and Borel set $B \subset \mathbb{R}$

$$\mathbb{P}(X_{t_2+s} - X_{t_1+s} \in B) = \mathbb{P}(X_{t_2} - X_{t_1} \in B)$$

then the process X_t has stationary independent increments.

Definition 3.24.

- *A one dimensional standard* Brownian motion $W(t) : \mathbb{R}^+ \to \mathbb{R}$ *is a real valued stochastic process such that*

  i. $W(0) = 0$.

  ii. $W(t)$ *has independent increments.*

  iii. For every $t > s \geqslant 0$ $W(t) - W(s)$ *has a Gaussian distribution with mean* 0 *and variance* $t - s$*. That is, the density of the random variable* $W(t) - W(s)$ *is*

  $$g(x; t, s) = \Big(2\pi(t-s)\Big)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2(t-s)}\right); \tag{3.16}$$

- *A* d*–dimensional standard Brownian motion* $W(t) : \mathbb{R}^+ \to \mathbb{R}^d$ *is a collection of* d *independent one dimensional Brownian motions:*

  $$W(t) = (W_1(t), \dots, W_d(t)),$$

  where $W_i(t), i = 1, \dots, d$ *are independent one dimensional Brownian motions. The density of the Gaussian random vector* $W(t) - W(s)$ *is thus*

  $$g(x; t, s) = \Big(2\pi(t-s)\Big)^{-d/2} \exp\left(-\frac{\|x\|^2}{2(t-s)}\right).$$

Brownian motion is sometimes referred to as the *Wiener process*.

Brownian motion has continuous paths. More precisely, it has a continuous modification.

Definition 3.25. *Let* X_t *and* $Y_t,\ t \in T$*, be two stochastic processes defined on the same probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. The process* Y_t *is said to be a modification of* X_t *if* $\mathbb{P}(X_t = Y_t) = 1\ \forall t \in T$.

Lemma 3.26. *There is a continuous modification of Brownian motion.*

This follows from a theorem due to Kolmogorov.

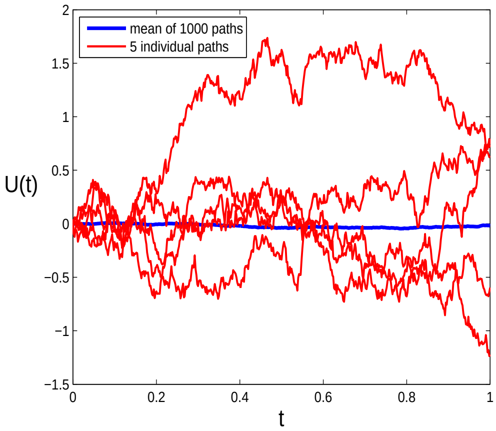

Figure 3.1: Brownian sample paths

Theorem 3.27. *(Kolmogorov) Let X_t, $t \in [0, \infty)$ be a stochastic process on a probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$. Suppose that there are positive constants α and β, and for each $T \geqslant 0$ there is a constant $C(T)$ such that*

$$\mathbb{E}|X_t - X_s|^{\alpha} \leqslant C(T)|t-s|^{1+\beta}, \quad 0 \leqslant s, t \leqslant T. \tag{3.17}$$

Then there exists a continuous modification Y_t of the process X_t.

The proof of Lemma 3.26 is left as an exercise.

Remark 3.28. *Equivalently, we could have defined the one dimensional standard Brownian motion as a stochastic process on a probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ with continuous paths for almost all $\omega \in \Omega$, and Gaussian finite dimensional distributions with zero mean and covariance $\mathbb{E}(W_{t_i} W_{t_j}) = \min(t_i, t_j)$. One can then show that Definition 3.24 follows from the above definition.*

It is possible to prove rigorously the existence of the Wiener process (Brownian motion):

Theorem 3.29. *(Wiener) There exists an almost-surely continuous process W_t with independent increments such and $W_0 = 0$, such that for each $t \geqslant 0$ the random variable W_t is $\mathcal{N}(0, t)$. Furthermore, W_t is almost surely locally Hölder continuous with exponent α for any $\alpha \in (0, \frac{1}{2})$.*

Notice that Brownian paths are not differentiable.

We can also construct Brownian motion through the limit of an appropriately rescaled random walk: let $X_1, X_2, \dots$ be iid random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with mean 0 and variance 1. Define the discrete time stochastic process S_n with $S_0 = 0$, $S_n = \sum_{j=1} X_j$, $n \geqslant 1$. Define now a continuous time stochastic process with continuous paths as the linearly interpolated, appropriately rescaled random walk:

$$W_t^n = \frac{1}{\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sqrt{n}} X_{[nt]+1},$$

where $[\cdot]$ denotes the integer part of a number. Then W_t^n converges weakly, as $n \to +\infty$ to a one dimensional standard Brownian motion.

Brownian motion is a Gaussian process. For the d–dimensional Brownian motion, and for I the $d \times d$ dimensional identity, we have (see (2.7) and (2.8))

$$\mathbb{E} W(t) = 0 \quad \forall t \geqslant 0$$

and

$$\mathbb{E}\Big((W(t) - W(s)) \otimes (W(t) - W(s)) \Big) = (t-s) I. \tag{3.18}$$

Moreover,

$$\mathbb{E}\Big(W(t) \otimes W(s) \Big) = \min(t,s) I. \tag{3.19}$$

From the formula for the Gaussian density $g(x, t-s)$, eqn. (3.16), we immediately conclude that $W(t) - W(s)$ and $W(t+u) - W(s+u)$ have the same pdf. Consequently, Brownian motion has stationary increments. Notice, however, that Brownian motion itself is not a stationary process. Since $W(t) = W(t) - W(0)$, the pdf of $W(t)$ is

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

We can easily calculate all moments of the Brownian motion:

$$\begin{aligned} \mathbb{E}(x^n(t)) &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2t}\, dx \\ &= \begin{cases} 1.3 \dots (n-1) t^{n/2}, & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases} \end{aligned}$$

Brownian motion is invariant under various transformations in time.

Theorem 3.30. *Let W_t denote a standard Brownian motion in $\mathbb{R}$. Then, W_t has the following properties:*

i. (Rescaling). For each $c > 0$ define $X_t = \frac{1}{\sqrt{c}}W(ct)$. Then $(X_t,\, t \geqslant 0) = (W_t,\, t \geqslant 0)$ in law.

ii. (Shifting). For each $c > 0$ $W_{c+t} - W_c, t \geqslant 0$ is a Brownian motion which is independent of $W_u,\ u \in [0, c]$.

iii. (Time reversal). Define $X_t = W_{1-t} - W_1,\ t \in [0, 1]$. Then $(X_t,\, t \in [0, 1]) = (W_t,\, t \in [0, 1])$ in law.

iv. (Inversion). Let $X_t, t \geqslant 0$ defined by $X_0 = 0,\ \ X_t = tW(1/t)$. Then $(X_t,\, t \geqslant 0) = (W_t,\, t \geqslant 0)$ in law.

We emphasize that the equivalence in the above theorem holds in law and not in a pathwise sense.

Proof. See Exercise 13. □

We can also add a drift and change the diffusion coefficient of the Brownian motion: we will define a Brownian motion with drift μ and variance σ^2 as the process

$$X_t = \mu t + \sigma W_t.$$

The mean and variance of X_t are

$$\mathbb{E}X_t = \mu t, \quad \mathbb{E}(X_t - \mathbb{E}X_t)^2 = \sigma^2 t.$$

Notice that X_t satisfies the equation

$$dX_t = \mu\, dt + \sigma\, dW_t.$$

This is the simplest example of a *stochastic differential equation*.

We can define the OU process through the Brownian motion via a time change.

Lemma 3.31. *Let $W(t)$ be a standard Brownian motion and consider the process*

$$V(t) = e^{-t}W(e^{2t}).$$

Then $V(t)$ is a Gaussian stationary process with mean 0 and correlation function

$$R(t) = e^{-|t|}. \tag{3.20}$$

For the proof of this result we first need to show that time changed Gaussian processes are also Gaussian.

Lemma 3.32. *Let $X(t)$ be a Gaussian stochastic process and let $Y(t) = X(f(t))$ where $f(t)$ is a strictly increasing function. Then $Y(t)$ is also a Gaussian process.*

Proof. We need to show that, for all positive integers N and all sequences of times $\{t_1,\, t_2, \dots t_N\}$ the random vector

$$\{Y(t_1),\, Y(t_2), \dots Y(t_N)\} \tag{3.21}$$

is a multivariate Gaussian random variable. Since $f(t)$ is strictly increasing, it is invertible and hence, there exist $s_i,\, i = 1, \dots N$ such that $s_i = f^{-1}(t_i)$. Thus, the random vector (3.21) can be rewritten as

$$\{X(s_1),\, X(s_2), \dots X(s_N)\},$$

which is Gaussian for all N and all choices of times $s_1,\, s_2, \dots s_N$. Hence $Y(t)$ is also Gaussian. □

Proof of Lemma 3.31. The fact that $V(t)$ is mean zero follows immediately from the fact that $W(t)$ is mean zero. To show that the correlation function of $V(t)$ is given by (3.20), we calculate

$$\begin{aligned}\mathbb{E}(V(t)V(s)) &= e^{-t-s}\mathbb{E}(W(e^{2t})W(e^{2s})) = e^{-t-s}\min(e^{2t}, e^{2s}) \\ &= e^{-|t-s|}.\end{aligned}$$

The Gaussianity of the process $V(t)$ follows from Lemma 3.32 (notice that the transformation that gives $V(t)$ in terms of $W(t)$ is invertible and we can write $W(s) = s^{1/2}V(\frac{1}{2}\ln(s))$). □

3.4 Other Examples of Stochastic Processes

Brownian Bridge Let $W(t)$ be a standard one dimensional Brownian motion. We define the Brownian bridge (from 0 to 0) to be the process

$$B_t = W_t - tW_1, \quad t \in [0,1]. \tag{3.22}$$

Notice that $B_0 = B_1 = 0$. Equivalently, we can define the Brownian bridge to be the continuous Gaussian process $\{B_t : 0 \leqslant t \leqslant 1\}$ such that

$$\mathbb{E}B_t = 0, \quad \mathbb{E}(B_tB_s) = \min(s,t) - st, \quad s,t \in [0,1]. \tag{3.23}$$

Another, equivalent definition of the Brownian bridge is through an appropriate time change of the Brownian motion:

$$B_t = (1-t)W\left(\frac{t}{1-t}\right), \quad t \in [0,1). \tag{3.24}$$

Conversely, we can write the Brownian motion as a time change of the Brownian bridge:

$$W_t = (t+1)B\left(\frac{t}{1+t}\right), \quad t \geqslant 0.$$

Fractional Brownian Motion

Definition 3.33. *A (normalized) fractional Brownian motion W_t^H, $t \geqslant 0$ with Hurst parameter $H \in (0,1)$ is a centered Gaussian process with continuous sample paths whose covariance is given by*

$$\mathbb{E}(W_t^H W_s^H) = \frac{1}{2}\big(s^{2H} + t^{2H} - |t-s|^{2H}\big). \tag{3.25}$$

Proposition 3.34. *Fractional Brownian motion has the following properties.*

i. When $H = \frac{1}{2}$, $W_t^{\frac{1}{2}}$ becomes the standard Brownian motion.

ii. $W_0^H = 0$, $\mathbb{E}W_t^H = 0$, $\mathbb{E}(W_t^H)^2 = |t|^{2H}$, $t \geqslant 0$.

iii. It has stationary increments, $\mathbb{E}(W_t^H - W_s^H)^2 = |t-s|^{2H}$.

iv. It has the following self similarity property

$$(W_{\alpha t}^H,\, t \geqslant 0) = (\alpha^H W_t^H,\, t \geqslant 0), \quad \alpha > 0, \tag{3.26}$$

where the equivalence is in law.

Proof. See Exercise 19. ☐

The Poisson Process

Definition 3.35. *The Poisson process with intensity λ, denoted by $N(t)$, is an integer-valued, continuous time, stochastic process with independent increments satisfying*

$$\mathbb{P}[(N(t) - N(s)) = k] = \frac{e^{-\lambda(t-s)}\big(\lambda(t-s)\big)^k}{k!}, \quad t > s \geqslant 0,\ k \in \mathbb{N}.$$

The Poisson process does not have a continuous modification. See Exercise 20.

3.5 The Karhunen-Loéve Expansion

Let $f \in L^2(\Omega)$ where Ω is a subset of $\mathbb{R}^d$ and let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis in $L^2(\Omega)$. Then, it is well known that f can be written as a series expansion:

$$f = \sum_{n=1}^{\infty} f_n e_n,$$

where

$$f_n = \int_{\Omega} f(x) e_n(x)\, dx.$$

The convergence is in $L^2(\Omega)$:

$$\lim_{N \to \infty} \left\| f(x) - \sum_{n=1}^{N} f_n e_n(x) \right\|_{L^2(\Omega)} = 0.$$

It turns out that we can obtain a similar expansion for an L^2 mean zero process which is continuous in the L^2 sense:

$$\mathbb{E}X_t^2 < +\infty, \quad \mathbb{E}X_t = 0, \quad \lim_{h \to 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0. \tag{3.27}$$

For simplicity we will take $T = [0, 1]$. Let $R(t, s) = \mathbb{E}(X_t X_s)$ be the autocorrelation function. Notice that from (3.27) it follows that $R(t, s)$ is continuous in both t and s (exercise 21).

Let us assume an expansion of the form

$$X_t(\omega) = \sum_{n=1}^{\infty} \xi_n(\omega) e_n(t), \quad t \in [0, 1] \tag{3.28}$$

where $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis in $L^2(0, 1)$. The random variables ξ_n are calculated as

$$\begin{aligned}
\int_0^1 X_t e_k(t)\, dt &= \int_0^1 \sum_{n=1}^{\infty} \xi_n e_n(t) e_k(t)\, dt \\
&= \sum_{n=1}^{\infty} \xi_n \delta_{nk} = \xi_k,
\end{aligned}$$

where we assumed that we can interchange the summation and integration. We will assume that these random variables are orthogonal:

$$\mathbb{E}(\xi_n \xi_m) = \lambda_n \delta_{nm},$$

where $\{\lambda_n\}_{n=1}^{\infty}$ are positive numbers that will be determined later.

Assuming that an expansion of the form (3.28) exists, we can calculate

$$\begin{aligned} R(t,s) = \mathbb{E}(X_t X_s) &= \mathbb{E}\left(\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\xi_k e_k(t)\xi_\ell e_\ell(s)\right) \\ &= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\mathbb{E}\left(\xi_k\xi_\ell\right)e_k(t)e_\ell(s) \\ &= \sum_{k=1}^{\infty}\lambda_k e_k(t)e_k(s). \end{aligned}$$

Consequently, in order to the expansion (3.28) to be valid we need

$$R(t,s) = \sum_{k=1}^{\infty}\lambda_k e_k(t)e_k(s). \tag{3.29}$$

From equation (3.29) it follows that

$$\begin{aligned} \int_0^1 R(t,s)e_n(s)\,ds &= \int_0^1\sum_{k=1}^{\infty}\lambda_k e_k(t)e_k(s)e_n(s)\,ds \\ &= \sum_{k=1}^{\infty}\lambda_k e_k(t)\int_0^1 e_k(s)e_n(s)\,ds \\ &= \sum_{k=1}^{\infty}\lambda_k e_k(t)\delta_{kn} \\ &= \lambda_n e_n(t). \end{aligned}$$

Hence, in order for the expansion (3.28) to be valid, $\{\lambda_n,\, e_n(t)\}_{n=1}^{\infty}$ have to be the eigenvalues and eigenfunctions of the integral operator whose kernel is the correlation function of X_t:

$$\int_0^1 R(t,s)e_n(s)\,ds = \lambda_n e_n(t). \tag{3.30}$$

Hence, in order to prove the expansion (3.28) we need to study the eigenvalue problem for the integral operator $\mathcal{R} : L^2[0,1] \mapsto L^2[0,1]$. It easy to check that this operator is self-adjoint ($(\mathcal{R}f,h) = (f,\mathcal{R}h)$ for all $f,\, h \in L^2(0,1)$) and nonnegative ($\mathcal{R}f,f \geqslant 0$ for all $f \in L^2(0,1)$). Hence, all its eigenvalues are real and nonnegative. Furthermore, it is a compact operator (if $\{\phi_n\}_{n=1}^{\infty}$ is a bounded sequence in $L^2(0,1)$, then $\{\mathcal{R}\phi_n\}_{n=1}^{\infty}$ has a convergent subsequence). The spectral theorem for compact, self-adjoint operators implies that $\mathcal{R}$ has a countable sequence of eigenvalues tending to 0. Furthermore, for every $f \in L^2(0,1)$ we can write

$$f = f_0 + \sum_{n=1}^{\infty} f_n e_n(t),$$

where $\mathcal{R} f_0 = 0$, $\{e_n(t)\}$ are the eigenfunctions of $\mathcal{R}$ corresponding to nonzero eigenvalues and the convergence is in L^2. Finally, Mercer's Theorem states that for $R(t,s)$ continuous on $[0,1] \times [0,1]$, the expansion (3.29) is valid, where the series converges absolutely and uniformly.

Now we are ready to prove (3.28).

Theorem 3.36. *(Karhunen-Loéve). Let $\{X_t,\, t \in [0,1]\}$ be an L^2 process with zero mean and continuous correlation function $R(t,s)$. Let $\{\lambda_n,\, e_n(t)\}_{n=1}^{\infty}$ be the eigenvalues and eigenfunctions of the operator $\mathcal{R}$ defined in* (3.36)*. Then*

$$X_t = \sum_{n=1}^{\infty} \xi_n e_n(t), \quad t \in [0,1], \tag{3.31}$$

where

$$\xi_n = \int_0^1 X_t e_n(t)\, dt, \quad \mathbb{E}\xi_n = 0, \quad \mathbb{E}(\xi_n \xi_m) = \lambda \delta_{nm}. \tag{3.32}$$

The series converges in L^2 to $X(t)$, uniformly in t.

Proof. The fact that $\mathbb{E}\xi_n = 0$ follows from the fact that X_t is mean zero. The orthogonality of the random variables $\{\xi_n\}_{n=1}^{\infty}$ follows from the orthogonality of the eigenfunctions of $\mathcal{R}$:

$$\begin{aligned}
\mathbb{E}(\xi_n \xi_m) &= \mathbb{E} \int_0^1 \int_0^1 X_t X_s e_n(t) e_m(s)\, dt ds \\
&= \int_0^1 \int_0^1 R(t,s) e_n(t) e_m(s)\, ds dt \\
&= \lambda_n \int_0^1 e_n(s) e_m(s)\, ds \\
&= \lambda_n \delta_{nm}.
\end{aligned}$$

Consider now the partial sum $S_N = \sum_{n=1}^N \xi_n e_n(t)$.

$$\begin{aligned}
\mathbb{E}|X_t - S_N|^2 &= \mathbb{E}X_t^2 + \mathbb{E}S_N^2 - 2\mathbb{E}(X_t S_N) \\
&= R(t,t) + \mathbb{E}\sum_{k,\ell=1}^N \xi_k \xi_\ell e_k(t) e_\ell(t) - 2\mathbb{E}\left(X_t \sum_{n=1}^N \xi_n e_n(t)\right) \\
&= R(t,t) + \sum_{k=1}^N \lambda_k |e_k(t)|^2 - 2\mathbb{E}\sum_{k=1}^N \int_0^1 X_t X_s e_k(s) e_k(t)\, ds \\
&= R(t,t) - \sum_{k=1}^N \lambda_k |e_k(t)|^2 \to 0,
\end{aligned}$$

by Mercer's theorem. □

Remark 3.37. *Let X_t be a Gaussian second order process with continuous covariance $R(t,s)$. Then the random variables $\{\xi_k\}_{k=1}^\infty$ are Gaussian, since they are defined through the time integral of a Gaussian processes. Furthermore, since they are Gaussian and orthogonal, they are also independent. Hence, for Gaussian processes the Karhunen-Loéve expansion becomes:*

$$X_t = \sum_{k=1}^{+\infty} \sqrt{\lambda_k}\xi_k e_k(t), \tag{3.33}$$

where $\{\xi_k\}_{k=1}^\infty$ are independent $\mathcal{N}(0,1)$ random variables.

Example 3.38. *The Karhunen-Loéve Expansion for Brownian Motion. The correlation function of Brownian motion is $R(t,s) = \min(t,s)$. The eigenvalue problem $\mathcal{R}\psi_n = \lambda_n \psi_n$ becomes*

$$\int_0^1 \min(t,s)\psi_n(s)\, ds = \lambda_n \psi_n(t).$$

Let us assume that $\lambda_n > 0$ (it is easy to check that 0 is not an eigenvalue). Upon setting $t = 0$ we obtain $\psi_n(0) = 0$. The eigenvalue problem can be rewritten in the form

$$\int_0^t s\psi_n(s)\, ds + t\int_t^1 \psi_n(s)\, ds = \lambda_n \psi_n(t).$$

We differentiate this equation once:

$$\int_t^1 \psi_n(s)\, ds = \lambda_n \psi_n'(t).$$

We set $t = 1$ in this equation to obtain the second boundary condition $\psi_n'(1) = 0$. A second differentiation yields;

$$-\psi_n(t) = \lambda_n \psi_n''(t),$$

where primes denote differentiation with respect to t. Thus, in order to calculate the eigenvalues and eigenfunctions of the integral operator whose kernel is the covariance function of Brownian motion, we need to solve the Sturm-Liouville problem

$$-\psi_n(t) = \lambda_n \psi_n''(t), \quad \psi(0) = \psi'(1) = 0.$$

It is easy to check that the eigenvalues and (normalized) eigenfunctions are

$$\psi_n(t) = \sqrt{2} \sin\left(\frac{1}{2}(2n-1)\pi t\right), \quad \lambda_n = \left(\frac{2}{(2n-1)\pi}\right)^2.$$

Thus, the Karhunen-Loéve expansion of Brownian motion on $[0,1]$ is

$$W_t = \sqrt{2} \sum_{n=1}^{\infty} \xi_n \frac{2}{(2n-1)\pi} \sin\left(\frac{1}{2}(2n-1)\pi t\right). \tag{3.34}$$

We can use the KL expansion in order to study the L^2-regularity of stochastic processes. First, let R be a compact, symmetric positive definite operator on $L^2(0,1)$ with eigenvalues and normalized eigenfunctions $\{\lambda_k,\, e_k(x)\}_{k=1}^{+\infty}$ and consider a function $f \in L^2(0,1)$ with $\int_0^1 f(s)\,ds = 0$. We can define the one parameter family of Hilbert spaces H^α through the norm

$$\|f\|_\alpha^2 = \|R^{-\alpha} f\|_{L^2}^2 = \sum_k |f_k|^2 \lambda^{-\alpha}.$$

The inner product can be obtained through polarization. This norm enables us to measure the regularity of the function $f(t)$.[3] Let X_t be a mean zero second order (i.e. with finite second moment) process with continuous autocorrelation function. Define the space $\mathcal{H}^\alpha := L^2((\Omega, P), H^\alpha(0,1))$ with (semi)norm

$$\|X_t\|_\alpha^2 = \mathbb{E}\|X_t\|_{H^\alpha}^2 = \sum_k |\lambda_k|^{1-\alpha}. \tag{3.35}$$

Notice that the regularity of the stochastic process X_t depends on the decay of the eigenvalues of the integral operator $\mathcal{R}\cdot := \int_0^1 R(t,s)\cdot\, ds$.

[3] Think of R as being the inverse of the Laplacian with periodic boundary conditions. In this case H^α coincides with the standard fractional Sobolev space.

As an example, consider the L^2-regularity of Brownian motion. From Example 3.38 we know that $\lambda_k \sim k^{-2}$. Consequently, from (3.35) we get that, in order for W_t to be an element of the space $\mathcal{H}^\alpha$, we need that

$$\sum_k |\lambda_k|^{-2(1-\alpha)} < +\infty,$$

from which we obtain that $\alpha < 1/2$. This is consistent with the Hölder continuity of Brownian motion from Theorem 3.29. [4]

3.6 Discussion and Bibliography

The Ornstein-Uhlenbeck process was introduced by Ornstein and Uhlenbeck in 1930 as a model for the velocity of a Brownian particle [25].

The kind of analysis presented in Section 3.2.3 was initiated by G.I. Taylor in [24]. The proof of Bochner's theorem 3.16 can be found in [14], where additional material on stationary processes can be found. See also [11].

The spectral theorem for compact, self-adjoint operators which was needed in the proof of the Karhunen-Loéve theorem can be found in [21]. The Karhunen-Loeve expansion is also valid for random fields. See [22] and the reference therein.

3.7 Exercises

1. Let $Y_0, Y_1, \dots$ be a sequence of independent, identically distributed random variables and consider the stochastic process $X_n = Y_n$.

   (a) Show that X_n is a strictly stationary process.

   (b) Assume that $\mathbb{E}Y_0 = \mu < +\infty$ and $\mathbb{E}Y_0^2 = sigma^2 < +\infty$. Show that

   $$\lim_{N\to+\infty} \mathbb{E}\left|\frac{1}{N}\sum_{j=0}^{N-1} X_j - \mu\right| = 0.$$

   (c) Let f be such that $\mathbb{E}f^2(Y_0) < +\infty$. Show that

   $$\lim_{N\to+\infty} \mathbb{E}\left|\frac{1}{N}\sum_{j=0}^{N-1} f(X_j) - f(Y_0)\right| = 0.$$

[4] Notice, however, that Wiener's theorem refers to a.s. Hölder continuity, whereas the calculation presented in this section is about L^2-continuity.

2. Let Z be a random variable and define the stochastic process $X_n = Z$, $n = 0, 1, 2, \dots$. Show that X_n is a strictly stationary process.

3. Let $A_0, A_1, \dots A_m$ and $B_0, B_1, \dots B_m$ be uncorrelated random variables with mean zero and variances $\mathbb{E}A_i^2 = \sigma_i^2$, $\mathbb{E}B_i^2 = \sigma_i^2$, $i = 1, \dots m$. Let $\omega_0, \omega_1, \dots \omega_m \in [0, \pi]$ be distinct frequencies and define, for $n = 0, \pm 1, \pm 2, \dots$, the stochastic process

$$X_n = \sum_{k=0}^{m} \Big(A_k \cos(n\omega_k) + B_k \sin(n\omega_k) \Big).$$

Calculate the mean and the covariance of X_n. Show that it is a weakly stationary process.

4. Let $\{\xi_n : n = 0, \pm 1, \pm 2, \dots\}$ be uncorrelated random variables with $\mathbb{E}\xi_n = \mu$, $\mathbb{E}(\xi_n - \mu)^2 = \sigma^2$, $n = 0, \pm 1, \pm 2, \dots$. Let $a_1, a_2, \dots$ be arbitrary real numbers and consider the stochastic process

$$X_n = a_1 \xi_n + a_2 \xi_{n-1} + \dots a_m \xi_{n-m+1}.$$

   (a) Calculate the mean, variance and the covariance function of X_n. Show that it is a weakly stationary process.

   (b) Set $a_k = 1/\sqrt{m}$ for $k = 1, \dots m$. Calculate the covariance function and study the cases $m = 1$ and $m \to +\infty$.

5. Let $W(t)$ be a standard one dimensional Brownian motion. Calculate the following expectations.

   (a) $\mathbb{E}e^{iW(t)}$.

   (b) $\mathbb{E}e^{i(W(t)+W(s))}$, $t, s, \in (0, +\infty)$.

   (c) $\mathbb{E}(\sum_{i=1}^n c_i W(t_i))^2$, where $c_i \in \mathbb{R}$, $i = 1, \dots n$ and $t_i \in (0, +\infty)$, $i = 1, \dots n$.

   (d) $\mathbb{E}e^{\left[i\left(\sum_{i=1}^n c_i W(t_i)\right)\right]}$, where $c_i \in \mathbb{R}$, $i = 1, \dots n$ and $t_i \in (0, +\infty)$, $i = 1, \dots n$.

6. Let W_t be a standard one dimensional Brownian motion and define

$$B_t = W_t - tW_1, \quad t \in [0, 1].$$

(a) Show that B_t is a Gaussian process with

$$\mathbb{E}B_t = 0, \quad \mathbb{E}(B_t B_s) = \min(t,s) - ts.$$

(b) Show that, for $t \in [0,1)$ an equivalent definition of B_t is through the formula

$$B_t = (1-t)W\left(\frac{t}{1-t}\right).$$

(c) Calculate the distribution function of B_t.

7. Let X_t be a mean-zero second order stationary process with autocorrelation function

$$R(t) = \sum_{j=1}^{N} \frac{\lambda_j^2}{\alpha_j} e^{-\alpha_j |t|},$$

where $\{\alpha_j,\, \lambda_j\}_{j=1}^N$ are positive real numbers.

(a) Calculate the spectral density and the correlaction time of this process.

(b) Show that the assumptions of Theorem 3.21 are satisfied and use the argument presented in Section 3.2.3 (i.e. the Green-Kubo formula) to calculate the diffusion coefficient of the process $Z_t = \int_0^t X_s \, ds$.

(c) Under what assumptions on the coefficients $\{\alpha_j,\, \lambda_j\}_{j=1}^N$ can you study the above questions in the limit $N \to +\infty$?

8. Prove Lemma 3.20.

9. Let $a_1, \dots a_n$ and $s_1, \dots s_n$ be positive real numbers. Calculate the mean and variance of the random variable

$$X = \sum_{i=1}^{n} a_i W(s_i).$$

10. Let $W(t)$ be the standard one-dimensional Brownian motion and let $\sigma,\, s_1,\, s_2 > 0$. Calculate

(a) $\mathbb{E}e^{\sigma W(t)}$.

(b) $\mathbb{E}\big(\sin(\sigma W(s_1)) \sin(\sigma W(s_2))\big)$.

11. Let W_t be a one dimensional Brownian motion and let $\mu, \sigma > 0$ and define

$$S_t = e^{t\mu + \sigma W_t}.$$

   (a) Calculate the mean and the variance of S_t.

   (b) Calculate the probability density function of S_t.

12. Use Theorem 3.27 to prove Lemma 3.26.

13. Prove Theorem 3.30.

14. Use Lemma 3.31 to calculate the distribution function of the stationary Ornstein-Uhlenbeck process.

15. Calculate the mean and the correlation function of the integral of a standard Brownian motion

$$Y_t = \int_0^t W_s \, ds.$$

16. Show that the process

$$Y_t = \int_t^{t+1} (W_s - W_t) \, ds, \quad t \in \mathbb{R},$$

   is second order stationary.

17. Let $V_t = e^{-t} W(e^{2t})$ be the stationary Ornstein-Uhlenbeck process. Give the definition and study the main properties of the Ornstein-Uhlenbeck bridge.

18. The autocorrelation function of the velocity $Y(t)$ a Brownian particle moving in a harmonic potential $V(x) = \frac{1}{2}\omega_0^2 x^2$ is

$$R(t) = e^{-\gamma|t|} \Big(cos(\delta|t|) - \frac{1}{\delta} \sin(\delta|t|) \Big),$$

   where γ is the friction coefficient and $\delta = \sqrt{\omega_0^2 - \gamma^2}$.

   (a) Calculate the spectral density of $Y(t)$.

   (b) Calculate the mean square displacement $\mathbb{E}(X(t))^2$ of the position of the Brownian particle $X(t) = \int_0^t Y(s) \, ds$. Study the limit $t \to +\infty$.

19. Show the scaling property (3.26) of the fractional Brownian motion.

20. Use Theorem (3.27) to show that there does not exist a continuous modification of the Poisson process.

21. Show that the correlation function of a process X_t satisfying (3.27) is continuous in both t and s.

22. Let X_t be a stochastic process satisfying (3.27) and $R(t, s)$ its correlation function. Show that the integral operator $\mathcal{R} : L^2[0, 1] \mapsto L^2[0, 1]$

$$\mathcal{R}f := \int_0^1 R(t, s)f(s)\, ds \tag{3.36}$$

is self-adjoint and nonnegative. Show that all of its eigenvalues are real and nonnegative. Show that eigenfunctions corresponding to different eigenvalues are orthogonal.

23. Let H be a Hilbert space. An operator $\mathcal{R} : H \to H$ is said to be Hilbert–Schmidt if there exists a complete orthonormal sequence $\{\phi_n\}_{n=1}^{\infty}$ in H such that

$$\sum_{n=1}^{\infty} \|\mathcal{R}e_n\|^2 < \infty.$$

Let $\mathcal{R} : L^2[0, 1] \mapsto L^2[0, 1]$ be the operator defined in (3.36) with $R(t, s)$ being continuous both in t and s. Show that it is a Hilbert-Schmidt operator.

24. Let X_t a mean zero second order stationary process defined in the interval $[0, T]$ with continuous covariance $R(t)$ and let $\{\lambda_n\}_{n=1}^{+\infty}$ be the eigenvalues of the covariance operator. Show that

$$\sum_{n=1}^{\infty} \lambda_n = T\, R(0).$$

25. Calculate the Karhunen-Loeve expansion for a second order stochastic process with correlation function $R(t, s) = ts$.

26. Calculate the Karhunen-Loeve expansion of the Brownian bridge on $[0, 1]$.

27. Let $X_t,\ t \in [0, T]$ be a second order process with continuous covariance and Karhunen-Loéve expansion

$$X_t = \sum_{k=1}^{\infty} \xi_k e_k(t).$$

Define the process

$$Y(t) = f(t) X_{\tau(t)}, \quad t \in [0, S],$$

where $f(t)$ is a continuous function and $\tau(t)$ a continuous, nondecreasing function with $\tau(0) = 0$, $\tau(S) = T$. Find the Karhunen-Loéve expansion of $Y(t)$, in an appropriate weighted L^2 space, in terms of the KL expansion of X_t. Use this in order to calculate the KL expansion of the Ornstein-Uhlenbeck process.

28. Calculate the Karhunen-Loéve expansion of a centered Gaussian stochastic process with covariance function $R(s,t) = \cos(2\pi(t-s))$.

29. Use the Karhunen-Loeve expansion to generate paths of the

    (a) Brownian motion on $[0,1]$.

    (b) Brownian bridge on $[0,1]$.

    (c) Ornstein-Uhlenbeck on $[0,1]$.

    Study computationally the convergence of the KL expansion for these processes. How many terms do you need to keep in the KL expansion in order to calculate accurate statistics of these processes?

Index

Bibliography

[1] R. Balescu. *Statistical dynamics. Matter out of equilibrium*. Imperial College Press, London, 1997.

[2] L. Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Corrected reprint of the 1968 original.

[3] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15(1):1–89, Jan 1943.

[4] A.J. Chorin and O.H. Hald. *Stochastic tools in mathematics and science*, volume 1 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2006.

[5] J. L. Doob. The Brownian movement and stochastic equations. *Ann. of Math. (2)*, 43:351–369, 1942.

[6] N. Wax (editor). *Selected Papers on Noise and Stochastic Processes*. Dover, New York, 1954.

[7] A. Einstein. *Investigations on the theory of the Brownian movement*. Dover Publications Inc., New York, 1956. Edited with notes by R. Fürth, Translated by A. D. Cowper.

[8] W. Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.

[9] W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.

[10] J. Jacod and A.N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2003.

[11] S. Karlin and H.M. Taylor. *A first course in stochastic processes*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[12] L. B. Koralov and Y. G. Sinai. *Theory of probability and random processes*. Universitext. Springer, Berlin, second edition, 2007.

[13] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.

[14] N. V. Krylov. *Introduction to the theory of diffusion processes*, volume 142 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1995.

[15] P. D. Lax. *Linear algebra and its applications*. Pure and Applied Mathematics (Hoboken). Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2007.

[16] M. Loève. *Probability theory. I*. Springer-Verlag, New York, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.

[17] M. Loève. *Probability theory. II*. Springer-Verlag, New York, fourth edition, 1978. Graduate Texts in Mathematics, Vol. 46.

[18] R.M. Mazo. *Brownian motion*, volume 112 of *International Series of Monographs on Physics*. Oxford University Press, New York, 2002.

[19] E. Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, Princeton, N.J., 1967.

[20] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1992.

[21] Frigyes Riesz and Béla Sz.-Nagy. *Functional analysis*. Dover Publications Inc., New York, 1990. Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original.

[22] C. Schwab and R.A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.

[23] D.W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, 1993.

[24] G. I. Taylor. Diffusion by continuous movements. *London Math. Soc.*, 20:196, 1921.

[25] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the Brownian motion. *Phys. Rev.*, 36(5):823–841, Sep 1930.