Introduction to Stochastic Calculus

Nagoya University, Spring 2024

Lecturer: Serge Richard

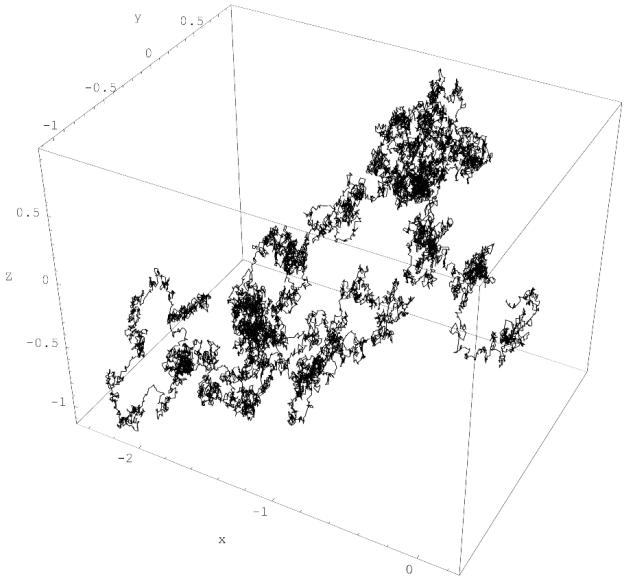

Figure 1: Brownian path in 3D

Website for this course:
http://www.math.nagoya-u.ac.jp/~richard/Stochastic.html

Comments or corrections are welcome:
richard@math.nagoya-u.ac.jp

Contents

Motivation

Our aim is to develop tools for the study of very irregular functions or curves and for their predictions. The two leading applications are the Brownian motion and the Black-Scholes model. Figures 1 and 2 are typical examples of curves that should be understood and described with the content of this course.

These notes and the corresponding course have been mainly inspired by the book [1], with additional material borrowed from [2], [6], and [13]. Other references will be mentioned on due time. The probability part is based on [8].

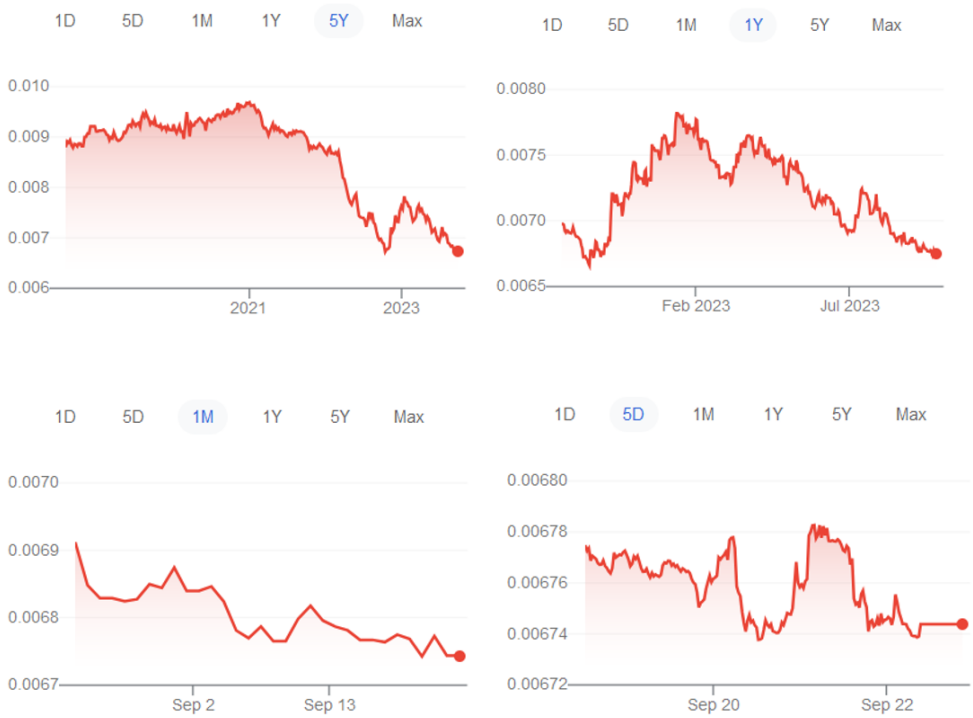

Figure 2: Exchange rate: Japanese yen - US dollar (y-axis gives the value in dollar of 1 yen, as a function of time) over different periods of time

Chapter 1

Mathematical Background

This chapter contains the mathematical background necessary for the understanding of concepts in stochastic calculus.

1.1 Probability and Random Variables

The aim of this section is to describe and quantify any non-predictable experiment. We give a framework suitable for many applications.

Definition 1.1.1 (Measurable space)**.** *A* measurable space $(\Omega,\mathcal{F})$ *is a set* Ω *together with a collection of subsets* $\mathcal{F}$ *closed under complement, countable unions and countable intersections: if* $A \in \mathcal{F}$*,* $A^c := \Omega\backslash A \in \mathcal{F}$*, if* $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{F}$*, then* $\cup_j A_j \in \mathcal{F}$ *and* $\cap_j A_j \in \mathcal{F}$*. One also says that* $\mathcal{F}$ *is a* σ-algebra*.*

Note that we always assume $\mathcal{F}$ to be non-empty, which means that there exists at least one element $A \in \mathcal{F}$. In this case, A^c also belongs to $\mathcal{F}$, and $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$ are also elements of $\mathcal{F}$.

Exercise 1.1.2. *Prove this statement: if* $\mathcal{F}$ *is a collection of subsets which is closed under complement and countable unions, then it is closed under countable intersections.*

An example of a measurable space is the usual space $\mathbb{R}^N$ together with the family of sets generated by intervals by considering countable unions, intersections, and complements. In this case, one speaks about the Borel σ-algebra σ_B. Thus, $(\mathbb{R}^N, \sigma_B)$ is the most common measurable space, and one usually denotes it simply by $\mathbb{R}^N$. An other example of a measurable space is provided by $\Omega = \{\lambda_1, \ldots, \lambda_N\}$ a finite set and $\mathcal{F}$ the *power set* of Ω consisting of all subsets of Ω. Two standard examples are

$$\Omega = \{\text{heads, tails}\} \quad \text{or} \quad \Omega = \Big\{ \boxdot, \boxdot, \boxdot, \boxdot, \boxdot, \boxdot \Big\}.$$

This second example can also be extended to an infinite set, like for example $\Omega = \mathbb{N}$ or $\Omega = \mathbb{Z}$, also endowed with their respective power set.

Exercise 1.1.3. *If* Ω *contains N elements, how many elements does its power set contain ? Provide an easy and understandable description of this power set.*

Definition 1.1.4 (Measurable function)**.** *For two measurable spaces* $(\Omega, \mathcal{F})$ *and* $(\Lambda, \mathcal{E})$*, a function* $f : \Omega \to \Lambda$ *is* measurable *if for any* $A \in \mathcal{E}$*, the set*

$$f^{-1}(A) := \{\omega \in \Omega \mid f(\omega) \in A\}$$

belongs to $\mathcal{F}$.

In simpler words, the function f is measurable if it transports back the structure of $(\Lambda, \mathcal{E})$ to the structure of $(\Omega, \mathcal{F})$. The set $f^{-1}(A)$ is called the *pre-image* of A by f, and it does not mean that f is injective. The measurability requirement is a very weak assumption, but it is usually the minimum requirement for being able to do anything with a function between two measurable spaces.

Let us now add one more structure on the measurable space $(\Omega, \mathcal{F})$.

Definition 1.1.5 (Probability space)**.** *A* probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *consists of a measurable space* $(\Omega, \mathcal{F})$ *and a function* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *satisfying* $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$ *and*

$$\mathbb{P}\Big(\bigcup_{j\in\mathbb{N}} A_j\Big) = \sum_{j\in\mathbb{N}} \mathbb{P}(A_j)$$

whenever $A_j \cap A_k = \emptyset \ \forall j \neq k$. *We call* Ω the sample space, $\mathcal{F}$ the event space, $\omega \in \Omega$ an elementary event *and* $A \in \mathcal{F}$ an event, *and finally* $\mathbb{P}$ the probability measure.

The function $\mathbb{P}$ should be thought at a way to measure the size of the elements of $\mathcal{F}$, or as a way to weight them.

Exercise 1.1.6. *If* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and if* $A, B \in \mathcal{F}$*, check that*

1) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*, where* $A^c := \Omega \setminus A$,

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

3) If $A \subset B$*, then* $\mathbb{P}(A) \leq \mathbb{P}(B)$.

The following statement will often be used, and is related to the monotone convergence theorem.

Lemma 1.1.7 (Continuity of probability)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and consider* $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{F}$*. If* $A_j \subset A_{j+1}$ *for any* j*, then*

$$\mathbb{P}\Big(\bigcup_{j\in\mathbb{N}} A_j\Big) = \lim_{j\to\infty} \mathbb{P}(A_j),$$

while if $A_j \supset A_{j+1}$ *for any* j*, then*

$$\mathbb{P}\Big(\bigcap_{j\in\mathbb{N}} A_j\Big) = \lim_{j\to\infty} \mathbb{P}(A_j).$$

The proof is left as an exercise, or can be found in [1, Lem. 1.4].

Usually, Ω is very complicated or unknown. Functions defined on Ω are more important than Ω itself. The following definition is a very general one, but in applications the measurable space $(\Lambda, \mathcal{E})$ will be chosen conveniently.

Definition 1.1.8 (Random variable)**.** *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a measurable space* $(\Lambda, \mathcal{E})$*.* A random variable X *is a measurable function from* $(\Omega, \mathcal{F})$ *to* $(\Lambda, \mathcal{E})$*. Namely,* X *is a function satisfying, for any* $A \in \mathcal{E}$*,*

$$\{\omega \in \Omega \mid X(\omega) \in A\} \equiv X^{-1}(A) \in \mathcal{F}. \tag{1.1.1}$$

Usually, we choose a measurable space $(\Lambda, \mathcal{E})$ which is much simpler than the initial measurable space $(\Omega, \mathcal{F})$. Note that in the special case $\Lambda = \mathbb{R}^N$, or more precisely if we consider $(\Lambda, \mathcal{E}) = (\mathbb{R}^N, \sigma_{\mathrm{B}})$, then (1.1.1) is satisfied if $X = (X_1, \ldots, X_N)$ verifies

$$\{\omega \in \Omega \mid X_j(\omega) \leq x_j \ \forall j = 1, \ldots, N\} \in \mathcal{F}$$

for any $(x_1, \dots, x_N) \in \mathbb{R}^N$. The special case $N = 1$ corresponds to a *univariate* random variables, while $N > 1$ corresponds to a *multivariate* random variables, also called *random vectors* or *vector valued random variables*.

So far we have not used probability measure $\mathbb{P}$ in the previous definition. The interest of the previous definition is coming in the notion of induced probability measure, a new measure on Λ.

Definition 1.1.9 (Induced probability measure)**.** *When $X : \Omega \to \Lambda$ is a random variable from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(\Lambda, \mathcal{E})$, the map $\mu_X : \mathcal{E} \to [0, 1]$ defined for any $A \in \mathcal{E}$ by*

$$\mu_X(A) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}) = \mathbb{P}\left(X^{-1}(A)\right) \equiv \mathbb{P}(X \in A)$$

is called the induced probability measure. *μ_X is also called* the law *of X, and we write $X \sim \mu_X$ for this correspondence.*

It is important to observe that μ_X is defined on the image of Ω, which can be considered as the set of outcomes of X. Usually, μ_X is much simpler than X, since it is defined on $(\Lambda, \mathcal{E})$, as for example on $(\mathbb{R}^N, \sigma_{\mathrm{B}})$, and not on $(\Omega, \mathcal{F})$. The term *probability distribution* is often used instead of probability measure, but this terminology is less precise since it also refers to other concepts. Let us mention that there exist two principal types of random variables (but others also exist). In the sequel, we shall write $\mathbb{R}_+$ for $[0, \infty)$.

Definition 1.1.10 (Absolutely continuous random variable)**.** *The random variable $X : \Omega \to \mathbb{R}^N$ is* absolutely continuous *if the induced probability measure is absolutely continuous with respect to the Lebesgue measure, namely if there exists a (measurable) function $\Pi_X : \mathbb{R}^N \to \mathbb{R}_+$ satisfying for any $A \in \sigma_{\mathrm{B}}$*

$$\mu_X(A) = \int_A \Pi_X(x) \, \mathrm{d}x.$$

The function Π_X is called the probability density function*, or simply the* pdf.

Definition 1.1.11 (Discrete valued random variable)**.** *The random variable $X : \Omega \to \Lambda$ is* discrete valued *if $X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$ is finite or countable. In this case, we define the function $p_X : X(\Omega) \to [0, 1]$ by*

$$p_X(x) := \mathbb{P}\left(X^{-1}(\{x\})\right)$$

for any $x \in X(\Omega)$. The function p_X is called the the probability mass function*, or simply* pmf.

In these two situations, we still write $X \sim \Pi_X$ or $X \sim p_X$. It is clear that the following properties hold: $\int_{\mathbb{R}^N} \Pi_X(x) \, \mathrm{d}x = 1$ and $\sum_{x \in X(\Omega)} p_X(x) = 1$. Observe also that for any absolutely continuous random variable X, one has $\mu_X(x) = 0$ for any $x \in \mathbb{R}^N$ while $\Pi_X(x) \in \mathbb{R}_+$ for (almost every) $x \in \mathbb{R}^N$.

Remark 1.1.12. *Any function $\Pi : \mathbb{R}^N \to \mathbb{R}_+$ satisfying $\int \Pi(x) \, \mathrm{d}x = 1$, or any function p from a finite set or a countable set Λ to $[0, 1]$ satisfying $\sum_x p(x) = 1$, defines the pdf or the pmf of a random variable. However, in such a situation we don't have the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we just have the law. It means that we have a rule to associate a weight to any subset of the set of outcomes, but we don't know the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (if necessary, such a probability space can be constructed, but it is somewhat artificial). Nevertheless, we shall say that Π or p define a random variable X with induced probability measure Π or p, and refer to Π and p as a* probability distribution*.*

Exercise 1.1.13 (Classical probability distributions)**.** *Recall the definition of a few classical probability distributions, and recast them in the framework and with the terminology introduced above. For example, consider the Bernoulli distribution, the binomial distribution, the Poisson distribution, the uniform distribution, the exponential distribution, etc.*

In the important case of univariate random variable, namely when $(\Lambda, \mathcal{E}) = (\mathbb{R}, \sigma_{\mathrm{B}})$, one more function can be defined.

Definition 1.1.14 (Cumulative distribution function)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $(\mathbb{R}, \sigma_{\mathrm{B}})$. The* cumulative distribution function *F_X is defined for any $x \in \mathbb{R}$ by*

$$F_X(x) := \mathbb{P}(X \leq x) \equiv \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq x\}).$$

One easily observes that $\lim_{x \to -\infty} F_X(x) = 0$ while $\lim_{x \to \infty} F_X(x) = 1$. The function F_X is also increasing and right-continuous, meaning that $\lim_{\varepsilon \searrow 0} F_X(x + \varepsilon) = F_X(x)$, but it is not left-continuous in general. Note also that this function can be defined because there exists an order on $\mathbb{R}$, which means that the notation $X(\omega) \leq x$ is meaningful. In an arbitrary measurable space $(\Lambda, \mathcal{E})$, this notion would be meaningless.

Exercise 1.1.15. *Prove the above statements for the cumulative distribution function, and provide a example of a cumulative distribution function which is not left-continuous.*

Exercise 1.1.16. *For the classical probability distributions mentioned in Exercise 1.1.13, determine the cumulative distribution function.*

1.2 Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(\Lambda, \mathcal{E})$ and $(\Xi, \mathcal{G})$ be two measurable spaces. Let $X : \Omega \to \Lambda$ be a random variable, and consider $f : \Lambda \to \Xi$ be a measurable function, as introduced in Definition 1.1.4. Then it is easy to check that the composition $f \circ X : \Omega \to \Xi$ defines a new random variable, simply denoted by $f(X)$.

For the following definition, we shall consider only some special instances of measurable spaces. More precisely let us call *standard* a measurable space $(\Xi, \mathcal{G})$ with Ξ either a finite or a countable subset of $\mathbb{R}^N$ endowed with their power set, or $(\mathbb{R}, \sigma_{\mathrm{B}})$, or $(\mathbb{R}^N, \sigma_{\mathrm{B}})$, or the set $M_{n \times m}(\mathbb{R})$ of $n \times m$ matrices with entries in $\mathbb{R}$ (since $M_{n \times m}(\mathbb{R})$ can be identified with $\mathbb{R}^{nm}$, it is also a measurable space with a suitable Borel σ-algebra).

Definition 1.2.1 (Expectation)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(\Lambda, \mathcal{E})$ and $(\Xi, \mathcal{G})$ be measurable spaces and assume $(\Xi, \mathcal{G})$ be standard, let $X : \Omega \to \Lambda$ be a random variable, and let $f : \Lambda \to \Xi$ be a measurable function. The* expectation *of $f(X)$ is defined by (the Lebesgue type integral)*

$$\mathbb{E}(f(X)) := \int_{\Lambda} f(x)\, \mu_X(\mathrm{d}x). \tag{1.2.1}$$

Note that when writing such an expression, we assume that it exists even with f replaced by $|f|$ (absolute convergence of the integral). If the measurable space $(\Lambda, \mathcal{E})$ is standard and equal to $(\Xi, \mathcal{G})$, and if f denotes the identity function id *with* $\mathrm{id}(x) = x$*, then we simply write $\mathbb{E}(X)$ for $\mathbb{E}(\mathrm{id}(X))$, and call it the* mean value *of X, of the expectation of X.*

Exercise 1.2.2. *Specialize the formula (1.2.1) for any absolutely continuous random variable, as presented in Definition 1.1.10, or for a discrete valued random variable, as presented in Definition 1.1.11, when $\Lambda \subset \mathbb{R}$.*

Exercise 1.2.3. *Why is $\mathbb{E}(X)$ not well defined if X is a random variable from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to an arbitrary measurable space $(\Lambda, \mathcal{E})$, why do we consider only standard measurable spaces ? Is there a more general framework ?*

In addition to the expectation, many standard quantities can be associated with a random variable taking values in $\mathbb{R}$.

Definition 1.2.4 (Variance, standard deviation, moments, moment generating function)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let X be a univariate random variable defined on it. The* variance *of X is defined by*

$$\mathrm{Var}(X) := \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big),$$

the standard deviation of *X is defined by the square root of the variance, the n-moment of X are defined by* $\mathbb{E}(X^n)$*, and the* moment generating function *by the function* $a \mapsto \mathbb{E}(\mathrm{e}^{aX})$.

In the framework of Definition 1.2.1, the variance corresponds to the function f given by $x \mapsto (x - \mathbb{E}(X))^2$, the moments to the functions $x \mapsto x^n$, and the moment generation function to the function $x \mapsto \mathrm{e}^{ax}$. Obviously, these quantities exist only if the corresponding integrals (or sums) converge absolutely. For completeness, we also recall that if the moment generating function $\mathbb{E}(\mathrm{e}^{aX})$ exists for all a with $|a| < \delta$ for some $\delta > 0$, then this function defines uniquely the univariate random variable X, see for example [8, Thm. 7.55].

For any measurable space $(\Lambda, \mathcal{E})$ and for any $A \in \mathcal{E}$, we define the *indicator function* $\mathbf{1}_A$ by $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ if $x \notin A$. Thus, $\mathbf{1}_A : \Lambda \to \mathbb{R}$ is a measurable function, and one observes that the following equalities hold:

$$\mathbb{P}(X \in A) \equiv \mathbb{P}\Big(\{\omega \in \Omega \mid X(\omega) \in A\}\Big) = \mu_X(A) = \int_A \mu_X(\mathrm{d}x) = \int_\Lambda \mathbf{1}_A(x)\, \mu_X(\mathrm{d}x) = \mathbb{E}(\mathbf{1}_A(X)).$$

Exercise 1.2.5. *For* $\sigma > 0$ *and* $\bar{x} \in \mathbb{R}$ *set* $\Pi : \mathbb{R} \to \mathbb{R}_+$ *by*

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

Check that $\int \Pi(x)\,\mathrm{d}x = 1$. *In the framework of Reminder 1.1.12 we write* $X = N(\bar{x}, \sigma^2)$ *for the corresponding univariate random variable, called* Gaussian random variable. *Check that* $\mathbb{E}(X) = \bar{x}$*, and* $\mathrm{Var}(X) = \sigma^2$.

More generally, for $\bar{x} \in \mathbb{R}^N$ *and* $P \in M_{N\times N}(\mathbb{R})$ *with* $P > 0$*, set* $\Pi : \mathbb{R}^N \to \mathbb{R}_+$ *with*

$$\Pi(x) := \frac{1}{(2\pi)^{N/2}|P|^{1/2}} \exp\left(-\frac{1}{2}(x - \bar{x})^T P^{-1}(x - \bar{x})\right),$$

with $|P| := \det(P)$. *Check that* $\int \Pi(x)\,\mathrm{d}x = 1$. *We write* $X = N(\bar{x}, P)$ *for the corresponding multivariate random variable, called* N-dim Gaussian random variable *or* Gaussian vector. *Check that* $\mathbb{E}(X) = \bar{x}$*, and that* $P = \mathbb{E}\left((X - \bar{x})(X - \bar{x})^T\right)$. *Here, P is called the* covariance matrix.

Exercise 1.2.6. *If* $X : \Omega \to \mathbb{R}^N$ *is absolutely continuous with pdf* Π_X *and if* $\phi : \mathbb{R}^N \to \mathbb{R}^N$ *is bijective and* C^∞*, show that* $Y := \phi(X) : \Omega \to \mathbb{R}^N$ *is a new absolutely continuous random variable, with pdf* Π_Y *given by* $\Pi_Y(y) = \Pi_X\left(\phi^{-1}(y)\right)|J_{\phi^{-1}}(y)|$. *Here,* $|J_{\phi^{-1}}(y)|$ *denotes the determinant of the Jacobian matrix of* ϕ^{-1}.

Let us close this section with a few inequalities computed with the expectation. These inequalities hold for univariate random variables.

Lemma 1.2.7 (Markov's inequality)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $X : \Omega \to \mathbb{R}$ *be a non-negative random variable (meaning that* $X(\omega) \geq 0$ *for all* $\omega \in \Omega$*). Then for any* $a > 0$ *the following inequality holds:*

$$\mathbb{P}(X > a) \equiv \mathbb{P}(\{\omega \in \Omega \mid X(\omega) > a\}) \leq \frac{1}{a}\mathbb{E}(X).$$

The proof of this inequality is left as an exercise. Also the following two inequalities can be deduced from it.

Corollary 1.2.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. Then for any $a > 0$ the following inequalities hold:*

$$\mathbb{P}(|X| > a) \leq \frac{1}{a^2}\mathbb{E}(|X|^2) \qquad \textit{Chebyshev's inequality,}$$

and for any $\lambda > 0$

$$\mathbb{P}(X > a) \leq \mathrm{e}^{-\lambda a}\mathbb{E}(\mathrm{e}^{\lambda X}) \qquad \textit{Chernoff's bound.}$$

1.3 Independence

In this section, we consider families of random variables. These random variables will be denoted generically by $\{X^j\}_j$ since each of them could be a multivariate random variable, and therefore have N components: $X^j = (X_1^j, X_2^j, \ldots, X_N^j)^T$. When each random variable is univariate (which means that it takes values in $\mathbb{R}$) then we shall simply write X_j for X^j.

Consider now two measurable spaces $(\Lambda^1, \mathcal{E}^1)$ and $(\Lambda^2, \mathcal{E}^2)$, and two random variables $X^1 : \Omega \to \Lambda^1$ and $X^2 : \Omega \to \Lambda^2$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The induced probability measures are denoted by μ_{X^1} and μ_{X^2}. Set $Z = (X^1, X^2) : \Omega \to \Lambda^1 \times \Lambda^2$ with $\Lambda^1 \times \Lambda^2 = \{(x^1, x^2) \mid x^1 \in \Lambda^1, x^2 \in \Lambda^2\}$. The set $\Lambda^1 \times \Lambda^2$ is endowed with the σ-algebra generated by boxes $A^1 \times A^2 = \{(x^1, x^2) \mid x^1 \in A^1 \text{ and } x^2 \in A^2\}$ for any $A^1 \in \mathcal{E}^1$ and $A^2 \in \mathcal{E}^2$. This σ-algebra is denoted by $\mathcal{E}^1 \times \mathcal{E}^2$. The induced probability measure μ_Z is called the *joint probability measure*. By definition, for any set $A \in \mathcal{E}^1 \times \mathcal{E}^2$, one has

$$\mathbb{P}\left(\left\{\omega \in \Omega \mid (X^1(\omega), X^2(\omega)) \in A\right\}\right) = \mu_Z(A) = \int_A \mu_Z(\mathrm{d}x^1 \times \mathrm{d}x^2) = \int_{\Lambda^1 \times \Lambda^2} \mathbf{1}_A(x^1, x^2)\mu_Z(\mathrm{d}x^1 \times \mathrm{d}x^2).$$

The following equalities then hold:

$$\mu_{X^1}(A^1) = \mu_Z(A^1 \times \Lambda^2) \quad \text{and} \quad \mu_{X^2}(A^2) = \mu_Z(\Lambda^1 \times A^2)$$

for any $A^1 \in \mathcal{E}^1$ and $A^2 \in \mathcal{E}^2$. The probability measures μ_{X^1} and μ_{X^2} are called the *marginal measures* of μ_Z.

Remark 1.3.1. *If $X^j : \Omega \to \mathbb{R}^{N_j}$ for $j \in \{1, 2\}$ and if we assume that the joint probability measure is absolutely continuous, with pdf denoted by $\Pi_{(X^1, X^2)}$, then the marginal pdfs are given by*

$$\Pi_{X^1}(x^1) = \int_{\mathbb{R}^{N_2}} \Pi_{(X^1,X^2)}(x^1, x^2)\,\mathrm{d}x^2 \quad \textit{and} \quad \Pi_{X^2}(x^2) = \int_{\mathbb{R}^{N_1}} \Pi_{(X^1,X^2)}(x^1, x^2)\,\mathrm{d}x^1.$$

Still for $X^j : \Omega \to \mathbb{R}^{N_j}$ but without assuming the absolute continuity we define the *cross-covariance matrix*

$$\mathrm{Cov}(X^1, X^2) := \mathbb{E}\left(\left(X^1 - \mathbb{E}(X^1)\right)\left(X^2 - \mathbb{E}(X^2)\right)^T\right) \in M_{N_1 \times N_2}(\mathbb{R}).$$

In particular, for $X : \Omega \to \mathbb{R}^N$ the *covariance matrix* is given by $\mathrm{Cov}(X) := \mathrm{Cov}(X, X) \in M_{N \times N}(\mathbb{R})$. In the special case $N_1 = N_2 = 1$ (a univariate random variable), the *correlation coefficient* is defined by

$$\mathrm{Corr}\,(X_1, X_2) := \frac{\mathbb{E}\Big((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\Big)}{\sqrt{\mathbb{E}\left((X_1 - \mathbb{E}(X_1))^2\right) \cdot \mathbb{E}\left((X_2 - \mathbb{E}(X_2))^2\right)}} \in [-1, 1]. \tag{1.3.1}$$

When $\mathrm{Corr}(X_1, X_2) = 0$ we say the two univariate random variables are *uncorrelated*. Observe that even if it is not written explicitly, all these expressions are computed with the joint probability measure.

Exercise 1.3.2. *Check that the covariance matrix* $\mathrm{Cov}(X)$ *is symmetric and positive semi-definite, namely it satisfies* $a^T \mathrm{Cov}(X)\, a \geq 0$ *for any* $a \in \mathbb{R}^N$ *with* $a \neq 0$.

The covariance matrix is also playing an important role when families of univariate random variables are considered. More precisely, let us consider N univariate random variables $X_1, \ldots, X_N$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. One then observes that any linear combination $a_1 X_1 + a_2 X_2 + \cdots + a_N X_N$ for $a_j \in \mathbb{R}$ also defines a univariate random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If we set $X := (X_1, \ldots, X_N)^T$ for the vector valued random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$, then this linear combination is nothing but $f(X)$ for $f(x) := \sum_{j=1}^N a_j x_j$ for $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$. For the following statement we assume that all computed quantities exist:

Proposition 1.3.3. *Let* $\{X_j\}_{=1}^N$ *be a family of univariate random variables defined on the same probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and let* $a = (a_1, \ldots, a_N)^T \in \mathbb{R}^N$*. Then the following equalities hold:*

$$\mathbb{E}\Big(\sum_{j=1}^N a_j X_j \Big) = \sum_{j=1}^N a_j \mathbb{E}(X_j), \tag{1.3.2}$$

$$\mathrm{Var}\Big(\sum_{j=1}^N a_j X_j \Big) = a^T \mathrm{Cov}(X)\, a, \tag{1.3.3}$$

where $X = (X_1, \ldots, X_N)^T$ *denotes the vector valued random variable made of* $X_1, \ldots, X_N$.

The proof of this proposition is left as an easy exercise. Still in the above framework, the *joint moment generating function* is defined by the expression $\mathbb{E}(\mathrm{e}^{a \cdot X})$ if it exists for all $a \in \mathbb{R}^N$ satisfying $\|a\| < \delta$ for some $\delta > 0$. The uniqueness result mentioned in Section 1.2 extends to this context:

Theorem 1.3.4. *Let* $\{X_j\}_{j=1}^N$ *be a family of univariate random variables* $X_j : \Omega \to \mathbb{R}$*, set* $X = (X_1, \ldots, X_N)^T$*, and assume that the joint moment generating function* $\mathbb{E}(\mathrm{e}^{a \cdot X})$ *exists for all* $a \in \mathbb{R}^N$ *satisfying* $\|a\| < \delta$ *for some* $\delta > 0$*. Then this function defines uniquely the multivariate random variable* X.

Exercise 1.3.5. *Look for a reference book where the proof is given, and study this proof.*

We now introduce the main definition of this section.

Definition 1.3.6 (Independence)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and consider two random variables* $X^1 : \Omega \to \Lambda^1$ *and* $X^2 : \Omega \to \Lambda^2$*. These random variables are* independent *if for any* $A^1 \in \mathcal{E}^1$*,* $A^2 \in \mathcal{E}^2$ *one has*

$$\mu_{(X^1, X^2)}(A^1 \times A^2) = \mu_{X^1}(A^1)\, \mu_{X^2}(A^2)$$

or equivalently

$$\mathbb{P}(X^1 \in A^1, X^2 \in A^2) = \mathbb{P}(X^1 \in A^1)\, \mathbb{P}(X^2 \in A^2).$$

In the special case of absolutely continuous random variables $X^1 : \Omega \to \mathbb{R}^{N_1}$ and $X^2 : \Omega \to \mathbb{R}^{N_2}$, the independence of X^1 and X^2 is equivalent to the condition $\Pi_{(X^1, X^2)} = \Pi_{X^1} \Pi_{X^2}$, which means that the joint measure is also absolutely continuous. The following notion plays a fundamental role in many applications of probability.

Definition 1.3.7 (IID)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $(\Lambda, \mathcal{E})$ *be a measurable space. A family of random variables* $\{X^j\}_j$ *with* $X^j : \Omega \to \Lambda$ *is said to be* independent and identically distributed *(in short IID) if they are all all independent and equally distributed (they share the same induced probability measure* μ_X*).*

Observe that if the family is finite, namely $\{X^j\}_{j=1}^N$, then this family is IID if the joint probability measure $\mu_{(X^1,\dots,X^N)}$ satisfies for any $A^j \in \mathcal{F}$

$$\mu_{(X^1,\dots,X^N)}(A^1 \times \dots \times A^N) = \mu_{X^1}(A^1)\dots\mu_{X^N}(A^N) \tag{1.3.4}$$

with all μ_{X^j} equal. If the family is infinite, the equality (1.3.4) must hold for any finite subfamily of random variables.

1.4 Univariate random variables

In this section, we consider univariate random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This means that the random variables X we shall consider are measurable functions on Ω taking values on $\mathbb{R}$. The subsequent definitions complement the various notions introduced in Definition 1.2.4. We state the following definition in the greatest generality, but the cases $p = 1$ and $p = 2$ will mainly be considered in the sequel.

Definition 1.4.1 (L^p-spaces on $(\Omega, \mathcal{F}, \mathbb{P})$)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $p \geq 1$. We set $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ for the set of all univariate random variables X on Ω satisfying $\mathbb{E}(|X|^p) < \infty$. For any $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ we also set $\|X\|_p := \left(\mathbb{E}(|X|^p)\right)^{1/p}$. The set of equivalent classes of elements of $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ are denoted by $L^p(\Omega, \mathcal{F}, \mathbb{P})$.*

Note that the random variables in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ are precisely the ones for which $\mathbb{E}(X)$ is well defined, as mentioned in Definition 1.2.1. The ones in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ correspond to the random variables for which the variance is well defined. For X in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, we say that X is square-integrable.

Exercise 1.4.2. *Show that $L^p(\Omega, \mathcal{F}, \mathbb{P})$ are vector spaces, that $\|\cdot\|_p$ defines a norm on $L^p(\Omega, \mathcal{F}, \mathbb{P})$, and that $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is complete with this norm. Show also that $L^{p_2}(\Omega, \mathcal{F}, \mathbb{P}) \subset L^{p_1}(\Omega, \mathcal{F}, \mathbb{P})$ whenever $p_2 \geq p_1$.*

The space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ has a nice geometric property: it is endowed with the scalar produced defined by the map

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) \times L^2(\Omega, \mathcal{F}, \mathbb{P}) \ni (X, Y) \mapsto \mathbb{E}(XY) \in \mathbb{R}.$$

This scalar product satisfies the Cauchy-Schwarz inequality $|\mathbb{E}(XY)| \leq \|X\|_2\|Y\|_2$, which leads directly to the inequality

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)}\,\sqrt{\mathrm{Var}(Y)}.$$

Clearly, this also leads to the correlation coefficient already introduced in (1.3.1).

Let us introduce a few more geometric concepts. For $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ we set

$$X = X - \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}Y + \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}Y = X^\perp + \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}Y, \tag{1.4.1}$$

with $X^\perp := X - \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}Y$, and easily observe that $\mathbb{E}(X^\perp Y) = 0$. The decomposition (1.4.1) can be seen as the decomposition of X in a component parallel to Y, and a component perpendicular to Y. The random variable $\frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}Y$ is called the *orthogonal projection* of X onto Y.

As usual, any norm on a vector space allows us to define the distance between two elements with the expression $\|X - Y\|_p$ for any $X, Y \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. One can also define the notion of convergence.

Definition 1.4.3 (L^p-convergence of random variables)**.** *A sequence $\{X_j\}_{j\in\mathbb{N}} \subset L^p(\Omega, \mathcal{F}, \mathbb{P})$ of univariate random variables on the same probability space* converges in the L^p-sense *to the random variable $X_\infty \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ if $\|X_j - X_\infty\|_p \to 0$ as $j \to \infty$.*

Note that $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is a Banach space, which ensures that any Cauchy sequence in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ has its limit in this space. When $p = 2$, the L^2-convergence is also called the *mean square convergence*, or the convergence in *quadratic mean*. The following exercise uses the notion of convergence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Exercise 1.4.4 (L^2-version of the weak law of large numbers)**.** *Consider a family* $\{X_j\}_{j\in\mathbb{N}} \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$ *of univariate IID random variable with* 0 *mean value (or in other words with expectation* 0*). Show that the* empirical mean

$$S_N := \frac{1}{N} \sum_{j=1}^{N} X_j$$

converges to 0 *in the* L^2*-sense, or more precisely* $\|S_N\|_2 \to 0$ *as* $N \to \infty$.

Chapter 2

Stochastic processes

In this chapter, we provide the definition of general stochastic processes. However, we shall start with the special instance of a Gaussian process.

2.1 Gaussian vectors

In order to define Gaussian processes, let us start by recalling the Gaussian probability distribution, and then move to the notion of Gaussian vectors.

The Gaussian probability distribution has been introduced in Exercise 1.2.5, and the corresponding random variable is denoted by $N(\overline{x}, \sigma^2)$ in the univariate case, and $N(\overline{x}, P)$ in the N-dimensional case. Here $\overline{x}$ stands for the value in $\mathbb{R}$ or the vector in $\mathbb{R}^N$ given by the expectation value of the corresponding random variable, $\sigma^2 > 0$ corresponds to the variance of $N(\overline{x}, \sigma^2)$, and the Hermitian matrix $P > 0$ is the covariance matrix of $N(\overline{x}, P)$. When $\overline{x} = 0$ and $\sigma = 1$ (or P is the $N \times N$ identity matrix) one speaks about standard Gaussian probability distribution, or standard Gaussian random variable. This distribution is going to play a central role in this chapter. For simplicity, we simply say Gaussian random variable for univariate Gaussian random variable, and say N-dimensional Gaussian random variable when $N > 1$. For convenience, we also introduce an extension of the usual univariate Gaussian random variables, namely $N(\overline{x}, 0)$. This random variable is a point mass on its mean $\overline{x}$, or in other words it corresponds to the random variable taking the value $\overline{x}$ with a probability 1. For example $\mathbf{0} = N(0, 0)$ is the random variable taking the value 0 with probability 1.

Definition 2.1.1 (Gaussian vector)**.** *A N-dimensional random vector $X = (X_1, \ldots, X_N)^T$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a* Gaussian vector *if for any $a = (a_1, \ldots, a_N)^T \in \mathbb{R}^N$ the random variable $a \cdot X := \sum_{j=1}^N a_j X_j$ is a Gaussian random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. In this case, we also say that the univariate random variables $X_1, \ldots, X_N$ are* jointly Gaussian*.*

Starting from a Gaussian vector X, we observe that each of its component X_j is a Gaussian random variable, by choosing $a = e_j$ with $\{e_j\}_{j=1}^N$ the standard basis of $\mathbb{R}^N$. Conversely, let us check that IID standard Gaussian random variables lead to a Gaussian vector.

Exercise 2.1.2. *Check that if X_1, X_2 are independent and standard Gaussian random variables, then $(X_1, X_2)^T$ is a Gaussian vector. Show that the random variable $a_1 X_1 + a_2 X_2$ is a Gaussian random variable with mean 0 and variance $a_1^2 + a_2^2$. Generalize your result for N independent and standard Gaussian random variables.*

The following result can also be proved as an exercise, see also [1, Ex. 2.8].

Lemma 2.1.3. *If X is a N-dimensional Gaussian vector and if $M \in M_{N\times N}(\mathbb{R})$, check that the new vector MX is also a N-dimensional random vector.*

The next statement provides an equivalent definition for a Gaussian vector. Its proof is based on Proposition 1.3.3 and on Theorem 1.3.4. Can you write it ?

Proposition 2.1.4. *A N-dimensional random vector X is Gaussian if and only if its moment generating function $\mathbb{E}(e^{a\cdot X})$ exists for all $a \in \mathbb{R}^N$ and satisfies*

$$\mathbb{E}(e^{a\cdot X}) = \exp\Big(a\cdot\mathbb{E}(X) + \tfrac{1}{2}a^T \operatorname{Cov}(X)\, a\Big).$$

As already mentioned, each component of a Gaussian vector is a Gaussian distribution. It has also been shown in Exercise 2.1.2 that N independent and standard Gaussian random variables lead to a Gaussian vector. The following statement provides a criterion for the independence of the components of an arbitrary N-dimensional Gaussian vector. Its proof is left as an exercise, see also [1, Prop. 2.10] for inspiration.

Proposition 2.1.5. *Let $X = (X_1, \dots, X_N)^T$ be a N-dimensional Gaussian vector. Its covariance matrix $\operatorname{Cov}(X)$ is diagonal if and only if the Gaussian random variables $X_1, \dots, X_N$ are independent.*

Before studying further the link between Gaussian vectors and the N-dimensional Gaussian random variables introduced in Exercise 1.2.5, let us recall a result of linear algebra about positive definite and positive semi-definite matrices. This result can be applied to the covariance matrix, see also Exercise 1.3.2.

Lemma 2.1.6 (Cholesky decomposition)**.** *For any positive definite matrix $A \in M_{N\times N}(\mathbb{R})$, there exists a lower triangular matrix $L \in M_{N\times N}(\mathbb{R})$ with strictly positive diagonal entries satisfying*

$$A = LL^T.$$

If A is positive semi-definite, the entries on the diagonal can be 0. In the former case, the matrix L is invertible, while in the latter case the matrix L is not invertible in general.

Let us emphasize that the matrix L is usually not unique, and that there exist several algorithm for computing the matrix L. For example, one algorithm is based on a Gram-Schmidt type procedure, as for $\mathbb{R}^N$.

We shall now gather a few useful results about Gaussian vectors. Proofs are not difficult and can be found in [1, Sec. 2.2] and worked out as exercises. They rely on Cholesky decomposition mentioned above. For their statement, we need one more notion: A N-dimensional Gaussian vector X is *non-degenerate* if its covariant matrix is invertible, namely if $\det(\operatorname{Cov}(X)) \neq 0$. Conversely, if $\det(\operatorname{Cov}(X)) = 0$, we say that the Gaussian vector X is *degenerate*. Recall that $\mathbf{0}$ is the random variable taking the value 0 with probability 1.

Lemma 2.1.7. *Let $X = (X_1, \dots, X_N)$ be a N-dimensional Gaussian vector with mean value $\mathbb{E}(X) = 0 \in \mathbb{R}^N$. Then X is degenerate if and only if the Gaussian random variables $X_1, \dots, X_N$ are linearly dependent, namely if and only if there exists $a \in \mathbb{R}^n$, $a \neq 0$ such that $a \cdot X = \mathbf{0}$.*

Exercise 2.1.8. *Provide an example of a degenerate 3-dimensional Gaussian vector, and check the previous lemma on this example.*

Theorem 2.1.9 (Decomposition into IID Gaussian random variables)**.** *Let X be a N-dimensional and non-degenerate Gaussian vector satisfying $\mathbb{E}(X) = 0 \in \mathbb{R}^N$. Then there exists an invertible matrix $L \in M_{N\times N}(\mathbb{R})$ and N IID standard Gaussian random variables $Z_1, \dots, Z_N$ such that $X = LZ$, with $Z = (Z_1, \dots, Z_N)^T$.*

As seen in Section 1.3, a family of univariate random variables defines a joint probability measure. If this measure is absolutely continuous with respect to the Lebesgue measure, it defines a probability density function, as emphasized in Definition 1.1.10. In these notes, we use the notation $\mathbb{R}_+$ for $[0, \infty)$.

Corollary 2.1.10 (Joint probability density function)**.** *Let X be a N-dimensional and non-degenerate Gaussian vector with expectation $\mathbb{E}(X) = \bar{x} \in \mathbb{R}^N$ and covariance matrix $\mathrm{Cov}(X) = P \in M_{N\times N}(\mathbb{R})$. Then the joint probability measure of $X_1, \dots, X_N$ is absolutely continuous with respect to the Lebesgue measure, and the corresponding probability density function $\Pi : \mathbb{R}^N \to \mathbb{R}_+$ is given for $x \in \mathbb{R}^N$ by*

$$\Pi(x) := \frac{1}{(2\pi)^{N/2}|P|^{1/2}} \exp\left(-\frac{1}{2}(x-\bar{x})^T P^{-1}(x-\bar{x})\right).$$

Exercise 2.1.11. *Provide the joint density measure for three Gaussian random variables which are not linearly independent.*

2.2 Gaussian processes

In this section, we briefly introduce some of the main Gaussian processes. The presentation is slightly formal since the precise definition of a stochastic process is not given yet. This section can be considered as a motivation for the subsequent developments.

We consider a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a fixed measurable space $(\Lambda, \mathcal{E})$, and let t denote the time. If $t \in \mathbb{N}$, we speak about a discrete time model, while if $t \in [0, T]$ with $T > 0$, or if $t \in \mathbb{R}_+$, we speak about a continuous time model. In general, we shall simply consider $t \in \mathcal{T}$, with $\mathcal{T}$ a subset of $\mathbb{R}$. The main object for describing the evolution of a system consists in the family

$$X := (X_t)_{t\in\mathcal{T}} \quad \text{with} \quad X_t : \Omega \to \Lambda \quad \text{a random variable.} \tag{2.2.1}$$

Clearly, X can also be seen as a function of two variables, namely

$$X : \mathcal{T} \times \Omega \ni (t, \omega) \mapsto X(t, \omega) := X_t(\omega) \in \Lambda.$$

We shall come back to this notion in the following sections, and the resulting object will be a *stochastic process*. Still, one can have another look at the previous object: For fixed $\omega \in \Omega$ the map $\mathcal{T} \ni t \mapsto X(t, \omega) \in \Lambda$ can be seen as a *trajectory* or as a *path* in Λ. These different points of view will complement each other.

In the next definition we consider a finite family $\{t_1, t_2, \dots, t_N\} \subset \mathcal{T}$. For this definition it does not matter if $\mathcal{T}$ is a discrete set or a continuous set. We can also observe that the regularity condition not specified above does not play any role. Note that the following examples are all taking place in a 1-dimensional space, mainly for simplicity.

Definition 2.2.1 (Gaussian process)**.** *The family $X := (X_t)_{t\in\mathcal{T}}$ with each X_t a univariate random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a* Gaussian process *if for any finite family $\{t_1, t_2, \dots, t_N\} \subset \mathcal{T}$ with $t_j < t_{j+1}$ the N-dimensional vector $(X_{t_1}, X_{t_2}, \dots, X_{t_N})^T$ is a Gaussian vector [1] (in the sense of Definition 2.1.1).*

In particular, as seen in Proposition 2.1.4, the Gaussian vector $(X_{t_1}, X_{t_2}, \dots, X_{t_N})$ is uniquely defined by its mean value and by its covariance matrix. In the sequel, we list some famous Gaussian processes. As already mentioned, the presentation is slightly informal, but nevertheless informative. Nice illustrations of these processes can be found in [1, Sec. 2.3], or can be realized as exercises, see [1, Numerical projects p. 45].

[1] Some authors require the vector to be non-degenerate, some not. If non-degeneracy is imposed, a characterization of the Gaussian process in terms of the joint probability measure of Corollary 2.1.10 is possible.

Example 2.2.2 (Brownian process[2])**.** *The Brownian process* $B := (B_t)_{t\in\mathbb{R}_+}$ *corresponds to a Gaussian process with* $\mathbb{E}(B_t) = 0$ *and* $\mathrm{Cov}(B_t, B_s) \equiv \mathbb{E}(B_t B_s) = t \wedge s$*, where* $t \wedge s$ *stands for the minimum between* s *and* t*. Additional properties will be imposed and studied subsequently.*

Example 2.2.3 (Brownian process with drift)**.** *Let* $\sigma > 0$ *be called the* volatility *or the* diffusion coefficient*, and let* $\mu \in \mathbb{R}$ *be called the* drift*. Let* $X_t := \sigma B_t + \mu t$ *be the random variable based on the Brownian process introduced in Example 2.2.2. Then* $X := (X_t)_{t\in\mathbb{R}_+}$ *is a Gaussian process which satisfies* $\mathbb{E}(X_t) = \mu t$ *and* $\mathrm{Cov}(X_t, X_s) = \sigma^2(t \wedge s)$*.*

Example 2.2.4 (Brownian bridge)**.** *The Gaussian process* $Z := (Z_t)_{t\in[0,1]}$ *whose mean value satisfies* $\mathbb{E}(Z_t) = 0$ *and whose covariance satisfies* $\mathrm{Cov}(Z_t, Z_s) = s(1-t)$ *for* $0 \le s \le t \le 1$*, is called the* Brownian bridge*. By construction one has* $Z_0 = \mathbf{0}$ *and* $Z_1 = \mathbf{0}$*. This process can be realized by setting for* $t \in [0, 1]$

$$Z_t := B_t - tB_1$$

where $(B_t)_{t\in[0,1]}$ *is (part of) the Brownian process introduced in Example 2.2.2.*

Example 2.2.5 (Fractional Brownian process)**.** *For* $H \in (0, 1)$*, called* Hurst index*, the* fractional Brownian process *is the Gaussian process* $B^H := (B^H_t)_{t\in\mathbb{R}_+}$ *satisfying* $\mathbb{E}(B^H_t) = 0$ *and*

$$\mathrm{Cov}(B^H_t, B^H_s) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}).$$

Note that the special case $H = \frac{1}{2}$ *corresponds to the Brownian process.*

Example 2.2.6 (Ornstein-Uhlenbeck process)**.** *The Ornstein-Uhlenbeck process* $Y := (Y_t)_{t\in\mathbb{R}_+}$ *corresponds to the Gaussian process starting at* $Y_0 = \mathbf{0}$ *with mean value satisfying* $\mathbb{E}(Y_t) = 0$ *and covariance given by* $\mathrm{Cov}(Y_t, Y_s) = \frac{\mathrm{e}^{-(t-s)}}{2}(1 - \mathrm{e}^{-2s})$ *for* $s \le t$*. If* Y_0 *is random and satisfies* $Y_0 = N(0, \frac{1}{2})$*, then* $\mathrm{Cov}(Y_t, Y_s) = \frac{\mathrm{e}^{-(t-s)}}{2}$*.*

Later, we shall see that these processes are solutions of some stochastic differential equations.

2.3 Stochastic processes

In this section we provide the precise definition of a stochastic process, and some general related notions. Note that this section is more general than the previous one, since Gaussian processes are special instances of stochastic processes.

Before the main definition of this section, we start with a slightly technical notion. Throughout the section, the set $\mathcal{T}$ denotes a subset of $\mathbb{R}$, and $t \in \mathcal{T}$ is used for a parameter representing the time.

Definition 2.3.1 (Filtration)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. A* filtration $(\mathcal{F}_t)_{t\in\mathcal{T}}$ *is a family of* σ*-subalgebras of* $\mathcal{F}$ *satisfying* $\mathcal{F}_s \subset \mathcal{F}_t$ *whenever* $s \le t$*.*

In other words, $\mathcal{F}_t$ is a subset of $\mathcal{F}$ which also satisfies the condition of Definition 1.1.1, and $\mathcal{F}_t$ contains more elements than $\mathcal{F}_s$ whenever $s \le t$. Observe that $\mathcal{F}_t$ containing more subsets of Ω than $\mathcal{F}_s$ means also that $\mathcal{F}_t$ is more "precise" than $\mathcal{F}_s$, or can provide more accurate information. Later on, $\mathcal{F}_t$ will be thought as *the amount of information available at time* t.

[2]The Gaussian process is also called the *Wiener process* (mainly by mathematicians) in honor of American mathematician Norbert Wiener for his investigations on the mathematical properties of the one-dimensional Brownian motion. This process is usually called Brownian motion (mainly by physicists) due to its historical connection with the physical process of the same name originally observed by Scottish botanist Robert Brown.

Definition 2.3.2 (Stochastic process)**.** *A* stochastic process *consists of the tuple*

$$X := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (X_t)_{t\in\mathcal{T}})$$

with $(\Omega, \mathcal{F}, \mathbb{P})$ *a probability space,* $(\mathcal{F}_t)_{t\in\mathcal{T}}$ *a filtration, and* $(X_t)_{t\in\mathcal{T}}$ *a family of random variables on* Ω*, taking values in a measurable space* $(\Lambda, \mathcal{E})$*, and* X_t *is measurable with respect to* $\mathcal{F}_t$*.*

Let us recall that the notion of measurable function has been provided in Definition 1.1.4: $X : \Omega \to \Lambda$ is measurable if $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{E}$. Thus, one observes that the measurability of X_t with respect to $\mathcal{F}_t$ is a stronger assumption: since $\mathcal{F}_t$ contains less elements than $\mathcal{F}$, having $X_t^{-1}(A)$ in $\mathcal{F}_t$ is less likely than having $X_t^{-1}(A)$ in $\mathcal{F}$. In the present situation, we also say that $(X_t)_{t\in\mathcal{T}}$ is *adapted* to the filtration $(\mathcal{F}_t)_{t\in\mathcal{T}}$. Note that there exists a minimal filtration, also called *natural filtration*. Namely, let us set $\sigma(X_s)$ for the smallest σ-subalgebra of $\mathcal{F}$ which contains $X_s^{-1}(A)$ for all $A \in \mathcal{E}$. We can then define $\mathcal{G}_t$ as the smallest σ-subalgebra of $\mathcal{F}$ containing $\sigma(X_s)$ for all $s \in \mathcal{T}$ with $s \leq t$. Equivalently, $\mathcal{G}_t$ corresponds to the smallest σ-subalgebra of $\mathcal{F}$ containing all elements $X_s^{-1}(A)$ for all $A \in \mathcal{E}$ and $s \in \mathcal{T}$ with $s \leq t$. Then $(\mathcal{G}_t)_{t\in\mathcal{T}}$ is adapted to $(X_t)_{t\in\mathcal{T}}$ and is called the natural filtration. Any other adapted filtration must contain it.

Note that in general $\cup_{t\in\mathcal{T}}\mathcal{F}_t$ is not a σ-algebra. If necessary, we can consider $\mathcal{F}_\mathcal{T}$, the smallest σ-subalgebra of $\mathcal{F}$ containing all $\mathcal{F}_t$ for $t \in \mathcal{T}$. Usually, $\mathcal{F}_\mathcal{T}$ is strictly contained in $\mathcal{F}$. There is still one filtration which might be useful, the *augmented natural filtration* $(\overline{\mathcal{G}}_t)_{t\in\mathcal{T}}$ and defined for any $t \in \mathcal{T}$ as the smallest σ-subalgebra of $\mathcal{F}$ containing $\mathcal{G}_t$ and all negligible events of $\mathcal{F}$, namely any $B \in \mathcal{F}$ satisfying $\mathbb{P}(B) = 0$. The addition of the negligible events is an important trick for the various notions of equivalences of stochastic processes[3]. For example, the stochastic processes $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (X_t)_{t\in\mathcal{T}})$ and $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (Y_t)_{t\in\mathcal{T}})$ are *versions* or *modifications* of one another if for any $t \in \mathcal{T}$, $\mathbb{P}(X_t = Y_t) = 1$, or more precisely if for any $t \in \mathcal{T}$

$$\mathbb{P}(\{\omega \in \Omega \mid X_t(\omega) = Y_t(\omega)\}) = 1.$$

Equivalently, if we set $N_t := \{\omega \in \Omega \mid X_t(\omega) \neq Y_t(\omega)\}$, then the previous condition reads $\mathbb{P}(N_t) = 0$. However, if we set $N := \cup_{t\in\mathcal{T}} N_t$, then it may happen in the continuous time setting that $\mathbb{P}(N) \neq 0$, since N is then given by an uncountable union. With this notation we say that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (X_t)_{t\in\mathcal{T}})$ and $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (Y_t)_{t\in\mathcal{T}})$ are *indistinguishable* if $\mathbb{P}(N) = 0$. Equivalently, this condition reads

$$\mathbb{P}(\{\omega \in \Omega \mid \exists t \in \mathcal{T} \text{ with } X_t(\omega) \neq Y_t(\omega)\}) = 0.$$

Note that in the discrete time setting, these two notions coincide.

So far, we have not imposed any regularity condition for the stochastic process X. By considering again the function of two variables

$$\mathcal{T} \times \Omega \ni (s, \omega) \mapsto X_s(\omega) \in \Lambda$$

we could require the measurability of this map, from $(\mathcal{T} \times \Omega, \sigma(\mathcal{T}) \times \mathcal{F})$ to $(\Lambda, \mathcal{E})$, where we denote by $\sigma(\mathcal{T})$ the σ-subalgebra of σ_B generated by the open sets of $\mathcal{T}$. However, this notion does not use the filtration. The right notion is the following:

Definition 2.3.3 (Progressively measurable)**.** *The stochastic process* X *is* progressively measurable *if for any* $t \in \mathcal{T}$ *the map*

$$\mathcal{T} \cap [0, t] \times \Omega \ni (s, \omega) \mapsto X_s(\omega) \in \Lambda$$

is measurable from $\left(\mathcal{T} \cap [0, t] \times \Omega, \sigma(\mathcal{T} \cap [0, t]) \times \mathcal{F}_t\right)$ *to* $(\Lambda, \mathcal{E})$*.*

[3] More precisely, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *complete* if for any $B \in \mathcal{F}$ with $\mathbb{P}(B) = 0$ and for any $A \subset B$, one has $A \in \mathcal{F}$ (and then $\mathbb{P}(A) = 0$). Accordingly, a filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ is *complete* if $\mathcal{F}_0$ contains the negligible sets

The previous notion is rather complicated. Fortunately, in most of the cases a stronger regularity property holds. Its statement contains the notion of continuity or a limit, which are not defined in arbitrary measurable spaces. For simplicity, we concentrate on the standard measurable spaces introduced in Section 1.2 but generalizations are possible.

Definition 2.3.4 (Continuous stochastic process)**.** *The stochastic process X taking values in a standard measurable space $(\Lambda, \mathcal{E})$ is* continuous *if for every $\omega \in \Omega$, the map $\mathcal{T} \ni t \mapsto X_t(\omega) \in \Lambda$ is continuous. It is* a.s. continuous *if for almost every ω, the map $\mathcal{T} \ni t \mapsto X_t(\omega) \in \Lambda$ is continuous. In other words, X is a.s. continuous if*

$$\mathbb{P}\Big(\{\omega \in \Omega \mid \lim_{s \to t} X_s(\omega) = X_t(\omega)\}\Big) = 1.$$

It turns out that any continuous stochastic process is progressively measurable, as shown in [2, Prop. 2.1]

Observe that we haven't impose any condition on the filtration yet. Some regularity on it are often necessary. For this we set for any $t \in \mathcal{T}$

$$\mathcal{F}_{t+} = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}. \tag{2.3.1}$$

It turns out that the intersection of any σ-algebras is also a σ-algebra, and therefore $\mathcal{F}_{t+}$ is also a σ-subalgebra of $\mathcal{F}$, with $\mathcal{F}_t \subset \mathcal{F}_{t+}$. With this notation, we say that the filtration is *right-continuous* if $\mathcal{F}_t = \mathcal{F}_{t+}$. One way to think about this condition is that any information known right after t is also known at t. Observe also that given a family $(X_t)_{t \in \mathcal{T}}$ adapted to the filtration $(\mathcal{F}_t)_{t \in \mathcal{T}}$, we can also define a new filtration $\{\mathcal{G}_t\}_{t \in \mathcal{T}}$ by setting $\mathcal{G}_t := \cap_{s > t} \mathcal{F}_s$. Then $\{\mathcal{G}_t\}_{t \in \mathcal{T}}$ is a right-continuous filtration, and $(X_t)_{t \in \mathcal{T}}$ is adapted to it. In applications it often required that the filtration is right-continuous and contains the negligible events.

2.4 Brownian motion

In this section we provide a brief description of the Brownian motion, also called Brownian process or Wiener process. The first definition is quite general, but later on we shall stick to a simpler presentation. In the sequel, a $\mathbb{R}^N$-valued stochastic process $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathcal{T}}, (X_t)_{t \in \mathcal{T}})$ means that each random variable X_t takes values in the measurable space $(\mathbb{R}^N, \sigma_{\mathrm{B}})$. Equivalently, we can also speak about multivariate stochastic process, or univariate stochastic process in the special case $N = 1$. We shall also use the short notation *a.s.* for *almost surely* meaning that an equality holds with probability 1. Note that we start with the Brownian motion in dimension 1 because any Brownian motion in dimension N decomposes in N independent Brownian motions of dimension 1. We also use the letter B instead of X, since this notation is commonly used for the Brownian motion. In the statement, B_t can be interpreted as the random variable corresponding to the *position* at time t, while for $t > s$ the difference $B_t - B_s$ should be interpreted as the *difference of position* or as the *increment* between the position at time s and position at time t.

Definition 2.4.1 (1-dimensional Brownian motion)**.** *A Stochastic process $B := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, (B_t)_{t \in \mathbb{R}_+})$ taking values in $\mathbb{R}$ is a* 1-dimensional Brownian motion *if*

1. *$B_0 = \mathbf{0}$ a.s.,*
2. *For any $0 \le s \le t$ the random variable $B_t - B_s$ is independent of $\mathcal{F}_s$,*
3. *For any $0 \le s < t$ the random variable $B_t - B_s$ is a Gaussian random variable $N(0, t - s)$.*

Condition 1. can be thought as an initial condition. The condition 2. needs an explanation and an interpretation. The notion of independence of random variables has been introduced in Definition 1.3.6. A family $\mathcal{F}_1, \dots, \mathcal{F}_m$

of σ-subalgebras of $\mathcal{F}$ are called *independent* if for any $A^1 \in \mathcal{F}_1, \dots, A^m \in \mathcal{F}_m$ one has

$$\mathbb{P}(A^1 \cap A^2 \cap \cdots \cap A^m) = \mathbb{P}(A^1)\mathbb{P}(A^2)\dots\mathbb{P}(A^m).^4$$

One then infers that the random variables $X^1, X^2, \dots, X^m$ on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if and only if the σ-algebras $\sigma(X^1), \sigma(X^2), \dots, \sigma(X^m)$ they generate are independent. Finally, a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is *independent* of a σ-subalgebra $\mathcal{G}$ of $\mathcal{F}$ if and only if the σ-algebras $\mathcal{G}$ and $\sigma(X)$ are independent. Note that this happens if and only if X is independent of every $\mathcal{G}$-measurable random variables. It is this latter notion of independence which is used in condition 2. of the previous definition. Intuitively, this condition means that the increment in the process after time s are independent of the process up to time s.

Exercise 2.4.2. *Prove the statements of the previous paragraph, namely:* The random variables X^1, X^2 on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if and only if the σ-algebras $\sigma(X^1), \sigma(X^2)$ they generate are independent, *and* A random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is *independent* of a σ-subalgebra $\mathcal{G}$ of $\mathcal{F}$ if and only X is independent of every $\mathcal{G}$-measurable random variables.

Exercise 2.4.3. *Show that the Brownian process is a Gaussian process, see also [2, Remarks 3.1].*

By a rather deep result about continuity (Kolmogorov's continuity theorem) it turns out that there exists a modification of $(B_t)_{t\in\mathbb{R}_+}$ for which the stochastic process is continuous. Thus, we shall assume from now on that the Brownian motion is continuous. In addition the following properties of the Brownian motion provides a new characterization of it, see [2, Prop. 3.1]:

Proposition 2.4.4. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (B_t)_{t\in\mathbb{R}_+})$ be a 1-dimensional Brownian motion. Then*

1. *$B_0 = \mathbf{0}$ a.s.,*
2. *For every $0 \le t_1 < t_2 < \cdots < t_N$, the N-dimensional vector $B := (B_{t_1}, B_{t_2}, \dots, B_{t_N})^T$ is a Gaussian vector with $\mathbb{E}(B) = 0$,*
3. *$\mathbb{E}(B_t B_s) = t \wedge s$.*

Conversely, properties 1. to 3. define a 1-dimensional Brownian motion with the natural filtration. It is called the natural Brownian motion.

Observe that choosing any $t \in \mathbb{R}_+$, one infers from 2. that $\mathbb{E}(B_t) = 0$. Thus, the converse statement corresponds to what was given in Example 2.2.2, when the notion of filtration was not introduced (see Remark below). Let us finally mention another consequence of Definition 2.4.1: for any sequence $\le t_0 < t_1 < \cdots < t_N$, the family of random variables $\{B_{t_j} - B_{t_{j-1}}\}_{j=1}^N$ are independent random variables. This property is sometimes used for (partially) defining the Brownian motion.

Remark 2.4.5. *The last part of Proposition 2.4.4 means that whenever the Brownian motion is introduced without any information about the filtration, then the natural filtration is implicitly used. Since the natural filtration is the minimal filtration, having the Brownian motion with the natural filtration is the weakest version (minimal amount of information available at any time t). In fact, it can be shown that any Brownian motion is a Brownian motion with respect to the augmented natural filtration, obtained by adding all negligible events to the natural filtration, and that this filtration is right-continuous [2, Prop. 4.3]. Thus, $(\Omega, \mathcal{F}, \mathbb{P}, (\overline{\mathcal{G}}_t)_{t\in\mathbb{R}_+}, (B_t)_{t\in\mathbb{R}_+})$ is a Brownian motion with a right-continuous filtration and with the negligible events contained in $\mathcal{F}_t$ for any $t \in \mathbb{R}_+$. A process with these two properties is called a* standard stochastic process.

[4] This is the common requirement for the independence of a family of events.

Let us now state additional properties of the Brownian motion with respect to certain transformations. The proof is interesting and can be studied as an exercise, see [2, Prop. 3.2].

Proposition 2.4.6. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (B_t)_{t\in\mathbb{R}_+})$ be a 1-dimensional Brownian motion. Then,*

1. *For any $s \geq 0$, $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_{t+s})_{t\in\mathbb{R}_+}, (B_{t+s} - B_s)_{t\in\mathbb{R}_+})$ is a 1-dimensional Brownian motion (time shift),*
2. *$(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (-B_t)_{t\in\mathbb{R}_+})$ is a 1-dimensional Brownian motion (mirror reflection),*
3. *For any $c > 0$, $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_{t/c^2})_{t\in\mathbb{R}_+}, (cB_{t/c^2})_{t\in\mathbb{R}_+})$ is a 1-dimensional Brownian motion (scaling),*
4. *The random variables defined by $Z_t := tB_{1/t}$ for $t > 0$ and $Z_0 = 0$ define a natural Brownian motion.*

We now discuss some properties of the paths, namely of the continuous functions $t \mapsto B_t(\omega)$ for fixed $\omega \in \Omega$. Recall that a *partition* $\mathcal{P}_\ell$ of an interval $[a, b]$ consists in a set $\mathcal{P}_\ell := \{t_0^\ell, t_1^\ell, \dots, t_{n_\ell}^\ell\}$ with $t_0^\ell = a$, $t_{n_\ell}^\ell = b$ and $t_j^\ell < t_{j+1}^\ell$. For a given partition $\mathcal{P}_\ell$, we set $|\mathcal{P}_\ell| := \max_{j\in\{1,\dots,n_\ell\}} |t_j^\ell - t_{j-1}^\ell|$ for its mesh. For any function $f : [a, b] \to \mathbb{R}$ we define the variation of f as

$$\mathrm{var}_{[a,b]}(f) := \sup_{\mathcal{P}_\ell} \sum_{j=1}^{n_\ell} |f(t_j^\ell) - f(t_{j-1}^\ell)| \tag{2.4.1}$$

where the supremum is taken over all partitions of $[a, b]$. If $\mathrm{var}(f) < \infty$, then f is said to be of *finite variation* or of *bounded variation*. If it is not bounded, we say that the function f has an *infinite variation* on $[a, b]$.

Exercise 2.4.7. *Let $f : [a, b] \to \mathbb{R}$. 1) If f is increasing, check that $\mathrm{var}_{[a,b]}(f) = f(b) - f(a)$. 2) If $f \in C^1([a, b])$, check that $\mathrm{var}_{[a,b]}(f) = \int_a^b |f'(t)|\mathrm{d}t$.*

The following statement is borrowed from [4, Prop. 3.6]. Recall that the notion of convergence in the L^2-sense has been introduced in Definition 1.4.3.

Theorem 2.4.8. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (B_t)_{t\in\mathbb{R}_+})$ be a 1-dimensional Brownian motion.*

1. *Almost every path has a infinite variation on any finite interval, namely for any $a, b \in \mathbb{R}_+$*

$$\mathbb{P}\Big(\big\{\omega \in \Omega \mid \mathrm{var}_{[a,b]}(t \mapsto B_t(\omega)) = \infty\big\}\Big) = 1.$$

2. *The quadratic variation of the Brownian motion converges in the L^2-sense, namely*

$$\lim_{|\mathcal{P}_\ell|\to 0} \mathbb{E}\Big(\Big[\sum_{j=1}^{n_\ell} (B_{t_j^\ell} - B_{t_{j-1}^\ell})^2 - (b - a)\Big]^2\Big) = 0.$$

3. *Almost every path is nowhere differentiable, namely*

$$\mathbb{P}\Big(\big\{\omega \in \Omega \mid t \mapsto B_t(\omega) \text{ is nowhere differentiable } \big\}\Big) = 1.$$

By restricting the partitions considered in 2., a different type of convergence can be used. Namely, let $\{\mathcal{P}_\ell\}_{\ell\in\mathbb{N}}$ be a sequence of partitions of $[a, b]$, and let us assume that $\sum_{\ell\in\mathbb{N}} |\mathcal{P}_\ell| < \infty$, which means that we consider a set of partition with rapidly decaying meshes.

Theorem 2.4.9. *Let $\{\mathcal{P}_\ell\}_{\ell\in\mathbb{N}}$ be a sequence of partitions of $[a,b]$ satisfying $\sum_{\ell\in\mathbb{N}}|\mathcal{P}_\ell|<\infty$. Then almost surely*

$$\lim_{\ell\to\infty}\sum_{j=1}^{n_\ell}(B_{t_j^\ell}-B_{t_{j-1}^\ell})^2=b-a,$$

or equivalently

$$\mathbb{P}\Big(\Big\{\omega\in\Omega\mid\lim_{\ell\to\infty}\sum_{j=1}^{n_\ell}(B_{t_j^\ell}(\omega)-B_{t_{j-1}^\ell}(\omega))^2=b-a\Big\}\Big)=1.$$

Let us mention that a proof of the statement 2. of Theorem 2.4.8 is provided in [1, Thm. 3.8] or in [14, Thm. 9.1], see also [2, Prop. 3.4]. A special case of Theorem 2.4.9 is also provided in [1, Corol. 3.16] while the general case is given in [14, Thm. 9.4]. From Theorem 2.4.9 it is possible to deduce statement 1. of Theorem 2.4.8. By a contradiction argument based on Exercise 2.4.7, one then deduces that the paths can not be continuously differentiable, but the statement 3. of Theorem 2.4.8 is much stronger, see for example [14, Thm. 10.3].

Before the end of this section, let us briefly mention the definition of a N-dimensional Brownian motion. The definition is completely similar to Definition 2.4.1.

Definition 2.4.10 (N-dimensional Brownian motion)**.** *A Stochastic process $(\Omega,\mathcal{F},\mathbb{P},(\mathcal{F}_t)_{t\in\mathbb{R}_+},(B_t)_{t\in\mathbb{R}_+})$ taking values in $\mathbb{R}^N$ is a* N-dimensional Brownian motion *if*

1. *$B_0=\mathbf{0}$ a.s.,*
2. *For any $0\le s\le t$ the random variable B_t-B_s is independent of $\mathcal{F}_s$,*
3. *For any $0\le s<t$ the random variable B_t-B_s is a Gaussian random variable $N(0,(t-s)I)$, where I denotes the $N\times N$ identity matrix.*

As for the 1-dimensional Brownian motion, we always assume the continuity of this stochastic process. Properties of the N-dimensional Brownian motion are similar to the 1-dimensional Brownian motion since each of its N components correspond to an independent 1-dimensional Brownian motion [2, Rem. 3.2].

Exercise 2.4.11. *Work on some of the exercises proposed in [1, Chap. 2 & 3] or in [2, Chap. 1 to 3].*

Chapter 3

Conditional expectation and martingales

In this chapter, we introduce the concept of martingales, which are special instances of stochastic processes. However, their definition involves the notion of conditional expectation and of conditional probability. We therefore start by introducing these notions.

3.1 Conditional expectation and conditional probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let A, B be two events, with $\mathbb{P}(B) \neq 0$. The *conditional probability* $\mathbb{P}(A|B)$ of A knowing B is defined by the formula

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

Clearly, if the events A and B are independent, one has

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

as it should be.

Whenever X, Y are two real and discrete valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, the notion of *conditional probability* of X given Y is also naturally defined: For any A included in the range of X and for any y belonging to the range of Y, we set

$$\mathbb{P}(X \in A|Y = y) := \frac{\mathbb{P}(X \in A, Y = y)}{\mathbb{P}(Y = y)} \tag{3.1.1}$$

assuming that $\mathbb{P}(Y = y) > 0$. This quantity corresponds to the probability of the random variable $X \in A$ knowing that the random variable Y takes the value y. In more precise terms, this reads

$$\mathbb{P}(X \in A|Y = y) = \frac{\mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A \text{ and } Y(\omega) = y\})}{\mathbb{P}(\{\omega' \in \Omega \mid Y(\omega') = y\})}.$$

Clearly, for B in the range of Y we can also define $\mathbb{P}(X \in A|Y \in B)$ by a similar formula:

$$\mathbb{P}(X \in A|Y \in B) = \frac{\mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A \text{ and } Y(\omega) \in B\})}{\mathbb{P}(\{\omega' \in \Omega \mid Y(\omega') \in B\})}.$$

Based on these formulas, the *conditional expectation* $\mathbb{E}(X|Y=y)$ of X given $Y=y$ can be computed by

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \sum_{x\in X(\Omega)} x\,\mathbb{P}(X=x|Y=y)\\ &= \sum_{x\in X(\Omega)} x\frac{\mathbb{P}(X=x \text{ and } Y=y)}{\mathbb{P}(Y=y)}\\ &= \frac{1}{\mathbb{P}(Y=y)}\sum_{x\in X(\Omega)} x\,\mathbb{P}(\{\omega\in\Omega \mid X(\omega)=x \text{ and } Y(\omega)=y)\\ &= \frac{\mathbb{E}(X\mathbf{1}_{Y=y})}{\mathbb{P}(Y=y)},\end{aligned} \tag{3.1.2}$$

where the last equality is left as an exercise. Note that $X\mathbf{1}_{Y=y}$ is indeed a new real and discrete valued random variable.

These expressions are well defined because the random variables (and in particular Y) were supposed to be discrete valued. Our aim is to generalize these concepts to arbitrary random variables. We start with a significant generalization of the conditional expectation given in (3.1.2). For it, observe that if X is a random variable on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with values in $(\Lambda,\mathcal{E})$, and if $D\in\mathcal{F}$, we can define a new induced measure $\mu_X^D:\mathcal{E}\to[0,1]$ given for any $A\in\mathcal{E}$ by

$$\mu_X^D(A) := \mathbb{P}\Big(X^{-1}(A)\cap D\Big) = \mathbb{P}\Big(\{\omega\in D \mid X(\omega)\in A\}\Big),$$

which is well defined since $(X^{-1}(A)\cap D)\in\mathcal{F}$. Observe that in general $\mu_X^D(\Lambda)\neq 1$, but nevertheless $\mu_X^D(\Lambda)\in[0,1]$. If $(\Lambda,\mathcal{E})$ is standard, it is possible to set

$$\int_D X\,\mathrm{d}\mathbb{P} = \int_D X(\omega)\,\mathbb{P}(\mathrm{d}\omega) := \int_\Lambda x\,\mu_X^D(\mathrm{d}x) = \mathbb{E}(\mathbf{1}_D X), \tag{3.1.3}$$

as long as the integral converges absolutely. In the sequel, we shall use the first notation in (3.1.3), as it often appears in the literature. Let us also observe that the notion of $L^1(\Omega,\mathcal{F},\mathbb{P})$ introduced in Section 1.4 is not only well defined for univariate random variables, but for any random variables taking values in a standard measurable space $(\Lambda,\mathcal{E})$. Therefore, the previous definition holds for any $X\in L^1(\Omega,\mathcal{F},\mathbb{P})$.

Definition 3.1.1 (Conditional expectation with respect to a σ-subalgebra)**.** *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let X be a random variable on $(\Omega,\mathcal{F},\mathbb{P})$, taking with values in a standard measurable space and belonging to $L^1(\Omega,\mathcal{F},\mathbb{P})$. Let also $\mathcal{G}$ be a σ-subalgebra of $\mathcal{F}$. The* conditional expectation of X given $\mathcal{G}$*, denoted by $\mathbb{E}(X|\mathcal{G})$, is the random variable taking values in $(\Lambda,\mathcal{E})$, measurable with respect to $\mathcal{G}$, belonging to $L^1(\Omega,\mathcal{G},\mathbb{P})$, and satisfying for any $D\in\mathcal{G}$*

$$\int_D \mathbb{E}(X|\mathcal{G})\,\mathrm{d}\mathbb{P} = \int_D X\,\mathrm{d}\mathbb{P}. \tag{3.1.4}$$

In this definition, the random variable $\mathbb{E}(X|\mathcal{G})$ is defined up to a set of $\mathbb{P}$-measure 0, which means that the *conditional expectation of X given $\mathcal{G}$* should be considered as an equivalence class of random variables. Clearly, we fix one in its representatives and consider always this one. Observe that in the l.h.s. of (3.1.4), only the *restricted measure $\mathbb{P}$ on $\mathcal{G}$* is involved, namely only the values of $\mathbb{P}$ on elements of $\mathcal{G}$ are playing a role in the integral. Nevertheless, we keep the same notation for this restricted measure.

It is the result of a theorem, see for example [2, Thm. 4.1], that there exists such a random variable $\mathbb{E}(X|\mathcal{G})$. Note that a rather easy consequence of this definition is that the equality

$$\mathbb{E}(W\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(WX) \tag{3.1.5}$$

holds for any bounded and $\mathcal{G}$-measurable univariate[5] random variable W on $(\Omega, \mathcal{F}, \mathbb{P})$. The property (3.1.5) is sometimes taken as part of the definition of the conditional expectation of X given $\mathcal{G}$, see [1, Def. 4.14]. By choosing the constant random variable $W = 1$ (function taking the constant value 1), one deduces from (3.1.5) that

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X). \tag{3.1.6}$$

Exercise 3.1.2. *Prove (3.1.5), or at least justifies it as precisely as possible, starting from Definition 3.1.1.*

The interest in the previous notion comes from the following framework. Let us consider a second random variable Y on $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $(\Lambda', \mathcal{E}')$, and recall that $\sigma(Y)$ denotes the smallest σ-subalgebra of $\mathcal{F}$ defined by Y. We can then consider the conditional expectation of X given $\sigma(Y)$, and denote it simply by $\mathbb{E}(X|Y)$ instead of $\mathbb{E}(X|\sigma(Y))$. Thus, if $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ takes values in the standard measurable space $(\Lambda, \mathcal{E})$, then the new random variable $\mathbb{E}(X|Y)$ belongs to $L^1(\Omega, \sigma(Y), \mathbb{P})$ and also takes values in $(\Lambda, \mathcal{E})$. By Doob-Dynkin lemma, one has automatically $\mathbb{E}(X|Y) = g(Y)$ for some measurable function $g : \Lambda' \to \Lambda$. Having this in mind, it follows that the univariate random variable W of (3.1.5) is always of the form $h(Y)$ for some measurable and bounded $h : \Lambda' \to \mathbb{R}$, where $\mathbb{R}$ is endowed with the σ-algebra σ_B of Borel sets on $\mathbb{R}$. In this case (3.1.5) reads

$$\mathbb{E}(h(Y)\mathbb{E}(X|Y)) = \mathbb{E}(h(Y)g(Y)) = \mathbb{E}(h(Y)X). \tag{3.1.7}$$

Let us now list a few properties of the conditional expectations. The proof of the next statement is left as an exercise. It is not so difficult, and inspiration can be obtained from [2, Prop. 4.1].

Proposition 3.1.3. *Let X, X^1, X^2 be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in a standard measurable space $(\Lambda, \mathcal{E})$, and assume that these random variables belong to $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a σ-subalgebra of $\mathcal{F}$, and let $\alpha, \beta \in \mathbb{R}$.*

1. $\mathbb{E}(\alpha X^1 + \beta X^2|\mathcal{G}) = \mathbb{E}(\alpha X^1|\mathcal{G}) + \mathbb{E}(\beta X^2|\mathcal{G})$,
2. *If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$,*
3. *If $X \geq 0$ a.s., then $\mathbb{E}(X|\mathcal{G}) \geq 0$ a.s.,*
4. *If W is an univariate bounded and $\mathcal{G}$-measurable random variable, then $\mathbb{E}(WX|\mathcal{G}) = W\mathbb{E}(X|\mathcal{G})$ a.s.,*
5. *If $\mathcal{G}'$ is another σ-subalgebra of $\mathcal{F}$ satisfying $\mathcal{G} \subset \mathcal{G}'$, then $\mathbb{E}(\mathbb{E}(X|\mathcal{G}')|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$ a.s.,*
6. *If X is independent of $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ a.s., where $\mathbb{E}(X)$ can be considered as a constant random variable,*
7. *If X is univariate and $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex lower semi-continuous function, then*

$$\mathbb{E}(\varphi(X)|\mathcal{G}) \geq \varphi(\mathbb{E}(X|\mathcal{G})). \qquad \textit{(Jensen's inequality)}$$

Exercise 3.1.4. *Study Examples 4.1, 4.2, and 4.3 on p. 92–93 of [2].*

[5] Why univariate only ?

Exercise 3.1.5 (♥). *In the framework of the previous proposition and for univariate random variables, show that the map $X \mapsto \mathbb{E}(X|\mathcal{G})$ is a bounded linear map from $L^p(\Omega, \mathcal{F}, \mathbb{P})$ to $L^p(\Omega, \mathcal{G}, \mathbb{P})$ with a norm smaller or equal to* 1*, for any $p \geq 1$. More explicitly, show the linearity and that $\mathbb{E}(|\mathbb{E}(X|\mathcal{G})|^p) \leq \mathbb{E}(|X|^p)$. In the proof, use Jensen's inequality for the function $x \mapsto |x|^p$.*

Let us now provide an interpretation of $\mathbb{E}(X|\mathcal{G})$ in the framework of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ for univariate random variables. Note firstly that $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, since it is a stable under addition and multiplication by scalars. In addition, for the univariate random variable $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ one infers from (3.1.5) that for any bounded and univariate $\mathcal{G}$-measurable function W on Ω one has

$$\mathbb{E}\Big(W(X - \mathbb{E}(X|\mathcal{G}))\Big) = \mathbb{E}(WX) - \mathbb{E}(W\mathbb{E}(X|\mathcal{G})) = 0. \tag{3.1.8}$$

Since bounded and $\mathcal{G}$-measurable functions are dense in $L^2(\Omega, \mathcal{G}, \mathbb{P})$, it follows that (3.1.8) holds for any $W \in L^2(\Omega, \mathcal{G}, \mathbb{P})$, when the expectation is understood as a scalar product, see section 1.4. Thus, $X - \mathbb{E}(X|\mathcal{G})$ (which belongs to $L^2(\Omega, \mathcal{F}, \mathbb{P})$) is orthogonal to all elements of the subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Equivalently, it means that $\mathbb{E}(X|\mathcal{G})$ is the orthogonal projection of X on $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Still in other words, $\mathbb{E}(X|\mathcal{G})$ is the element of the subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$ which minimizes the L^2-distance to X, or equivalently $\mathbb{E}(X|\mathcal{G})$ is the best approximation of X by elements of $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Later on and for $\mathcal{G} = \sigma(Y)$, we shall understand $\mathbb{E}(X|Y)$ as the best estimation of X given the information of Y.

The statements 2. and 6. of Proposition 3.1.3 are particularly simple. In the next statement we combine them, and the result turns out to be useful, see [2, Lem. 4.1] for a sketch of the proof, and [14, Lem. A.3] for more details.

Lemma 3.1.6 (Freezing lemma). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}_1, \mathcal{G}_2$ be two independent σ-subalgebras of $\mathcal{F}$. For $j \in \{1, 2\}$ let X^j be a $\mathcal{G}_1$-measurable random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ to a standard measurable space $(\Lambda^j, \mathcal{E}^j)$ and belonging to $L^1(\Omega, \mathcal{G}_j, \mathbb{P})$. Let $\Psi : \Lambda^1 \times \Lambda^2 \to \mathbb{R}$ be a measurable function, when $\Lambda^1 \times \Lambda^2$ is endowed with the σ-algebra $\mathcal{E}^1 \times \mathcal{E}^2$. Then the following equalities hold:*

$$\mathbb{E}(\Psi(X^1, X^2)|\mathcal{G}^1) = \Big(\mathbb{E}(\Psi(\cdot, X^2))\Big)(X^1) = \mathbb{E}(\Psi(X^1, X^2)|X^1)$$

wherever the map $\Omega \ni \omega \mapsto \Psi(X^1(\omega), X^2(\omega)) \in \mathbb{R}$ is absolutely integrable.

Exercise 3.1.7. *Check the statement for a function Ψ satisfying $\Psi(x^1, x^2) = \Psi_1(x^1)\Psi_2(x^2)$, for $x^1 \in \Lambda^1$ and $x^2 \in \Lambda^2$.*

Exercise 3.1.8 (♥). *Study the position of the Brownian motion at a random time, see Example 4.5 p. 95 of [2].*

Let us now move to the notion of conditional probability, as sketched in (3.1.1) in the discrete setting.

Definition 3.1.9 (Conditional probability). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X, Y be random variables on this space, with X taking values in a measurable space $(\Lambda, \mathcal{E})$ and Y taking values in a measurable space $(\Lambda', \mathcal{E}')$. A family $\{\nu_y\}_{y \in \Lambda'}$ of probability measures on $(\Lambda, \mathcal{E})$ is called a* conditional probability for X given Y *if*

1. *For every $A \in \mathcal{E}$ the map $\Lambda' \ni y \mapsto \nu_y(A) \in \mathbb{R}$ is measurable from $(\Lambda', \mathcal{E}')$ to $(\mathbb{R}_+, \sigma_B)$,*
2. *For any $A \in \mathcal{E}$ and $B \in \mathcal{E}'$ one has*

$$\mathbb{P}(X \in A, Y \in B) = \int_B \nu_y(A)\mu_Y(\mathrm{d}y), \tag{3.1.9}$$

where μ_Y is the induced probability measure of the random variable Y.

The probability measure ν_y on $(\Lambda, \mathcal{E})$ can be understood as a suitable law for the random variable X keeping into account the information $Y = y$. Since the l.h.s. of (3.1.9) can be understood as $\mathbb{E}(\mathbf{1}_A(X)\mathbf{1}_B(Y))$, one can extend this equality by linearity. More precisely if $(\Lambda, \mathcal{E})$ is a standard measurable space, if $f : \Lambda \to \Lambda$ is measurable and verifies $f(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and if $h : \Lambda' \to \mathbb{R}$ is measurable and bounded, then one obtains the equality

$$\mathbb{E}(h(Y)f(X)) = \int_{\Lambda'} \Big(\int_{\Lambda} f(x)\, \nu_y(\mathrm{d}x) \Big) h(y) \mu_Y(\mathrm{d}y). \tag{3.1.10}$$

Considering then $f = \mathrm{id}$, and setting $g(y) := \int_\Lambda x\, \nu_y(\mathrm{d}x)$ the previous equation reads

$$\mathbb{E}(h(Y)X) = \int_{\Lambda'} h(y)\, g(y) \mu_Y(\mathrm{d}y) = \mathbb{E}(h(Y)g(Y)).$$

By a comparison with (3.1.7), we observe that $\{\nu_y\}_{y\in\Lambda'}$ is the conditional probability of X given Y if and only if whenever $f(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ one has

$$\mathbb{E}(f(X)|Y) = g(Y) \text{ a.s.} \quad \text{with} \quad g(y) = \int_\Lambda f(x)\, \nu_y(\mathrm{d}x). \tag{3.1.11}$$

In particular, for $f = \mathrm{id}$ one infers that

$$\mathbb{E}(X|Y) = g(Y) \text{ a.s.} \quad \text{with} \quad g(y) = \int_\Lambda x\, \nu_y(\mathrm{d}x)$$

which means that the conditional expectation is the mean value of the conditional probability.

If the conditional probability $\{\nu_y\}_{y\in\Lambda'}$ exists, the following notations are used for any $A \in \mathcal{E}$:

$$\mathbb{P}(X \in A|Y = y) = \mathbb{E}(\mathbf{1}_A(X)|Y = y) = \nu_y(A).$$

However, let us stress that the conditional expectation $\mathbb{E}(X|Y)$ always exists, while nothing similar can be said about the conditional probability. Nevertheless, we shall see that it exists in some standard situations, as shown below.

Example 3.1.10. *Let X a $\mathbb{R}^m$-valued random variable and let Y be a $\mathbb{R}^n$-valued random variable, both defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that their joint probability measure $\mu_{(X,Y)}$ is absolutely continuous, meaning that there exists a probability density function $\Pi_{(X,Y)} : \mathbb{R}^{m+n} \to \mathbb{R}_+$, verifying $\int_{\mathbb{R}^{m+n}} \Pi_{(X,Y)}(x, y)\, \mathrm{d}x\, \mathrm{d}y < \infty$ and*

$$\mu_{(X,Y)}(A) = \int_A \Pi_{(X,Y)}(x, y)\, \mathrm{d}x\, \mathrm{d}y \qquad \forall A \subset \sigma_{\mathrm{B}}(\mathbb{R}^{m+n}).$$

Let $\Pi_Y : \mathbb{R}^n \to \mathbb{R}_+$ be the marginal density function of Y defined by $\Pi_Y(y) := \int_{\mathbb{R}^m} \Pi_{(X,Y)}(x, y)\, \mathrm{d}x$, and set $Q := \{y \in \mathbb{R}^n \mid \Pi_Y(y) = 0\}$. Clearly, $\mathbb{P}(Y \in Q) = \int_Q \Pi_Y(y)\, \mathrm{d}y = 0$. We then define for a.e. $x \in \mathbb{R}^m$

$$\overline{\Pi}_{(X,y)}(x) := \begin{cases} \frac{\Pi_{(X,Y)}(x,y)}{\Pi_Y(y)} & \text{if } y \notin Q, \\ \pi(x) & \text{if } y \in Q, \end{cases}$$

with $\pi : \mathbb{R}^m \to \mathbb{R}_+$ any density function. Then for any $y \in \mathbb{R}^n$ the function $\overline{\Pi}_{(X,y)}$ corresponds to the density function of the conditional probability of X given Y, previously denoted by ν_y. Indeed, one can check the conditions of Definition 3.1.9, with $\mu_y = \overline{\Pi}_{(X,y)}$:

1. *If $y \notin Q$:*
$$\int_{\mathbb{R}^m} \overline{\Pi}_{(X,y)}(x)\,\mathrm{d}x = \int_{\mathbb{R}^m} \frac{\Pi_{(X,Y)}(x,y)}{\Pi_Y(y)}\mathrm{d}x = 1,$$
and if $y \in Q$ the same result holds,

2. *For any $A \in \sigma_{\mathrm{B}}(\mathbb{R}^m)$, the function*
$$\mathbb{R}^n \ni y \mapsto \overline{\Pi}_{(X,y)}(A) = \begin{cases} \frac{1}{\Pi_Y(y)} \int_A \Pi_{(X,Y)}(x,y)\,\mathrm{d}x & \text{if } y \notin Q, \\ \int_A \pi(x)\,\mathrm{d}x & \text{if } y \in Q, \end{cases} \in \mathbb{R}_+$$
is measurable,

3. *If $A \in \sigma_{\mathrm{B}}(\mathbb{R}^m)$ and $B \in \sigma_{\mathrm{B}}(\mathbb{R}^m)$ with $B \subset Q^c$, then*
$$\mathbb{P}(X \in A, Y \in B) = \int_B \Big(\int_A \Pi_{(X,Y)}(x,y)\,\mathrm{d}x \Big)\mathrm{d}y = \int_B \Big(\int_A \overline{\Pi}_{(X,y)}(x)\,\mathrm{d}x \Big)\Pi_Y(y)\,\mathrm{d}y.$$
If $B \subset Q$, then the same equality holds, with both sides equal to 0.

Thus, the function $y \mapsto \overline{\Pi}_{(X,y)}$ corresponds to the density of the conditional probability of X given Y, or equivalently $\mathbb{E}(X|Y = y) = \overline{\Pi}_{(X,y)}\,\mathrm{d}x$.

It is also a good exercise to check that the formalism developed in this section can be applied to discrete valued random variables, and that it leads to some familiar expressions, as presented at the very beginning of this section. Alternatively, Gaussian random variables give other possible applications:

Exercise 3.1.11. *Study and report about the conditional probability for Gaussian vectors, as presented in [2, Sec. 4.4].*

3.2 Martingales

In this section we study a new type of stochastic processes, which are at the root of the subsequent developments.

Definition 3.2.1 (Martingale, supermartingale, submartingale)**.** *For $\mathcal{T} \subset \mathbb{R}_+$, a real valued stochastic process $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (M_t)_{t\in\mathcal{T}})$ satisfying $M_t \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ for any $t \in \mathcal{T}$ is a* martingale *if $\mathbb{E}(M_t|\mathcal{F}_s) = M_s$ for all $s \le t$. It is a* supermartingale *if $\mathbb{E}(M_t|\mathcal{F}_s) \le M_s$ or a* submartingale *if $\mathbb{E}(M_t|\mathcal{F}_s) \ge M_s$.*

Note that a martingale is a special instance of a supermartingale and of a submartingale. As already mentioned for the Brownian motion, if the filtration is not mentioned, it means that the natural one is considered. Note also that we consider only the univariate case for simplicity, but martingales can also take values in a standard measurable space, or have values in $\mathbb{C}$. Let us start by looking at some example of martingales.

Exercise 3.2.2. *Let $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let X be a univariate random variable on this space, with $\mathbb{E}(|X|) < \infty$. Set $X_t := \mathbb{E}(X|\mathcal{F}_t)$. Show that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (X_t)_{t\in\mathcal{T}})$ is a martingale.*

Exercise 3.2.3. *Consider $\mathcal{T} = \mathbb{N}$ and a sequence $(X_n)_{n\in\mathbb{N}}$ of independent and real valued random variables satisfying $\mathbb{E}(|X_n|) < \infty$ and $\mathbb{E}(X_n) = 0$. Set $Y_n := \sum_{j=1}^n X_j$. Show that $(Y_j)_{j\in\mathbb{N}}$ and the natural filtration define a martingale.*

Exercise 3.2.4. *Show that the standard*[6] 1*-dimensional Brownian motion is a martingale.*

Exercise 3.2.5. *Let* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (B_t)_{t\in\mathbb{R}_+})$ *be the standard* 1*-dimensional Brownian motion, and consider the* geometric Brownian motion *defined by* $S_t := S_0 \exp(\sigma B_t + \mu t)$*, with* $\sigma > 0$*,* $\mu \in \mathbb{R}$*, and* $S_0 \in \mathbb{R}$ *an arbitrary initial value. Show that this process is a martingale if and only if* $\mu = -\frac{1}{2}\sigma^2$*.*

Exercise 3.2.6. *Let* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (B_t)_{t\in\mathbb{R}_+})$ *be the standard* 1*-dimensional Brownian motion. Show that the new process defined by* $X_t := B_t^2$ *is a submartingale, but that the process defined by* $X_t := B_t^2 - t$ *is a martingale.*

Let us add one more exercise about the relation between supermartingale and martingale, see [2, Ex. 5.1].

Exercise 3.2.7 (Constant expectation)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (M_t)_{t\in\mathcal{T}})$ *be a supermartingale, and assume that* $\mathbb{E}(M_t)$ *is a constant independent of* t*. Then this stochastic process is a martingale.*

Exercise 3.2.8 (Alternative definition of a martingale)**.** *Show that the following definition is equivalent to Definition 3.2.1 : A real valued stochastic process* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (M_t)_{t\in\mathcal{T}})$ *satisfying* $M_t \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *for any* $t \in \mathcal{T}$ *is a* martingale *if* $\mathbb{E}((M_t - M_s)|\mathcal{F}_s) = \mathbf{0}$ *for all* $s \leq t$*.*

Let us now state one result for discrete time martingales, namely when $\mathcal{T} = \mathbb{N}$. Such stochastic processes have many applications, and are simpler than the continuous ones. As a starter, a discrete time stochastic process $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (A_n)_{n\in\mathbb{N}})$ is said to be an *increasing predictable process* if $A_0 = \mathbf{0}$, $A_n \leq A_{n+1}$, and A_{n+1} is $\mathcal{F}_n$-measurable, for all $n \in \mathbb{N}$. Note that the $\mathcal{F}_n$-measurability of A_{n+1} roughly means that at time n we know the value of the process at time $n+1$. This knowledge is responsible for the term "predictable". Such processes appear in the following statement about the decomposition of any submartingale. It proof can be found here or in [2, Thm. 5.1], and can be studied as an exercise.

Theorem 3.2.9 (Doob's decomposition theorem)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (X_n)_{n\in\mathbb{N}})$ *be a submartingale, then there exists a unique decomposition* $X_n = M_n + A_n$*, with* $(M_n)_{n\in\mathbb{N}}$ *a martingale and* $(A_n)_{n\in\mathbb{N}}$ *an increasing predictable process. The process* $(A_n)_{n\in\mathbb{N}}$ *is called the* compensator*.*

There exists a similar result for continuous time submartingales, but its precise statement is more delicate. However, observe that Exercise 3.2.6 is already an illustration of this result. We now introduce a new concept:

Definition 3.2.10 (Stopping time)**.** *Let* $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ *be a filtration on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. A random variable* $\tau : \Omega \to \mathcal{T} \cup \{+\infty\}$ *on this probability space is said to be a* stopping time *for this filtration if for any* $t \in \mathcal{T}$

$$\{\tau \leq t\} \equiv \{\omega \in \Omega \mid \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

For any stopping time τ *we set* $\mathcal{F}_\tau$ *for the* σ*-subalgebra of* $\mathcal{F}$ *defined by*

$$\mathcal{F}_\tau = \left\{A \in \mathcal{F}_\mathcal{T} \mid (A \cap \{\tau \leq t\}) \in \mathcal{F}_t \text{ for every } t \in \mathcal{T}\right\} \tag{3.2.1}$$

where $\mathcal{F}_\mathcal{T}$ *is the smallest* σ*-subalgebra of* $\mathcal{F}$ *containing all* $\mathcal{F}_t$ *for* $t \in \mathcal{T}$*.*

In less precise words, τ is a stopping time if we can decide if the events $\{\omega \in \Omega \mid \tau(\omega) \leq t\}$ occurred based on the information available at time t. Note that since $\mathcal{F}_t$ is stable under complement, the event $\{\omega \in \Omega \mid \tau(\omega) > t\}$ also belongs to $\mathcal{F}_t$. Typical examples of stopping times are entry and hitting times of a $\mathbb{R}^N$-valued process $(X_t)_{t\in\mathcal{T}}$ into a set $A \in \sigma_B(\mathbb{R}^N)$, as for example:

1. First *entry time* into A: $\tau_A^o(\omega) : \inf\{t \geq 0 \mid X_t(\omega) \in A\}$,

[6] *Standard Brownian motion* means the Brownian motion endowed with the right-continuous filtration generated by the augmented natural filtration.

2. First *hitting time* of A: $\tau_A(\omega) : \inf\{t > 0 \mid X_t(\omega) \in A\}$,

3. First *exit time* from A: $\tau_{A^c}(\omega) : \inf\{t > 0 \mid X_t(\omega) \notin A\}$,

with the convention that $\inf(\emptyset) := \infty$. Note that depending on the regularity of the process and of the filtration, and depending on the nature of A, it can be difficult to prove rigorously that these random variables are indeed stopping times. Two example of precise statements are indicated below. Recall that a right-continuous filtration $(\mathcal{F}_{t+})_{t\in\mathcal{T}}$ can be constructed from any filtration, see the last paragraph of Section 2.3.

Lemma 3.2.11. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (X_t)_{t\in\mathbb{R}_+})$ be a $\mathbb{R}^N$-valued stochastic process with right-continuous paths, and let $A \subset \mathbb{R}^N$ be an open set. Then the first hitting time τ_A is a stopping time if $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is the right-continuous filtration constructed from the natural filtration.*

Lemma 3.2.12. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (X_t)_{t\in\mathbb{R}_+})$ be a $\mathbb{R}^N$-valued stochastic process with continuous paths, and let $A \subset \mathbb{R}^N$ be a closed set. Then the first entry time τ_A^o is a stopping time for the natural filtration, while τ_A is a stopping time for the right-continuous filtration constructed from the natural filtration.*

We refer to [14, Sec. 5.2] for the proof of these statements. Note that any *last passage time* can not (in general) be a stopping time, since the future should be known in order to decide if it is a last passage or not. We now gather a few general results about stopping times. The easy proof is left as an exercise, see also [2, Prop. 3.5].

Lemma 3.2.13. *Let τ, η be two stopping times for the same filtration.*

1. *τ is $\mathcal{F}_\tau$-measurable,*

2. *$\tau \vee \eta := \max\{\tau, \eta\}$ and $\tau \wedge \eta := \min\{\tau, \eta\}$ are stopping times,*

3. *If $\eta \leq \tau$, then $\mathcal{F}_\eta \subset \mathcal{F}_\tau$,*

4. *$\mathcal{F}_{\eta\wedge\tau} = \mathcal{F}_\eta \cap \mathcal{F}_\tau$.*

Based on the notion of stopping time, let us state an important result for martingales, called *Stopping theorem*, see [2, Thm. 5.13].

Theorem 3.2.14. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (M_t)_{t\in\mathcal{T}})$ be a right-continuous martingale, and let τ_1, τ_2 be two a.s. bounded stopping times, with $\tau_1 \leq \tau_2$ a.s.. Then $\mathbb{E}(M_{\tau_2}|\mathcal{F}_{\tau_1}) = M_{\tau_1}$. The same statement holds for supermartingale with $\mathbb{E}(M_{\tau_2}|\mathcal{F}_{\tau_1}) \leq M_{\tau_1}$, and for submartingale with $\mathbb{E}(M_{\tau_2}|\mathcal{F}_{\tau_1}) \geq M_{\tau_1}$.*

A rather direct consequence of this statement is a statement about *stopped martingales*, see [1, Prop. 4.37] or [2, Thm. 5.14].

Proposition 3.2.15. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (M_t)_{t\in\mathcal{T}})$ be a right-continuous martingale, and let τ be a stopping time. Then $(M_{t\wedge\tau})_{t\in\mathcal{T}}$ defines also a martingale with the same filtration, where $M_{t\wedge\tau}(\omega) = M_t(\omega)$ if $t \leq \tau(\omega)$ and $M_{t\wedge\tau}(\omega) = M_{\tau(\omega)}(\omega)$ if $t \geq \tau(\omega)$.*

Exercise 3.2.16. *In the discrete case $\mathcal{T} = \mathbb{N}$, show that $(M_{n\wedge\tau})_{n\in\mathbb{N}}$ also also adapted to the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.*

Observe that the notion of right-continuity was used in the previous two statements. Unlike Brownian motions, martingales are not automatically continuous (through a modification). Nevertheless, the following statement provides a sufficient criterion for the right continuity, see [2, Thm. 5.14] and [3, Thm. 1.4.3 & Corollary].

Theorem 3.2.17. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\mathcal{T}}, (M_t)_{t\in\mathcal{T}})$ be a supermartingale, and assume that the filtration is standard (= right-continuous and containing the negligible sets). Then $(M_t)_{t\in\mathcal{T}}$ admits a right-continuous modification if and only if the map $t \mapsto \mathbb{E}(M_t)$ is continuous. In particular, if $(M_t)_{t\in\mathcal{T}}$ is a martingale, it has a right-continuous modification.*

Let us still mentioned another consequence of the stopping theorem, which is referred to as the *Optional Stopping Theorem*:

Theorem 3.2.18. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (M_n)_{n\in\mathbb{N}})$ be a discrete time martingale, and let τ be an a.s. bounded stopping time. Then $\mathbb{E}(M_\tau) = \mathbb{E}(M_0)$.*

Let us now state a result which makes discrete time martingales particularly attractive. For this, observe that any univariate random variable X can be written as $X = X^+ - X^-$ with $X^\pm \geq 0$ (decomposition into positive and negative part of any function).

Theorem 3.2.19. *$(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (M_n)_{n\in\mathbb{N}})$ be a discrete time supermartingale satisfying the uniform condition $\sup_{n\in\mathbb{N}} \mathbb{E}(M_n^-) < \infty$. Then $(M_n)_{n\in\mathbb{N}}$ converges almost surely to a finite limit, namely there exists $C < \infty$ such that*

$$\mathbb{P}(\{\omega \in \Omega \mid \lim_{n\to\infty} M_n(\omega) = C\}) = 1.$$

In particular, if $M_n \geq 0$, then $(M_n)_{n\in\mathbb{N}}$ converges almost surely to a finite limit.

For various exercise, it is necessary to use the *Dominated convergence theorem*, which is now recalled:

Theorem 3.2.20. *If $(X_n)_{n\in\mathbb{N}}$ is a family of univariate random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ converging almost surely to a random variable X_∞, and if there exists another random variable $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ verifying $|X_n| \leq Y$, then $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X_\infty)$.*

Let us mention one specific application of this theorem. In its framework, suppose that all univariate random variables are uniformly bounded by a constant, namely there exists $c > 0$ such that $|X_n| \leq c$ for all n. Since c, seen as a constant function, belongs to $L^1(\Omega, \mathcal{F}, \mathbb{P})$, then this constant function can play the role of the random variable Y and the theorem applies. It means that any uniformly bounded family of univariate random variables $\{X_n\}_{n\in\mathbb{N}}$ converging almost surely to a random variable X_∞ satisfies $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X_\infty)$. In particular, it implies that any convergence almost surely is also a convergence in probability: If $\{X_n\}_{n\in\mathbb{N}}$ converges almost surely to X_∞, for any $\varepsilon > 0$ one has

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X_\infty| > \varepsilon) = \lim_{n\to\infty} \mathbb{E}(\mathbf{1}_{|X_n - X_\infty|>\varepsilon}) = \mathbb{E}(\lim_{n\to\infty} \mathbf{1}_{|X_n - X_\infty|>\varepsilon}) = \mathbb{E}(\mathbf{1}_{|\mathbf{0}|>\varepsilon}) = 0 \tag{3.2.2}$$

where we have used the uniform bound $\mathbf{1}_{|X_n - X_\infty|>\varepsilon} \leq 1$ for the second equality.

Exercise 3.2.21. *Study and report on the gambler's ruin problem, see for example [1, Example 4.41 & 4.42].*

Exercise 3.2.22. *Study and report on the first passage time for the 1D Brownian motion, for example see [1, Example 4.43].*

Exercise 3.2.23. *Work on some problems proposed in [1, p. 93–97] or in [2, p. 75–85, 104–107, or 139–150]. The second book is more complicated, but solutions or hints are presented at the end of the book.*

Let us conclude this section with a few additional results about the 1D standard Brownian motion $(B_t)_{t\geq 0}$, mostly based on the notion of stopping time. For more details, we refer to [1, Sec. 4.5] and to [11, Chap. 3].

Let $a \in \mathbb{R}$ and let τ_a be the first hitting time, namely $\tau_a := \inf\{t > 0 \mid B_t = a\}$. For $a < 0 < b$ we also set $\tau_{ab} := \min\{\tau_a, \tau_b\}$, which corresponds to the time to exit the interval (a, b). Note that these random variables take values in $\mathbb{R}_+$.

Proposition 3.2.24. *For any $a < 0 < b$, one has $\mathbb{P}(\tau_{ab} < \infty) = 1$ and $\mathbb{E}(\tau_{ab}) < \infty$. Similarly, $\mathbb{P}(\tau_a < \infty) = 1$ and $\mathbb{P}(\tau_0 < \infty) = 1$.*

We refer to [11, Thm. 3.13 & 3.14] for the proof of the previous statement, and to [11, Thm. 3.18] for the following one.

Proposition 3.2.25. *For any $a \in \mathbb{R}$ the random variable τ_a corresponding to the hitting time is absolutely continuous, with probability density function $\Pi_{\tau_a} : \mathbb{R}_+ \to \mathbb{R}_+$ given for $t > 0$ by*

$$\Pi_{\tau_a}(t) := \frac{|a|}{\sqrt{2\pi}} t^{-\frac{3}{2}} \mathrm{e}^{-\frac{a^2}{2t}}.$$

Also, $\mathbb{E}(\tau_a) = \infty$.

As a final nice result, let us provide the probability that the standard Brownian motion starting at 0 comes back to 0 in a certain interval of time, see [11, Sec. 3.9].

Proposition 3.2.26. *For $0 < t < T$, the probability that the Brownian motion starting at 0 comes back to 0 at least once in the time interval (t, T) is given by $\frac{2}{\pi} \arccos\left(\frac{t}{T}\right)$. The probability that it does not come back to 0 in the time interval (t, T) is given by $\frac{2}{\pi} \arcsin\left(\frac{t}{T}\right)$.*

Many other quantities can be explicitly computed for the Brownian motion. You can enjoy reading various textbooks on the topic, and write any report on this topic. ☺

Bibliography

[1] J.-L. Arguin, *A first course in stochastic calculus*, Pure and Applied Undergraduate texts 53, American Mathematical Society, 2022.

[2] P. Baldi, *Stochastic calculus, an introduction through theory and exercises*, Universitext, Springer, 2017.

[3] K.L. Chung, *Lectures from Markov processes to Brownian motion*, Springer-Verlag, New York-Berlin, 1982.

[4] J. Duan, *An Introduction to Stochastic Dynamics*, Cambridge University Press, 2015.

[5] R. Durrett, *Stochastic calculus, a practical introduction*, CRC 1996.

[6] L.C. Evans, *An introduction to stochastic differential equations*, American Mathematical Society, 2013.

[7] A. Friedman, *Stochastic differential equations and applications (Volume 1)*, Academic Press, 1975.

[8] G. Grimmett, D. Welsh, *Probability: an introduction*, Second edition. Oxford University Press, 2014.

[9] M. Holmes-Cerfon, *Applied Stochastic Analysis Lecture Notes*, see this website (2023).

[10] I. Karatzas, S.E. Shreve, *Methods of Mathematical Finance*, Springer, 1998.

[11] F. Klebaner, *Introduction to stochastic calculus, with applications*, third edition, Imperial College Press, 2012.

[12] Hui-Hsiung Kuo, *Introduction to stochastic integration*, Springer, 2006.

[13] T. Mikosch, *Elementary stochastic calculus with finance in mind*, Advances Series on Statistical Science and Applied Probability 6, World Scientific, 1998.

[14] R. Schilling, L. Partzsch, *Brownian Motion: an introduction to stochastic processes*, De Gruyter, 2014.

[15] R.J. Williams, *Introduction to the mathematics of finance*, American Mathematical Society, Providence, 2006.