Special Mathematics Lecture

Introduction to Stochastic Calculus

Nagoya University, Fall 2023

Lecturer: Serge Richard
Teaching assistant: Vic Austen

Goals of these Lectures notes:

Provide the necessary background information for understanding the main ideas of stochastic calculus. These notes correspond to 14 lectures lasting 90 minutes each.

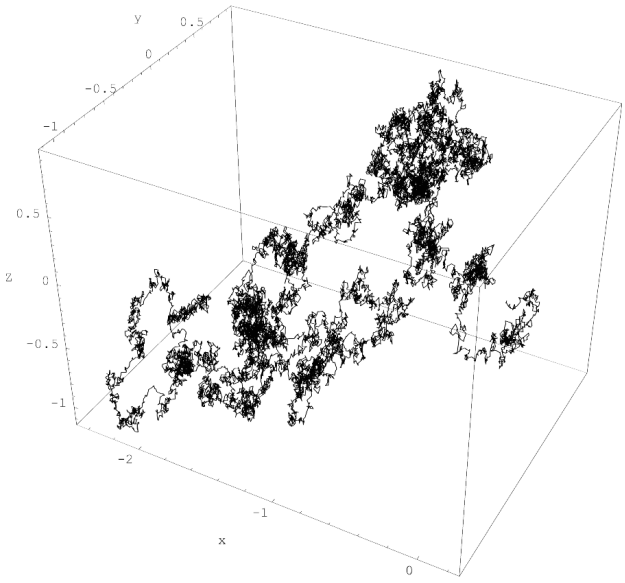

Figure 1: Brownian path in 3D

Website for this course:
http://www.math.nagoya-u.ac.jp/~richard/SMLfall2023.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].

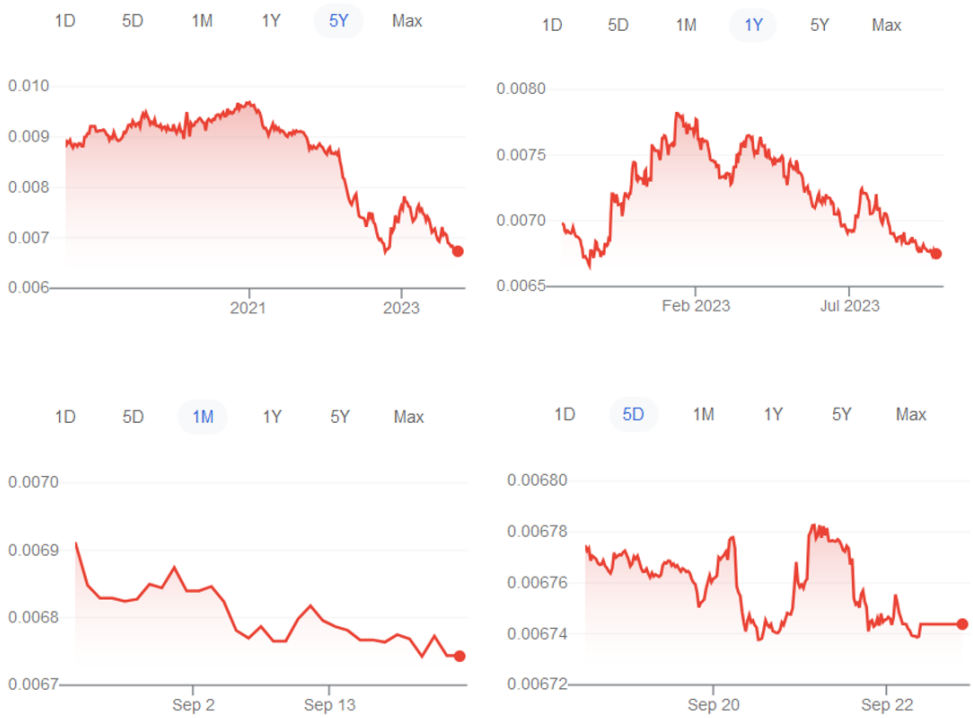

Figure 2: Exchange rate: Japanese yen - US dollar (y-axis gives the value in dollar of 1 yen, as a function of time) over different periods of time

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).

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.*

In many cases, the above integral reduces to a more standard Riemann integral. Thus, we can keep in mind a limit of Riemann sum for this r.h.s. of (1.2.1).

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 define 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, and that $\|\cdot\|_p$ defines a norm on $L^p(\Omega, \mathcal{F}, \mathbb{P})$. 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_∞ if $\|X_j - X_\infty\|_p \to 0$ as $j \to \infty$.*

Note that $L^p(\Omega, \mathcal{F}, \mathbb{P})$ are Banach spaces, which ensures that the limit X_∞ also belongs to $L^p(\Omega, \mathcal{F}, \mathbb{P})$ whenever the sequence converges (and even more generally when a Cauchy sequence is considered). 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

Gaussian processes

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.

2.1 Gaussian vectors

We start with the main definition of this section.

Definition 2.1.1 (Gaussian vector)**.** *A N-dimensional random vector $X = (X_1, \dots, X_N)^T$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a* Gaussian vector *if for any $a = (a_1, \dots, a_N)^T \in \mathbb{R}^N$ 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 $X_1, \dots, X_N$ 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 variable 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 exists 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 define 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].

Example 2.2.2 (Brownian process)**.** *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.*

[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 2.2.3 (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 notions. Note that this section is independent of the Gaussian processes mentioned in the title of this chapter, since stochastic processes more general than Gaussian 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*.

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[2]. 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 Borel 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 $\Big(\mathcal{T} \cap [0, t] \times \Omega, \sigma(\mathcal{T} \cap [0, t]) \times \mathcal{F}_t\Big)$ to $(\Lambda, \mathcal{E})$.

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.

[2]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

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 process

In this section we provide a brief description of the Brownian 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 process in dimension 1 because any Brownian process in dimension N decomposes in N independent Brownian processes of dimension 1. We also use the letter B instead of X, since this notation is commonly used for the Brownian process. 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 process)**.** *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 process *if*

1. *$B_0 = \mathbf{0}$ a.s.,*
2. *For any $0 \leq s \leq t$ the random variable $B_t - B_s$ is independent of $\mathcal{F}_s$,*
3. *For any $0 \leq s < t$ 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 $\mathcal{F}_1, \ldots, \mathcal{F}_m$ of σ-subalgebras of $\mathcal{F}$ are called *independent* if for any $A^1 \in \mathcal{F}_1, \ldots, A^m \in \mathcal{F}_m$ one has

$$\mathbb{P}(A^1 \cap A^2 \cap \cdots \cap A^m) = \mathbb{P}(A^1)\mathbb{P}(A^2) \ldots \mathbb{P}(A^m).$$[3]

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

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 process is continuous. In addition the following properties of the Brownian process 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 process. 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 process with the natural filtration. It is called the natural Brownian process.

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 process.

Remark 2.4.5. *The last part of Proposition 2.4.4 means that whenever the Brownian process 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 process 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 process is a Brownian process 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 process 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.

Let us now state additional properties of the Brownian process 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 process. Then,*

1. *For any $s \ge 0$, $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_{t+s})_{t\in\mathbb{R}_+}, (B_{t+s} - B_s)_{t\in\mathbb{R}_+})$ is a 1-dimensional Brownian process (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 process (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 process (scaling),*

4. *The random variables defined by* $Z_t := tB_{1/t}$ *for* $t > 0$ *and* $Z_0 = 0$ *define a natural Brownian process.*

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 process.*

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 process 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*

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

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 process. The definition is completely similar to Definition 2.4.1.

Definition 2.4.10 (N-dimensional Brownian process)**.** *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 process *if*

1. $B_0 = \mathbf{0}$ *a.s.,*
2. *For any* $0 \leq s \leq t$ *the random variable* $B_t - B_s$ *is independent of* $\mathcal{F}_s$*,*
3. *For any* $0 \leq s < t$ *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 process, we always assume the continuity of this stochastic process. Properties of the N-dimensional Brownian process are similar to the 1-dimensional Brownian process since each of its N components correspond to an independent 1-dimensional Brownian process [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 $Y=y$ *Conditional expectation* 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 $\in L^1(\Omega,\mathcal{G},\mathbb{P})$, 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[4] 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].*

[4] 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 process 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 process, 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*[5] 1*-dimensional Brownian process 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 process, and consider the* geometric Brownian process *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 process. 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: $\tau_A^o(\omega) : \inf\{t \geq 0 \mid X_t(\omega) \in A\}$,

[5] *Standard Brownian process* means the Brownian process 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 process, 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 $M_t \geq 0$, 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 process, 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 process $(B_t)_{t\geq 0}$, 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 process 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 process 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 process. You can enjoy reading various textbooks on the topic, and write any report on this topic. ☺

Chapter 4

Stochastic integrals

In this chapter, we give a meaning to a very new types of integrals, involving the Brownian process and martingales. Before this, we quickly recall the construction of the Riemann integrals, and present one first extension of these integrals.

4.1 Riemann and Riemann-Stieltjes integrals

For any interval $[a,b]$, we write $\mathcal{P}_\ell$ for the partition defined by $\mathcal{P}_\ell := \{t_0^\ell, t_1^\ell, \dots, t_{n_\ell}^\ell\}$ with $t_0^\ell = a$, $t_{n_\ell}^\ell = b$ and $t_j^\ell < t_{j+1}^\ell$, and set $|\mathcal{P}_\ell| := \max_{j\in\{1,\dots,n_\ell\}} |t_j^\ell - t_{j-1}^\ell|$ for its mesh. We also consider a family $C := \{c_1, c_2, \dots, c_{n_\ell}\}$ with $t_{j-1}^\ell \le c_j \le t_j^\ell$. For any bounded function $f : [a,b] \to \mathbb{R}$ we define the sum

$$\mathcal{R}(f, \mathcal{P}_\ell, C) := \sum_{j=1}^{n_\ell} f(c_j)(t_j^\ell - t_{j-1}^\ell)$$

and call it *the Riemann sum for* f, depending on the partition $\mathcal{P}_\ell$ and on the family C. In order to lose the dependence on C, we can introduce the Darboux sums defined by

$$\mathcal{R}_{\min}(f, \mathcal{P}_\ell) := \sum_{j=1}^{n_\ell} \inf_{c\in[t_{j-1}^\ell, t_j^\ell]} f(c)\,(t_j^\ell - t_{j-1}^\ell)$$

and

$$\mathcal{R}_{\max}(f, \mathcal{P}_\ell) := \sum_{j=1}^{n_\ell} \sup_{c\in[t_{j-1}^\ell, t_j^\ell]} f(c)\,(t_j^\ell - t_{j-1}^\ell).$$

Observe that if f is continuous, then the Darboux sums correspond to some Riemann sums, but for general bounded functions, the infimum or the supremum in each subinterval might not be realized by a point in the interval. Clearly, $\mathcal{R}_{\min}(f, \mathcal{P}_\ell) \le \mathcal{R}_{\max}(f, \mathcal{P}_\ell)$, and we say that the function is *Riemann integrable on* $[a,b]$ if the following equality holds

$$\sup_{\mathcal{P}_\ell} \Big(\mathcal{R}_{\min}(f, \mathcal{P}_\ell)\Big) = \inf_{\mathcal{P}_\ell} \Big(\mathcal{R}_{\max}(f, \mathcal{P}_\ell)\Big).$$

In this case, we simply write $\int_a^b f(t)\,\mathrm{d}t$ for the resulting number. Let us finally emphasize two important properties of the Riemann integral: it is linear and additive, namely

$$\int_a^b (f + \lambda g)(t)\,\mathrm{d}t = \int_a^b f(t)\,\mathrm{d}t + \lambda \int_a^b g(t)\,\mathrm{d}t$$

if f and g are Riemann integrable on $[a,b]$ and $\lambda \in \mathbb{R}$, and

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t$$

for any $c \in (a,b)$, whenever f is Riemann integrable on $[a,b]$.

Let us now introduce the Riemann-Stieltjes integral. For this, we consider two functions $F, g : [a,b] \to \mathbb{R}$ and set

$$\mathcal{S}(g,F,\mathcal{P}_\ell,C) := \sum_{j=1}^{n_\ell} g(c_j)(F(t_j^\ell) - F(t_{j-1}^\ell)).$$

Clearly, if $F = \mathrm{id}$, then this expression corresponds to the Riemann sum. The function g is called the *integrand* while the function F is called the *integrator*. We can then look for sufficient conditions on g and F such that this expression admits a limit as $|\mathcal{P}_\ell| \to 0$ (independent of $\mathcal{P}_\ell$ and of C), in which case we write

$$\lim_{|\mathcal{P}_\ell|\to 0} \sum_{j=1}^{n_\ell} g(c_j)(F(t_j^\ell) - F(t_{j-1}^\ell)) =: \int_a^b g(t)\,\mathrm{d}F(t) \tag{4.1.1}$$

and call it the *Riemann-Stieltjes integral* of g with respect to F. For example, it is known that if g is continuous and if F is of bounded variation, as introduced in (2.4.1), then the Riemann-Stieltjes integral exists. However, for future use of this integral, this set of conditions is not suitable. For this we introduce the notion of *p-variation*: for $p \geq 1$ the p-variation of a function $f : [a,b] \to \mathbb{R}$ is defined by

$$\mathrm{var}^p_{[a,b]}(f) := \left(\sup_{\mathcal{P}_\ell} \sum_{j=1}^{n_\ell} |f(t_j^\ell) - f(t_{j-1}^\ell)|^p \right)^{\frac{1}{p}} \tag{4.1.2}$$

where the supremum is taken over all partitions of $[a,b]$. Clearly, $\mathrm{var}^1_{[a,b]}(f) = \mathrm{var}_{[a,b]}(f)$. Then, it turns out that if

1. g has bounded p-variation on $[a,b]$, F has bounded q-variation on $[a,b]$, and $\frac{1}{p} + \frac{1}{q} > 1$,
2. g and F have no common discontinuity on $[a,b]$,

then the limit in (4.1.1) exists.

As a possible application of this result, expressions of the form $\int_a^b g(t)\,\mathrm{d}B_t(\omega)$ can be defined, where $(B_t)_{t\geq 0}$ denotes the standard 1-dimensional Brownian process. Indeed, the function $t \mapsto B_t(\omega)$ has a bounded q-variation for any $q > 2$, and it is then possible to choose any function g which has bounded p-variation for some $p < 2$.

Exercise 4.1.1. *Let $f : [a,b] \to \mathbb{R}$ be differentiable, with bounded derivative. Show that f has bounded variation. Note that the statement is slightly different compared to Exercise 2.4.7.*

However, the previous construction is not sufficient, since expressions of the form $\int_a^b B_t(\omega)\,\mathrm{d}B_t(\omega)$ or simply $\int_a^b g(t)\,\mathrm{d}B_t(\omega)$ with g continuous, can not be considered. In other words, since such expressions can not hold for individual paths, a different approach is necessary.

4.2 Itô integral

In this section, we shall give a meaning to integrals of the form $\int_0^T X_t \, \mathrm{d}B_t$ for some stochastic process $(X_t)_{t\in[0,T]}$ for some $T > 0$. If $X_t = 1$, this integral should satisfy $\int_0^T X_t \, \mathrm{d}B_t = B_T - B_0$. As before, $(B_t)_{t\geq 0}$ denotes the standard 1-dimensional Brownian process. More precisely, when dealing with two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and endowed with the same filtration $(\mathcal{F}_t)_{t\in\mathcal{T}}$, we shall carefully introduced the first one as a stochastic process, and say that the second one is adapted to its filtration.

For simplicity, we start by introducing the notion of stochastic integral in discrete time. As already mentioned on page 27, a discrete time stochastic process $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (X_n)_{n\in\mathbb{N}})$ is *predictable* if X_n is $\mathcal{F}_{n-1}$-measurable, for all $n \geq 1$. As an example, the amount of money a gambler will play at time n is predictable based on the outcomes at time $1, 2, \ldots, n-1$, but not on the outcome at time n.

Definition 4.2.1 (Discrete time stochastic integral)**.** *Let $M := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (M_n)_{n\in\mathbb{N}})$ be a univariate stochastic process, and let $X := (X_n)_{n\in\mathbb{N}}$ be an adapted and predictable univariate stochastic process. The* discrete time stochastic integral of X with respect to the process M *is defined for any $n \in \mathbb{N}$ by*

$$(X \cdot M)_n := X_0 M_0 + \sum_{j=1}^{n} X_j (M_j - M_{j-1}).$$

If the stochastic process $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (M_n)_{n\in\mathbb{N}})$ is a martingale, the above expression is called a *the martingale transform*. Let us immediately state two easy results. Proofs can be studied as an exercise, see also [11, Thm. 3.36 & 3.37].

Theorem 4.2.2. *Let $M := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (M_n)_{n\in\mathbb{N}})$ be a martingale, and let $X := (X_n)_{n\in\mathbb{N}}$ be an adapted and predictable univariate stochastic process, with $(X \cdot M)_n$ belonging to $L^1(\Omega, \mathcal{F}, \mathbb{P})$ for any $n \in \mathbb{N}$. Then $((X \cdot M)_n)_{n\in\mathbb{N}}$ defines a martingale with the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.*

Corollary 4.2.3. *Let $M := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (M_n)_{n\in\mathbb{N}})$ be a martingale, and let $X := (X_n)_{n\in\mathbb{N}}$ be an adapted, predictable and bounded univariate stochastic process, then $((X \cdot M)_n)_{n\in\mathbb{N}}$ defines a martingale with the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. If $X_n, M_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ for any $n \in \mathbb{N}$, then the same result hold.*

As an example of the previous construction, let $(B_t)_{t\geq 0}$ be the standard 1-dimensional standard Brownian process, and consider $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$. One can then check that the sequence $M := (M_n)_{n\in\mathbb{N}}$, defined by $M_0 := 0$ and $M_j := B_{t_j}$ for $j \in 1, 2, \ldots, n$, is a martingale for the filtration given by $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_j$ the smallest σ-algebra generated by $\sigma(B_{t_i})$ for $i \in \{1, \ldots, j\}$. By setting $X := (0, B_{t_0}, B_{t_1}, \ldots, B_{t_{n-1}})$, one has for any $k \in \{1, \ldots, n\}$

$$(X \cdot M)_k = \sum_{j=1}^{k} B_{t_{j-1}} (B_{t_j} - B_{t_{j-1}}). \tag{4.2.1}$$

Observe that this process is a discrete time analogue of the integral $\int_a^b B_t(\omega) \, \mathrm{d}B_t(\omega)$ mentioned before. However, in the above formula, there is no evaluation on $\omega \in \Omega$, which means that the resulting sum is a random variable, not a number.

Exercise 4.2.4. *Check that the content of (4.2.1) corresponds to a discrete time stochastic integral.*

Let us now move to the notion of stochastic integral in continuous time. For that, recall that the notion of progressively measurable has been introduced in Definition 2.3.3, and that continuous process are automatically progressively measurable, see last part of Section 2.3. In fact, left-continuous or right-continuous stochastic

processes are also automatically progressively measurable. We start with the definition of a large class of processes that will play the role of the integrand.

Definition 4.2.5 (L^p-progressively integrable process)**.** *For $0 \leq a < b < \infty$ and $p \geq 1$, a progressively measurable and univariate process $X = (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[a,b]}, (X_t)_{t\in[a,b]})$ belongs to $M^p_{\text{loc}}([a,b])$ if $\int_a^b |X_t|^p \, \mathrm{d}t < \infty$ almost surely, or equivalently if*

$$\mathbb{P}\Big(\Big\{\omega \in \Omega \mid \int_a^b |X_t(\omega)|^p \, \mathrm{d}t < \infty\Big\}\Big) = 1.$$

Such process belongs to $M^p([a,b])$ if

$$\mathbb{E}\Big(\int_a^b |X_t|^p \, \mathrm{d}t\Big) < \infty.$$

Let us observe that if a stochastic process is continuous, it automatically belong to $M^p_{\text{loc}}([a,b])$. In order to prove if it is also in $M^p([a,b])$, one can use Fubini's theorem, namely

$$\mathbb{E}\Big(\int_a^b |X_t|^p \, \mathrm{d}t\Big) = \int_a^b \mathbb{E}(|X_t|^p) \, \mathrm{d}t.$$

Thus, in order to be in $M^p([a,b])$ it is sufficient that the map $t \mapsto \mathbb{E}(|X_t|^p)$ belongs to $L^1([a,b])$.

Among these processes, some are particularly simple, namely the *elementary processes*. Namely, consider a partition $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$, then the stochastic process X is elementary if

$$X_t(\omega) = \sum_{j=0}^{n-1} X_j(\omega) \mathbf{1}_{(t_j, t_{j+1}]}(t). \tag{4.2.2}$$

Note that X_j has to be $\mathcal{F}_{t_j}$-measurable, and the elementary process (4.2.2) is clearly left-continuous, and therefore progressively measurable. Also, it automatically belongs to $M^p_{\text{loc}}([a,b])$. In addition, observe that

$$\mathbb{E}\Big(\int_a^b |X_t|^p \, \mathrm{d}t\Big) = \mathbb{E}\Big(\sum_{j=0}^{n-1} |X_j|^p \, (t_{j+1} - t_j)\Big) = \sum_{j=0}^{n-1} \mathbb{E}(|X_j|^p)(t_{j+1} - t_j),$$

which means that elementary process $X \in M^p([a,b])$ if and only $X_j \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for all j.

Recall that in our definition of a stochastic process, the family of random variables is automatically adapted to the filtration. For the elementary processes we define their stochastic integrals as follows:

Definition 4.2.6 (Stochastic integral of elementary process)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ be the standard 1-dimensional Brownian process, and let $X := (X_t)_{t\in[a,b]}$ be an adapted elementary process as in (4.2.2). The stochastic integral of X with respect to the Brownian process is defined by*

$$\int_a^b X_t \, \mathrm{d}B_t := \sum_{j=0}^{n-1} X_j (B_{t_{j+1}} - B_{t_j}).$$

Let us immediately state some properties of these integrals. Proofs can be found in [2, Lem. 7.1] or in [11, p. 93].

Proposition 4.2.7. *In the framework of the previous definition, the integral is linear in the integrand X, and if $X_t = \mathbf{1}_{(a,b]}$ it satisfies $\int_a^b X_t \,\mathrm{d}B_t = B_b - B_a$. In addition, if X belongs to $M^2([a,b])$ one has*

1. $\mathbb{E}\Big(\int_a^b X_t \,\mathrm{d}B_t \,\Big|\, \mathcal{F}_a \Big) = \mathbf{0}$,
2. $\mathbb{E}\Big(\Big(\int_a^b X_t \,\mathrm{d}B_t \Big)^2 \,\Big|\, \mathcal{F}_a \Big) = \mathbb{E}\Big(\int_a^b X_t^2 \,\mathrm{d}t \,\Big|\, \mathcal{F}_a \Big)$ *(isometry property).*

In particular $\mathbb{E}\Big(\int_0^T X_t \,\mathrm{d}B_t \Big) = 0$ *and* $\mathbb{E}\Big(\Big(\int_0^T X_t \,\mathrm{d}B_t \Big)^2 \Big) = \mathbb{E}\Big(\int_0^T X_t^2 \,\mathrm{d}t \Big) = \int_0^T \mathbb{E}(X_t^2)\,\mathrm{d}t$, *for any* $T > 0$.

In order to define stochastic integrals for more general processes, one needs to approximate them by elementary processes. The next technical statement provides the information about such approximations.

Lemma 4.2.8. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[a,b]}, (X_t)_{t\in[a,b]})$ be a stochastic process belonging to $M^p_{\mathrm{loc}}([a,b])$ for some $p \geq 1$. Then there exists a sequence of adapted elementary processes $((X_{n,t})_{t\in[a,b]})_{n\in\mathbb{N}}$ belonging to $M^p_{\mathrm{loc}}([a,b])$ such that*

$$\lim_{n\to\infty} \int_a^b |X_t - X_{n,t}|^p \mathrm{d}t = \mathbf{0} \qquad a.s. \tag{4.2.3}$$

If $X \in M^p([a,b])$, then there exists a sequence of adapted elementary processes $((X_{n,t})_{t\in[a,b]})_{n\in\mathbb{N}}$ belonging to $M^p([a,b])$ such that

$$\lim_{n\to\infty} \mathbb{E}\Big(\int_a^b |X_t - X_{n,t}|^p \mathrm{d}t \Big) = 0. \tag{4.2.4}$$

As a consequence of these approximation procedures, it is possible to give a definition of a stochastic integral for all elements of $M^2_{\mathrm{loc}}([a,b])$ and of $M^2([a,b])$. Recall that the L^p-convergence has been introduced in Definition 1.4.3 while the convergence in probability[6] has been mentioned in (3.2.2).

Definition 4.2.9 (Itô stochastic integral)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ be the standard 1-dimensional Brownian process, and let $(X_t)_{t\in[a,b]}$ be an adapted stochastic process belonging to $M^2_{\mathrm{loc}}([a,b])$. Then one sets*

$$\int_a^b X_t \,\mathrm{d}B_t := \lim_{n\to\infty} \int_a^b X_{n,t} \,\mathrm{d}B_t, \tag{4.2.5}$$

with $((X_{n,t})_{t\in[a,b]})_{n\in\mathbb{N}}$ an approximating sequence by elementary processes in $M^2_{\mathrm{loc}}([a,b])$ and satisfying (4.2.3), and where the convergence is in probability. If $(X_t)_{t\in[a,b]}$ belongs to $M^2([a,b])$, then the approximating sequence can be taken in $M^2([a,b])$ and satisfying (4.2.4) and the convergence holds in the L^2-sense.

Clearly, these definitions are meaningful if and only if the expression for $\int_a^b X_t \,\mathrm{d}B_t$ does not depend on the sequence chosen for defining it. This property holds, as shown in [11, Thm. 4.3] but it is not trivial and it is based on the completeness of Hilbert spaces. Let us now state some properties of the Itô stochastic integrals. They clearly follows from Proposition 4.2.7.

Proposition 4.2.10. *In the framework of the previous definition and if $X := (X_t)_{t\in[a,b]}$ belongs to $M^2_{\mathrm{loc}}([a,b])$, then the integral defined in (4.2.5) is linear in the integrand X. If $(X_t)_{t\in[a,b]}$ belongs to $M^2([a,b])$, then one has*

1. $\mathbb{E}\Big(\int_a^b X_t \,\mathrm{d}B_t \,\Big|\, \mathcal{F}_a \Big) = \mathbf{0}$,
2. $\mathbb{E}\Big(\Big(\int_a^b X_t \,\mathrm{d}B_t \Big)^2 \,\Big|\, \mathcal{F}_a \Big) = \mathbb{E}\Big(\int_a^b X_t^2 \,\mathrm{d}t \,\Big|\, \mathcal{F}_a \Big)$.

[6] A sequence of univariate random variables $(X_n)_{n\in\mathbb{N}}$ converges to the random variable X_∞ in probability if for any $\varepsilon > 0$ one has $\lim_{n\to\infty} \mathbb{P}(|X_n - X_\infty| \geq \varepsilon) = 0$.

In particular $\mathbb{E}\Big(\int_0^T X_t\,\mathrm{d}B_t\Big) = 0$ *and* $\mathbb{E}\Big(\Big(\int_0^T X_t\,\mathrm{d}B_t\Big)^2\Big) = \mathbb{E}\Big(\int_0^T X_t^2\,\mathrm{d}t\Big) = \int_0^T \mathbb{E}(X_t^2)\,\mathrm{d}t$, *for any* $T > 0$.

It is interesting to observe that Itô stochastic integrals need not have a mean or a variance, but when they do, the mean is zero and the variance is given by 2. of Proposition 4.2.10.

Corollary 4.2.11. *Let* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ *be the standard* 1*-dimensional Brownian process, and let* $(X_t)_{t\in[a,b]}$ *be an adapted and continuous stochastic process. Then the stochastic integral* $\int_a^b X_t\,\mathrm{d}B_t$ *exists. In particular, if* f *is any continuous real function on* $\mathbb{R}$*, then the stochastic integral* $\int_a^b f(B_t)\,\mathrm{d}B_t$ *is well defined.*

Exercise 4.2.12. *In the framework of the previous corollary, study the examples 4.3 to 4.5 of [11].*

Exercise 4.2.13. *Study the equality*

$$\int_a^b B_t\,\mathrm{d}B_t = \frac{1}{2}\Big(B_b^2 - B_a^2 - (b-a)\Big). \tag{4.2.6}$$

The proof appears at several places, see for example [13, Sec. 2.2.1] or [12, Sec. 4.1].

Let us state one more result about the product of Itô stochastic integrals. The proof can be studied as an exercise, see [2, Rem. 7.1 p. 189] of [11, Thm. 4.5].

Theorem 4.2.14. *Let* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ *be the standard* 1*-dim. Brownian process, and let* $(X_t)_{t\in[0,T]}$, $(Y_t)_{t\in[0,T]}$ *be adapted stochastic processes belonging to* $M^2([0,T])$*. Then*

$$\mathbb{E}\Big(\Big(\int_0^T X_t\,\mathrm{d}B_t\Big)\Big(\int_0^T Y_t\,\mathrm{d}B_t\Big)\Big) = \int_0^T \mathbb{E}(X_t Y_t)\,\mathrm{d}t.$$

4.3 Martingale property

For any stochastic process $(X_t)_{t\in[0,T]}$ adapted to the standard 1-dimensional Brownian filtration and belonging to $M^2_{\mathrm{loc}}([0,T])$, one can consider the stochastic integral

$$Y_t := \int_0^t X_u\,\mathrm{d}B_u, \qquad t \in [0,T]$$

and the corresponding map $t \mapsto Y_t$. Clearly, the following equality holds for $0 \leq s < t \leq T$:

$$Y_t - Y_s = \int_s^t X_u\,\mathrm{d}B_u.$$

By coming back to the definition of Itô integral as a limit of elementary functions, it can also be shown that Y_t is $\mathcal{F}_t$-measurable. and therefore that this new stochastic process $Y := (Y_t)_{t\in[0,T]}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ of the standard 1-dimensional Brownian process.

Assume now that $(X_t)_{t\in[0,T]}$ belongs to $M^2([0,T])$. It then follows from Proposition 4.2.10 that for $0 < s < t \leq T$

$$\mathbb{E}(Y_t|\mathcal{F}_s) = \mathbb{E}\Big(\int_0^t X_u\,\mathrm{d}B_u \,\Big|\, \mathcal{F}_s\Big) = \int_0^s X_u\,\mathrm{d}B_u + \mathbb{E}\Big(\int_s^t X_u\,\mathrm{d}B_u \,\Big|\, \mathcal{F}_s\Big) = Y_s + \mathbf{0} = Y_s.$$

Therefore, $(Y_t)_{t\in[0,T]}$ is a martingale. It also follows from Proposition 4.2.10 that $\mathbb{E}(Y_t^2) = \int_0^t \mathbb{E}(X_u^2)\,\mathrm{d}u$. As a consequence, one has

$$\sup_{t\in[0,T]} \mathbb{E}(Y_t^2) = \int_0^T \mathbb{E}(X_u^2)\,\mathrm{d}u < \infty.$$

A martingale satisfying this property is said to be *square integrable on* $[0,T]$. Let us also mention that this martingale has a continuous version, which is always chosen. Let us summarize these properties in one single statement, see [2, Thm. 7.3]:

Theorem 4.3.1. *Let $(\Omega,\mathcal{F},\mathbb{P},(\mathcal{F}_t)_{t\geq 0},(B_t)_{t\geq 0})$ be the standard 1-dim. Brownian process, and let $(X_t)_{t\in[0,T]}$ be an adapted stochastic process belonging to $M^2([0,T])$. Then the stochastic process $Y:=(Y_t)_{t\in[0,T]}$ given by $Y_t:=\int_0^t X_u\,\mathrm{d}B_u$ defines a mean zero and square integrable continuous martingale.*

Let us mention that if $(X_t)_{t\in[0,T]}$ is an adapted stochastic process belonging to $M^2_{\mathrm{loc}}([0,T])$ only, the corresponding process $(Y_t)_{t\in[0,T]}$ is continuous but may fail to be a martingale: it is only a local martingale, see [2, Sec. 7.5]. For completeness we provide the definition (note that martingales also belong to the set of local martingales).

Definition 4.3.2 (Local martingale)**.** *A real valued stochastic process $M:=(\Omega,\mathcal{F},\mathbb{P},(\mathcal{F}_t)_{t\geq 0},(M_t)_{t\geq 0})$ is a* local martingale *if there exists an increasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of stopping times such that*

1. *$\tau_n\to\infty$ as $n\to\infty$ a.s.,*
2. *$(M_{t\wedge\tau_n})_{t\geq 0}$ is a martingale for the filtration $(\mathcal{F}_t)_{t\geq 0}$, for every n.*

The sequence $(\tau_n)_{n\in\mathbb{N}}$ is said to reduce *the local martingale M.*

Exercise 4.3.3. *Study the examples 5.19 and 5.20 of [1], showing that the Ornstein-Uhlenbeck process and the Brownian bridge can be obtained as Itô integrals.*

Let us provide one result about stochastic integrals when the integrand in not random:

Theorem 4.3.4 (Wiener integral)**.** *Let $f:[0,T]\to\mathbb{R}$ be square integrable, namely $\int_0^T f(s)^2\,\mathrm{d}s<\infty$. Then the function*

$$[0,T]\ni t\mapsto\int_0^t f(u)\,\mathrm{d}B_u$$

is a Gaussian process with zero mean value and covariance function given by

$$\mathrm{Cov}(Y_t,Y_s)=\int_0^{t\wedge s} f(u)^2\,\mathrm{d}u.$$

Exercise 4.3.5. *Study the proof of the above theorem, looking either at [1, Cor. 5.18] or at [11, Thm. 4.9].*

Finally, let us state some results about the regularity of the stochastic process $(Y_t)_{t\in[0,T]}$. For that purpose and by analogy to (4.1.2) we define the *quadratic variation* of a stochastic process $Y=(Y_t)_{t\in[0,T]}$ by

$$[Y]_t:=\lim_{|\mathcal{P}_\ell|\to 0}\sum_{j=1}^{n_\ell}|Y_{t_j^\ell}-Y_{t_{j-1}^\ell}|^2 \tag{4.3.1}$$

where the limit is in probability over all partitions $\mathcal{P}_\ell$ of the interval $[0,t]$ with mesh $|\mathcal{P}_\ell|=\delta_n\to 0$ as $n\to\infty$. Note that this quadratic variation is a random variable, and that this notation looks rather standard (and awkward ☹).

Proposition 4.3.6. *Let $(\Omega,\mathcal{F},\mathbb{P},(\mathcal{F}_t)_{t\geq 0},(B_t)_{t\geq 0})$ be the standard 1-dimensional Brownian process, and let $(X_t)_{t\in[0,T]}$ be an adapted stochastic process belonging to $M^2_{\mathrm{loc}}([0,T])$. Then the quadratic variation of the local martingale $Y:=(Y_t)_{t\in[0,T]}$ defined by $Y_t:=\int_0^t X_u\,\mathrm{d}B_u$ satisfies*

$$[Y]_t=\int_0^t X_u^2\,\mathrm{d}u. \tag{4.3.2}$$

We refer to [11, Thm. 4.9] for a sketch of the proof, and to [5, Thm. 3.8] for a more general result.

Before the next statement, let us state a relatively easy property for real functions on an interval $[a, b]$. Recall that p-variations were introduced in (4.1.2).

Exercise 4.3.7. *Let $f : [a, b] \to \mathbb{R}$ be continuous. If $\mathrm{var}^1_{[a,b]}(f) < \infty$, then $\mathrm{var}^2_{[a,b]}(f) = 0$.*

The next statement is quite general, and hold not only for the martingale process Y, see [2, Thm. 5.15].

Theorem 4.3.8. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (M_t)_{t\geq 0})$ be a continuous martingale process, square integrable on $[0, T]$. If the process has finite variation, then it is a.s. constant. In particular, if $(X_t)_{t\in[0,T]}$ is a stochastic process belonging to $M^2([0, T])$ and adapted to the standard 1-dimensional Brownian process and with $\int_0^t X_u^2 \, \mathrm{d}u > 0$ for all $t \in (0, T]$, then the martingale $Y := (Y_t)_{t\in[0,T]}$ defined by $Y_t := \int_0^t X_u \, \mathrm{d}B_u$ has infinite variation on $[0, t]$ for all $t \in (0, T]$.*

4.4 Itô's lemma

In this section we introduce Itô's lemma which is of crucial importance for the computation of several stochastic integrals. Several versions of this lemma exist, and some of them are presented subsequently.

The starting point for these developments is the usual formula for the derivative of the composition of two functions, namely

$$\frac{\mathrm{d}}{\mathrm{d}t} f(g(t)) = f'(g(t))\, g'(t)$$

whenever this formula makes sense. Note that an equivalent formulation of this equality can be written as

$$f(g(b)) = f(g(a)) + \int_a^b f'(g(t))\, g'(t)\, \mathrm{d}t.$$

This equality is valid for the usual Riemann integral, but what is then its analog for the Itô integral ? The answer corresponds precisely to Itô's lemma. For its statement, we recall that $C^k(\mathbb{R})$ denotes the set of k-times continuously differentiable functions on $\mathbb{R}$.

Theorem 4.4.1 (Itô's lemma)**.** *Let $B := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ be the standard 1-dimensional Brownian process, and let $f \in C^2(\mathbb{R})$. Then the following equality holds:*

$$f(B_t) = f(\mathbf{0}) + \int_0^t f'(B_s)\, \mathrm{d}B_s + \tfrac{1}{2} \int_0^t f''(B_s)\, \mathrm{d}s. \tag{4.4.1}$$

Let us sketch the proof, and refer to a [11, Thm. 4.14] for the final technical (but essential) argument. A proof of Itô's lemma can be found in almost all books on stochastic calculus. An accessible one is for example provided in [12, Sec. 7.2].

Proof. First of all, observe that the integrals in (4.4.1) are well defined, by Corollary 4.2.11. Then, let $\mathcal{P}_\ell$ be a partition of $[0, t]$. By the telescopic formula, one has

$$f(B_t) = f(\mathbf{0}) + \sum_{j=0}^{n_\ell - 1} \Big(f(B_{t^\ell_{j+1}}) - f(B_{t^\ell_j}) \Big).$$

By Taylor formula for each term in the sum one has

$$f(B_{t^\ell_{j+1}}) - f(B_{t^\ell_j}) = f'(B_{t^\ell_j})\,(B_{t^\ell_{j+1}} - B_{t^\ell_j}) + \tfrac{1}{2} f''(\theta^\ell_j)\,(B_{t^\ell_{j+1}} - B_{t^\ell_j})^2$$

for some $\theta^\ell_j \in (B_{t^\ell_j}, B_{t^\ell_{j+1}})$. As a consequence, one infers that

$$f(B_t) = f(\mathbf{0}) + \sum_{j=0}^{n_\ell - 1} f'(B_{t^\ell_j})\,(B_{t^\ell_{j+1}} - B_{t^\ell_j}) + \tfrac{1}{2} \sum_{j=0}^{n_\ell - 1} f''(\theta^\ell_j)\,(B_{t^\ell_{j+1}} - B_{t^\ell_j})^2.$$

By definition of the Itô integral, the second term on the r.h.s. converges to $\int_0^t f'(B_s)\mathrm{d}B_s$ as $|\mathcal{P}_\ell| \to 0$. Depending on f this convergence is either taking place in the L^2-sense, or in probability, see [11, Examples 4.3 to 4.5] for various examples. The type of convergence depends if $f(B) \in M^2([0,t])$ or if $f(B) \in M^2_{\mathrm{loc}}([0,T])$. Then, the third term on the r.h.s converges in probability to $\frac{1}{2}\int_0^t f''(B_s)\mathrm{d}s$. This technical part is shown for example in [11, Thm. 4.14]. □

As a first application of this result, observe that if we consider $f : \mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$, then one gets

$$B_t^2 = 2\int_0^t B_s\,\mathrm{d}B_s + \tfrac{1}{2}\int_0^t 2\,\mathrm{d}s$$

which easily leads to (4.2.6). More generally, observe that the equality (4.4.1) can be rewritten as

$$\int_0^t f'(B_s)\mathrm{d}B_s = f(B_t) - f(\mathbf{0}) - \tfrac{1}{2}\int_0^t f''(B_s)\mathrm{d}s$$

which is helpful for the computation of the l.h.s. of this equality.

Remark 4.4.2. *The fact that for any $g \in C(\mathbb{R})$ one has*

$$\lim_{|\mathcal{P}_\ell|\to 0} \sum_{j=0}^{n_\ell - 1} g(\theta^\ell_j)\,(B_{t^\ell_{j+1}} - B_{t^\ell_j})^2 = \int_0^t g(B_s)\mathrm{d}s$$

for any $\theta^\ell_j \in (B_{t^\ell_j}, B_{t^\ell_{j+1}})$ and with a convergence in probability is sometimes explained by heuristic arguments. Some authors say that $(\Delta_j B)^2$ is of order $\Delta_j t$, with $\Delta_j B := B_{t^\ell_{j+1}} - B_{t^\ell_j}$ and $\Delta_j t := t^\ell_{j+1} - t^\ell_j$. Other authors write $\mathrm{d}t = \mathrm{d}B_t \cdot \mathrm{d}B_t$, but we shall refrain from using such notations. In fact, a very much related result has already been mentioned in Theorem 2.4.8 when mentioning the convergence of the quadratic variation of the 1-dimensional Brownian process.

Exercise 4.4.3. *1) Show that the following formula holds*

$$B_t^3 = 3\int_0^t B_s^2\,\mathrm{d}B_s + 3\int_0^t B_s\,\mathrm{d}s.$$

2) Set $Y_t := \int_0^t B_s\,\mathrm{d}s$, and consider the process $Y := (Y_t)_{t\geq 0}$. Show that Y is a Gaussian process with $\mathbb{E}(Y_t) = 0$ for any $t \geq 0$ and that $\mathbb{E}(Y_s Y_t) = \frac{s^2 t}{2} - \frac{s^3}{6}$ for $0 \leq s < t$. In particular $\mathrm{Var}(T_t) = \frac{t^3}{3}$. We refer to [1, Example 5.26] for hints and for an illustration of this process.

Let us directly provide a slightly generalized version of Itô's lemma, when the function f depends of two variables, namely for functions of the type $(t,x) \mapsto f(t,x) \in \mathbb{R}$. The proof is quite similar but involves Taylor expansion for functions of 2 variables, and therefore partial derivatives are involved. For that purpose, let us introduce the notation $\partial_t f$ for the derivative of f with respect to its time variable, and $\partial_x f$ for its derivative with respect to its space variable.

Proposition 4.4.4. *Let $B := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ be the standard 1-dimensional Brownian process, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuous and with $\partial_t f$, $\partial_x f$, and $\partial_x^2 f$ (second derivative with respect to its space variable) also continuous. Then the following equality holds:*

$$f(t, B_t) = f(0, \mathbf{0}) + \int_0^t [\partial_x f](s, B_s)\, \mathrm{d}B_s + \int_0^t \left\{ [\partial_t f](s, B_s) + \tfrac{1}{2}[\partial_x^2 f](s, B_s) \right\} \mathrm{d}s, \tag{4.4.2}$$

Note that the continuity requirements on the derivatives of f ensure that the conditions of Definition 4.2.9 are satisfied for the term involving a stochastic integral. Clearly, if f does not depend explicitly on the variable t, this statement corresponds to Itô's lemma.

Remark 4.4.5. *According to Theorem 4.3.1 and to the subsequent paragraph, the Itô integral $\int_0^t X_u \, \mathrm{d}B_u$ defines a martingale if $X \in M^2([0,T])$ or a local martingale if $X \in M^2_{\mathrm{loc}}([0,T])$. Thus, by rewriting (4.4.2) as*

$$\int_0^t [\partial_x f](s, B_s)\, \mathrm{d}B_s = f(t, B_t) - f(0, \mathbf{0}) - \int_0^t \left\{ [\partial_t f](s, B_s) + \tfrac{1}{2}[\partial_x^2 f](s, B_s) \right\} \mathrm{d}s,$$

it is sometimes possible to deduce when the r.h.s. is a martingale or a local martingale.

The following exercises are applications of Itô's lemma or its slightly generalized version, see Theorem 4.4.1 or Proposition 4.4.4.

Exercise 4.4.6. *By considering the function $f(t,x) = tx^2$, write an expression (as simple as possible) for the Itô integral $\int_0^t sB_s \, \mathrm{d}B_s$, see also [12, Example 7.3.2].*

Exercise 4.4.7 (The Itô exponential)**.** *Consider the function $f(t,x) = \mathrm{e}^{x - \frac{1}{2}t}$ and show that*

$$\int_0^t f(s, B_s)\, \mathrm{d}B_s = f(t, B_t) - f(0, \mathbf{0}).$$

Observe that for the usual exponential function one has $\int_0^t \mathrm{e}^s \, \mathrm{d}s = \mathrm{e}^t - \mathrm{e}^0$, and for this reason the function f is sometimes called the Itô exponential*.*

Exercise 4.4.8. *Set $f(t,x) := \mathrm{e}^{(c - \frac{1}{2}\sigma^2)t + \sigma x}$ for $c \in \mathbb{R}$ and $\sigma > 0$, and consider the process $X_t := f(t, B_t)$. Show that*

$$X_t = X_0 + c \int_0^t X_s \, \mathrm{d}s + \sigma \int_0^t X_s \, \mathrm{d}B_s.$$

Exercise 4.4.9 (♥). *Study and report on the gambler's ruin for Brownian process with drift, as presented in [1, Sec. 5.5].*

4.5 N-dimensional Itô integral

As seen in Definition 2.4.10 the N-dimensional Brownian process consists of N independent 1-dimensional Brownian processes. For that reason, the Itô integral can be generalized "straightforwardly" to the N-dimensional case. Let us concentrate on the analog of Itô's lemma in this framework [1, Thm. 6.6] :

Theorem 4.5.1. *Let $B := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ be the standard N-dimensional Brownian process, and let $f \in C^2(\mathbb{R}^N)$. Then the following equality holds:*

$$f(B_t) = f(\boldsymbol{0}) + \sum_{j=1}^{N} \int_0^t [\partial_j f](B_s)\, \mathrm{d}B_s^j + \tfrac{1}{2} \int_0^t \sum_{j=1}^{N} [\partial_j^2 f](B_s)\, \mathrm{d}s, \tag{4.5.1}$$

where $B_s = (B_s^1, B_s^2, \ldots, B_s^N)^T$ are the components of B_s.

Let us immediately observe that the previous equality can be rewritten with shorter notations:

$$f(B_t) = f(\mathbf{0}) + \int_0^t ([\nabla f](B_s))^T\, \mathrm{d}B_s + \tfrac{1}{2} \int_0^t [\Delta f](B_s)\, \mathrm{d}s.$$

It is interesting to observe that the mixed partial derivatives of f do not appear in this formula. In addition, if the function f also depends explicitly on the time parameter and satisfies the necessary regularity conditions on the partial derivatives, then the following equality [1, Thm. 6.11] also holds

$$f(t, B_t) = f(0, \mathbf{0}) + \int_0^t ([\nabla f](s, B_s))^T\, \mathrm{d}B_s + \int_0^t \left\{ [\partial_t f](s, B_s) + \tfrac{1}{2}[\Delta f](s, B_s) \right\} \mathrm{d}s. \tag{4.5.2}$$

Based on these formulas and on the notion of harmonic functions, the following important result can be proved. It corresponds to the N-dimensional generalization of the 1-dimensional result provided in Proposition 3.2.24.

Theorem 4.5.2. *Let $B := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ be the standard N-dimensional Brownian process starting at $x \in \mathbb{R}^N$ at time $t = 0$. Let $r < \|x\|$ and define the stopping time τ_r by*

$$\tau_r(\omega) := \min\{t \geq 0 \mid \|B_t(\omega)\| \leq r\}$$

which corresponds to the first hitting time of the ball centered at 0 and of radius r. Then one has $\mathbb{P}(\tau_r < \infty) = 1$ if $N = 2$, and $\mathbb{P}(\tau_r < \infty) = (\frac{r}{\|x\|})^{N-2}$ if $N \geq 3$. In particular, for $N \leq 2$ the paths are recurrent*, meaning that they come back to a neighborhood of the origin infinitely many times, while for $N \geq 3$ each path will eventually never come back to a neighborhood of the origin (*transient behavior*).*

Exercise 4.5.3 (Recurrence and transience of Brownian process (♥))**.** *Study and report on the proof of the previous theorem. It is provided in many textbooks, as for example in [1, Sec. 6.3].*

Chapter 5

Itô processes and stochastic differential equations

If we look back at Itô's lemma, the main equality can be written in a *differential form* as

$$\mathrm{d}f(B_t) = f'(B_t)\,\mathrm{d}B_t + \tfrac{1}{2}f''(B_t)\,\mathrm{d}t.$$

The meaning of this notation is the one given by (4.4.1), but this formulation is convenient and short. Similarly, observe that (4.4.2) can be rewritten as

$$\mathrm{d}f(t, B_t) = [\partial_x f](t, B_t)\,\mathrm{d}B_t + \left\{[\partial_t f](t, B_t) + \tfrac{1}{2}[\partial_x^2 f](t, B_t)\right\}\mathrm{d}t.$$

and the N-dimensional version of (4.5.2) as

$$\mathrm{d}f(B_t) = ([\nabla f](t, B_t))^T\,\mathrm{d}B_t + \left\{[\partial_t f](t, B_t) + \tfrac{1}{2}[\Delta f](t, B_t)\right\}\mathrm{d}t.$$

In this chapter, we shall constantly use these notations, and develop the notion of stochastic differential equations. However, the true meaning of these equations is always their integral analogs.

5.1 Itô processes

We start with the main definition of this section.

Definition 5.1.1 (Itô process, stochastic differential)**.** *Let $B := (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (B_t)_{t\geq 0})$ be the standard 1-dimensional Brownian process, let $V = (V_t)_{t\in[0,T]}$ and $D = (D_t)_{t\in[0,T]}$ be adapted and univariate stochastic processes, with $V \in M^2_{\mathrm{loc}}([0,T])$ and $D \in M^1_{\mathrm{loc}}([0,T])$, and let X_0 is a $\mathcal{F}_0$-measurable random variable. The process $X := (X_t)_{t\in[0,T]}$ defined by*

$$X_t = X_0 + \int_0^t V_s\,\mathrm{d}B_s + \int_0^t D_s\,\mathrm{d}s, \tag{5.1.1}$$

is called a Itô process *or an* integral process*. The relation (5.1.1) is also written*

$$\mathrm{d}X_t = V_t\,\mathrm{d}B_t + D_t\,\mathrm{d}t. \tag{5.1.2}$$

It is said that the stochastic process X has the stochastic differential (5.1.2)

Note that the letters used for defining this process are not at all uniform in the literature. For the notations currently used, V is called the *local volatility* and D is called the *local drift*. Let us stress that (5.1.2) has no meaning in itself, it is a short notation of (5.1.1). In addition, an initial value has to be prescribed, since (5.1.2) does not contain any information for the time $t = 0$.

The simplest case of (5.1.1) is when V_t and D_t are real constants. Another special instance of Itô processes can defined when the two processes V and D depend on X only.

Definition 5.1.2 (Time-homogeneous / time-inhomogeneous diffusion process)**.** *The Itô process satisfying*

$$\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}B_t + \mu(X_t)\,\mathrm{d}t$$

with $\sigma, \mu : \mathbb{R} \to \mathbb{R}$ is called a time-homogeneous diffusion process*, while the Itô process satisfying*

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mu(t, X_t)\,\mathrm{d}t$$

is called a time-inhomogeneous diffusion process*. For these definitions, the stochastic processes must be adapted, $(\sigma(X_t))_{t\in[0,T]}$ and $(\sigma(t, X_t))_{t\in[0,T]}$ must belong to $M^2_{\mathrm{loc}}([0, T])$, while $(\mu(X_t))_{t\in[0,T]}$ and $(\mu(t, X_t))_{t\in[0,T]}$ must belong to $M^1_{\mathrm{loc}}([0, T])$,*

Exercise 5.1.3. *Write formulas similar to (5.1.2) for the stochastic processes mentioned in Examples 2.2.3 to 2.2.6. Discuss also the content of Examples 7.3 and 7.4 of [1].*

The notion of quadratic variation for a general process has been introduced in (4.3.1), and the quadratic variation for process of the form $Y_t := \int_0^t X_u\,\mathrm{d}B_u$ has been computed in Proposition 4.3.6. It turns out that for more general Itô processes the result is still valid. Namely, if X is a Itô process of the form (5.1.1), then its quadratic variation satisfies

$$[X]_t = \int_0^t V_s^2\,\mathrm{d}s. \tag{5.1.3}$$

We provide below the analog of Itô's lemma for general Itô processes. Note that we directly state the version with a function f which can also have an explicit dependence on time. It thus corresponds to a generalization of Proposition 4.4.4, see [12, Thm. 7.4.3]. As before, the proof relies essentially on a Taylor expansion, but technicalities are also necessary [1, Prop. 7.7 & Thm. 7.8].

Proposition 5.1.4 (Itô's lemma for Itô process)**.** *Let X be a Itô process of the form $\mathrm{d}X_t = V_t\,\mathrm{d}B_t + D_t\,\mathrm{d}t$, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuous and with $\partial_t f$, $\partial_x f$, and $\partial_x^2 f$ also continuous. Then the following equality hold for any $s \le t$:*

$$\begin{aligned}
f(t, X_t) &= f(s, X_s) + \int_s^t V_u[\partial_x f](u, X_u)\,\mathrm{d}B_u + \int_s^t \Big\{[\partial_t f](u, X_u) + D_u[\partial_x f](u, X_u) + \tfrac{1}{2}(V_u)^2[\partial_x^2 f](u, X_u)\Big\}\mathrm{d}u \\
&= f(s, X_s) + \int_s^t \Big\{[\partial_t f](u, X_u) + \tfrac{1}{2}(V_u)^2[\partial_x^2 f](u, X_u)\Big\}\mathrm{d}u + \int_s^t [\partial_x f](u, X_u)\Big(V_u\,\mathrm{d}B_u + D_u\,\mathrm{d}u\Big) \\
&= f(s, X_s) + \int_s^t [\partial_x f](u, X_u)\,\mathrm{d}X_u \int_s^t \Big\{[\partial_t f](u, X_u) + \tfrac{1}{2}(V_u)^2[\partial_x^2 f](u, X_u)\Big\}\mathrm{d}u.
\end{aligned}$$

Thus, in this statement we have naturally introduced a Itô integral with respect to the stochastic process X instead of the Brownian process B. In fact we can even generalize this notation by considering a second stochastic process $Y = (Y_t)_{t\in[0,T]}$, adapted to the Brownian process and satisfying $\int_0^T Y_t^2 V_t^2\,\mathrm{d}t < \infty$ and $\int_0^T |Y_t D_t|\,\mathrm{d}t < \infty$, then one sets for $t \in (0, T]$

$$\int_0^t Y_u\,\mathrm{d}X_u := \int_0^t Y_u V_u\,\mathrm{d}B_u + \int_0^t Y_u D_u\,\mathrm{d}u.$$

Exercise 5.1.5. *By using the previous statement, check that the quadratic variation provided in* (5.1.3) *can also be expressed as*

$$[X]_t = X_t^2 - X_0^2 - 2\int_0^t X_u \,\mathrm{d}X_u.$$

Exercise 5.1.6. *Consider the Itô process satisfying*

$$\mathrm{d}X_t = X_t \,\mathrm{d}B_t + \tfrac{1}{2} X_t \,\mathrm{d}t, \qquad X_0 = x_0,$$

and assume that $X_t \geq 0$ for all $t \geq 0$. By applying Proposition 5.1.4 to the function $t \mapsto \ln(X_t)$, show that one solution of this equation is $X_t = x_0 \,\mathrm{e}^{B_t}$.

Exercise 5.1.7 (Langevin equation)**.** *Consider the Itô process satisfying*

$$\mathrm{d}X_t = -\beta X_t \,\mathrm{d}t + \alpha \,\mathrm{d}B_t, \qquad X_0 = x_0, \tag{5.1.4}$$

for $\alpha \in \mathbb{R}$ and $\beta > 0$. This equation is called the Langevin equation*. Note that this equation can be written equivalently*

$$X_t = x_0 + \alpha B_t - \beta \int_0^t X_s \,\mathrm{d}s.$$

Show that the solution of this equation reads

$$X_t = \mathrm{e}^{-\beta t} x_0 + \alpha \int_0^t \mathrm{e}^{-\beta(t-u)} \,\mathrm{d}B_u. \tag{5.1.5}$$

Inspiration can be obtained from [1, Example 7.9], [11, Example 5.6], or from [12, Example 7.4.5]. Reference [13, Example 3.2.5] also discusses this problem and its solution in details. The process defined by (5.1.5) *is also called the* Ornstein-Uhlenbeck process*, see also Example 2.2.6.*

Let us still consider the case when two Itô processes are involved, both constructed with respect to the same Brownian process.

Lemma 5.1.8. *For $j \in \{1, 2\}$, let X^j be a Itô process of the form $\mathrm{d}X_t^j = V_t^j \,\mathrm{d}B_t + D_t^j \,\mathrm{d}t$, and let $f : \mathbb{R}^3 \to \mathbb{R}$ be continuous, with $\partial_t f$ continuous, and with continuous partial derivatives with respect to the space variables up to order two. Then the following equality hold for any $s \leq t$:*

$$\begin{aligned} f(t, X_t^1, X_t^2) =& f(s, X_s^1, X_s^2) + \int_s^t [\partial_t f](u, X_u^1, X_u^2) \,\mathrm{d}u + \sum_{j=1}^2 \int_s^t [\partial_j f](u, X_u^1, X_u^2) \,\mathrm{d}X_u^j \\ &+ \tfrac{1}{2} \sum_{j=1}^2 \sum_{k=1}^2 \int_s^t [\partial_j \partial_k f](u, X_u^1, X_u^2) V_u^1 V_u^2 \,\mathrm{d}u. \end{aligned}$$

From the previous formula, if we consider $f(t, x^1, x^2) = x^1 x^2$, then we get:

Lemma 5.1.9 (Integration by parts)**.** *For $j \in \{1, 2\}$, let X^j be a Itô process of the form $\mathrm{d}X_t^j = V_t^j \,\mathrm{d}B_t + D_t^j \,\mathrm{d}t$. Then one has*

$$\mathrm{d}(X_t^1 X_t^2) = X_t^2 \,\mathrm{d}X_t^1 + X_t^1 \,\mathrm{d}X_t^2 + V_t^1 V_t^2 \,\mathrm{d}t. \tag{5.1.6}$$

Exercise 5.1.10. *Rewrite the content of* (5.1.6) *in an integral form, as explicitly as possible.*

Exercise 5.1.11. *For $f \in C^1([0,T])$ and for $t \in [0,T]$ prove the following equality:*

$$\int_0^t f(s)\,\mathrm{d}B_s = f(t)\,B_t - \int_0^t f'(s)\,B_s\,\mathrm{d}s.$$

Exercise 5.1.12. *Consider N Itô processes of the form $\mathrm{d}X_t^j = V_t^j\,\mathrm{d}B_t + D_t^j\,\mathrm{d}t$, and let $f : \mathbb{R}^{N+1} \to \mathbb{R}$ be continuous, with $\partial_t f$ continuous, and with continuous partial derivatives with respect to the space variables up to order two. Write the analog of Lemma 5.1.8 in this framework.*

Exercise 5.1.13. *Study some of the examples 7.10 to 7.13 provided in [1].*

5.2 Stochastic differential equations

Let $f : \mathbb{R}^{m+2} \to \mathbb{R}$ be a sufficiently regular function, and consider $x : \mathbb{R} \to \mathbb{R}$ also sufficiently regular. A *deterministic differential equation* for x is an equation of the form

$$f(t, x(t), x'(t), x''(t), \ldots, x^{(m)}(t)) = 0 \tag{5.2.1}$$

for $t \in [0,T]$, where $x^{(m)}$ denotes the m^{th}-derivative of x. The function x is called the *unknown*. Usually, equation (5.2.1) is provided with a set of initial conditions, as for example $x(0) = x_0$, $x'(0) = x'_0$, etc. Other types of initial conditions are also possible. Well-known examples of such equations are

$$c\,x(t) - x'(t) = 0, \qquad c \in \mathbb{R},$$

or

$$mx''(t) + kx(t) = 0, \qquad m, k > 0,$$

or

$$x''(t) - mg = 0, \qquad m > 0, g = 9.81...$$

Note that the first example is a *first order differential equation*, while the second and the third examples are *second order differential equations*. There exist various techniques for solving these equations, but quite often only numerical solutions can be exhibited.

The simplest examples of differential equations are first order differential equations of the form

$$x'(t) = f(t, x(t)), \qquad x(0) = x_0$$

for some $f : \mathbb{R}^2 \to \mathbb{R}$ sufficiently regular. Note that this equation does not cover all first order differential equations, but only a very specific subfamily. Quite informally, this equation can be rewritten as

$$\mathrm{d}x(t) = f(t, x(t))\,\mathrm{d}t, \qquad x(0) = x_0, \tag{5.2.2}$$

and its solution can be written as

$$x(t) = x_0 + \int_0^t f(s, x(s))\,\mathrm{d}s.$$

Let us now introduce some randomness in this formalism. The easiest way is to add some randomness in the initial conditions. For example, one can modify (5.2.2) and get the equation

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t, \qquad X_0 = Y,$$

with Y a random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$[7]. Such an equation is called *random differential equation* and can be solved separately for each initial condition $Y(\omega)$, $\omega \in \Omega$. Such an approach is interesting when the initial condition is not precisely known, or for the study of the sensitivity to the initial conditions.

A more interesting way to add randomness in the deterministic equations can be performed by adding a random noise term, at any time. With this approach, (5.2.2) becomes

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \qquad X_0 = Y, \tag{5.2.3}$$

still with Y a random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and where B denotes the 1-dimensional standard Brownian process. For this equation, the randomness is due to the initial condition and to the noise generated by the Brownian process, at any time. The equation (5.2.3) is called a *Itô stochastic differential equation*, and the Brownian process is called the *driving process* in this equation. If a solution of this equation exists, observe that it corresponds to a time-inhomogeneous diffusion process, as introduced in Definition 5.1.2. Note that it is possible to replace the Brownian process by other driving processes, but this necessitates a more advanced theory. However, this is necessary for modelling jumps, since the Brownian process is continuous.

As emphasized at the beginning of this chapter, (5.2.3) should be interpreted through its integral version, namely through the equation

$$X_t = X_0 + \int_0^t f(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s. \tag{5.2.4}$$

However, it is then necessary to ensure that some minimal regularity conditions are satisfied, as for example $s \mapsto f(s, X_s) \in M^1_{\mathrm{loc}}([0,T])$ and $s \mapsto \sigma(s, X_s) \in M^2_{\mathrm{loc}}([0,T])$ for some $T > 0$. Note that since the process X is not known yet, it is not possible to check *a priori* these conditions. Before begin able to check this regularity condition, it is even more fundamental to know if (5.2.4) admits any solution ? If it exists, is this solution unique or is there a family of solutions ? And in which sense can one speak about a solution ?

Definition 5.2.1 (Strong solution)**.** *A* strong solution *to the Itô stochastic differential equation* (5.2.3) *is a stochastic process $X = (X_t)_{t\in[0,T]}$ for some $T > 0$ such that X is adapted to the standard Brownian process B and such that $s \mapsto f(s, X_s) \in M^1_{\mathrm{loc}}([0,T])$ and $s \mapsto \sigma(s, X_s) \in M^2_{\mathrm{loc}}([0,T])$.*

Let us stress that all solutions of the exercises considered so far are strong solutions. Subsequently, we shall mention another type of solutions, the weak solutions.

Exercise 5.2.2. *This exercise is a continuation of Exercise 5.1.6. Consider the stochastic differential equation*

$$\mathrm{d}X_t = \sigma X_t\,\mathrm{d}B_t + rX_t\,\mathrm{d}t, \qquad X_0 = 1. \tag{5.2.5}$$

Show that the solution of this equation is given by the geometric Brownian process

$$X_t = \mathrm{e}^{(r-\sigma^2/2)t+\sigma B_t}.$$

This stochastic process was already mentioned in Exercise 3.2.5, see also [11, Examples 5.1 & 5.5], and has application in finance, see [11, Example 5.9]. One pedestrian solution of this problem is also proposed in [13, p. 140].

[7] Observe that we moved the time dependence to an index, since X_t is now a function defined on Ω.

Exercise 5.2.3. *Study the stochastic differential equation*

$$\mathrm{d}X_t = \sigma(t)\,\mathrm{d}B_t$$

with a non-random continuously differentiable function σ, see [11, Example 5.4]. Make the link with Wiener integral, see Theorem 4.3.4.

Remark 5.2.4. *One method for solving stochastic differential equations is to use* Stratonovich calculus. *It essentially consists in replacing the left rule of Itô integral by a mid-interval rule. More explicitly, the Stratonovich approximating sum for $\int_0^t X_s\,\mathrm{d}B_s$ is given for a partition $\mathcal{P}_\ell$ by*

$$\tfrac{1}{2}\sum_{j=1}^{n_\ell}(X_{t_j^\ell}+X_{t_{j+1}^\ell})(B_{t_{j+1}^\ell}-B_{t_j^\ell})$$

and the general case is obtained by a limiting process in the L^2-sense. Stratonovich calculus has rules which are slightly different from Itô calculus, but these two calculus complement each other. We do not develop this topic, but refer to [13, Sec. 2.4] for an introduction to this calculus and to [13, Sec. 3.2.3] for its use. Another precise construction is provided in [11, Sec. 5.9], with the link between Stratonovich stochastic differential equations and Itô stochastic differential equations.

Let us now look at a few examples of stochastic differential equations, starting with the stochastic exponential, see also Exercise 4.4.7. Consider a stochastic differential X, see Definition 5.1.1, and let U satisfy

$$\mathrm{d}U_t = U_t\,\mathrm{d}X_t, \qquad U_0 = 1. \tag{5.2.6}$$

Equivalently, this equation reads $U_t = 1 + \int_0^t U_s\,\mathrm{d}X_s$. We say in this case that U is the *stochastic exponential* of X. In this setting one has:

Lemma 5.2.5. *If X is a Itô process, the solution of (5.2.6) is given by*

$$U_t = \mathrm{e}^{X_t - X_0 - \frac{1}{2}[X]_t},$$

where $[X]_t$ denotes the quadratic variation of X defined in (4.3.1).

The proof of this statement is given in [11, Thm. 5.2] and can be studied as an exercise. Note that for general initial condition U_0 the solution is given by $U_t = U_0\mathrm{e}^{X_t - X_0 - \frac{1}{2}[X]_t}$. Quite interestingly and unlike the usual exponential, the expression for U_t requires the knowledge of the full process up to time t, since the quadratic variation $[X]_t$ requires such a knowledge. Note that there also exists a *stochastic logarithm*, which is presented in [11, Thm. 5.3].

The set of *linear stochastic differential equations* also admit explicit solutions. More precisely, consider the linear equation

$$\mathrm{d}X_t = (\alpha_t + \beta_t X_t)\,\mathrm{d}t + (\gamma_t + \delta_t X_t)\,\mathrm{d}B_t, \tag{5.2.7}$$

with $\alpha, \beta, \gamma, \delta$ continuous and univariate stochastic processes adapted to the standard 1-dimensional Brownian . The general solution takes the form

$$X_t := U_t\Big(X_0 + \int_0^t \frac{\alpha_s - \delta_s\gamma_s}{U_s}\,\mathrm{d}s + \int_0^t \frac{\gamma_s}{U_s}\,\mathrm{d}B_s\Big), \tag{5.2.8}$$

where $U = (U_s)_{s\geq 0}$ is the solution of the initial problem when $\alpha = 0$ and $\gamma = 0$ and is given by

$$U_s := U_0 \exp\Big(\int_0^s (\beta_u - \tfrac{1}{2}\delta_u^2)\,\mathrm{d}u + \int_0^s \delta_u\,\mathrm{d}B_u\Big).$$

Observe that Exercises 5.1.6, 5.1.7, 5.2.2, and 5.2.3 correspond to linear stochastic differential equations. The Brownian bridge, already mentioned in Example 2.2.4 is also the solution of a linear stochastic differential equation, see [11, p. 133–134] for the details.

Exercise 5.2.6. *Study and report on the general solution* (5.2.8)*, getting inspiration from [11, Sec. 5.3]. Detailed explanations are also provided in [13, Sec. 3.3].*

Let us now state a sufficient condition for the existence of a strong solution of (5.2.3). Clearly, providing a necessary and sufficient condition is often impossible. The following statement is proved in [7, Thm. 5.1.1], see also [2, Thm. 9.4].

Theorem 5.2.7. *Let B be the standard 1-dimensional Brownian process, and assume that the random variable Y is independent of B_t for any $t \in [0,T]$ and satisfies $\mathbb{E}(Y^2) < \infty$. Let $f, \sigma : [0,T]\times\mathbb{R} \to \mathbb{R}$ be* locally Lipschitz *in x uniformly in t, namely for any $N > 0$ there exists $K > 0$ such that for $|x| \leq N$, $|y| \leq N$, and all $t \in [0,T]$ one has*

$$|f(t,x) - f(t,y)| + |\sigma(t,x) - \sigma(t,y)| < K|x-y|.$$

In addition, assume that the functions f and σ satisfy the sublinear growth condition

$$|f(t,x)| + |\sigma(t,x)| \leq K(1+|x|) \qquad \forall x \in \mathbb{R}.$$

Then the Itô stochastic differential equation

$$\mathrm{d}X_t = f(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}B_t, \qquad X_0 = Y, \tag{5.2.9}$$

has a unique strong solution X on $[0,T]$, which is continuous and belong to $M^2([0,T])$.

Observe that the local Lipschitz condition is satisfies if the functions f and g are differentiable with respect to x, with bounded derivatives for any $|x| \leq N$ and $|y| \leq N$ and all $t \in [0,T]$. Obviously, this is satisfies if the partial derivatives are continuous. Let us stress that other sets of conditions also exist, leading also to existence and uniqueness of the solution of (5.2.9). For example, [1, Thm. 7.21] contains a set of conditions for the existence and uniqueness of the solution of a simplified stochastic differential equation under stronger regularity conditions but a weaker growth condition.

Remark 5.2.8. *Quite interestingly, it is also possible to consider a stochastic differential equation with two (or more) independent driving Brownian processes. We do not develop the theory, but refer to [13, Example 3.2.8] for an example. Note that the related construction is different from the one sketched in Lemma 5.1.8.*

As already mentioned, there also exists a second type of solutions to the stochastic differential equation (5.2.3) : the weak solutions. The main difference with the strong solutions is that the setting is not defined *a priori*.

Definition 5.2.9 (Weak solution)**.** *Consider the stochastic differential equation*

$$\mathrm{d}X_t = f(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}B_t. \tag{5.2.10}$$

Given an initial condition Y, a weak solution *of* (5.2.10) *consists of a 1-dimensional Brownian process $\hat{B} = \left(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (\hat{B}_t)_{t\geq 0}\right)$ and a stochastic process $\hat{X}$ adapted to $\hat{B}$ such that the following integrals are well defined and satisfy the equality*

$$\hat{X}_t = Y + \int_0^t f(s,\hat{X}_s)\,\mathrm{d}s + \int_0^t \sigma(s,\hat{X}_s)\,\mathrm{d}\hat{B}_s$$

for all $t \in [0,T]$ and some $T > 0$.

Let us emphasize that in the definition of weak solutions, the filtered probability space and the Brownian process are not specified *a priori*, they are part of the solution. Clearly, any strong solution is also a weak solution, but the converse is not true. For weak solutions, the notion of uniqueness can only be defined with the induced probability measure, since the underlying probability spaces can be different for two weak solutions: two weak solutions are equal if they define the same induced probability distribution.

Let us still mention a theorem about the existence of weak solutions, which complements Theorem 5.2.7. Again, the content is a sufficient condition, not a necessary one. We refer to [11, Thm. 5.11] for the statement, and for the references mentioned there for its proof.

Theorem 5.2.10. *Consider the stochastic differential equation*

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t. \tag{5.2.11}$$

Assume that σ is positive and continuous, and for any $T > 0$ that there exists $K_T > 0$ such that

$$|f(y, x)| + |\sigma(t, x)| \leq K_T(1 + |x|) \qquad \forall x \in \mathbb{R} \text{ and } t \in [0, T].$$

Then there exists a unique weak solution to (5.2.11) *starting at any $x_0 \in \mathbb{R}$ and at any time $t \geq 0$.*

Observe that this statement corresponds to a deterministic initial condition, since the initial condition is given for any fixed $x_0 \in \mathbb{R}$. Additional information about weak solutions can be found in [11, Sec. 5.7].

Chapter 6

Markov processes

In this chapter, we provide several definitions and results related to Markov processes. We also relate these processes to the solutions of the differential equations seen in the previous chapter.

6.1 Markov property

Stochastic processes with the Markov property are processes with no memory ! Their theory is well-developed and these processes play an enormous role in many applications. We provide here the main definitions and a few results. Note that the discrete time version is rather easy to understand and has several applications as well. They are often part of a course on probability. Here, we focus on the continuous time version.

Definition 6.1.1 (Markov property)**.** *Let* $X = (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (X_t)_{t\geq 0})$ *be a stochastic process taking values in a standard measurable space* $(\Lambda, \mathcal{E})$*. Then* X *possesses the* Markov property *if for any* $A \in \mathcal{E}$ *and any* $0 \leq s < t$

$$\mathbb{E}(\boldsymbol{1}_A(X_t)|\mathcal{F}_s) = \mathbb{E}(\boldsymbol{1}_A(X_t)|X_s) \quad a.s. \tag{6.1.1}$$

We recall that conditional expectation $\mathbb{E}(\mathbf{1}_A(X_t)|X_s)$ means $\mathbb{E}(\mathbf{1}_A(X_t)|\sigma(X_s))$ with $\sigma(X_s)$ the σ-algebra generated by X_s. There exist several reformulations of the above condition. For example, some authors write

$$\mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{P}(X_t \in A|X_s) \quad a.s.$$

instead of (6.1.1). It is also equivalent to require that

$$\mathbb{E}(f(X_t)|\mathcal{F}_s) = \mathbb{E}(f(X_t)|X_s) \quad a.s.$$

for all bounded and measurable functions $f : \Lambda \to \mathbb{R}$. The meaning of these requirements is that the information about the system at the time t given $\mathcal{F}_s$ does not depend on the history before time s but only on the state of the system at time s.

A special instance of the Markov property is very useful and convenient:

Definition 6.1.2 (Homogeneous Markov property)**.** *Let* $X = (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (X_t)_{t\geq 0})$ *be a stochastic process taking values in a standard measurable space* $(\Lambda, \mathcal{E})$*. Then* X *possesses the* time-homogeneous Markov property*, or simply the* homogeneous Markov property *if for any* $A \in \mathcal{E}$ *and any* $0 \leq s < t$

$$\mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{P}(X_{t-s} \in A|X_0) \quad a.s. \tag{6.1.2}$$

Remark 6.1.3. *Some authors call* Markov process *any stochastic process with the Markov property, and* time-homogeneous Markov process*, or simply* homogeneous Markov process *any process with the homogeneous Markov property. Other authors also assume some implicit regularity (right continuity, continuity, etc) when defining Markov processes. We shall refrain assuming any such additional condition and stick to the simplest definitions.*

Exercise 6.1.4. *Show that the standard* 1*-dimensional Brownian process has the time-homogeneous Markov property, see for example [1, p. 177–178] and [11, Thm. 3.9]. Note that the same result holds for the standard N-dimensional Brownian process as well.*

Exercise 6.1.5. *Recall the definition of the Markov property for discrete time stochastic processes.*

Another concept is deeply related to the Markov property:

Definition 6.1.6 (Markov transition function)**.** *Let* $(\Lambda, \mathcal{E})$ *be a measurable space. A function*

$$p : \mathbb{R}_+ \times \mathbb{R}_+ \times \Lambda \times \mathcal{E} \to [0, 1]$$

satisfying

1. *For any fixed* $s, t \in \mathbb{R}_+$ *and* $A \in \mathcal{E}$*, the map* $p(s, t, \cdot, A) : \Lambda \to [0, 1]$ *is measurable,*
2. *For any fixed* $s, t \in \mathbb{R}_+$ *and* $y \in \Lambda$*, the map* $p(s, t, y, \cdot) : \mathcal{E} \to [0, 1]$ *is a probability measure,*
3. *The Chapman-Kolmogorov equation*

$$p(s, t, y, A) = \int_\Lambda p(u, t, z, A)\, p(s, u, y, \mathrm{d}z) \tag{6.1.3}$$

for every $s < u < t$*, is called a* Markov transition function *on* $(\Lambda, \mathcal{E})$*.*

The exact interpretation of the Chapman-Kolmogorov equation will be discussed soon. Let us now link with the initial Markov process.

Definition 6.1.7 (Markov process associated to a transition function)**.** *Let* $(\Lambda, \mathcal{E})$ *be a standard measurable space, and let* p *be a Markov transition function on* $(\Lambda, \mathcal{E})$*. A stochastic process* $X = (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (X_t)_{t\geq 0})$ *taking values in* Λ *is said to be a* Markov process associated to p *if for any bounded and measurable function* $f : \Lambda \to \mathbb{R}$ *and any* $0 \leq s < t$ *one has*

$$\mathbb{E}(f(X_t)|\mathcal{F}_s) = \int_\Lambda f(x)\, p(s, t, X_s, \mathrm{d}x), \quad a.s. \tag{6.1.4}$$

Observe that if $f = \mathbf{1}_A$ for some $A \in \mathcal{E}$, then the equality (6.1.4) reads

$$\mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{E}(\mathbf{1}_A(X_t)|\mathcal{F}_s) = p(s, t, X_s, A) \quad a.s. \tag{6.1.5}$$

Therefore, by Proposition 3.1.3 (5. and then 2.) one has

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A(X_t)|X_s) &\equiv \mathbb{E}(\mathbf{1}_A(X_t)|\sigma(X_s)) = \mathbb{E}(\mathbb{E}(\mathbf{1}_A(X_t)|\mathcal{F}_s)\big|\sigma(X_s)) \\ &= \mathbb{E}(p(s, t, X_s, A)\big|\sigma(X_s)) = p(s, t, X_s, A) = \mathbb{E}(\mathbf{1}_A(X_t)|\mathcal{F}_s) \quad a.s.,\end{aligned}$$

which means that any process associated to a Markov transition function is a Markov process.

Remark 6.1.8. *It is shown in [7, Thm. 2.1.1] that for any Markov transition function p on $\mathbb{R}^N$ it is possible to construct a Markov process associated to it. This result clearly extends to the set of all Markov transition function on standard measurable spaces. In [2, p.155–157] the same construction is provided for any complete, separable metric space Λ, endowed with its σ-algebra of Borel sets. On the other hand, the Markov property seems not to be sufficient for guarantying the existence of a Markov transition function.*

Let us now discuss the meaning of the formalism introduced so far. First of all, let us rewrite (6.1.4) as

$$\mathbb{E}(f(X_t)|X_s) = \int_\Lambda f(x)\, p(s,t,X_s,\mathrm{d}x), \quad a.s.$$

By comparing this equality with (3.1.11), one infers that for any $A \in \mathcal{E}$

$$p(s,t,y,A) = \mathbb{P}(X_t \in A|X_s = y) = \mathbb{E}(\mathbf{1}_A(X_t)|X_s = y)$$

with the notation introduced in the paragraph following (3.1.11). Thus, $(p(s,t,y,\cdot))_{y\in\Lambda}$ corresponds to the conditional probability of X_t given X_s, see Definition 3.1.9. It then follows that $p(s,t,y,A)$ represents the probability that the process, being at position y at time s, will move to a position in the set A at time t. Similarly, the Chapman-Kolmogorov equation (6.1.3) means that if $s < u < t$ the probability of moving from position y at time s to a position in A at time t is equal to the probability of moving to a position z at the intermediate time u, and then from z to A at time t, integrated over all possible positions z.

Exercise 6.1.9. *Describe the transition function for stochastic processes in discrete time and taking values in a discrete (or finite) set.*

The existence of a transition function makes most of the computations relatively easy and explicit. For example, let us compute $\mathbb{E}(f_1(X_t)\, f_2(X_s))$ for f_1, f_2 two bounded and measurable real functions on Λ, and for $s < t$. Then, by using the intermediate conditioning formula (3.1.6) one gets

$$\begin{aligned}
\mathbb{E}(f_1(X_s)\, f_2(X_t)) &= \mathbb{E}\Big(f_1(X_s)\,\mathbb{E}(f_2(X_t)|\mathcal{F}_s)\Big) \\
&= \mathbb{E}\Big(f_1(X_s)\int_\Lambda f_2(x)p(s,t,X_s,\mathrm{d}x)\Big) \\
&= \int_\Lambda f_1(y)\Big[\int_\Lambda f_2(x)p(s,t,y,\mathrm{d}x)\Big]\mu_s(\mathrm{d}y),
\end{aligned}$$

where μ_s is the induced probability measure of X_s on Λ. If $s = 0$, observe that this formula depends only on the initial distribution μ_0 and on the transition function p

Exercise 6.1.10. *Generalize the previous formula for a sequence $t_0 < t_1 < t_2 < \cdots < t_n$ and for a family of bounded and measurable functions $f_j : \Lambda \to \mathbb{R}$. More precisely, compute*

$$\mathbb{E}(f_0(X_{t_0})\, f_1(X_{t_1}) \dots f_n(t_n)).$$

If $t_0 = 0$, observe that this formula depends only on the initial distribution μ_0 and on the transition function p.

Exercise 6.1.11. *Study the Markov transition function based on the Gaussian distribution, as presented in [2, Example 6.1].*

Let us finally come back to the time-homogeneous Markov property introduced in Definition 6.1.2. We consider a homogeneous Markov process X associated to a Markov transition function p. By using the time-homogeneity condition (6.1.2) and relation (6.1.5) one infers that for $0 \le s < t$

$$p(s,t,X_s,A) = \mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{P}(X_{t-s} \in A|X_0) = p(0,t-s,X_0,A) \quad a.s.$$

Thus, if we set $p(h, y, A) := p(0, h, y, A)$ for any $h \geq 0$, $y \in \Lambda$ and $A \in \mathcal{E}$, then we infer that a Markov process with Markov transition probability p is time-homogeneous if

$$p(s, t, X_s, A) = p(0, t - s, X_0, A) = p(t - s, X_0, A). \tag{6.1.6}$$

In this case, the Chapman-Kolmogovov equation takes the form

$$p(t, y, A) = \int_\Lambda p(t - s, z, A)\, p(s, y, \mathrm{d}z) \tag{6.1.7}$$

for any $0 \leq s < t$.

In the final exercise, the link between stochastic differential equations and Markov processes is established. Namely, it is shown that diffusion processes with regular functions possess the Markov property.

Exercise 6.1.12. *Under sufficient regularity, show that a time-homogeneous diffusion process, as introduced in Definition 5.1.2, is a time-homogeneous Markov process. Inspiration can be obtained from [1, Thm. 8.4]. More generally, show that any time-inhomogeneous diffusion process is a Markov process, see [11, Thm. 5.6] and [2, Sec. 9.7]. Note that a Markov transition function can be defined for these processes.*

6.2 Feller and strong Markov properties

Let us start by recalling that the notion of stopping time has been introduced in Definition 3.2.10 and that the σ-algebra $\mathcal{F}_\tau$ has been defined in (3.2.1). This σ-algebra contains the events which took place before or up to time τ. A stronger notion of Markov property is expressed in terms of stopping time.

Definition 6.2.1 (Strong Markov property)**.** *Let $X = (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (X_t)_{t\geq 0})$ be a stochastic process with values in a standard measurable space $(\Lambda, \mathcal{E})$. This process possesses the* strong Markov property *if for any finite stopping time τ for the filtration $(\mathcal{F}_t)_{t\geq 0}$ and for any $t \geq 0$*

$$\mathbb{E}(f(X_{t+\tau})|\mathcal{F}_\tau) = \mathbb{E}(f(X_{t+\tau})|X_\tau) \quad a.s.$$

for all bounded and measurable functions $f : \Lambda \to \mathbb{R}$.

As before, this condition means that the conditional expectation depends only on X_τ, and not on the full history encoded in $\mathcal{F}_\tau$. Also, observe that the above condition is equivalent to

$$\mathbb{E}(\mathbf{1}_A(X_{t+\tau})|\mathcal{F}_\tau) = \mathbb{E}(\mathbf{1}_A(X_{t+\tau})|X_\tau) = p(\tau, t + \tau, X_\tau, A) \quad a.s.$$

for all $A \in \mathcal{E}$, if X is associated to the Markov transition function p.

Exercise 6.2.2. *Show that the standard 1-dimensional Brownian process has the strong Markov property, see for example [2, Example 6.2] or [1, Thm. 8.6 & Example 8.7].*

Let us now move to the Feller property. It corresponds to a regularity property of the Markov transition function.

Definition 6.2.3 (Feller property, Feller process)**.** *A Markov transition function p on a standard measurable space $(\Lambda, \mathcal{E})$ possesses the* Feller property *if for any bounded and continuous function $f : \Lambda \to \mathbb{R}$ and any $h > 0$, the map*

$$\mathbb{R}_+ \times \Lambda \ni (t, y) \mapsto \int_\Lambda f(z)\, p(t, t + h, y, \mathrm{d}z)$$

is continuous. A Markov process associated to a Markov transition function p is a Feller process *if p satisfies the Feller property.*

Observe that if the Markov process is homogeneous, then the Feller condition reduces to the continuity of the map

$$\Lambda \ni y \mapsto \int_\Lambda f(z)\, p(h, y, \mathrm{d}z)$$

with p the function introduced before (6.1.6). Let us also mention two rather technical results: Firstly, for any right continuous Feller process, we can choose the right continuous filtration $(\mathcal{F}_{t+})_{t\geq 0}$ introduced in (2.3.1). The resulting stochastic process is still a Markov process associated with the same Markov transition function, see [7, Thm. 2.1]. Thus, for right-continuous Feller process, there is no restriction for considering it endowed with $(\mathcal{F}_{t+})_{t\geq 0}$. Secondly, any right-continuous Feller process endowed with the filtration $(\mathcal{F}_{t+})_{t\geq 0}$ possesses the strong Markov property. A proof is provided in [7, Thm. 2.4] or in [2, Thm. 6.1].

Let us now assume consider the Markov transition function associated to a homogeneous Markov process, with the notation introduced before (6.1.6). For the standard measurable space $(\Lambda, \mathcal{E})$, let us denote by $M_b(\Lambda)$ the set of bounded and measurable functions $f : \Lambda \to \mathbb{R}$, and denote by $C_b(\Lambda)$ the subset of continuous and bounded functions. For any $f \in M_b(\Lambda)$, any $t \geq 0$, and any $y \in \Lambda$ let us set

$$[U_t f](y) := \int_\Lambda f(z) p(t, y, \mathrm{d}z).$$

With the notation of conditional probability, this also corresponds to $\mathbb{E}(f(X_t)|X_0 = y)$. The following properties for the operator hold:

1. $\|U_t f\|_\infty \leq \|f\|_\infty$, with $\|g\|_\infty := \sup_{y\in\Lambda} |g(y)|$ for any $g \in M_b(\Lambda)$,
2. $U_s U_t = U_{s+t}$ for any $s, t \geq 0$,
3. If p has the Feller property, $U_t f$ belongs to $C_b(\Lambda)$ whenever $f \in C_b(\Lambda)$.

Let us mention that property 1. means that U_t is a contraction, property 2. means that the family $(U_t)_{t\geq 0}$ is a semi-group, while property 3. says that the operator U_l leaves $C_b(\Lambda)$ invariant, if the Markov transition function has the Feller property.

Exercise 6.2.4. *Prove the above properties, using the Chapman-Kolmogorov equation for the second one.*

In the above framework and whenever the Feller property holds, let us still define the *infinitesimal generator* of the semi-group $(U_t)_{t\geq 0}$, namely for $f \in C_b(\Lambda)$ we set

$$Af := \lim_{t\searrow 0} \tfrac{1}{t}(U_t f - f) \tag{6.2.1}$$

with this limit taken in $C_b(\Lambda)$. The set on which this limit exists is called the *domain of A* and is denoted by $\mathcal{D}(A)$.

If the Markov process is not homogeneous, one can still define the family of operators $(U_{s,t})_{t\geq s\geq 0}$ by

$$[U_{s,t} f](y) := \int_\Lambda f(z) p(s, t, y, \mathrm{d}z) = \mathbb{E}(f(X_t)|X_s = y).$$

This family of operators satisfies $U_{s,t} = U_{s,u} U_{u,t}$ for any $s \leq u \leq t$. For the generator, there are two natural definitions, leading to two different operators:

$$A_s^+ f := \lim_{h\searrow 0} \tfrac{1}{h}(U_{s,s+h} f - f) \tag{6.2.2}$$

with this limit taken in $C_b(\Lambda)$, and

$$A_s^- f := \lim_{h \searrow 0} \tfrac{1}{h}(U_{s-h,s} f - f) \tag{6.2.3}$$

with this limit taken in $C_b(\Lambda)$. Their natural domain are denoted by $\mathcal{D}(A_s^+)$ and $\mathcal{D}(A_s^-)$, respectively. We shall come back to these operators in the next section.

6.3 SDE and PDE

In this section, we establish the link between stochastic differential equations (SDE) and partial differential equations (PDE). This link is very successful and there exist many important results based on the intersection of these two subjects. Here we only scratch the surface, and consider only processes with values in $\mathbb{R}$. The generalization to processes with values in $\mathbb{R}^N$ is "straightforward" (which is always a lie ☹).

Let us come back to the framework of Definition 5.1.2 on time-inhomogeneous diffusion processes of the form

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mu(t, X_t)\,\mathrm{d}t, \quad t \geq 0.$$

with B the standard 1-dimensional Brownian process. Simultaneously, we define the second order differential operator

$$L_t := \tfrac{1}{2}\sigma^2(t, x)\partial_x^2 + \mu(t, x)\partial_x. \tag{6.3.1}$$

With this notation, and for any continuous $f : \mathbb{R}^2 \to \mathbb{R}$, with $\partial_t f$, $\partial_x f$, and $\partial_x^2 f$ also continuous, Itô's lemma for Itô processes takes the form

$$\mathrm{d}f(t, X_t) = \Big\{[\partial_t f](t, X_t) + [L_t f](t, X_t)\Big\}\mathrm{d}t + \sigma(t, X_t)[\partial_x f](t, X_t)\,\mathrm{d}B_t$$

which is equivalent to

$$\int_0^t \sigma(u, X_u)[\partial_x f](u, X_u)\,\mathrm{d}B_u = f(t, X_t) - f(0, X_0) - \int_0^t [\partial_t f + L_u f](u, X_u)\,\mathrm{d}u.$$

As mentioned in Theorem 4.3.1, the l.h.s. defines a mean zero and continuous martingale if the map $u \mapsto \sigma(u, X_u)[\partial_x f](u, X_u)$ belongs to $M^2([0, T])$ for some $T > 0$. Thus, under sufficient conditions on f, μ, and σ, one can ensure that this condition holds. Note that this result is based on existence theorems similar to Theorem 5.2.7, see [11, Thm .6.2 & 6.3]. Under such assumptions, and if f solves the *backward equation*

$$\partial_t f + L_t f = 0 \tag{6.3.2}$$

then one infers that $(f(t, X_t) - f(0, X_0))_{t\in[0,T]}$ is a mean zero and continuous martingale. However, even if (6.3.2) does not hold, the following statement takes place. Note that we provide a set of sufficient conditions, but other conditions are available in the specialized literature, see for example [11, Thm. 6.3].

Theorem 6.3.1 (Dynkin's formula)**.** *Let $f : [0, T] \times \mathbb{R}$, and assume that f, $\partial_t f$, $\partial_x f$, and $\partial_x^2 f$ are continuous, Assume also that σ, μ are Lipschitz*[8] *and have sublinear growth. Then, the map $u \mapsto \sigma(u, X_u)[\partial_x f](u, X_u)$ belongs to $M^2([0, T])$, with X_t defined by*

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mu(t, X_t)\,\mathrm{d}t, \quad t \geq 0,$$

[8] In this context, it means that $|\sigma(t, x) - \sigma(t, y)| + |\mu(t, x) - \mu(t, y)| \leq K|x - y|$ for all $x, y \in \mathbb{R}$ and for a constant K independent of $t \in [0, T]$.

and for any $t \in [0,T]$ one has

$$\mathbb{E}(f(t,X_t)) = \mathbb{E}(f(0,X_0)) + \mathbb{E}\Big(\int_0^t [\partial_t f + L_u f](u,X_u)\,\mathrm{d}u\Big).$$

The same equality also holds if t is replaced by a bounded stopping time τ, if $0 \le \tau \le T$.

The simple proof of this statement is based on Theorem 4.3.1 and on the Optional Stopping Theorem 3.2.18.

Let us now establish more clearly a link between SDE and PDE. The next statement provides a method for the computation of the expectation of a function of a stochastic process, namely the aim is to compute $\mathbb{E}(g(X_T)|X_t = y)$ for some bounded function g.

Theorem 6.3.2 (Backward equation with terminal value)**.** *Let X be solution for $t > s \ge 0$ of*

$$\mathrm{d}X_t = \sigma(t,X_t)\,\mathrm{d}B_t + \mu(t,X_t)\,\mathrm{d}t, \quad X_s = y$$

for some $y \in \mathbb{R}$ and with σ, μ Lipschitz and with sublinear growth. Let $f : [0,T]\times\mathbb{R}$ with f, $\partial_t f$, $\partial_x f$, and $\partial_x^2 f$ continuous, and with $\partial_x f$ bounded. Assume also that f is a solution of the backward equation $\partial_t f + L_t f = 0$ with L_t defined by (6.3.1), and also that $f(T,y) = g(y)$ for some bounded function g. Then the equality

$$f(t,y) = \mathbb{E}(g(X_T)|X_t = y) = \int_{\mathbb{R}} g(z)p(t,T,y,\mathrm{d}z) = [U_{t,T}g](y) \tag{6.3.3}$$

holds for any $t \in [0,T]$ and $y \in \mathbb{R}$.

The proof of this statement can be studied as an exercise, see [11, Corol. 6.4 & Thm. 6.6]. Clearly, the equality (6.3.3) makes a link between PDE and SDE, since the l.h.s. is a solution of a PDE, while the r.h.s. involves the solution of a SDE. Note also that a converse statement also holds, under suitable conditions. Namely, if we set $f(t,y) := \mathbb{E}(g(X_T)|X_t = y)$ for g, σ, μ verifying certain regularity and growth properties, then f is a solution of the backward equation (6.3.2) with boundary condition $f(T,y) = \lim_{t\nearrow T} f(t,y) = g(y)$. Such a statement is called *Kolmogorov's equation*, see [7, Thm. 5.6.1] and also [11, Thm. 6.9].

In the previous statement, the terminal condition was provided by the system at time T, and this is apparently the most useful version of the backward equation. Nevertheless, a similar statement also holds for an initial condition at time 0. The following statement is borrowed from [1, Thm. 8.11]. Note that the setting is less general and the conditions stronger.

Theorem 6.3.3 (Backward equation with initial value)**.** *Let X satisfy the time-homogeneous diffusion's condition*

$$\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}B_t + \mu(X_t)\,\mathrm{d}t, \quad t \ge 0$$

with σ, μ differentiable with bounded derivative. Set $L := \frac{1}{2}\sigma^2(x)\partial_x^2 + \mu(x)\partial_x$, and let $g \in C^2(\mathbb{R})$ be vanishing outside of a bounded interval. Then the solution of the initial value problem

$$\partial_t f = Lf, \qquad \text{with } f(0,y) = g(y)$$

satisfies the equality

$$f(t,y) = \mathbb{E}(g(X_t)|X_0 = y) = \int_{\mathbb{R}} g(z)p(t,y,\mathrm{d}z) = [U_t g](y) \tag{6.3.4}$$

for any $t \ge 0$ and $y \in \mathbb{R}$.

Exercise 6.3.4. *Illustrate the previous result with the heat equation. In this case, $L := \frac{1}{2}\partial_x^2$, and one ends up with the heat equation $\partial_t f = \frac{1}{2}\partial_x^2 f$, see Example [1, 8.10] and [11, Example 6.2].*

Let us now establish the link with the infinitesimal generator introduced at the end of Section 6.2, at least formally. By relation (6.2.1) and by considering the derivative with respect to the variable t of the equation (6.3.4), one gets

$$Ag = \lim_{t \searrow 0} \tfrac{1}{t}(U_t g - g) = \partial_t U_t g|_{t=0} = L U_t g|_{t=0} = Lg,$$

meaning that L is the infinitesimal generator. Similarly, by (6.2.3) and by considering the derivative with respect to the variable t of the equation (6.3.3), one gets

$$A_t^- f = \lim_{h \searrow 0} \tfrac{1}{h}(U_{t-h,t} - 1)U_{t,T} g = \lim_{h \searrow 0} \tfrac{1}{h}(U_{t-h,T} - U_{t,T})g = -\partial_t U_{t,T} g = L_t U_{t,T} g = L_t f,$$

meaning that L_t is the infinitesimal generator A_t^-.

As an example of the previous observation, let us look at the infinitesimal generator of a few processes already mentioned in these notes: The infinitesimal generator of the Ornstein-Uhlenbeck process can be deduced from (5.1.4), namely $L = \frac{1}{2}\alpha^2 \partial_x^2 - \beta x \partial_x$, the infinitesimal generator of the geometric Brownian process can be be inferred from (5.2.5), namely $L = \frac{1}{2}\sigma^2 x^2 \partial_x^2 + rx\partial_x$. The infinitesimal generator of any linear SDE can also be written based on (5.2.7).

Let us state one more relation between PDE and SDE, a simplified version of the celebrated Feynman-Kac formula. It corresponds to an extension of the previous statement, when the backward equation contains an additional term. The following statement is borrowed from [11, Thm. 6.8], but [2, Thm. 10.5 & Thm. 10.6] contain more general statements.

Theorem 6.3.5 (Feynman-Kac formula with terminal value)**.** *Consider a Itô process satisfying the differential stochastic equation* $\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mu(t, X_t)\,\mathrm{d}t$ *and let* $r : [0,T] \times \mathbb{R} \to \mathbb{R}$ *and* $g : \mathbb{R} \to \mathbb{R}$ *be measurable and bounded functions. Assume that* $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ *is a solution of the equation*

$$[\partial_t f](t,x) + [L_t f](t,x) = r(t,x) f(t,x), \qquad f(T,y) = g(y).$$

Then, f *is unique and satisfies the relation*

$$f(t,y) = \mathbb{E}\Big(\exp\Big(- \int_t^T r(u, X_u)\mathrm{d}u \Big) g(X_T) \,\Big|\, X_t = y \Big).$$

Exercise 6.3.6. *Give a probabilistic representation of the solution of the equation*

$$\partial_t f + \tfrac{1}{2}\sigma^2 x^2 \partial_x^2 f + \mu x \partial_x f = rf, \qquad f(T,y) = y^2$$

for $r, \sigma, \mu > 0$*, see also [11, Example 6.5].*

Let us now have a quick look at the notion of *forward equations*. Its theory is slightly more demanding since solutions are often considered in the weak sense only. First of all, we define the formal adjoint of an operator. Namely, consider f, g smooth real functions on $\mathbb{R}$ vanishing outside of a bounded set, and let $\langle f, g \rangle$ denotes their (real) scalar product: $\langle f, g \rangle = \int_{\mathbb{R}} f(x)\, g(x)\, \mathrm{d}x$. Then the formal adjoint of an operator A is defined by the relation

$$\langle f, Ag \rangle = \langle A^* f, g \rangle.$$

In particular if we consider the operator L_t defined in (6.3.1), then its formal adjoint acts on f as

$$L_t^* f := \tfrac{1}{2}\partial_x^2(\sigma^2(t,x) f) - \partial_x(\mu(t,x) f). \tag{6.3.5}$$

Note that there already appears an ambiguity: In the theory developed so far, the function $x \mapsto \sigma(t,x)$ has not been assumed to be two times differentiable, and the function $x \mapsto \mu(t,x)$ has not been assumed to be once differentiable. Thus, the derivatives in (6.3.5) have to be understood in a weak sense, which we do not develop here.

Exercise 6.3.7. *Compute the adjoint of the generators of the Brownian process, of the geometric Brownian process, and of the Ornstein-Uhlenbeck process, see for example [1, Examples 8.19 to 8.21].*

The adjoint operator can also be used for describing the evolution of diffusion process X. The forward equation is also called Fokker-Planck equation, or *Forward Kolmogorov equation*. In the next statement we provide two versions of forward equation, one for the induced probability measure, one for the Markov transition function. For the former, we assume that the induced probability measure is induced by a probability density function, while for the latter, we assume that there exists a density for the Markov transition function, namely that $p(s,t,y,A) = \int_A p(s,t,y,x)\,\mathrm{d}x$. The following statement is borrowed from [9, Lecture 10], and we refer to [1, Thm. 8.24] and to [12, Thm. 10.9.10] for the time-homogeneous version of these results.

Theorem 6.3.8 (Fokker-Planck equation)**.** *Assume the existence of the univariate stochastic process defined by*

$$\mathrm{d}X_t = \sigma(t,X_t)\,\mathrm{d}B_t + \mu(t,X_t)\,\mathrm{d}t, \quad X_0 = Y.$$

1. *Let μ_t be the induced probability probability density function of X_t, with μ_0 the induced probability density function of Y. Then the map $[0,T]\times\mathbb{R} \ni (t,x) \mapsto \mu_t(x) \in \mathbb{R}$ solves the equation*

$$\partial_t \mu_t = L_t^* \mu_t.$$

2. *Suppose that X is associated to a Markov transition function with a density denoted by p. Then the map $\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}\times\mathbb{R} \ni (s,t,y,x) \mapsto p(s,t,y,x) \in \mathbb{R}_+$ satisfies the following equality in the weak sense:*

$$\partial_t p = L_t^* p, \quad p(s,s,y,x) = \delta_0(x-y).$$

where L_t^* acts on the x-variable of the function p.

As a final remark for this chapter, let us mention that the processes considered so far had no restriction in the space variable, they took values in $\mathbb{R}$ or in $\mathbb{R}^N$. When spatial constraints are added, new phenomena take place, and can be encoded in the PDE setting. Namely, various boundary conditions are possible, like reflecting conditions, absorbing conditions, periodic boundary conditions, etc. Depending on these conditions, the behavior of the stochastic processes can change drastically. For that reason, the correct choices of the boundary conditions play a crucial role in applications. We leave this for your future...

Chapter 7

Applications in finance

In this chapter, we present some applications of concepts introduced in the previous chapters in mathematical finance. In fact, these applications have partially triggered the development of stochastic calculus, after the seminal works of Black and Scholes (1973) and of Merton (1973). Let us emphasize that the content of this chapter is based on models, which are often either too simple or not accurate enough. They can often be thought as a first approximation of a more complicated reality. However, once in a model, the mathematical developments should be as precise as possible.

7.1 Options

We firstly introduce a few definitions related to finance. The following definition is borrowed from [15, p. 3].

Definition 7.1.1 (Options)**.** *An* option *is a contract which gives the holder of the option the right, but not the obligation to buy or to sell an asset at a given price K, sometimes called* exercise price *or* strike price*, either at a fixed time T (called the* maturity time *or* expiration time*) or any time t within a fixed interval of time* $[0, T]$*. More precisely, a* call option *gives the holder of the option the right to buy at the given price, while a* put option *gives the holder the right to sell at a given price. A* European option *is an option for which the holder can exercise their right only at time T, while an* American option *is an option for which the holder can exercise their right at any time t in* $[0, T]$*.*

For example, a European call option is intended to guarantee its holder of being able to acquire the underlying asset at a price that is not larger than K (and therefore being safe from the market fluctuations). If at maturity the underlying asset has a price greater than K, then the holder of the option will exercise their right and obtain the asset at the price K. Otherwise, the option will be dropped, and the asset can be bought on the market (if available) at a price lower than K. More concretely, suppose that the price of the asset is represented by a stochastic process S. At time T, if $S_T > K$, then the holder of the European call option will probably exercise their right, buy the asset at the price K and sell it at the price S_T. The gain is then $S_T - K$. On the other hand, if $S_T \leq K$, then the option will expire as a worthless contract. Thus, the holder of a European call option will be able to earn at time T the amount given by

$$(S_T - K)^+ := \max\{0, S_T - K\} = \begin{cases} S_T - K & \text{if } S_T > K, \\ 0 & \text{if } S_T \leq K. \end{cases} \tag{7.1.1}$$

Observe that the same gain holds for an American call option, with the major difference that the time T has to

be replaced by any time $t \in [0, T]$. Clearly, it is then more difficult to decide when the holder will exercise its right, since the opportunities where $S_t > K$ can appear at various times.

For a European put option, a similar description holds: at time T if $S_T < K$, then it is advantageous to exercise the right to sell the asset at the price K and to buy it again at the market price S_T, while if $S_T > K$, it is better not to exercise the option. Thus, the holder of the option can earn $(K - S_T)^+$ at time T. Similarly, the holder of an American put option can earn $(K - S_t)^+$ at time $t \in [0, T]$. In these four situations, the amount of money earned by the holder of an option is often referred to as the *payoff*.

Note that in Definition 7.1.1, only two types of options (European or American) are described, but other options exist, often with more complicated rules. For examples, some options have a payoff which is "path-dependent" (meaning that payoff depend on the whole price process S), or whose payoff at time T is a more complicated function of S_T. We refer to [1, Sec. 10.7] or to [2, Example 13.3] for some examples. In addition and as already suggested, American options are more difficult to tackle than European options, since the payoff is time-dependent. Henceforth, we shall concentrate on the European option only, as it is done in several introductory books.

For any European option, since the price S_T is not known at time $t = 0$, the main question is *how much would anyone be willing to pay for this option ?*, or alternatively, *what is a rational price for this option at time $t = 0$?* These are the typical questions for the potential purchaser of an option. From the perspective of the issuer, the question becomes *how much should the issuer be paid in order to compensate the risk of having to pay the payoff ?* A related question is *what could be the strategy of the issuer in order to protect himself from a loss ?* We shall try to answer these questions in the following sections, under a few simplifying assumptions: 1) There are no transaction costs for buying or selling assets, and no taxes, 2) It is possible to buy or sell an arbitrary quantity of an asset, meaning that they can be multiplied by any real number and not only by integers or specific fractions. In this context, a negative number corresponds to the sale of the asset. It is also possible to sell assets which are not owned by the vendor (or equivalently to possess a negative amount of this asset).

7.2 Market model, trading strategies, arbitrage

In this section and in the following ones, we consider the standard N-dimensional Brownian process $B = (\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\in[0,T]}, \{B_t\}_{t\in[0,T]})$ for some $T > 0$. In particular, it means that the filtration is right continuous $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$ and that $\mathcal{F}_0$ contains the negligible sets of Ω. With no loss of generality, we can also choose $\mathcal{F} = \mathcal{F}_T$. The components of the Brownian process are denoted by $B^1, \dots, B^N$.

Let us start by introducing the framework: A *market model* consists of $n+1$ stochastic processes $S^0, S^1, \dots, S^n$, all taking values in $\mathbb{R}_+$ and defined by the stochastic differential equations:

$$\mathrm{d}S_t^i = A_t^i \,\mathrm{d}t + \sum_{j=1}^{N} G_t^{ij} \,\mathrm{d}B_t^j, \qquad i \in \{1, \dots, n\} \tag{7.2.1}$$

where A^i, G^{ij} are continuous and adapted to the filtration provided by the standard Brownian process. These equations correspond to Itô processes, as introduced in Definition 5.1.1 for the special case $N = 1$. The process S^0 satisfies the simpler stochastic equation

$$\mathrm{d}S_t^0 = r_t S_t^0 \,\mathrm{d}t \tag{7.2.2}$$

where the *spot rate* $t \mapsto r_t \geq 0$ is progressively measurable (see Definition 2.3.3) and bounded. These $n + 1$ stochastic processes represent the prices of some assets. The process S^0 represents a *risk-free asset*, which is going to play the role of a reference in the model, while the processes $S^1, \dots, S^n$ represent *risky assets*.

Example 7.2.1 (Black-Scholes model)**.** *One of the simplest examples of a market model is given for $N = 1$ and $n = 1$ by the stochastic differential equation*

$$\mathrm{d}S_t^1 = \mu S_t^1 \,\mathrm{d}t + \sigma S_t^1 \,\mathrm{d}B_t \tag{7.2.3}$$

with $\mu \in \mathbb{R}$ and $\sigma > 0$. The constant μ is called the mean rate of return *and the constant σ the* volatility*. On the other hand, the stochastic process for S^0 is still given by (7.2.2). Note that the function r can also be deterministic, which is independent of the underlying probability space associated with the Brownian process.*

The first equation corresponds to time-homogeneous diffusion processes, as introduced in Definition 5.1.2, while the second equation is a time-inhomogeneous diffusion process. The solutions of these equations are known: the first one has already been mentioned in Exercise 5.2.2 and reads

$$S_t^1 = S_0^1 \,\mathrm{e}^{(\mu-\sigma^2/2)t+\sigma B_t}, \tag{7.2.4}$$

with S_0^1 an initial condition independent of the Brownian process, while

$$S_t^0 = S_0^0 \exp\left(\int_0^t r_s \,\mathrm{d}s\right). \tag{7.2.5}$$

In the special case of a constant parameter r, then the solution is simply $S_t^0 = S_0^0 \,\mathrm{e}^{rt}$ for $t \in [0, T]$.

For completeness, let us describe how the equation (7.2.3) can be heuristically introduced. We denote by "$\mathrm{d}t$" a very small interval of time, and assume that the *relative return* from the asset S^1 and between time t and time $t + \mathrm{d}t$ is given by the relation

$$\frac{S_{t+\mathrm{d}t}^1 - S_t^1}{S_t^1} = \mu \,\mathrm{d}t + \sigma \,\mathrm{d}B_t.$$

In this relation, the term $\mu \,\mathrm{d}t$ can be interpreted as the *linear trend* while the term $\sigma \,\mathrm{d}B_t$ corresponds to the *stochastic noise*. Observe also that this relation can be rewritten as

$$S_{t+\mathrm{d}t}^1 - S_t^1 = \mu S_t^1 \,\mathrm{d}t + \sigma S_t^1 \,\mathrm{d}B_t$$

which corresponds to a heuristic version of (7.2.3).

A *trading strategy* or a *portfolio* for our assets consists in a stochastic process H with values in $\mathbb{R}^{n+1}$, namely a map

$$H : [0, T] \times \Omega \ni (t, \omega) \mapsto H_t(\omega) \equiv (H_t^0(\omega), H_t^1(\omega), \ldots, H_t^n(\omega)) \in \mathbb{R}^{n+1},$$

which is progressively measurable, see Definition 2.3.3, and such that the following technical conditions $H^0 \in M_{\mathrm{loc}}^1([0, T])$ and $H^i \in M_{\mathrm{loc}}^2([0, T])$ are satisfied for $i \in \{1, \ldots, n\}$, see Definition 4.2.5. Note that these regularity conditions ensure the following necessary properties

$$H^0 r \in M_{\mathrm{loc}}^1([0, T]), \quad H^i A^i \in M_{\mathrm{loc}}^1([0, T]), \quad H^i G^{ij} \in M_{\mathrm{loc}}^2([0, T])$$

for any $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, N\}$.

In the sequel, we shall interpret H_t^i as the number of risky asset S_t^i possessed at time t, while H_t^0 will represent the number of risk-free asset S_t^0 possessed at time t. Thus, given the assets S^i and S^0 defined by (7.2.1) and (7.2.2), the *value of the portfolio* or simply the *portfolio* $V(H)$ associated with the trading strategy H is given for $t \in [0, T]$ by

$$V_t(H) = H_t^0 S_t^0 + \sum_{i=1}^n H_t^i S_t^i \tag{7.2.6}$$

and is a measure of wealth. Note that H_t^i can take negative values, meaning a sale of the corresponding asset.

A trading strategy is *self-financing* if the stochastic process $V(H) \equiv \{V_t(H)\}_{t\in[0,T]}$ satisfies

$$V_t(H) = V_0(H) + \sum_{i=0}^{n} \int_0^t H_u^i \, \mathrm{d}S_u^i.$$

Equivalently, this relation reads $\mathrm{d}V_t = \sum_{i=0}^n H_t^i \, \mathrm{d}S_t^i$ with initial condition $V_0(H)$, or more explicitly:

$$\mathrm{d}V_t(H) = H_t^0 r_t S_t^0 \, \mathrm{d}t + \sum_{i=1}^{n} H_t^i A_t^i \, \mathrm{d}t + \sum_{i=1}^{n} H_t^i \sum_{j=1}^{N} G_t^{ij} \, \mathrm{d}B_t^j,$$

which is again a Itô process, see Definition 5.1.1 for the special case of $N = 1$. The meaning of this condition is that the changes in the value of the portfolio are due to capital losses or gains and not to increase or decrease of the invested funds. In other words, there are no external infusion of capital and no spending.

Let us now introduce some relative quantities. Indeed, the value of one unit of money changes over time because of the risk-free asset: 1¥ at time 0 is worth $\mathrm{e}^{\int_0^t r_s \, \mathrm{d}s}$¥ at time t, or e^{rt}¥ in the simplest case of a constant parameter r. This process is sometimes referred to as the *discounting process*. Then the *discounted price processes* and the *discounted portfolio* associated with the trading strategy H are defined for $t \in [0, T]$ and $i \in \{1, \dots, n\}$ by

$$\tilde{S}_t^i := \frac{S_t^i}{S_t^0} = \mathrm{e}^{-\int_0^t r_s \, \mathrm{d}s} S_t^i \qquad \text{and} \qquad \tilde{V}_t(H) := \frac{V_t(H)}{S_t^0} = \mathrm{e}^{-\int_0^t r_s \, \mathrm{d}s} V_t(H).$$

In the simplest case of a constant parameter r these expressions takes the form

$$\tilde{S}_t^i := \mathrm{e}^{-rt} S_t^i \qquad \text{and} \qquad \tilde{V}_t(H) := \mathrm{e}^{-rt} V_t(H).$$

Intuitively, $\tilde{S}_t^i$ corresponds to the amount of money which has to be invested at time $t = 0$ in the risk-free asset i in order to have the amount of S_t^i at time t.

Let us provide a statement which links the notion of self-financing with the discounted portfolio. The proof can be found in [2, Prop. 13.1] and looking at its proof is an instructive exercise.

Proposition 7.2.2. *The trading strategy H is self-financing if and only if for any $t \in [0, T]$*

$$\tilde{V}_t(H) = V_0(H) + \sum_{i=1}^{n} \int_0^t H_u^i \, \mathrm{d}\tilde{S}_u^i.$$

Among the family of all self-financing trading strategies, let us call *admissible* the ones satisfying $V_t(H) \geq 0$ a.s. for all $t \in [0, T]$. This condition means that the value of the portfolio remains non-negative at any time and almost surely, or in other words that the investor is solvent at any time[9]. We now define a special class of trading strategies. A self-financing trading strategy H over $[0, T]$ is said to be an *arbitrage strategy* or an *arbitrage opportunity* if the associated portfolio satisfies the three conditions $V_0(H) = 0$, $V_t(H) \geq 0$ for any $t \in [0, T]$, and $\mathbb{P}(V_T(H) > 0) > 0$. In other words, an arbitrage strategy does not require any initial capital, is admissible, and produces a gain with a strictly positive probability. Clearly, an arbitrage strategy is very seldom, and if it takes place, it is during a very short period of time.

[9] Some authors call *admissible* all strategies leading to $V_t(H) \geq -C$ for some $C > 0$, meaning that a certain debt can be tolerated.

Remark 7.2.3. *The condition $V_t(H) \geq 0$ for all $t \in [0,T]$ is very strong, and some authors require only that $V_T(H) \geq 0$, meaning that the investor might not be solvable at all time. In the framework of the Black-Scholes model, examples of arbitrage strategy can be exhibited, as for example when $r = 0$, $\mu = 0$ and $\sigma = 1$, see for example [15, Example p. 60] or [10, Example p. 9]. In these examples, the portfolio $V_T(H)$ at time T can take an arbitrary positive value, but this portfolio is not bounded from below for $t \in (0,T)$, which is not realistic. In addition, the number of jumps for the functions H^0 and H^1 might be infinite. Additional conditions are then imposed for avoiding these situations, like the condition $V_t(H) \geq 0$ for all $t \in [0,T]$.*

Exercise 7.2.4. *Describe an example of an arbitrage strategy. Can you find any which really took place and got famous ?*

Since arbitrage strategies are very rare, one often assumes that they don't exist. More precisely, a market model is said to be *arbitrage-free* if any admissible trading strategy H with $V_0(H) = 0$ is such that $\mathbb{P}(V_T(H) > 0) = 0$. This condition will be rephrased below in terms of an equivalent measure.

7.3 Equivalent martingale measures

Given two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$, we say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$ if $\mathbb{P}(A) = 0$ for $A \in \mathcal{F}$ implies $\mathbb{Q}(A) = 0$. This relation is often denoted by $\mathbb{P} \gg \mathbb{Q}$. The Radon-Nikodym theorem states that if $\mathbb{P} \gg \mathbb{Q}$ then there exists a measurable function $\Pi : \Omega \to \mathbb{R}_+$ such that

$$\mathbb{Q}(A) = \int_A \Pi \, \mathrm{d}P = \int_A \Pi(\omega) \mathbb{P}(\mathrm{d}\omega).$$

If both $\mathbb{P} \gg \mathbb{Q}$ and $\mathbb{Q} \gg \mathbb{P}$, then the two measures are said to be *equivalent*. This notion means that the two measures have the same negligible sets. Observe that it is the notion which has already been introduced in Definition 1.1.10, when one of the two measures is the Lebesgue measure on $\mathbb{R}^N$.

In relation with the previous section, one notion of equivalent measure is important. Before introducing it, recall that a process $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[0,T]}, (M_t)_{t\in[0,T]})$ is a martingale if $\mathbb{E}(M_t|\mathcal{F}_s) = M_s$ for all $s \leq t$, see Definition 3.2.1. Let us emphasize that the conditional expectation $\mathbb{E}(M_t|\mathcal{F}_s)$ has been defined in Section 3.1 and is based on the measure $\mathbb{P}$. Clearly, a change of this measure would give a different conditional expectation.

Definition 7.3.1 (Equivalent martingale measure)**.** *In the framework of the previous section, a probability measure $\mathbb{P}^*$ on $(\Omega, \mathcal{F})$ is called an* equivalent martingale measure *if $\mathbb{P}^*$ is equivalent to $\mathbb{P}$ and if the discounted price processes $\tilde{S}^1, \ldots, \tilde{S}^n$ are martingales with the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ and under $\mathbb{P}^*$. An equivalent martingale measure is also called a* risk-neutral probability.

It turns out that if there exists an equivalent martingale measure $\mathbb{P}^*$ and if H is an admissible trading strategy, then the discounted portfolio $\tilde{V}(H)$ is a supermartingale with the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ and under $\mathbb{P}^*$, see [2, Prop. 13.2]. However, it is often important to consider trading strategies leading to a discounted portfolio which is a true martingale under $\mathbb{P}^*$. Thus, whenever an equivalent martingale measure exists, we write $\mathcal{M}_T(\mathbb{P}^*)$ for the set of admissible trading strategies leading to a discounted portfolio $\tilde{V}(H)$ which is a martingale with the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ and under $\mathbb{P}^*$.

Let us now state the notion of arbitrage-free in terms of equivalent measure, see [2, Prop. 13.3] for the proof. It shows the importance of the notion of martingale equivalent measure. A more explicit statement can also be found in [12, Thm. 11.2.2] in a slightly

Proposition 7.3.2. *If there exists an equivalent martingale measure $\mathbb{P}^*$, then the market model is arbitrage-free.*

Let us mention that the converse statement also holds, namely the absence of arbitrage implies the existence of an equivalent martingale measure (under suitable assumptions). This statement corresponds to the first half of the fundamental theorem of asset pricing and plays a key role in mathematical finance, see also [11, Thm. 11.6].

We illustrate the previous definition and proposition with the Black-Scholes model introduced in Example 7.2.1 for $N = 1$ and $n = 1$. Consider the stochastic processes exhibited in (7.2.4) and (7.2.5), and let us rewrite the corresponding discounted price process:

$$\tilde{S}_t^1 = \mathrm{e}^{-\int_0^t r_s \,\mathrm{d}s} S_t^1 = S_0^1 \mathrm{e}^{(\mu-\sigma^2/2)t+\sigma B_t - \int_0^t r_s \,\mathrm{d}s} = S_0^1 \mathrm{e}^{-\sigma^2/2 t + \sigma\left(B_t + \frac{\mu}{\sigma}t - \int_0^t \frac{r_s}{\sigma}\,\mathrm{d}s\right)}.$$

Then, the new process $\tilde{B}_t := B_t + \frac{\mu}{\sigma}t - \int_0^t \frac{r_s}{\sigma}\,\mathrm{d}s$ can be seen as a Brownian process with a *drift*. It turns out that there exists an explicit equivalent martingale measure on $(\Omega, \mathcal{F})$ under which $\tilde{B}$ is distributed as a standard Brownian process (which is a martingale). Such a result is called Girsanov's theorem and its statement and proof can be found in [1, Thm. 9.11], [2, Thm. 12.1], or [11, Thm. 10.16], see also [12, Sec. 8.9] for a thorough presentation. Note that the idea of changing the underlying probability measure on $(\Omega, \mathcal{F})$ and studying its effect on random variables or on random processes is a well developed theory and can be found for example in [1, Chap. 9], [11, Chap. 10].

7.4 Replicating strategies and market completeness

Let us now come back to the notion of European options introduced in Section 7.1. Any of them is attached with the two quantities (Z, T), namely the time of maturity T and the payoff Z, which is a $\mathcal{F}_T$-measurable and non-negative random variable, see (7.1.1) for a European call option, and the similar formula for a European put option. Once again, our aim is to find a suitable price for acquiring this option.

Assume that we have a market model and that an equivalent martingale measure $\mathbb{P}^*$ exists. Let (Z, T) be a European option with $Z \in L^1(\Omega, \mathcal{F}_T, \mathbb{P}^*)$, see Definition 1.4.1. The option (Z, T) is *attainable* if there exists an admissible trading strategy $H \in \mathcal{M}_T(\mathbb{P}^*)$ such that $V_T(H) = Z$. Such a strategy H is said to *replicate* the option (Z, T) in $\mathcal{M}_T(\mathbb{P}^*)$. Note that the condition $Z \in L^1(\Omega, \mathcal{F}_T, \mathbb{P}^*)$ is in fact necessary for an option to be attainable (which means that we don't have to require it *a priori*). In this context the following statement plays an important role, see [2, Prop. 13.4].

Proposition 7.4.1. *Assume that an equivalent martingale measure $\mathbb{P}^*$ exists for a given market model, and let (Z, T) be an attainable European option. Then for any trading strategy H replicating Z in $\mathcal{M}_T(\mathbb{P}^*)$ the corresponding portfolio is given by*

$$V_t(H) = \mathbb{E}^*\left(\mathrm{e}^{-\int_t^T r_s \,\mathrm{d}s} Z \,|\, \mathcal{F}_t\right), \tag{7.4.1}$$

where $\mathbb{E}^*$ is the expectation computation with respect to $\mathbb{P}^*$.

As one can expect from the l.h.s., this result is independent of any specific choice of an equivalent martingale measure, if more than one exist. In the framework of the previous statement, let us stress that the amount of money defined by $V_t(H)$ is the right price for the European option at time t. Indeed, it can be shown that if the price is either higher or lower than $V_t(H)$, then there exists an arbitrage strategy, as introduced before. It means that with a different price for the option at time t, there would be a strategy leading to a strictly positive gain without any initial capital.

Exercise 7.4.2. *Explain how an arbitrage strategy can be established if the price of the option at time t is not given by $V_t(H)$, see for example [2, p. 405].*

Since for any attainable European option one can fix a no-arbitrage price, it is natural to wonder if all European options are attainable ? Recall that any good question is related to a definition ☺. In the present situation: A market model is *complete* if 1) there exists an equivalent martingale measure $\mathbb{P}^*$, 2) for any equivalent martingale measure $\mathbb{P}^*$, any European options (Z, T), with $Z \in L^1(\Omega, \mathcal{F}_T, \mathbb{P}^*)$, is attainable. Otherwise the market model is said to be *incomplete*. In the complete case, one has:

Proposition 7.4.3. *If the market model is complete, then an equivalent martingale measure is unique.*

The previous statement, with a simple proof can be found in [2, Thm. 13.1]. A converse statement also holds and corresponds to the second half of the fundamental theorem of asset pricing, see also [11, Thm. 11.15]. A more elaborated approach can be found in [15, Sec. 5.3 & 5.5].

7.5 Generalized Black-Scholes models

Let us start by reducing the generality of the stochastic processes introduced in Section 7.2. For that purpose, we assume that the stochastic differential equation (7.2.1) is of the form

$$\mathrm{d}S_t^i = S_t^i \mathrm{b}_i(t, S_t)\,\mathrm{d}t + \sum_{j=1}^{N} S_t^i \sigma_{ij}(t, S_t)\,\mathrm{d}B_t^j, \qquad i \in \{1, \dots, n\} \tag{7.5.1}$$

for $S_t = (S_t^1, \dots, S_t^n)$, the *drift* $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, and the *volatility* $\sigma : [0, T] \times \mathbb{R}^n \to M_{n\times N}(\mathbb{R})$. Observe that factorizing S_t^i in both terms of the r.h.s. is suggested by the expression presented in (7.2.3) for the Black-Scholes model, see Example 7.2.1. It addition, we assume that the functions b and σ are bounded and locally Lipschitz continuous, as introduced in the statement of Theorem 5.2.7. With these assumptions, there exists a solution of (7.5.1) which belongs to $M^2([0, T])$, see the mentioned theorem and [2, Thm. 9.1].

We shall now state a few results which are valid in this framework. For this, let us firstly recall that a matrix $a(t, x) \in M_n(\mathbb{R})$ is *positive definite* if $\langle a(t, x)\xi, \xi\rangle > 0$ for every $\xi \in \mathbb{R}^n$ with $\xi \neq 0$. Note that we use the notation $\langle \zeta, \xi\rangle$ for the scalar product between the two elements $\zeta, \xi \in \mathbb{R}^n$. We also say that matrix valued function a is *uniformly positive definite* or *uniformly elliptic* if there exists $\lambda > 0$ such that $\langle a(t, x)\xi, \xi\rangle \geq \lambda\|\xi\|^2$ for all $\xi \in \mathbb{R}^n$, $\xi \neq 0$, and all t and x.

Proposition 7.5.1. *Assume that $N \geq n$, that the matrix $\sigma(t, x)$ in (7.5.1) is of rank n for any t and x, and that the matrix valued function $\sigma\sigma^T$ is uniformly elliptic. Then there exists at least one equivalent martingale measure $\mathbb{P}^*$. If $N = n$, then this measure is unique, while if $N > n$ it is not.*

The previous statement can be found in [2, Prop. 13.6]. The next one, which is the main result for the generalized Black-Scholes models, corresponds to [2, Thm. 13.3].

Theorem 7.5.2. *If $N = n$ and if the matrix valued function $\sigma\sigma^T$ is uniformly elliptic, then the generalized Black-Scholes models defined in (7.5.1) is complete.*

It is interesting to mention that if $N < n$, then an equivalent martingale measure might not exist, and arbitrages might exist. Such an example is provided in [2, Exercise 13.5]. From now, we assume that $n = N$ and that the matrix valued function $\sigma\sigma^T$ is uniformly elliptic. By the completeness of the market model, any European options (Z, T), with $Z \in L^1(\Omega, \mathcal{F}_T, \mathbb{P}^*)$, is attainable, and the porfolio is given by (7.4.1). However, how can we construct explicitly the corresponding trading strategy H ?

For answering this question, let us impose two more technical assumptions: The function σ is Lipschitz continuous in both variables, see footnote 8 or here, and the function $t \mapsto r_t$ is driven by the risky asset, namely $r_t = r(t, S_t)$. We also assume that this function is Lipschitz continuous. Observe that a deterministic function (independent of any randomness) fits into this framework.

Let us also assume that the payoff of the European option is of the form $h(S_T)$ for some function $h : \mathbb{R}^n \to \mathbb{R}_+$ satisfying $h(x) \leq C(1 + \|x\|^\alpha)$ for some $C, \alpha > 0$. By Doob-Dynkin lemma mentioned on page 23, it follows that[10]

$$V_t(H) = \mathbb{E}^*\Big(\mathrm{e}^{-\int_t^T r(s,S_s)\,\mathrm{d}s} h(S_T)\,|\,\mathcal{F}_t\Big) = P(t, S_t)$$

where the function P belongs to $C^{1,2}([0,T] \times \mathbb{R}^n)$ and satisfies the PDE

$$\begin{cases} \partial_t P(t,x) + [L_t P](t,x) - r(t,x)\,P(t,x) = 0 \quad \text{for } (t,x) \in [0,T) \times \mathbb{R}^n, \\ P(T,x) = h(x), \end{cases} \tag{7.5.2}$$

with L_t the second order differential operator defined by

$$L_t := r(t,x) \sum_{i=1}^n x_i \partial_{x_i} + \tfrac{1}{2} \sum_{i,j=1}^n (\sigma\sigma^T)_{ij} x_i x_j \partial_{x_i} \partial_{x_j}.$$

The proof of the above statement can be found in [2, Thm. 13.4], and additional relations between SDE and PDE were already introduced in Section 6.3. Note that the equation (7.5.2) is sometimes called the *fundamental PDE following from the no-arbitrage approach*. Based on the function P, it is possible to compute the price of a European option with payoff $h(S_T)$ by solving a PDE. Quite surprisingly, observe that the equation (7.5.2) does not depend on the drift term appearing in (7.5.1), namely the term involving the function b.

Let us now state some relation between the function P and the trading strategy replicating the European option $(h(S_T), T)$. These relations read

$$\begin{cases} H_t^i = [\partial_{x_i} P](t, S_t) \quad \text{for } i \in \{1, \dots, n\}, \\ H_t^0 = \mathrm{e}^{-\int_0^t r(s,S_s)\,\mathrm{d}s} \Big(P(t,S_t) - \sum_{i=1}^n H_t^i S_t^i\Big). \end{cases} \tag{7.5.3}$$

In the financial jargon, the quantities H_t^i are also called the *deltas* of the option, and one has

$$\Delta_i(t, S_t) := [\partial_{x_i} P](t, S_t)$$

for $i \in \{1, \dots, n\}$. Additional funny names are given to other partial derivatives of P, namely:

$$\begin{aligned} &\Gamma_{ij}(t,S_t) := [\partial_{x_i}\partial_{x_j} P](t,S_t) && \text{gammas}, \\ &\Theta(t,S_t) := [\partial_t P](t,S_t) && \text{theta}, \\ &\rho(t,S_t) := [\partial_r P](t,S_t) && \text{rho}, \\ &\mathrm{Vega}(t,S_t) := [\partial_\sigma P](t,S_t) && \text{vega}. \end{aligned}$$

Clearly, the last two expressions have to be better defined, if these variables are not constants. Altogether, these expressions are called the *Greek*, for a quite obvious reason. And additional expressions apparently exist, called *vanna*, *charm*, *speed*, *vomma*, *veta*, *zomma*, *ultima*, *charm*, *vera*, *color*, see [1, Sec. 10.5] and here.

[10] The second equality holds because the uniform ellipticity of $\sigma\sigma^T$ implies the equality between the original filtration $\mathcal{F}_t$ and the augmented natural filtration generated by S_t.

7.6 The Black-Scholes model

In this final section, we further develop some the previous results in one of the simplest market model, the Black-Scholes model already introduced in Example 7.2.1. For the presentation, we mainly follow [2, Sec. 13.7] but another pedestrian approach can be found in [13, Chap. 4]. In this model, there is only one risky asset, whose evolution is denoted by the stochastic process S, and which satisfies the stochastic differential equation

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t, \tag{7.6.1}$$

where μ and the volatility σ are constant. The spot rate of the risk-free asset is also assumed to be constant. If we consider the simple payoff introduced in (7.1.1), then one has

$$P(t, S_t) = \mathbb{E}^*\Big(\mathrm{e}^{-r(T-t)}(S_T - K)^+ | \mathcal{F}_t\Big)$$

with K the strike price of a European call option with maturity time T. Note that the equivalent martingale measure is explicit for this example, which means that the above expression can be explicitly computed, and one gets

$$P(t, x) = \frac{\mathrm{e}^{-r(T-t)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (x\mathrm{e}^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z} - K)^+ \mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

Observe that the integrand vanishes if $x\mathrm{e}^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z} \leq K$, or equivalently if $z \leq d_0(T-t, x)$ with

$$d_0(t, x) := \frac{1}{\sigma\sqrt{t}}\Big(-\ln(x/K) - (r - \sigma^2/2)t\Big).$$

As a consequence, one infers that

$$P(t, x) = \frac{\mathrm{e}^{-r(T-t)}}{\sqrt{2\pi}} \int_{d_0(T-t,x)}^{\infty} (x\mathrm{e}^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z} - K)\mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

If we introduce function $\Phi : \mathbb{R} \to [0, 1]$ by $\Phi(y) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \mathrm{e}^{-z^2/2}\,\mathrm{d}z$ (the cumulative distribution function), then one gets (after a few lines of computations)

$$P(t, x) = x\Phi\Big(-d_0(T-t, x) + \sigma\sqrt{T-t}\Big) - K\mathrm{e}^{-r(T-t)}\Phi(-d_0(T-t, x)).$$

Thus, if we set

$$d_1(t, x) := -d_0(t, x) + \sigma\sqrt{t} = \frac{1}{\sigma\sqrt{t}}\Big(\ln(x/K) + (r + \sigma^2/2)t\Big)$$

and

$$d_2(t, x) = -d_0(t, x) = \frac{1}{\sigma\sqrt{t}}\Big(\ln(x/K) + (r - \sigma^2/2)t\Big)$$

then one finally obtains

$$P(t, x) = x\Phi(d_1(T-t, x)) - K\mathrm{e}^{-r(T-t)}\Phi(d_2(T-t, x)).$$

For different values of x, K and σ, this function is represented in [2, p. 424].

The previous result is valid for the simple payoff introduced in (7.1.1). For completeness, let us state a more general result which can be found in [1, Prop. 10.13].

Proposition 7.6.1. *For the Black-Scholes model and for a payoff given by $h(x) \leq C(1+|x|^{\alpha})$ for some $C, \alpha > 0$. Then, the function P is given by*

$$P(t,x) = \frac{\mathrm{e}^{-r(T-t)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(x\mathrm{e}^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z})\,\mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

Based on the above expression for the simple payoff function, everything can be computed for the Black-Scholes model. We list a few results, most of them can be obtained by straightforward computations. Note that these expressions are computed for a European call option, similar computations also hold for a European put option. We set $\phi := \Phi'$ and get (see also [2, p. 425–426]) :

$$\begin{aligned} H_t^1 &= [\partial_{x_i} P](t, S_t) = \Phi(d_1(T-t, S_t)), \\ H_t^0 &= \mathrm{e}^{-rt}\Big(P(t, S_t) - \Phi(d_1(T-t, S_t))\Big), \end{aligned}$$

and also

$$\begin{aligned} \Gamma(t,x) &= \frac{\phi(d_1(T-t,x))}{x\sigma\sqrt{T-t}}, \\ \Theta(t,x) &= -\frac{x\phi(d_1(T-t,x))\sigma}{2\sqrt{T-t}} - rK\mathrm{e}^{-r(T-t)}\phi(d_2(T-t,x)), \\ \rho(t,x) &= K(T-t)\mathrm{e}^{-r(T-t)}\Phi(d_2(T-t,x)), \\ \mathrm{Vega}(t,x) &= x\sqrt{T-t}\phi(d_1(T-t,x)). \end{aligned}$$

Exercise 7.6.2. *Check these expressions, based on the definitions given in the previous section.*

Let us finally mention that an example of a *path-dependent option* is thoroughly presented in [2, Example 13.3] and various additional examples are available in [1, Chap. 10]. These examples can be understood with the material introduced so far. For more advanced examples, and also for the study of American options, we recommend [15, Chap. 4]. However, additional efforts are necessary for understanding these developments. No such thing as a free lunch !

Index

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 process*, 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 process: an introduction to stochastic processes*, De Gruyter, 2014.

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