Stochastic Calculus
Financial Derivatives
and PDE's

Simone Calogero

January 11, 2021

Preface

Financial derivatives, such as stock options for instance, are indispensable instruments in modern financial markets. The introduction of (call) options markets in the early 70's, and the continuous appearance of new types of derivative contracts, gave impulse to the birth of what is now known as *options pricing theory*. This theory is the main subject of these notes, together with the required background on probability theory, stochastic calculus and partial differential equations which are essential mathematical tools in modern options pricing theory. The main part of this text dealing with applications to finance is Chapter 6, but several important financial concepts are scattered in the previous chapters as well. It is strongly recommended to complement the reading of these notes with the book by Shreve [26], which is by now a standard reference on the subject.

The solutions of some selected exercises can be found in Appendix B. Exercises marked with the symbol $(\star)$ are left to the students as assignments (see the course homepage for the submission deadline).

Contents

Chapter 1

Probability spaces

1.1 σ-algebras and information

We begin with some notation and terminology. The symbol Ω denotes a generic non-empty set; the **power of Ω**, denoted by 2^Ω, is the set of all subsets of Ω. If the number of elements in the set Ω is $M \in \mathbb{N}$, we say that Ω is **finite**. If Ω contains an infinite number of elements and there exists a bijection $\Omega \leftrightarrow \mathbb{N}$, we say that Ω is **countably infinite**. If Ω is neither finite nor countably infinite, we say that it is **uncountable**. An example of uncountable set is the set $\mathbb{R}$ of real numbers. When Ω is finite we write $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$, or $\Omega = \{\omega_k\}_{k=1,\dots,M}$. If Ω is countably infinite we write $\Omega = \{\omega_k\}_{k\in\mathbb{N}}$. For a finite set Ω with M elements, the power set contains 2^M elements. For instance, if $\Omega = \{\heartsuit, 1, \$\}$, then

$$2^\Omega = \{\emptyset, \{\heartsuit\}, \{1\}, \{\$\}, \{\heartsuit, 1\}, \{\heartsuit, \$\}, \{1, \$\}, \{\heartsuit, 1, \$\} = \Omega\},$$

which contains $2^3 = 8$ elements. Here $\emptyset$ denotes the **empty set**, which by definition is a subset of all sets.

Within the applications in probability theory, the elements $\omega \in \Omega$ are called **sample points** and represent the possible outcomes of a given experiment (or trial), while the subsets of Ω correspond to **events** which may occur in the experiment. For instance, if the experiment consists in throwing a dice, then $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $A = \{2, 4, 6\}$ identifies the event that the result of the experiment is an even number. Now let $\Omega = \Omega_N$,

$$\Omega_N = \{(\gamma_1, \dots, \gamma_N),\ \gamma_k \in \{H, T\}\} = \{H, T\}^N,$$

where H stands for "head" and T stands for "tail". Each element $\omega = (\gamma_1, \dots, \gamma_N) \in \Omega_N$ is called a **N-toss** and represents a possible outcome for the experiment "tossing a coin N consecutive times". Evidently, Ω_N contains 2^N elements and so 2^{Ω_N} contains 2^{2^N} elements. We show in Section 1.4 that Ω_∞—the sample space for the experiment "tossing a coin infinitely many times"—is uncountable.

A collection of events, e.g., $\{A_1, A_2, \dots\} \subset 2^\Omega$, is also called **information**. The power set of the sample space provides the total accessible information and represents the collection of all the events that can be resolved (i.e., whose occurrence can be inferred) by knowing the outcome of the experiment. For an uncountable sample space, the total accessible information is huge and it is typically replaced by a subclass of events $\mathcal{F} \subset 2^\Omega$, which is imposed to form a σ-algebra.

Definition 1.1. *A collection $\mathcal{F} \subseteq 2^\Omega$ of subsets of Ω is called a* **σ-algebra** *(or* **σ-field***) on Ω if*

(i) $\emptyset \in \mathcal{F}$;

(ii) $A \in \mathcal{F} \Rightarrow A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{F}$;

(iii) $\bigcup_{k=1}^\infty A_k \in \mathcal{F}$, *for all* $\{A_k\}_{k\in\mathbb{N}} \subseteq \mathcal{F}$.

If $\mathcal{G}$ is another σ-algebra on Ω and $\mathcal{G} \subset \mathcal{F}$, we say that $\mathcal{G}$ is a **sub-σ-algebra** *of $\mathcal{F}$.*

Exercise 1.1. *Let $\mathcal{F}$ be a σ-algebra. Show that $\Omega \in \mathcal{F}$ and that $\cap_{k\in\mathbb{N}} A_k \in \mathcal{F}$, for all countable families $\{A_k\}_{k\in\mathbb{N}} \subset \mathcal{F}$ of events.*

Exercise 1.2. *Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ be the sample space of a dice roll. Which of the following sets of events are σ-algebras on Ω?*

1. $\{\emptyset, \{1\}, \{2, 3, 4, 5, 6\}, \Omega\}$,
2. $\{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{1, 3, 4, 5, 6\}, \{2, 3, 4, 5, 6\}, \{3, 4, 5, 6\}, \Omega\}$,
3. $\{\emptyset, \{1, 3, 4\}, \{5, 6\}, \{1, 3, 4, 5, 6\}, \Omega\}$.

Exercise 1.3 (Sol. 1)**.** *Prove that the intersection of any number of σ-algebras (including uncountably many) is a σ-algebra. Show with a counterexample that the union of two σ-algebras is not necessarily a σ-algebra.*

Remark 1.1 (Notation)**.** The letter A is used to denote a generic event in the σ-algebra. If we need to consider two such events, we denote them by A, B, while N generic events are denoted $A_1, \dots, A_N$.

Let us comment on Definition 1.1. The empty set represents the "nothing happens" event, while A^c represents the "A does not occur" event. Given a finite number $A_1, \dots, A_N$ of events, their union is the event that at least one of the events $A_1, \dots, A_N$ occurs, while their intersection is the event that all events $A_1, \dots, A_N$ occur. The reason to include the countable union/intersection of events in our analysis is to make it possible to "take limits" without crossing the boundaries of the theory. Of course, unions and intersections of infinitely many sets only matter when Ω is not finite.

The smallest σ-algebra on Ω is $\mathcal{F} = \{\emptyset, \Omega\}$, which is called the **trivial σ-algebra**. There is no relevant information contained in the trivial σ-algebra. The largest possible σ-algebra is $\mathcal{F} = 2^\Omega$, which contains the full amount of accessible information. When Ω is countable, it is common to pick 2^Ω as σ-algebra of events. However, as already mentioned, when Ω is uncountable this choice is unwise. A useful procedure to construct a σ-algebra of events when Ω is uncountable is the following. First we select a collection of events (i.e., subsets of Ω), which for some reason we regard as fundamental. Let $\mathcal{O}$ denote this collection of events. Then we introduce the smallest σ-algebra containing $\mathcal{O}$, which is formally defined as follows.

Definition 1.2. *Let $\mathcal{O} \subset 2^\Omega$. The* **$\sigma$-algebra generated by** *$\mathcal{O}$ is*

$$\mathcal{F}_\mathcal{O} = \bigcap \left\{ \mathcal{F} : \mathcal{F} \subset 2^\Omega \text{ is a } \sigma\text{-algebra and } \mathcal{O} \subseteq \mathcal{F} \right\},$$

i.e., $\mathcal{F}_\mathcal{O}$ is the smallest σ-algebra on Ω containing $\mathcal{O}$.

Recall that the intersection of any number of σ-algebras is still a σ-algebra, see Exercise 1.3, hence $\mathcal{F}_\mathcal{O}$ is a well-defined σ-algebra. For example, let $\Omega = \mathbb{R}^d$ and let $\mathcal{O}$ be the collection of all open balls:

$$\mathcal{O} = \{B_x(R)\}_{R>0, x\in\mathbb{R}^d}, \quad \text{where } B_x(R) = \{y \in \mathbb{R}^d : |x - y| < R\}.$$

The σ-algebra generated by $\mathcal{O}$ is called **Borel σ-algebra** and denoted $\mathcal{B}(\mathbb{R}^d)$. The elements of $\mathcal{B}(\mathbb{R}^d)$ are called **Borel sets**.

Remark 1.2 (Notation)**.** The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ plays an important role in these notes, so we shall use a specific notation for its elements. A generic event in the σ-algebra $\mathcal{B}(\mathbb{R})$ will be denoted U; if we need to consider two such events we denote them by U, V, while N generic Borel sets of $\mathbb{R}$ will be denoted $U_1, \dots U_N$. Recall that for general σ-algebras we use the notation indicated in Remark 1.1.

The σ-algebra generated by $\mathcal{O}$ has a particular simple form when $\mathcal{O}$ is a partition of Ω.

Definition 1.3. *Let $I \subseteq \mathbb{N}$. A collection $\mathcal{O} = \{A_k\}_{k\in I}$ of non-empty subsets of Ω is called a* **partition** *of Ω if*

(i) the events $\{A_k\}_{k\in I}$ are **disjoint***, i.e., $A_j \cap A_k = \emptyset$, for $j \neq k$;*

(ii) $\bigcup_{k\in I} A_k = \Omega$.

If I is a finite set we call $\mathcal{O}$ a **finite partition** *of Ω.*

For example any countable sample space $\Omega = \{\omega_k\}_{k\in\mathbb{N}}$ is partitioned by the **atomic events** $A_k = \{\omega_k\}$, where $\{\omega_k\}$ identifies the event that the result of the experiment is exactly ω_k.

Exercise 1.4. *Show that when $\mathcal{O}$ is a partition of Ω, the σ-algebra generated by $\mathcal{O}$ is given by the set of all subsets of Ω which can be written as the union of sets in the partition $\mathcal{O}$ (plus the empty set, of course).*

Exercise 1.5. *Find the partition of $\Omega = \{1, 2, 3, 4, 5, 6\}$ that generates the σ-algebra 2 in Exercise 1.2.*

1.2 Probability measure

To any event $A \in \mathcal{F}$ we want to assign a probability that A occurred.

Definition 1.4. *Let $\mathcal{F}$ be a σ-algebra on Ω. A* **probability measure** *is a function*

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$

such that

(i) $\mathbb{P}(\Omega) = 1$*;*

(ii) for any countable collection of disjoint events $\{A_k\}_{k\in\mathbb{N}} \subseteq \mathcal{F}$*, we have*

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

A triple $(\Omega, \mathcal{F}, \mathbb{P})$ *is called a* **probability space***.*

The quantity $\mathbb{P}(A)$ is called **probability of the event A**; if $\mathbb{P}(A) = 1$ we say that the event A occurs **almost surely**, which is sometimes shortened by **a.s.**; if $\mathbb{P}(A) = 0$ we say that A is a **null set**. In general, the elements of $\mathcal{F}$ with probability zero or one will be called **trivial events** (as trivial is the information that they provide). For instance, $\mathbb{P}(\Omega) = 1$, i.e., the probability that "something happens" is one, and $\mathbb{P}(\emptyset) = \mathbb{P}(\Omega^c) = 1 - \mathbb{P}(\Omega) = 0$, i.e., the probability the "nothing happens" is zero.

Exercise 1.6 (Sol. 2)**.** *Prove the following properties:*

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*;*
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$*;*
3. *If* $A \subset B$*, then* $\mathbb{P}(A) < \mathbb{P}(B)$*.*

Exercise 1.7 (Continuity of probability measures ($\star$))**.** *Let* $\{A_k\}_{k\in\mathbb{N}} \subseteq \mathcal{F}$ *such that* $A_k \subseteq A_{k+1}$*, for all* $k \in \mathbb{N}$*. Let* $A = \cup_k A_k$*. Show that*

$$\lim_{k\to\infty} \mathbb{P}(A_k) = \mathbb{P}(A).$$

Similarly, if now $\{A_k\}_{k\in\mathbb{N}} \subseteq \mathcal{F}$ *such that* $A_{k+1} \subseteq A_k$*, for all* $k \in \mathbb{N}$ *and* $A = \cap_k A_k$*, show that*

$$\lim_{k\to\infty} \mathbb{P}(A_k) = \mathbb{P}(A).$$

Let us see some examples of probability space.

- There is only one probability measure defined on the trivial σ-algebra, namely $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.

- In this example we describe the general procedure to construct a probability space on a countable sample space $\Omega = \{\omega_k\}_{k\in\mathbb{N}}$. We pick $\mathcal{F} = 2^\Omega$ and let $0 < p_k < 1$, $k \in \mathbb{N}$, be real numbers such that
$$\sum_{k=1}^{\infty} p_k = 1.$$
We introduce a probability measure on $\mathcal{F}$ by first defining the probability of the atomic events $\{\omega_1\}, \{\omega_2\}, \dots$ as
$$\mathbb{P}(\{\omega_k\}) = p_k, \quad k \in \mathbb{N}.$$
Since every (non-empty) subset of Ω can be written as the disjoint union of atomic events, then the probability of any event can be inferred using the property (ii) in the definition of probability measure, e.g.,
$$\begin{aligned}\mathbb{P}(\{\omega_1, \omega_3, \omega_5\}) &= \mathbb{P}(\{\omega_1\} \cup \{\omega_3\} \cup \{\omega_5\}) \\ &= \mathbb{P}(\{\omega_1\}) + \mathbb{P}(\{\omega_3\}) + \mathbb{P}(\{\omega_5\}) = p_1 + p_3 + p_5.\end{aligned}$$
In general we have
$$\mathbb{P}(A) = \sum_{k:\omega_k\in A} p_k, \quad A \in 2^\Omega,$$
while $\mathbb{P}(\emptyset) = 0$. In countable sample spaces the empty set is the only event with zero probability.

- As a special case of the previous example we now introduce a probability measure on the sample space Ω_N of the N-coin tosses experiment. Given $0 < p < 1$ and $\omega \in \Omega_N$, we define the probability of the atomic event $\{\omega\}$ as
$$\mathbb{P}(\{\omega\}) = p^{N_H(\omega)}(1-p)^{N_T(\omega)}, \tag{1.1}$$
where $N_H(\omega)$ is the number of H in ω and $N_T(\omega)$ is the number of T in ω ($N_H(\omega) + N_T(\omega) = N$). We say that the coin is **fair** if $p = 1/2$. The probability of a generic event $A \in \mathcal{F} = 2^{\Omega_N}$ is obtained by adding up the probabilities of the atomic events whose disjoint union forms the event A. For instance, assume $N = 3$ and consider the event
"The first and the second toss are equal".
Denote by $A \in \mathcal{F}$ the set corresponding to this event. Since A is the (disjoint) union of the atomic events
$$\{(H, H, H)\},\ \{(H, H, T)\},\ \{(T, T, T)\},\ \{(T, T, H)\},$$
then
$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(\{(H, H, H)\}) + \mathbb{P}(\{(H, H, T)\}) + \mathbb{P}(\{(T, T, T)\}) + \mathbb{P}(\{(T, T, H)\}) \\ &= p^3 + p^2(1-p) + (1-p)^3 + (1-p)^2 p = 2p^2 - 2p + 1.\end{aligned}$$

In Section 1.4 it is shown how to extend the probability measure (1.1) to Ω_∞ using Caratheódory's theorem.

- Let $f : \mathbb{R} \to [0, \infty)$ be a measurable function[1] such that

$$\int_{\mathbb{R}} f(x)\, dx = 1.$$

Then

$$\mathbb{P}(U) = \int_U f(x)\, dx, \tag{1.2}$$

defines a probability measure on $\mathcal{B}(\mathbb{R})$.

Remark 1.3 (Riemann vs. Lebesgue integral)**.** The integral in (1.2) must be understood in the Lebesgue sense, since we are integrating a general measurable function over a general Borel set. If f is sufficiently regular (say, continuous), and $U = (a, b) \subset \mathbb{R}$ is an interval, then the integral in (1.2) can be understood in the Riemann sense. Although the latter case is often sufficient in many applications (including in finance), all integrals in these notes should be understood in the Lebesgue sense, unless otherwise stated. The knowledge of Lebesgue integration theory is however not required for our purposes.

Exercise 1.8 (Sol. 3)**.** *Prove that* $\sum_{\omega\in\Omega_N} \mathbb{P}(\{\omega\}) = 1$*, where* $\mathbb{P}(\{\omega\})$ *is given by* (1.1).

Equivalent probability measures

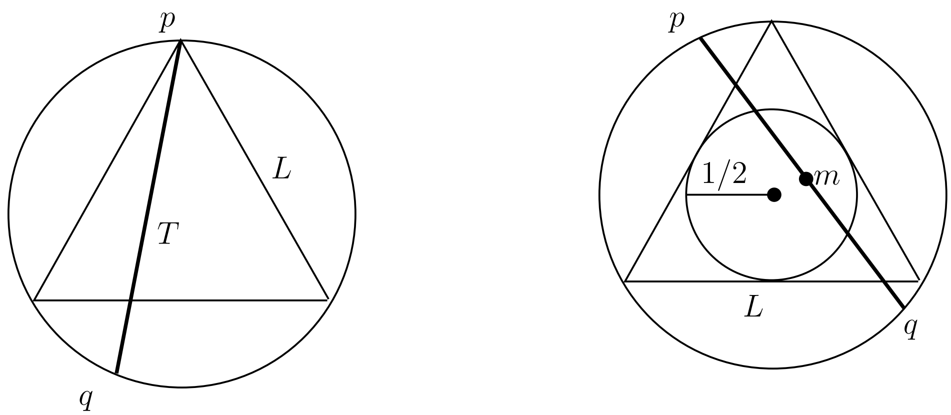
A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ and if we change one element of this triple we get a different probability space. The most interesting case is when a new probability measure is introduced. Let us first show with an example (known as **Bertrand's paradox**) that there might not be just one "reasonable" definition of probability measure associated to a given experiment. Suppose we perform an experiment whose result is a pair of points p, q on the unit circle C (e.g., throw two balls in a *roulette*). The sample space for this experiment is $\Omega = \{(p, q) : p, q \in C\}$. Let T be the length of the chord joining p and q. Now let L be the length of the side of a equilateral triangle inscribed in the circle C. Note that all such triangles are obtained one from another by a rotation around the center of the circle and all have the same sides length L. Consider the event $A = \{(p, q) \in \Omega : T > L\}$. What is a reasonable definition for $\mathbb{P}(A)$? From one hand we can suppose that one vertex of the triangle is p, and thus T will be greater than L if and only if the point q lies on the arch of the circle between the two vertexes of the triangle different from p, see Figure 1.1(a). Since the length of such arc is 1/3 the perimeter of the circle, then it is reasonable to define $\mathbb{P}(A) = 1/3$. On the other hand, it is simple to see that $T > L$ whenever the midpoint m of the chord lies within a circle of radius 1/2 concentric to C, see Figure 1.1(b). Since the area of the interior circle is 1/4 the area of C, we are led to define $\mathbb{P}(A) = 1/4$.

[1]See Section 2.1 for the definition of measurable function.

Whenever two probabilities are defined for the same experiment, it is of particular interest to determine whether they are equivalent, in the following sense.

Definition 1.5. *Given two probability spaces* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$*, the probability measures* $\mathbb{P}$ *and* $\widetilde{\mathbb{P}}$ *are said to be* **equivalent** *if* $\mathbb{P}(A) = 0 \Leftrightarrow \widetilde{\mathbb{P}}(A) = 0$.

Hence equivalent probability measures agree on which events are impossible. A complete characterization of the probability measures $\widetilde{\mathbb{P}}$ equivalent to a given $\mathbb{P}$ will be given in Theorem 3.3.



(a) $\mathbb{P}(A) = 1/3$

(b) $\mathbb{P}(A) = 1/4$

Figure 1.1: The Bertrand paradox. The length T of the cord $\overline{pq}$ is greater then L.

Conditional probability. Independent events

It might be that the occurrence of an event B makes the occurrence of another event A more or less likely. For instance, the probability of the event $A = \{$the first two tosses of a fair coin are both head$\}$ is $1/4$; however if know that the first toss is a tail, then $\mathbb{P}(A) = 0$, while $\mathbb{P}(A) = 1/2$ if we know that the first toss is a head. This leads to the important definition of conditional probability.

Definition 1.6. *Given two events* A, B *such that* $\mathbb{P}(B) > 0$*, the* **conditional probability** *of* A *given* B *is defined as*

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

To justify this definition, let $\mathcal{F}_B = \{A \cap B\}_{A \in \mathcal{F}}$, and set

$$\mathbb{P}_B(\cdot) = \mathbb{P}(\cdot|B).$$

Then $(B, \mathcal{F}_B, \mathbb{P}_B)$ is a probability space in which the events that cannot occur simultaneously with B are null events. Therefore it is natural to regard $(B, \mathcal{F}_B, \mathbb{P}_B)$ as the restriction of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ when B has occurred.

If $\mathbb{P}(A|B) = \mathbb{P}(A)$, the two events are said to be independent. The interpretation is the following: if two events A, B are independent, then the occurrence of the event B does not change the probability that A occurred. By Definition 1.6 we obtain the following equivalent characterization of independent events.

Definition 1.7. *Two events A, B are said to be* **independent** *if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. In general, the events $A_1, \dots, A_N$ ($N \geq 2$) are said to be independent if, for all $1 \leq k_1 < k_2 < \dots < k_m \leq N$, we have*

$$\mathbb{P}(A_{k_1} \cap \dots \cap A_{k_m}) = \prod_{j=1}^{m} \mathbb{P}(A_{k_j}).$$

Two σ-algebras $\mathcal{F}, \mathcal{G}$ are said to be independent if A and B are independent, for all $A \in \mathcal{G}$ and $B \in \mathcal{F}$. In general the σ-algebras $\mathcal{F}_1, \dots, \mathcal{F}_N$ ($N \geq 2$) are said to be independent if $A_1, A_2, \dots, A_N$ are independent events, for all $A_1 \in \mathcal{F}_1, \dots, A_N \in \mathcal{F}_N$.

The independence property of events is connected to the probability measure, i.e., two events may be independent in the probability $\mathbb{P}$ and *not* independent in the probability $\widetilde{\mathbb{P}}$, even if $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent. Moreover if $\mathcal{F}, \mathcal{G}$ are two independent σ-algebras and $A \in \mathcal{F} \cap \mathcal{G}$, then A is a trivial event. In fact, if $A \in \mathcal{F} \cap \mathcal{G}$, then $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$. Hence $\mathbb{P}(A) = 0$ or 1. The interpretation of this simple remark is that independent σ-algebras carry separate information.

Exercise 1.9 (Sol. 4)**.** *Given a fair coin and assuming N is odd, consider the following two events $A, B \in \Omega_N$:*

$$A = \text{“the number of heads is greater than the number of tails”},$$

$$B = \text{“The first toss is a head”}.$$

Use your intuition to guess whether the two events are independent; then verify your answer numerically (e.g., using Mathematica*).*

1.3 Filtered probability spaces

Consider again the N-coin tosses probability space. Let A_H be the event that the first toss is a head and A_T the event that it is a tail. Clearly $A_T = A_H^c$ and the σ-algebra $\mathcal{F}_1$ generated by the partition $\{A_H, A_T\}$ is $\mathcal{F}_1 = \{A_H, A_T, \Omega, \emptyset\}$. Now let A_{HH} be the event that the first 2 tosses are heads, and similarly define A_{HT}, A_{TH}, A_{TT}. These four events form a partition of Ω_N and they generate a σ-algebra $\mathcal{F}_2$ as indicated in Exercise 1.4. Clearly, $\mathcal{F}_1 \subset \mathcal{F}_2$. Going on with three tosses, four tosses, and so on, until we complete the N-toss, we construct a sequence

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_N = 2^{\Omega_N}$$

of σ-algebras. The σ-algebra $\mathcal{F}_k$ contains all the events of the experiment that depend on (i.e., which are resolved by) the first k tosses. The family $\{\mathcal{F}_k\}_{k=1,\dots,N}$ of σ-algebras is an example of filtration.

Definition 1.8. *A* **filtration** *is a one parameter family* $\{\mathcal{F}(t)\}_{t\geq 0}$ *of* σ*-algebras such that* $\mathcal{F}(t) \subseteq \mathcal{F}$ *for all* $t \geq 0$ *and* $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ *for all* $s \leq t$*. A quadruple* $(\Omega, \mathcal{F}, \{\mathcal{F}(t)\}_{t\geq 0}, \mathbb{P})$ *is called a* **filtered probability space***.*

In our applications t stands for the time variable and filtrations are associated to experiments in which "information accumulates with time". For instance, in the example given above, the more times we toss the coin, the higher is the number of events which are resolved by the experiment, i.e., the more information becomes accessible.

1.4 The "infinite-coin tosses" probability space

In this section we outline the construction of the probability space for the ∞-coin tosses experiment using Carathéodory's theorem. The sample space is

$$\Omega_\infty = \{\omega = (\gamma_n)_{n\in\mathbb{N}},\ \gamma_n \in \{H, T\}\}.$$

Let us show first that Ω is uncountable. We use the well-known **Cantor diagonal argument**. Suppose that Ω_∞ is countable and write

$$\Omega_\infty = \{\omega_k\}_{k\in\mathbb{N}}. \tag{1.3}$$

Each $\omega_k \in \Omega_\infty$ is a sequence of infinite tosses, which we write as $\omega_k = (\gamma_j^{(k)})_{j\in\mathbb{N}}$, where $\gamma_j^{(k)}$ is either H or T, for all $j \in \mathbb{N}$ and for each fixed $k \in \mathbb{N}$. Note that $(\gamma_j^{(k)})_{j,k\in\mathbb{N}}$ is an "$\infty \times \infty$" matrix. Now consider the ∞-toss corresponding to the diagonal of this matrix, that is

$$\bar{\omega} = (\bar{\gamma}_m)_{m\in\mathbb{N}}, \quad \bar{\gamma}_m = \gamma_m^{(m)}, \text{ for all } m \in \mathbb{N}.$$

Finally consider the ∞-toss ω which is obtained by changing each single toss of $\bar{\omega}$, that is to say

$$\omega = (\gamma_m)_{m\in\mathbb{N}}, \quad \text{where } \gamma_m = H \text{ if } \bar{\gamma}_m = T, \text{ and } \gamma_m = T \text{ if } \bar{\gamma}_m = H, \text{ for all } m \in \mathbb{N}.$$

It is clear that the ∞-toss ω does not belong to the set (1.3). In fact, by construction, the first toss of ω is different from the first toss of ω_1, the second toss of ω is different from the second toss of ω_2, ..., the n^{th} toss of ω is different from the n^{th} toss of ω_n, and so on, so that each ∞-toss in (1.3) is different from ω. We conclude that the elements of Ω_∞ cannot be listed as they were comprising a countable set.

Now, let $N \in \mathbb{N}$ and $\Omega_N = \{H, T\}^N$ be the sample space for the N-tosses experiment. For each $\omega^* = (\gamma_1^*, \dots, \gamma_N^*) \in \Omega_N$ we define the event $A_{\omega^*} \subset \Omega_\infty$ by

$$A_{\omega^*} = \{\omega = (\gamma_n)_{n\in\mathbb{N}} : \gamma_j = \gamma_j^*,\ j = 1, \dots, N\},$$

i.e., the event that the first N tosses in a ∞-toss be equal to $(\gamma_1^*, \dots, \gamma_N^*)$. Define the probability of this event as the probability of the N-toss ω^*, that is

$$\mathbb{P}_0(A_{\omega^*}) = p^{N_H(\omega^*)}(1-p)^{N_T(\omega^*)},$$

where $0 < p < 1$, $N_H(\omega^*)$ is the number of heads in the N-toss ω^* and $N_T(\omega^*) = N - N_H(\omega^*)$ is the number of tails in ω^*, see (1.1). Next consider the family of events

$$\mathcal{U}_N = \{A_{\omega^*}\}_{\omega^*\in\Omega_N} \subset 2^{\Omega_\infty}.$$

Note that $\mathcal{U}_N$ is, for each fixed $N \in \mathbb{N}$, a partition of Ω_∞. Hence the σ-algebra $\mathcal{F}_N = \mathcal{F}_{\mathcal{U}_N}$ is generated according to Exercise 1.4 and contains all events of Ω_∞ that are resolved by the first N tosses. Moreover $\mathcal{F}_N \subset \mathcal{F}_{N+1}$, that is to say, $\{\mathcal{F}_N\}_{N\in\mathbb{N}}$ is a filtration. Since $\mathbb{P}_0$ is defined for all $A_{\omega^*} \in \mathcal{U}_N$, then it can be extended uniquely to the entire $\mathcal{F}_N$, because each element $A \in \mathcal{F}_N$ is the disjoint union of events of $\mathcal{U}_N$ (see again Exercise 1.4) and therefore the probability of A can be inferred by the property (ii) in the definition of probability measure, see Definition 1.4. But then $\mathbb{P}_0$ extends uniquely to

$$\mathcal{F}_\infty = \bigcup_{N\in\mathbb{N}} \mathcal{F}_N.$$

Hence we have constructed a triple $(\Omega_\infty, \mathcal{F}_\infty, \mathbb{P}_0)$. Is this triple a probability space? The answer is *no*, because $\mathcal{F}_\infty$ is *not* a σ-algebra. To see this, let A_k be the event that the k^{th} toss in a infinite sequence of tosses is a head. Clearly $A_k \in \mathcal{F}_k$ for all k and therefore $\{A_k\}_{k\in\mathbb{N}} \subset \mathcal{F}_\infty$. Now assume that $\mathcal{F}_\infty$ is a σ-algebra. Then the event $A = \cup_k A_k$ would belong to $\mathcal{F}_\infty$ and therefore also $A^c \in \mathcal{F}_\infty$. The latter holds if and only if there exists $N \in \mathbb{N}$ such that $A^c \in \mathcal{F}_N$. But A^c is the event that all tosses are tails, which of course cannot be resolved by the information $\mathcal{F}_N$ accumulated after just N tosses. We conclude that $\mathcal{F}_\infty$ is not a σ-algebra. In particular, we have shown that $\mathcal{F}_\infty$ is not in general closed with respect to the countable union of its elements. However it is easy to show that $\mathcal{F}_\infty$ is closed with respect to the *finite* union of its elements, and in addition it satisfies the properties (i), (ii) in Definition 1.4. This set of properties makes $\mathcal{F}_\infty$ an **algebra**. To complete the construction of the probability space for the ∞-coin tosses experiment, we need the following deep result.

Theorem 1.1 (Carathéodory's theorem). *Let $\mathcal{U}$ be an algebra of subsets of Ω and $\mathbb{P}_0 : \mathcal{U} \to [0, 1]$ be a map satisfying $\mathbb{P}_0(\Omega) = 1$ and $\mathbb{P}_0(\cup_{i=1}^N A_i) = \sum_{i=1}^N \mathbb{P}_0(A_i)$, for every finite collection $\{A_1, \dots, A_N\} \subset \mathcal{U}$ of disjoint sets*[2]*. Then there exists a unique probability measure $\mathbb{P}$ on $\mathcal{F}_\mathcal{U}$ such that $\mathbb{P}(A) = \mathbb{P}_0(A)$, for all $A \in \mathcal{U}$.*

[2] $\mathbb{P}_0$ is called a **pre-measure**.

Hence the map $\mathbb{P}_0 : \mathcal{F}_\infty \to [0,1]$ defined above extends uniquely to a probability measure $\mathbb{P}$ defined on $\mathcal{F} = \mathcal{F}_{\mathcal{F}_\infty}$. The resulting triple $(\Omega_\infty, \mathcal{F}, \mathbb{P})$ defines the probability space for the ∞- coin tosses experiment.

Chapter 2

Random variables and stochastic processes

Throughout this chapter we assume that $(\Omega, \mathcal{F}, \{\mathcal{F}(t)\}_{t\geq 0}, \mathbb{P})$ is a given filtered probability space.

2.1 Random variables

In many applications of probability theory, and in financial mathematics in particular, one is more interested in knowing the value attained by quantities that depend on the outcome of the experiment, rather than knowing which specific events have occurred. Such quantities are called random variables.

Definition 2.1. *A map $X : \Omega \to \mathbb{R}$ is called a (real-valued)* **random variable** *if $\{X \in U\} \in \mathcal{F}$, for all $U \in \mathcal{B}(\mathbb{R})$, where $\{X \in U\} = \{\omega \in \Omega : X(\omega) \in U\}$ is the pre-image of the Borel set U. If there exists $c \in \mathbb{R}$ such that $X(\omega) = c$ almost surely, we say that X is* **a deterministic constant***.*

Occasionally we shall also need to consider complex-valued random variables. These are defined as the maps $Z : \Omega \to \mathbb{C}$ of the form $Z = X + iY$, where X, Y are real-valued random variables and i is the imaginary unit ($i^2 = -1$). Similarly a vector valued random variable $X = (X_1, \dots, X_N) : \Omega \to \mathbb{R}^N$ can be defined by simply requiring that each component $X_j : \Omega \to \mathbb{R}$ is a random variable in the sense of Definition 2.1.

Remark 2.1 (Notation)**.** A generic real-valued random variable will be denoted by X. If we need to consider two such random variables we will denote them by X, Y, while N real-valued random variables will be denoted by $X_1, \dots, X_N$. Equivalently $(X_1, \dots, X_N) : \Omega \to \mathbb{R}^N$ is a vector-valued random variable.

Remark 2.2. Equality among random variables is always understood to hold up to a null set. That is to say, $X = Y$ always means $X = Y$ almost surely (a.s.).

Random variables are also called **measurable functions**, but this terminology will be used in this text only when $\Omega = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}(R)$. Measurable functions will be denoted by small Latin letters (e.g., $f, g, \dots$). If X is a random variable and $Y = f(X)$ for some measurable function f, then Y is also a random variable. We denote $\mathbb{P}(X \in U) = \mathbb{P}(\{X \in U\})$ the probability that X takes value in $U \in \mathcal{B}(R)$. Moreover, given two random variables $X, Y : \Omega \to \mathbb{R}$ and the Borel sets U, V, we denote

$$\mathbb{P}(X \in U, Y \in V) = \mathbb{P}(\{X \in U\} \cap \{Y \in V\}),$$

which is the probability that the random variable X takes value in U and Y takes value in V. The generalization to an arbitrary number of random variables is straightforward.

As the value attained by X depends on the result of the experiment, random variables carry information, i.e., upon knowing the value attained by X we know something about the outcome ω of the experiment. For instance, if $X(\omega) = (-1)^\omega$, where ω is the result of tossing a dice, and if we are told that X takes value 1, then we infer immediately that the dice roll is even. The information carried by a random variable X forms the σ-algebra generated by X, whose precise definition is the following.

Definition 2.2. *Let $X : \Omega \to \mathbb{R}$ be a random variable. The* **σ-algebra generated by** X *is the collection $\sigma(X) \subseteq \mathcal{F}$ of events given by*

$$\sigma(X) = \{A \in \mathcal{F} : A = \{X \in U\},\ \textit{for some } U \in \mathcal{B}(\mathbb{R})\}.$$

If $\mathcal{G} \subseteq \mathcal{F}$ is another σ-algebra of subsets of Ω and $\sigma(X) \subseteq \mathcal{G}$, we say that X is **$\mathcal{G}$-measurable**. *If $Y : \Omega \to \mathbb{R}$ is another random variable and $\sigma(Y) \subseteq \sigma(X)$, we say that Y is* **X-measurable**

Exercise 2.1 (Sol. 5). *Prove that $\sigma(X)$ is indeed a σ-algebra, as claimed in Definition 2.2.*

Thus $\sigma(X)$ contains all the events that are resolved by knowing the value of X. The interpretation of X being $\mathcal{G}$-measurable is that the information contained in $\mathcal{G}$ suffices to determine the value taken by X in the experiment. Clearly the σ-algebra generated by a deterministic constant consists of trivial events only.

Definition 2.3. *The σ-algebra $\sigma(X, Y)$ generated by two random variables $X, Y : \Omega \to \mathbb{R}$ is the smallest σ-algebra containing $\sigma(X) \cup \sigma(Y)$, that is to say*[1] *$\sigma(X, Y) = \mathcal{F}_{\mathcal{O}}$, where $\mathcal{O} = \sigma(X) \cup \sigma(Y)$, and similarly for any number of random variables.*

If Y is X-measurable then $\sigma(X, Y) = \sigma(X)$, i.e., the random variable Y does not add any new information to the one already contained in X. Clearly, if $Y = f(X)$ for some measurable function f, then Y is X-measurable. It can be shown that the opposite is also true: if $\sigma(Y) \subseteq \sigma(X)$, then there exists a measurable function f such that $Y = f(X)$ (see Prop. 3 in [23]). The other extreme is when X and Y carry distinct information, i.e., when $\sigma(X) \cap \sigma(Y)$ consists of trivial events only. This occurs in particular when the two random variables are independent.

[1] See Definition 1.2.

Definition 2.4. *Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. We say that X is independent of $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent in the sense of Definition 1.7. Two random variables $X, Y : \Omega \to \mathbb{R}$ are said to be* **independent random variables** *if the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent. More generally, the random variables $X_1, \dots, X_N$ are independent if $\sigma(X_1), \dots, \sigma(X_N)$ are independent σ-algebras.*

In the intermediate case, i.e., when Y is neither X-measurable nor independent of X, it is expected that the knowledge on the value attained by X helps to derive information on the values attainable by Y. We shall study this case in the next chapter.

Exercise 2.2 (Sol. 6)**.** *Show that when X, Y are independent random variables, then $\sigma(X) \cap \sigma(Y)$ consists of trivial events only (i.e., events with probability zero or one). Show that two deterministic constants are always independent. Finally assume $Y = g(X)$ and show that in this case the two random variables are independent if and only if Y is a deterministic constant.*

Exercise 2.3. *Which of the following pairs of random variables $X, Y : \Omega_N \to \mathbb{R}$ are independent? (Use only the intuitive interpretation of independence and not the formal definition.)*

1. *$X(\omega) = N_T(\omega)$; $Y(\omega) = 1$ if the first toss is head, $Y(\omega) = 0$ otherwise.*
2. *$X(\omega) = 1$ if there exists at least a head in ω, $X(\omega) = 0$ otherwise; $Y(\omega) = 1$ if there exists exactly a head in ω, $Y(\omega) = 0$ otherwise.*
3. *$X(\omega) =$ number of times that a head is followed by a tail; $Y(\omega) = 1$ if there exist two consecutive tail in ω, $Y(\omega) = 0$ otherwise.*

The next theorem shows how to construct new independent random variables from a given sequence of independent random variables.

Theorem 2.1. *Let $X_1, \dots, X_N$ be independent random variables. Let us divide the set $\{X_1, \dots, X_N\}$ into m separate groups of random variables, namely, let*

$$\{X_1, \dots, X_N\} = \{X_{k_1}\}_{k_1 \in I_1} \cup \{X_{k_2}\}_{k_2 \in I_2} \cup \dots \cup \{X_{k_m}\}_{k_m \in I_m},$$

where $\{I_1, I_2, \dots I_m\}$ is a partition of $\{1, \dots, N\}$. Let n_i be the number of elements in the set I_i, so that $n_1 + n_2 + \dots + n_m = N$. Let $g_1, \dots, g_m$ be measurable functions such that $g_i : \mathbb{R}^{n_i} \to \mathbb{R}$. Then the random variables

$$Y_1 = g_1((X_{k_1})_{k_1 \in I_1}), \ Y_2 = g_2((X_{k_2})_{k_2 \in I_2}), \ \dots, Y_m = g_m((X_{k_m})_{k_m \in I_m})$$

are independent.

For instance, in the case of $N = 2$ independent random variables X_1, X_2, Theorem 2.1 asserts that $Y_1 = g(X_1)$ and $Y_2 = f(X_2)$ are independent random variables, for all measurable functions $f, g : \mathbb{R} \to \mathbb{R}$.

Exercise 2.4 (Sol. 7)**.** *Prove Theorem 2.1 for the case $N = 2$.*

Simple and discrete random variables

A special role is played by simple random variables. The simplest possible one is the **indicator function** of an event: Given $A \in \mathcal{F}$, the indicator function of A is the random variable that takes value 1 if $\omega \in A$ and 0 otherwise, i.e.,

$$\mathbb{I}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c. \end{cases}$$

Obviously, $\sigma(\mathbb{I}_A) = \{A, A^c, \emptyset, \Omega\}$.

Definition 2.5. *Let $N \in \mathbb{N}$, $\{A_k\}_{k=1,\dots,N} \subset \mathcal{F}$ be a (finite) partition of Ω and $a_1, \dots, a_N$ be distinct real numbers. The random variable*

$$X = \sum_{k=1}^{N} a_k \mathbb{I}_{A_k}$$

is called a **simple random variable**. *If $N \in \mathbb{N}$ is replaced by $N = \infty$ in this definition, we call X a* **discrete random** *variable.*

Thus a simple random variable X attains only a finite number of values, while a discrete random variable X attains countably infinite many values[2]. In both cases we have

$$\mathbb{P}(X = x) = \begin{cases} 0, & \text{if } x \notin \mathrm{Image}(X), \\ \mathbb{P}(A_k), & \text{if } x = a_k, \end{cases}$$

where $\mathrm{Image}(X) = \{x \in \mathbb{R} : X(\omega) = x, \text{ for some } \omega \in \Omega\}$ is the image of X. Moreover for a simple, or discrete, random variable X, $\sigma(X)$ is the σ-algebra generated by the partition $\{A_1, A_2, \dots\}$, which is constructed as stated in Exercise 1.4. Let us consider two examples of simple/discrete random variables that have applications in financial mathematics (and in many other fields).

A simple random variable X is called a **binomial** random variable if

- $\mathrm{Image}(X) = \{0, 1, \dots, N\}$;
- There exists $p \in (0, 1)$ such that $\mathbb{P}(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$, $k = 0, \dots, N$.

For instance, if we let X to be the number of heads in a N-toss, then X is binomial.

A discrete random variable X is called a **Poisson** variable if

- $\mathrm{Image}(X) = \mathbb{N} \cup \{0\}$;
- There exists $\mu > 0$ such that $\mathbb{P}(X = k) = \dfrac{\mu^k e^{-\mu}}{k!}$, $k = 0, 1, 2, \dots$

[2]Not all authors distinguish between simple and discrete random variables.

We denote by $\mathcal{P}(\mu)$ the set of all Poisson random variables with parameter $\mu > 0$.

The following important theorem shows that all non-negative random variables can be approximated by a sequence of simple random variables.

Theorem 2.2. *Let $X : \Omega \to [0, \infty)$ be a random variable and let $n \in \mathbb{N}$ be given. For $k = 0, 1, ..., n2^n - 1$, consider the sets*

$$A_{k,n} := \Big\{X \in \Big[\frac{k}{2^n}, \frac{k+1}{2^n}\Big)\Big\}$$

and for $k = n2^n$ let

$$A_{n2^n,n} = \{X \geq n\}.$$

(Note that $\{A_{k,n}\}_{k=0,\dots,n2^n}$ is a partition of Ω, for all fixed $n \in \mathbb{N}$.) Define the simple random variables

$$s_n^X(\omega) = \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbb{I}_{A_{k,n}}(\omega).$$

Then $0 \leq s_1^X(\omega) \leq s_2^X(\omega) \leq \cdots \leq s_n^X(\omega) \leq s_{n+1}^X(\omega) \leq \cdots \leq X(\omega)$, for all $\omega \in \Omega$ and

$$\lim_{n\to\infty} s_n^X(\omega) = X(\omega), \quad \text{for all } \omega \in \Omega.$$

(The limit exists because the sequence $\{s_n^X\}_{n\in\mathbb{N}}$ is non-decreasing.)

Exercise 2.5. *Prove Theorem 2.2.*

2.2 Distribution and probability density functions

Definition 2.6. *The* **(cumulative) distribution function** *of the random variable $X : \Omega \to \mathbb{R}$ is the non-negative function $F_X : \mathbb{R} \to [0, 1]$ given by $F_X(x) = \mathbb{P}(X \leq x)$. Two random variables X, Y are said to be* **identically distributed** *if $F_X = F_Y$.*

Exercise 2.6 (Sol. 8)**.** *Show that*

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

Show also that F_X is (1) right-continuous, (2) non-decreasing, (3) $\lim_{x\to+\infty} F_X(x) = 1$ and $\lim_{x\to-\infty} F_X(x) = 0$.

Exercise 2.7 (Sol. 9)**.** *Let $F : \mathbb{R} \to [0, 1]$ be a measurable function satisfying the properties (1)–(3) in Exercise 2.6. Show that there exists a probability space and a random variable X such that $F = F_X$.*

Definition 2.7. *A random variable $X : \Omega \to \mathbb{R}$ is said to admit the* **probability density function (pdf)** *$f_X : \mathbb{R} \to [0, \infty)$ if f_X is integrable on $\mathbb{R}$ and*

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy. \tag{2.1}$$

If f_X is the pdf of a random variable, then necessarily

$$\int_{\mathbb{R}} f_X(x)\,dx = \lim_{x\to\infty} F_X(x) = 1.$$

All probability density functions considered in these notes are almost everywhere continuous[3], and therefore the integral in (2.1) can be understood in the Riemann sense. Moreover in this case F_X is differentiable and we have

$$f_X = \frac{dF_X}{dx}.$$

If the integral in (2.1) is understood in the Lebesgue sense, then the density f_X can be a quite irregular function. In this case, the fundamental theorem of calculus for the Lebesgue integral entails that the distribution $F_X(x)$ satisfying (2.1) is absolutely continuous, and so in particular it is continuous. Conversely, if F_X is absolutely continuous, then X admits a density function.

We remark that, regardless of the notion of integral being used, a simple (or discrete) random variable X cannot admit a density in the sense of Definition 2.7. Suppose in fact that $X = \sum_{k=1}^N a_k \mathbb{I}_{A_k}$ is a simple random variable and assume $a_1 = \max(a_1, \dots, a_N)$. Then

$$\lim_{x\to a_1^-} F_X(x) = \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_N) < 1,$$

while

$$\lim_{x\to a_1^+} F_X(x) = 1 = F_X(a_1).$$

It follows that $F_X(x)$ is not continuous, and so in particular it cannot be written in the form (2.1). To define the pdf of the simple random variable $X = \sum_{k=1}^N a_k \mathbb{I}_{A_k}$, we observe first that its distribution function is

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{a_k \leq x} \mathbb{P}(X = a_k). \tag{2.2}$$

The probability density function $f_X(x)$ is defined as

$$f_X(x) = \begin{cases} \mathbb{P}(X = x), & \text{if } x = a_k \text{ for some } k \\ 0 & \text{otherwise} \end{cases}.$$

Thus with a slight abuse of notation we can rewrite (2.2) as

$$F_X(x) = \sum_{y \leq x} f_X(y), \tag{2.3}$$

[3] I.e., continuous everywhere except possibly in a real set of zero Lebesgue measure.

which extends (2.1) to simple random variables[4].

We shall see in the following chapters that when a random variable X admits a density, then all the relevant statistical information on X can be deduced by f_X. We also remark that often one can prove the existence of the pdf f_X without however being able to derive an explicit formula for it. For instance, f_X is often given as the solution of a partial differential equation, or through its (inverse) Fourier transform, which is called the characteristic function of X, see Section 3.1. Some examples of density functions, which have important applications in financial mathematics, are the following.

Examples of probability density functions

- A random variable $X : \Omega \to \mathbb{R}$ is said to be a **normal** (or **normally distributed**) random variable if it admits the density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-m)^2}{2\sigma^2}},$$

  for some $m \in \mathbb{R}$ and $\sigma > 0$, which are called respectively the **expectation** (or **mean**) and the **deviation** of the normal random variable X, while σ^2 is called the **variance** of X. The typical profile of a normal density function is shown in Figure 2.1(a). We denote by $\mathcal{N}(m, \sigma^2)$ the set of all normal random variables with expectation m and variance σ^2. If $m = 0$ and $\sigma^2 = 1$, $X \in \mathcal{N}(0,1)$ is said to be a **standard** normal random variable. The density function of standard normal random variables is denoted by ϕ, while their distribution is denoted by Φ, i.e.,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}}\, dy.$$

- A random variable $X : \Omega \to \mathbb{R}$ is said to be an **exponential** (or **exponentially distributed**) random variable if it admits the density

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{x \geq 0},$$

  for some $\lambda > 0$, which is called the **intensity** of the exponential random variable X. A typical profile is shown in Figure 2.1(b) . We denote by $\mathcal{E}(\lambda)$ the set of all exponential random variables with intensity $\lambda > 0$. The distribution function of an exponential random variable X with intensity λ is given by

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy = \lambda \int_{0}^{x} e^{-\lambda y}\, dy = 1 - e^{-\lambda x}.$$

[4] It is possible to unify the definition of pdf for continuum and discrete random variables by writing the sum (2.3) as an integral with respect to the Dirac measure, but we shall not do so.

- A random variable $X : \Omega \to \mathbb{R}$ is said to be **chi-squared distributed** if it admits the density

$$f_X(x) = \frac{x^{\delta/2-1}e^{-x/2}}{2^{\delta/2}\Gamma(\delta/2)}\mathbb{I}_{x>0},$$

  for some $\delta > 0$, which is called the **degree** of X. Here $\Gamma(t) = \int_0^\infty z^{t-1}e^{-z}dz$, $t > 0$, is the Gamma-function. Recall the relation

$$\Gamma(n) = (n-1)!$$

  for $n \in \mathbb{N}$. We denote by $\chi^2(\delta)$ the set of all chi-squared distributed random variables with degree δ. Three typical profiles of this density are shown in Figure 2.2(a).

- A random variable $X : \Omega \to \mathbb{R}$ is said to be **non-central chi-squared distributed** with **degree** $\delta > 0$ and **non-centrality parameter** $\beta > 0$ if it admits the density

$$f_X(x) = \frac{1}{2}e^{-\frac{x+\beta}{2}}\left(\frac{x}{\beta}\right)^{\frac{\delta}{4}-\frac{1}{2}} I_{\delta/2-1}(\sqrt{\beta x})\mathbb{I}_{x>0}, \tag{2.4}$$

  where $I_\nu(y)$ denotes the modified Bessel function of the first kind. We denote by $\chi^2(\delta,\beta)$ the random variables with density (2.4). It can be shown that $\chi^2(\delta,0) = \chi^2(\delta)$. Three typical profiles of the density (2.4) are shown in Figure 2.2(b).

- A random variable $X : \Omega \to \mathbb{R}$ is said to be **Cauchy distributed** if it admits the density

$$f_X(x) = \frac{\gamma}{\pi((x-x_0)^2+\gamma^2)}$$

  for $x_0 \in \mathbb{R}$ and $\gamma > 0$, called the **location** and the **scale** of X.

- A random variable $X : \Omega \to \mathbb{R}$ is said to be **Lévy distributed** if it admits the density

$$f_X(x) = \sqrt{\frac{c}{2\pi}}\frac{e^{-\frac{c}{2(x-x_0)}}}{(x-x_0)^{3/2}}\mathbb{I}_{x>x_0},$$

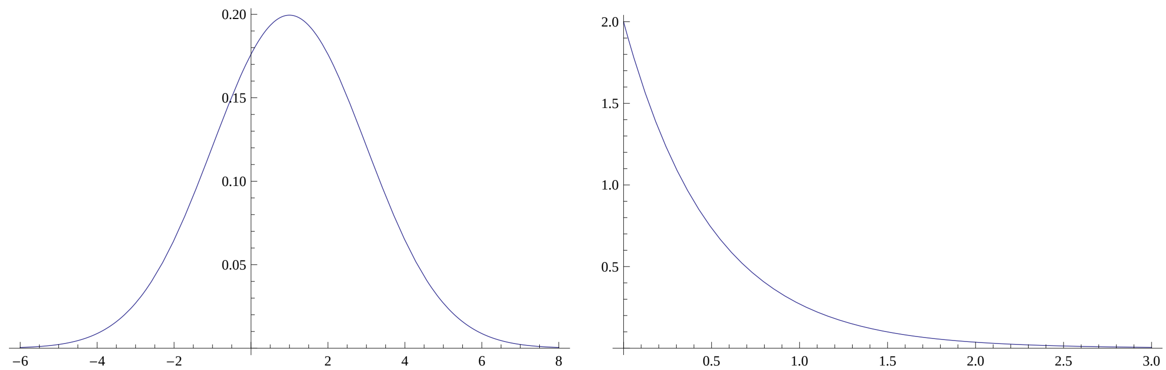
  for $x_0 \in \mathbb{R}$ and $c > 0$, called the **location** and the **scale** of X.

If a random variable X admits a density f_X, then for all (possibly unbounded) intervals $I \subseteq \mathbb{R}$ the result of Exercise 2.6 entails

$$\mathbb{P}(X \in I) = \int_I f_X(y)\,dy. \tag{2.5}$$

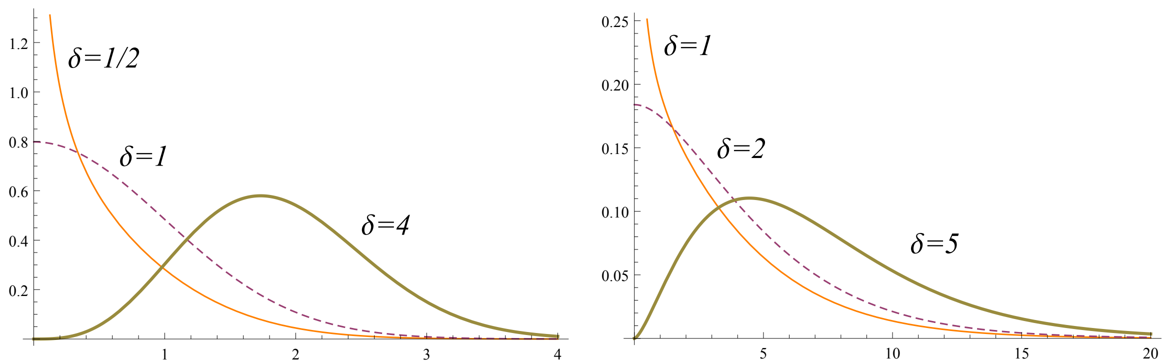
It can be shown that (2.5) extends to

$$\mathbb{P}(g(X) \in I) = \int_{x:g(x)\in I} f_X(x)\,dx, \tag{2.6}$$



(a) $X \in \mathcal{N}(1, 2)$ (b) $X \in \mathcal{E}(2)$

Figure 2.1: Densities of a normal random variable X and of an exponential random variable Y.



(a) $X \in \chi^2(\delta)$ (b) $X \in \chi^2(\delta, 2)$

Figure 2.2: Densities of (non-central) chi-squared random variables with different degree.

for all measurable functions $g : \mathbb{R} \to \mathbb{R}$. For example, if $X \in \mathcal{N}(0, 1)$, then

$$\mathbb{P}(X^2 \leq 1) = \mathbb{P}(-1 \leq X \leq 1) = \int_{-1}^{1} \phi(x)\, dx \approx 0.683,$$

which means that a standard normal random variable has about 68.3 % chances to take value in the interval $[-1, 1]$.

Exercise 2.8 (Sol. 10)**.** *Let $X \in \mathcal{N}(0, 1)$ and $Y = X^2$. Show that $Y \in \chi^2(1)$.*

Exercise 2.9. *Let $X \in \mathcal{N}(0, 1)$. Show that the random variable W defined by*

$$W = \begin{cases} 1/X^2 & \text{for } X \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

is Lévy distributed.

Exercise 2.10. *Let $X \in \mathcal{N}(m, \sigma^2)$ and $Y = X^2$. Show that*

$$f_Y(x) = \frac{\cosh(m\sqrt{x}/\sigma^2)}{\sqrt{2\pi x \sigma^2}} \exp\left(-\frac{x+m^2}{2\sigma^2}\right) \mathbb{I}_{x>0}.$$

Random variables with boundary values

Random variables in mathematical finance do not always admit a density in the classical sense described above (or in any other sense), and the purpose of this section is to present an example when one has to consider a generalized notion of density function. Suppose that X takes value on the semi-open interval $[0, \infty)$. Then clearly $F_X(x) = 0$ for $x < 0$, $F_X(0) = \mathbb{P}(X = 0)$, while for $x > 0$ we can write

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(0 \leq X \leq x) = \mathbb{P}(X = 0) + \mathbb{P}(0 < X \leq x).$$

Now assume that F_X is differentiable on the open set $x \in (0, \infty)$. Then there exists a function $f_X^+(x)$, $x > 0$, such that $F_X(x) - F_X(0) = \int_0^x f_X^+(t)\, dt$. Hence, for all $x \in \mathbb{R}$ we find

$$F_X(x) = p_0 H(x) + \int_{-\infty}^x f_X^+(t)\mathbb{I}_{t>0}\, dt,$$

where $p_0 = \mathbb{P}(X = 0)$ and $H(x)$ is the Heaviside function, i.e., $H(x) = 1$ if $x \geq 0$, $H(x) = 0$ if $x < 0$. By introducing the delta-distribution through the formal identity

$$H'(x) = \delta(x) \tag{2.7}$$

then we obtain, again formally, the following expression for the density function

$$f_X(x) = \frac{dF_X(x)}{dx} = p_0\delta(x) + f_X^+(x). \tag{2.8}$$

The formal identities (2.7)-(2.8) become rigorous mathematical expressions when they are understood in the sense of distributions. We shall refer to the term $p_0\delta(x)$ as the **discrete part** of the density and to the function f_X^+ as the **continuum part** (sometimes also called **defective density**). Note that

$$\int_0^\infty f_X^+(x)\, dx = 1 - p_0.$$

Hence f_+ is the actual pdf of X if and only if $p_0 = 0$.

The typical example of financial random variable whose pdf may have a discrete part is the stock price $S(t)$ at time t. For simple models (such us the geometric Browniam motion (2.14) defined in Section 2.4 below), the stock price is strictly positive a.s. at all finite times and the density has no discrete part. However for more sophisticated models the stock price can reach zero with positive probability at any finite time and so the pdf of the stock price admits a discrete part $\mathbb{P}(S(t) = 0)\delta(x)$, which is the probability of **default** of the stock. We shall see an example in Section 6.6.

Joint distribution

Definition 2.8. *The* **joint (cumulative) distribution** $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ *of two random variables* $X, Y : \Omega \to \mathbb{R}$ *is defined as*

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

It can be shown that two random variables are independent if and only if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$. In Theorem 2.3 below we prove a special case of this result assuming that the two random variables admit a joint pdf, defined as follows.

Definition 2.9. *The random variables* X, Y *are said to admit the* **joint (probability) density function** $f_{X,Y} : \mathbb{R}^2 \to [0,\infty)$ *if* $f_{X,Y}$ *is integrable in* $\mathbb{R}^2$ *and*

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

The joint density and distribution satisfy the formal identities

$$f_{X,Y} = \frac{\partial^2 F_{X,Y}}{\partial x\, \partial y}, \qquad \int_{\mathbb{R}^2} f_{X,Y}(x,y)\, dx\, dy = 1.$$

Moreover, if two random variables X, Y admit a joint density $f_{X,Y}$, then each of them admits a density (called **marginal density** in this context) which is given by

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy, \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx.$$

To see this we write

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \in \mathbb{R}) = \int_{-\infty}^{x}\int_{\mathbb{R}} f_{X,Y}(\eta,\xi)\, d\eta\, d\xi = \int_{-\infty}^{x} f_X(\eta)\, d\eta$$

and similarly for the random variable Y. If $W = g(X,Y)$, for some measurable function g, and $I \subseteq \mathbb{R}$ is an interval, the analogue of (2.6) in 2 dimensions holds, namely:

$$\mathbb{P}(g(X,Y) \in I) = \int_{x,y:g(x,y)\in I} f_{X,Y}(x,y)\, dx\, dy.$$

Example. Let $m = (m_1\ m_2)$ be a two dimensional row vector and $C = (C_{ij})_{i,j=1,2}$ be a $2{\times}2$ positive definite, symmetric matrix. Two random variables $X_1, X_2 : \Omega \to \mathbb{R}$ are said to be **jointly normally distributed** with **mean** m and **covariance matrix** C if they admit the joint density

$$f_{X_1,X_2}(x) = \frac{1}{\sqrt{(2\pi)^2 \det C}} \exp\left[-\frac{1}{2}(x-m)C^{-1}(x-m)^T\right], \tag{2.9}$$

where $x = (x_1\ x_2)$, C^{-1} is the inverse matrix of C and v^T is the transpose of the vector v. We denote by $\mathcal{N}(m, C)$ the set of jointly normally distributed random variables $X = (X_1, X_2)$ with mean $m \in \mathbb{R}^2$ and covariance matrix C.

Exercise 2.11 (Sol. 11)**.** *Show that two random variables X_1, X_2 are jointly normally distributed if and only if*

$$f_{X_1,X_2}(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times$$
$$\times \exp\left(-\frac{1}{2(1-\rho^2)}\Big[\frac{(x_1-m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\Big]\right), \quad (2.10)$$

where

$$\sigma_1^2 = C_{11}, \quad \sigma_2^2 = C_{22}, \quad \rho = \frac{C_{12}}{\sigma_1\sigma_2}.$$

Moreover show that $(X_1, X_2) \in \mathcal{N}(m, C)$ implies $X_1 \in \mathcal{N}(m_1, \sigma_1^2)$, $X_2 \in \mathcal{N}(m_2, \sigma_2^2)$.

By the previous exercise, when $\sigma_1 = \sigma_2 = 1$ and $m = (0\ 0)$, each random variable X_1, X_2 is a standard normal random variable. We denote by $\phi(x_1, x_2; \rho)$ the joint normal density in this case and call it the (2-dimensional) **standard joint normal density with correlation coefficient** ρ:

$$\phi(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right). \quad (2.11)$$

The corresponding cumulative distribution is given by

$$\Phi(x_1, x_2; \rho) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \phi(y_1, y_2; \rho)\, dy_1\, dy_2. \quad (2.12)$$

In the next theorem we establish a simple condition for the independence of two random variables which admit a joint density.

Theorem 2.3. *The following holds.*

(i) If two random variables X, Y admit the densities f_X, f_Y and are independent, then they admit the joint density

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

(ii) If two random variables X, Y admit a joint density $f_{X,Y}$ of the form

$$f_{X,Y}(x, y) = u(x)v(y),$$

for some functions $u, v : \mathbb{R} \to [0, \infty)$, then X, Y are independent and admit the densities f_X, f_Y given by

$$f_X(x) = cu(x), \quad f_Y(y) = \frac{1}{c} v(y),$$

where

$$c = \int_{\mathbb{R}} v(x)\, dx = \Big(\int_{\mathbb{R}} u(y)\, dy\Big)^{-1}.$$

Proof. As to (i) we have

$$\begin{aligned} F_{X,Y}(x,y) &= \mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x)\mathbb{P}(Y \le y) \\ &= \int_{-\infty}^{x} f_X(\eta)\, d\eta \int_{-\infty}^{y} f_Y(\xi)\, d\xi \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(\eta) f_Y(\xi)\, d\eta\, d\xi. \end{aligned}$$

To prove (ii), we first write

$$\{X \le x\} = \{X \le x\} \cap \Omega = \{X \le x\} \cap \{Y \le \mathbb{R}\} = \{X \le x, Y \le \mathbb{R}\}.$$

Hence,

$$\mathbb{P}(X \le x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(\eta, y)\, dy\, d\eta = \int_{-\infty}^{x} u(\eta)\, d\eta \int_{\mathbb{R}} v(y)\, dy = \int_{-\infty}^{x} cu(\eta)\, d\eta,$$

where $c = \int_{\mathbb{R}} v(y)\, dy$. Thus X admits the density $f_X(x) = c\, u(x)$. At the same fashion one proves that Y admits the density $f_Y(y) = c' v(y)$, where $c' = \int_{\mathbb{R}} u(x) dx$. Since

$$1 = \int_{\mathbb{R}}\int_{\mathbb{R}} f_{X,Y}(x,y)\, dx\, dy = \int_{\mathbb{R}} u(x)\, dx \int_{\mathbb{R}} v(y)\, dy = c'c,$$

then $c' = 1/c$. It remains to prove that X, Y are independent. This follows by

$$\begin{aligned} \mathbb{P}(X \in U, Y \in V) &= \int_U \int_V f_{X,Y}(x,y)\, dx\, dy = \int_U u(x)\, dx \int_V v(y)\, dy \\ &= \int_U cu(x)\, dx \int_V \frac{1}{c} v(y)\, dy = \int_U f_X(x)\, dx \int_V f_Y(y)\, dy \\ &= \mathbb{P}(X \in U)\mathbb{P}(Y \in V), \quad \text{for all } U, V \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

□

Remark 2.3. By Theorem 2.3 and the result of Exercise 2.11, we have that two jointly normally distributed random variables are independent if and only if $\rho = 0$ in the formula (2.10).

Exercise 2.12 (Sol. 12). *Let $X \in \mathcal{N}(0,1)$ and $Y \in \mathcal{E}(1)$ be independent. Compute $\mathbb{P}(X \le Y)$.*

Exercise 2.13. *Let $X \in \mathcal{E}(2)$, $Y \in \chi^2(3)$ be independent. Compute numerically (e.g., using* Mathematica*) the following probability*

$$\mathbb{P}(\log(1 + XY) < 2).$$

Result:$\approx$ 0.893.

In Exercise 3.23 we give another criterion to establish whether two random variables are independent, which applies also when the random variables do not admit a density.

2.3 Stochastic processes

A **stochastic process** is a one-parameter family of random variables, which we denote by $\{X(t)\}_{t\geq 0}$, or by $\{X(t)\}_{t\in[0,T]}$ if the parameter t is restricted to the interval $[0,T]$, $T>0$. Hence, for each $t\geq 0$, $X(t):\Omega\to\mathbb{R}$ is a random variable. We denote by $X(t,\omega)$ the value of $X(t)$ on the sample point $\omega\in\Omega$, i.e., $X(t,\omega)=X(t)(\omega)$. For each $\omega\in\Omega$ fixed, the curve $\gamma_X^\omega:\mathbb{R}\to\mathbb{R}$, $\gamma_X^\omega(t)=X(t,\omega)$ is called the **ω-path** of the stochastic process and is assumed to be a measurable function. If the paths of a stochastic process are all almost surely equal, we say that the stochastic process is a **deterministic function of time**.

The parameter t will be referred to as **time** parameter, since this is what it represents in the applications in financial mathematics. Examples of stochastic processes in financial mathematics are given in the next section.

Definition 2.10. *Two stochastic processes $\{X(t)\}_{t\geq 0}$, $\{Y(t)\}_{t\geq 0}$ are said to be independent if for all $m,n\in\mathbb{N}$ and $0\leq t_1<t_2<\dots<t_n$, $0\leq s_1<s_2<\dots<s_m$, the σ-algebras $\sigma(X(t_1),\dots,X(t_n))$, $\sigma(Y(s_1),\dots,Y(s_m))$ are independent.*

Hence two stochastic processes $\{X(t)\}_{t\geq 0}$, $\{Y(t)\}_{t\geq 0}$ are independent if the information obtained by "looking" at the process $\{X(t)\}_{t\geq 0}$ up to time T is independent of the information obtained by "looking" at the process $\{Y(t)\}_{t\geq 0}$ up to time S, for all $S,T>0$. Similarly one defines the notion of several independent stochastic processes.

Remark 2.4 (Notation)**.** If t runs over a countable set, i.e., $t\in\{t_k\}_{k\in\mathbb{N}}$, then a stochastic process is equivalent to a sequence of random variables $X_1,X_2,\dots$, where $X_k=X(t_k)$. In this case we say that the stochastic process is **discrete** and we denote it by $\{X_k\}_{k\in\mathbb{N}}$. An example of discrete stochastic process is the random walk defined below.

A special role is played by **step** processes: given $0=t_0<t_1<t_2<\dots$, a step process is a stochastic process $\{\Delta(t)\}_{t\geq 0}$ of the form

$$\Delta(t,\omega)=\sum_{k=0}^{\infty}X_k(\omega)\mathbb{I}_{[t_k,t_{k+1})}.$$

A typical path of a step process is depicted in Figure 2.3. Note that the paths of a step process are right-continuous, but in general they are not left-continuous. Moreover, since $X_k(\omega)=\Delta(t_k,\omega)$, we can rewrite $\Delta(t)$ as

$$\Delta(t)=\sum_{k}^{\infty}\Delta(t_k)\mathbb{I}_{[t_k,t_{k+1})}.$$

It will be shown in Theorem 4.2 that any sufficiently regular stochastic process can be approximated, in a suitable sense, by a sequence of step processes.

In the same way as a random variable generates a σ-algebra, a stochastic process generates a filtration. Informally, the filtration generated by the stochastic process $\{X(t)\}_{t\geq 0}$ contains the information accumulated by looking at the process for longer and longer periods of time.

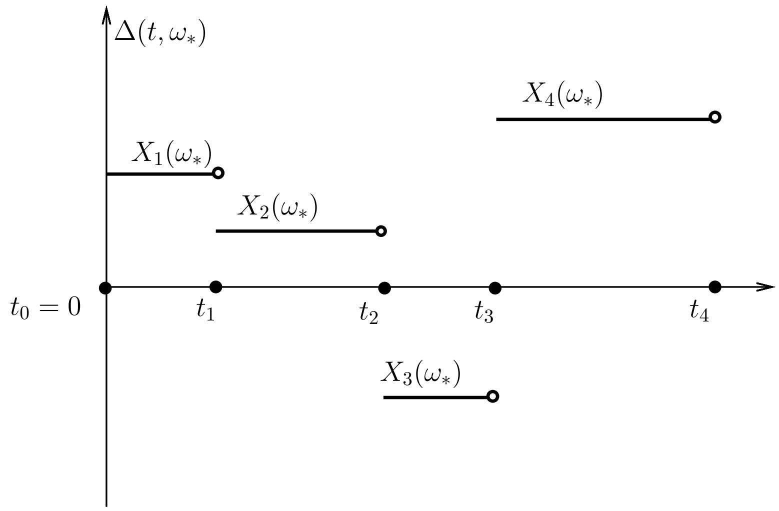


Figure 2.3: The path $\omega = \omega_*$ of a step process.

Definition 2.11. *The filtration generated by the stochastic process $\{X(t)\}_{t\geq 0}$ is given by $\{\mathcal{F}_X(t)\}_{t\geq 0}$, where*

$$\mathcal{F}_X(t) = \mathcal{F}_{\mathcal{O}(t)}, \quad \mathcal{O}(t) = \cup_{0\leq s\leq t}\sigma(X(s)).$$

Hence $\mathcal{F}_X(t)$ is the smallest σ-algebra containing $\sigma(X(s))$, for all $0 \leq s \leq t$, see Definition 1.2. Similarly one defines the filtration $\{\mathcal{F}_{X,Y}(t)\}_{t\geq 0}$ generated by two stochastic processes $\{X(t)\}_{t\geq 0}$, $\{Y(t)\}_{t\geq 0}$, as well as the filtration generated by any number of stochastic processes.

Definition 2.12. *If $\{\mathcal{F}(t)\}_{t\geq 0}$ is a filtration and $\mathcal{F}_X(t) \subseteq \mathcal{F}(t)$, for all $t \geq 0$, we say that the stochastic process $\{X(t)\}_{t\geq 0}$ is* **adapted** *to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$.*

The property of $\{X(t)\}_{t\geq 0}$ being adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$ means that the information contained in $\mathcal{F}(t)$ suffices to determine the value attained by the random variable $X(s)$, for all $s \in [0,t]$. Clearly, $\{X(t)\}_{t\geq 0}$ is adapted to its own generated filtration $\{\mathcal{F}_X(t)\}_{t\geq 0}$. Moreover if $\{X(t)\}_{t\geq 0}$ is adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$ and $Y(t) = f(X(t))$, for some measurable function f, then $\{Y(t)\}_{t\geq 0}$ is also adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$.

Next we give an example of (discrete) stochastic process. Let $\{X_t\}_{t\in\mathbb{N}}$ be a sequence of independent and identically distributed (**i.i.d**) random variables satisfying

$$X_t = 1 \quad \text{with probability } p, \qquad X_t = -1 \quad \text{with probability } 1-p,$$

for all $t \in \mathbb{N}$ and some $p \in (0,1)$. For a concrete realization of these random variables, we may think of X_t as being defined on the sample space Ω_∞ of the ∞-coin tosses experiment (see Section 1.4). In fact, letting $\omega = (\gamma_j)_{j\in\mathbb{N}} \in \Omega_\infty$, we may set

$$X_t(\omega) = \begin{cases} -1, & \text{if } \gamma_t = H, \\ 1, & \text{if } \gamma_t = T. \end{cases}$$

Hence $X_t : \Omega \to \{-1, 1\}$ is the simple random variable $X_t(\omega) = \mathbb{I}_{A_t} - \mathbb{I}_{A_t^c}$, where $A_t = \{\omega \in \Omega_\infty : \gamma_t = H\}$. Clearly, $\mathcal{F}_X(t)$ is the collection of all the events that are resolved by the first t-tosses, which is given as indicated at the beginning of Section 1.3.

Definition 2.13. *The stochastic process $\{M_t\}_{t\in\mathbb{N}}$ given by*

$$M_0 = 0, \quad M_t = \sum_{k=1}^{t} X_k,$$

is called **random walk**. *For $p = 1/2$, the random walk is said to be* **symmetric**.

To understand the meaning of the term "random walk", consider a particle moving on the real line in the following way: if $X_t = 1$ (i.e., if the toss number t is a head), at time t the particle moves one unit of length to the right, if $X_t = -1$ (i.e., if the toss number t is a head) it moves one unit of length to the left. Then M_t gives the total amount of units of length that the particle has travelled to the right or to the left up to time t.

Exercise 2.14. *Which of the following holds?*

$$\mathcal{F}_M(t) \subset \mathcal{F}_X(t), \quad \mathcal{F}_M(t) = \mathcal{F}_X(t), \quad \mathcal{F}_X(t) \subset \mathcal{F}_M(t).$$

Justify the answer.

The **increments** of the random walk are defined as follows. If $(k_1, \dots, k_N) \in \mathbb{N}^N$, such that $1 \leq k_1 < k_2 < \cdots < k_N$, we set

$$\Delta_1 = M_{k_1} - M_0 = M_{k_1}, \quad \Delta_2 = M_{k_2} - M_{k_1}, \dots, \quad \Delta_N = M_{k_N} - M_{k_{N-1}}.$$

Hence Δ_j is the total displacement of the particle from time k_{j-1} to time k_j.

Theorem 2.4. *The increments $\Delta_1, \dots, \Delta_N$ of the random walk are independent random variables.*

Proof. Since

$$\begin{aligned}
\Delta_1 &= X_1 + \cdots + X_{k_1} = g_1(X_1, \dots, X_{k_1}), \\
\Delta_2 &= X_{k_1+1} + \cdots + X_{k_2} = g_2(X_{k_1+1}, \dots, X_{k_2}), \\
&. \\
&. \\
\Delta_N &= X_{k_{N-1}+1} + \cdots + X_{k_N} = g_N(X_{k_{N-1}+1}, \dots, X_{k_N}),
\end{aligned}$$

the result follows by Theorem 2.1. $\square$

The interpretation of this result is that the particle has no memory of past movements: the distance travelled by the particle in a given interval of time is not affected by the motion of the particle at earlier times.

We may now define the most important of all stochastic processes.

Definition 2.14. *A* **Brownian motion** *(or* **Wiener process***) is a stochastic process $\{W(t)\}_{t\geq 0}$ such that*

(i) The paths are continuous and start from 0 *almost surely, i.e., the sample points* $\omega \in \Omega$ *such that* $\gamma_W^\omega(0) = 0$ *and* γ_W^ω *is a continuous function comprise a set of probability 1;*

(ii) The increments over disjoint time intervals are independent, i.e., for all $0 = t_0 < t_1 < \cdots < t_m$*, the random variables*

$$W(t_1) - W(t_0),\ W(t_2) - W(t_1),\ \ldots,\ W(t_m) - W(t_{m-1})$$

are independent;

(iii) For all $s < t$*, the increment* $W(t) - W(s)$ *belongs to* $\mathcal{N}(0, t-s)$*.*

Remark 2.5. The properties defining a Brownian motion depend on the probability measure $\mathbb{P}$. Thus a stochastic process may be a Brownian motion relative to a probability measure $\mathbb{P}$ and not a Brownian motion with respect to another (possibly equivalent) probability measure $\widetilde{\mathbb{P}}$. If we want to emphasize the probability measure $\mathbb{P}$ with respect to which a stochastic process is a Brownian motion we shall say that it is a $\mathbb{P}$-Brownian motion.

It can be shown that Brownian motions exist. In particular, it can be shown that the sequence of stochastic processes $\{W_n(t)\}_{t\geq 0}$, $n \in \mathbb{N}$, defined by

$$W_n(t) = \frac{1}{\sqrt{n}} M_{[nt]}, \tag{2.13}$$

where M_t is the symmetric random walk and $[z]$ denotes the integer part of z, converges (in distribution) to a Brownian motion. Therefore one may think of a Brownian motion as a time-continuum version of a symmetric random walk which runs for an infinite number of "infinitesimal time steps". In fact, provided the number of time steps is sufficiently large, the process $\{W_n(t)\}_{t\geq 0}$ gives a very good approximation of a Brownian motion, which is useful for numerical computations. An example of path to the stochastic process $\{W_n(t)\}_{t\geq 0}$, for $n = 1000$, is shown in Figure 2.4.

Once a Brownian motion is introduced it is natural to require that the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ be somehow related to it. For our financial applications in Chapter 6, the following class of filtrations will play a fundamental role.

Definition 2.15. *Let* $\{W(t)\}_{t\geq 0}$ *be a Brownian motion and denote by* $\sigma^+(W(t))$ *the* σ-

algebra generated by the increments $\{W(s) - W(t); s \geq t\}$, that is

$$\sigma^+(W(t)) = \mathcal{F}_{O(t)}, \ \mathcal{O}(t) = \cup_{s\geq t}\sigma(W(s) - W(t)).$$

A filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ is said to be a **non-anticipating** *filtration for the Brownian motion $\{W(t)\}_{t\geq 0}$ if $\{W(t)\}_{t\geq 0}$ is adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$ and if the σ-algebras $\sigma^+(W(t))$, $\mathcal{F}(t)$ are independent for all $t \geq 0$.*

The meaning is the following: the increments of the Brownian motion after time t are independent of the information available at time t in the σ-algebra $\mathcal{F}(t)$. Clearly $\{\mathcal{F}_W(t)\}_{t\geq 0}$ is a non-anticipating filtration for $\{W(t)\}_{t\geq 0}$. We shall see later that many properties of Brownian motions that depend on $\{\mathcal{F}_W(t)\}_{t\geq 0}$ also holds with respect to any non-anticipating filtration (e.g., the martingale property, see Section 3.4).

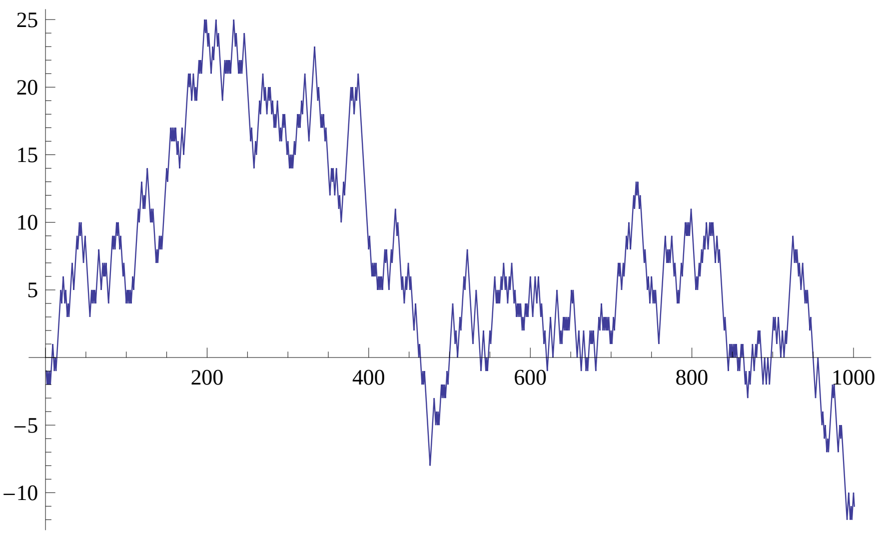


Figure 2.4: A path of the stochastic process (2.13) for $n = 1000$.

Another important example of stochastic process used in financial mathematics is the following.

Definition 2.16. *A* **Poisson process** *with* **rate** λ *is a stochastic process $\{N(t)\}_{t\geq 0}$ such that*

(i) $N(0) = 0$ *a.s.;*

(ii) The increments over disjoint time-intervals are independent;

(iii) For all $s < t$, the increment $N(t) - N(s)$ belongs to $\mathcal{P}(\lambda(t - s))$.

Note in particular that $N(t)$ is a discrete random variable, for all $t \geq 0$, and that, in contrast to the Brownian motion, the paths of a Poisson process are not continuous. The Poisson process is the building block to construct more general stochastic processes with jumps, which are popular nowadays as models for the price of certain financial assets, see [4].

2.4 Stochastic processes in financial mathematics

Remark 2.6. More information on the financial concepts introduced in this section can be found in the text *Basic financial concepts*, available on the course homepage. Unless otherwise stated, it is assumed throughout these notes that *assets pay no dividend.*

All variables in financial mathematics are represented by stochastic processes. The most obvious example is the **price** of financial assets. The stochastic process representing the price per share of a generic asset at different times will be denoted by $\{\Pi(t)\}_{t\geq 0}$. Depending on the type of asset considered, we use a different specific notation for the stochastic process modeling its price.

Remark 2.7. We always assume that $t = 0$ is earlier or equal to the present time. In particular, the value of all financial variables is known at time $t = 0$. Hence, if $\{X(t)\}_{t\geq 0}$ is a stochastic process modelling a financial variable, then $X(0)$ is a deterministic constant.

Stock price

The price per share a time t of a stock will be denoted by $S(t)$. Typically $S(t) > 0$, for all $t \geq 0$, however, as discussed in Section 2.2, some models allow for the possibility that $S(t) = 0$ with positive probability at finite times $t > 0$ (risk of default). Clearly $\{S(t)\}_{t\geq 0}$ is a stochastic process. If we have several stocks, we shall denote their price by $\{S_1(t)\}_{t\geq 0}$, $\{S_2(t)\}_{t\geq 0}$, etc.

A popular model for the price of stocks is the **geometric Brownian motion** stochastic process, which is given by

$$S(t) = S(0)\exp(\alpha t + \sigma W(t)). \tag{2.14}$$

Here $\{W(t)\}_{t\geq 0}$ is a Brownian motion, $\alpha \in \mathbb{R}$ is called the **mean of log return** (or **log-drift**) of the stock, while $\sigma > 0$ is called the **volatility** of the stock (α and σ are constant parameters in this model). Moreover, $S(0)$ is the price at time $t = 0$ of the stock, which, according to Remark 2.7, is a deterministic constant. The interpretation of the parameters α, σ is the following: If α is positive (resp. negative), the stock price as a tendency to increase (resp. decrease), while the larger is σ^2, the more wildly the stock price oscillates in time. In Chapter 4 we introduce a generalization of the geometric Brownian motion, in which the mean of log-return and the volatility of the stock are stochastic processes $\{\alpha(t)\}_{t\geq 0}$, $\{\sigma(t)\}_{t\geq 0}$ (generalized geometric Brownian motion).

Exercise 2.15 (Sol. 13)**.** *Derive the density of the geometric Brownian motion* (2.14) *and use the result to show that* $\mathbb{P}(S(t) = 0) = 0$*, i.e., a stock whose price is described by a geometric Brownian motion cannot default.*

Financial derivative

A **financial derivative** (or derivative security) is a contract whose value depends on the performance of one (or more) other asset(s), which is called the **underlying asset**. There exist various types of financial derivatives, the most common being options, futures, forwards and swaps. Financial derivatives can be traded **over the counter** (OTC), or in a regularized **exchange market**. In the former case, the contract is stipulated between two individual investors, who agree upon the conditions and the price of the contract. In particular, the same derivative (on the same asset, with the same parameters) can have two different prices over the counter. Derivatives traded in the market, on the contrary, are standardized contracts. Anyone, after a proper authorization, can make offers to buy or sell derivatives in the market, in a way much similar to how stocks are traded. Let us see some examples of financial derivatives (we shall introduce more in Chapter 6).

A **call option** is a contract between two parties, the buyer (or **owner**) of the call and the seller (or **writer**) of the call. The contract gives to the buyer the right, but not the obligation, to buy the underlying asset at some future time for a price agreed upon today, which is called **strike price** of the call. If the buyer can exercise this option only at some given time $t = T > 0$ (where $t = 0$ corresponds to the time at which the contract is stipulated) then the call option is called **European**, while if the option can be exercised at any time in the interval $(0, T]$, then the option is called **American**. The time $T > 0$ is called **maturity time**, or **expiration date** of the call. The seller of the call is obliged to sell the asset to the buyer (at the strike price) if the latter decides to exercise the option. If the option to buy in the definition of a call is replaced by the option to sell, then the option is called a **put option**.

In exchange for the option, the buyer must pay a **premium** to the seller. Suppose that the option is a European option with strike price K, maturity time T and premium Π_0 on a stock with price $S(t)$ at time t. In which case is it then convenient for the buyer to exercise the call? Let us define the **payoff** of a European call as

$$Y = (S(T) - K)_+ := \max(0, S(T) - K) \quad \text{(call)};$$

similarly for a European put we set

$$Y = (K - S(T))_+ \quad \text{(put)}.$$

Note that Y is a random variable, because it depends on the random variable $S(T)$. Clearly, if $Y > 0$ it is more convenient for the buyer to exercise the option rather than buying/selling the asset on the market. Note however that the real **profit** for the buyer is given by $N(Y - \Pi_0)$, where N is the number of option contracts owned by the buyer.

Let us introduce some further terminology. A call (resp. put) is said to be **in the money** at time t if $S(t) > K$ (resp. $S(t) < K$). The call (resp. put) is said to be **out of the money** if $S(t) < K$ (resp. $S(t) > K$). If $S(t) = K$, the (call or put) option is said to be **at the money** at time t. The meaning of this terminology is self-explanatory.

European call and put options are examples of more general contracts called **European derivatives**. Given a function $g : (0,\infty) \to \mathbb{R}$, a **standard European derivative** with pay-off $Y = g(S(T))$ and maturity time $T > 0$ is a contract that pays to its owner the amount Y at time $T > 0$. Here $S(T)$ is the price of the underlying asset (which we take to be a stock) at time T. The function g is called **pay-off function** of the derivative, while $Y(t) = g(S(t))$ is called **intrinsic value** of the derivative. The term "European" refers to the fact that the contract cannot be exercised before time T, while the term "standard" refers to the fact that the pay-off depends only on the price of the underlying at time T. The pay-off of non-standard (or **exotic**) European derivatives depends on the path of the asset price during the interval $[0,T]$. For example, the pay-off of an **Asian call** is given by $Y = (\frac{1}{T}\int_0^T S(t)\,dt - K)_+$.

The price at time t of a European derivative (standard or not) with pay-off Y and expiration date T will be denoted by $\Pi_Y(t)$. Hence $\{\Pi_Y(t)\}_{t\in[0,T]}$ is a stochastic process.

A **standard American derivative** with pay-off function g is a contract which can be exercised at any time $t \in (0,T]$ prior or equal to its maturity and that, upon exercise, pays the amount $g(S(t))$ (i.e., the intrinsic value) to the holder of the derivative. Non-standard American derivatives are defined similarly as the European ones but with the further option of earlier exercise. In these notes we are mostly concerned with European derivatives, but in Section 6.10 we also discuss briefly some properties of American call/put options.

Portfolio

The portfolio of an investor is the set of all assets in which the investor is trading. Mathematically it is described by a collection of N stochastic processes

$$\{h_1(t)\}_{t\geq 0}, \{h_2(t)\}_{t\geq 0}, \ldots, \{h_N(t)\}_{t\geq 0},$$

where $h_k(t)$ represents the number of shares of the asset k at time t in the investor portfolio. If $h_k(t)$ is positive, resp. negative, the investor has a long, resp. short, position on the asset k at time t. If $\Pi_k(t)$ denotes the value of the asset k at time t, then $\{\Pi_k(t)\}_{t\geq 0}$ is a stochastic process; the **portfolio value** is the stochastic process $\{V(t)\}_{t\geq 0}$ given by

$$V(t) = \sum_{k=1}^{N} h_k(t)\Pi_k(t).$$

Remark 2.8. For modeling purposes, it is convenient to assume that an investor can trade any fraction of shares of the assets, i.e., $h_k(t) : \Omega \to \mathbb{R}$, rather than $h_k(t) : \Omega \to \mathbb{Z}$.

A portfolio process is said to be **self-financing** in the interval $[0,T]$ if no cash is ever added or withdrawn from the portfolio during the interval $[0,T]$. In particular, in a self-financing portfolio, buying more shares of one asset is only possible by selling shares of another asset for an equivalent value. The owner of a self-financing portfolio makes a profit in the time interval $[0,T]$ if $V(T) > V(0)$, while if $V(T) < V(0)$ the investor incurs in a loss. We now introduce the important definition of arbitrage portfolio.

Definition 2.17. *A self-financing portfolio process in the interval* $[0,T]$ *is said to be an* **arbitrage portfolio** *if its value* $\{V(t)\}_{t\in[0,T]}$ *satisfies the following properties:*

(i) $V(0) = 0$ *almost surely;*

(ii) $V(T) \geq 0$ *almost surely;*

(iii) $\mathbb{P}(V(T) > 0) > 0$.

Hence a self-financing arbitrage portfolio is a risk-free investment in the interval $[0,T]$ which requires no initial wealth and with a positive probability to give profit. We remark that the arbitrage property depends on the probability measure $\mathbb{P}$. However, it is clear that if two measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, then the arbitrage property is satisfied with respect to $\mathbb{P}$ if and only if it is satisfied with respect to $\widetilde{\mathbb{P}}$. The guiding principle to devise theoretical models for asset prices in financial mathematics is to ensure that one cannot set-up an arbitrage portfolio by investing on these assets, in which case the market is said to be **arbitrage-free**.

We now show in an arbitrage free market there holds $\Pi_Y(T) = Y$, i.e., there exist no offers to buy or sell a derivative for less or more than Y at the time of maturity. In fact, if a derivative is sold for $\Pi_Y(T) < Y$ "just before" it expires at time T, then buyer would make the sure profit $Y - \Pi_Y(T)$ at time T. Conversely, if a derivative is sold "just before" maturity for more than Y, then the seller will make the sure profit $\Pi_Y(T) - Y$. Thus, in a arbitrage-free market, $\Pi_Y(T) = Y$ (or, more precisely, $\Pi_Y(t) \to Y$, as $t \to T$).

The discount process

Let $\{r(t)\}_{t\geq 0}$ be a stochastic process modeling the **risk-free** rate of the money market. Denote by $B(t)$ the value at time t of a **risk-free asset** with value $B(0)$ at time $t = 0$, that is

$$B(t) = B(0)\exp\left(-\int_0^t r(s)\,ds\right). \tag{2.15}$$

The stochastic process $\{D(t)\}_{t\geq 0}$ given by

$$D(t) = \frac{B(0)}{B(t)} = \exp\left(-\int_0^t r(s)\,ds\right) \tag{2.16}$$

is called the **discount process**. If $\tau < t$ and $X(t)$ denotes the price of an asset at time t, the quantity $D(t)X(t)/D(\tau)$, is called the t-price of the asset discounted at time τ. When

$\tau = 0$ we refer to $D(t)X(t)/D(0) = D(t)X(t) = X^*(t)$ simply as the **discounted price** of the asset. For instance, the discounted (at time $t = 0$) price of a stock with price $S(t)$ at time t is given by $S^*(t) = D(t)S(t)$ and has the following meaning: $S^*(t)$ is the amount that should be invested on the money market at time $t = 0$ in order that the value of this investment at time t replicates the value of the stock at time t. Notice that $S^*(t) < S(t)$ when $r(t) > 0$. The discounted price of the stock measures, roughly speaking, the loss in the stock value due to the "time-devaluation" of money expressed by the ratio $B(0)/B(t)$.

Markets

A market in which the objects of trading are N risky assets (e.g., stocks) and M risk-free assets in the money market is said to be "$N+M$ dimensional". Most of these notes focus on the case of **1+1 dimensional markets** in which we assume that the risky asset is a stock. A portfolio process invested in this market is a stochastic process $\{h_S(t), h_B(t)\}_{t\geq 0}$, where $h_S(t)$ is the number of shares of the stock and $h_B(t)$ the number of shares of the risk-free asset in the portfolio at time t. The value of such portfolio is given by

$$V(t) = h_S(t)S(t) + h_B(t)B(t),$$

where $S(t)$ is the price of the stock (given for instance by (2.14)), while $B(t)$ is the value at time t of the risk-free asset, which is given by (2.15).

Chapter 3

Expectation

Throughout this chapter we assume that $(\Omega, \mathcal{F}, \{\mathcal{F}(t)\}_{t\geq 0}, \mathbb{P})$ is a given filtered probability space.

3.1 Expectation and variance of random variables

Suppose that we want to estimate the value of a random variable X before the experiment has been performed. What is a reasonable definition for our "estimate" of X? Let us first assume that X is a simple random variable of the form

$$X = \sum_{k=1}^{N} a_k \mathbb{I}_{A_k},$$

for some finite partition $\{A_k\}_{k=1,\dots,N}$ of Ω and real distinct numbers $a_1, \dots, a_N$. In this case, it is natural to define the **expected value** (or **expectation**) of X as

$$\mathbb{E}[X] = \sum_{k=1}^{N} a_k \mathbb{P}(A_k) = \sum_{k=1}^{N} a_k \mathbb{P}(X = a_k).$$

That is to say, $\mathbb{E}[X]$ is a weighted average of all the possible values attainable by X, in which each value is weighted by its probability of occurrence. This definition applies also for $N = \infty$ (i.e., for discrete random variables) provided of course the infinite series converges. For instance, if $X \in \mathcal{P}(\mu)$ we have

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \sum_{k=0}^{\infty} k \frac{\mu^k e^{-\mu}}{k!} \\
&= e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} = e^{-\mu} \sum_{r=0}^{\infty} \frac{\mu^{r+1}}{r!} = e^{-\mu}\mu \sum_{r=0}^{\infty} \frac{\mu^r}{r!} = \mu.
\end{aligned}$$

Exercise 3.1 (Sol. 14). *Compute the expectation of binomial variables.*

Now let X be a non-negative random variable and consider the sequence $\{s_n^X\}_{n\in\mathbb{N}}$ of simple functions defined in Theorem 2.2. Recall that s_n^X converges pointwise to X as $n \to \infty$, i.e., $s_n^X(\omega) \to X(\omega)$, for all $\omega \in \Omega$ (see Exercise 2.5). Since

$$\mathbb{E}[s_n^X] = \sum_{k=1}^{n2^n-1} \frac{k}{2^n}\mathbb{P}(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) + n\mathbb{P}(X \geq n), \tag{3.1}$$

it is natural to introduce the following definition.

Definition 3.1. *Let $X : \Omega \to [0, \infty)$ be a non-negative random variable. We define the expectation of X as*

$$\mathbb{E}[X] = \lim_{n\to\infty} \sum_{k=1}^{n2^n-1} \frac{k}{2^n}\mathbb{P}(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) + n\mathbb{P}(X \geq n), \tag{3.2}$$

i.e., $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[s_n^X]$, where $s_1^X, s_2^X, \dots$ is the sequence of simple functions converging pointwise to X and defined in Theorem 2.2.

We remark that the limit in (3.2) exists, because (3.1) is an increasing sequence (see next exercise), although this limit could be infinity. When the limit is finite we say that X has finite expectation. This happens for instance when X is bounded, i.e., $0 \leq X \leq C$ a.s., for some positive constant C.

Exercise 3.2. *Show that $\mathbb{E}[s_n^X]$ is increasing in $n \in \mathbb{N}$. Show that the limit* (3.2) *is finite when the non-negative random variable X is bounded.*

Remark 3.1 (Monotone convergence theorem)**.** It can be shown that that the limit (3.2) is the same along any non-decreasing sequence of non-negative random variables that converge pointwise to X, hence we can use any such sequence to define the expectation of a non-negative random variable. This follows by the **monotone convergence theorem**, whose precise statement is the following: If $X_1, X_2, , \dots$ is a non-decreasing sequence of non-negative random variables such that $X_n \to X$ pointiwise a.s., then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Remark 3.2 (Dominated convergence theorem)**.** The sequence of simple random variables used to define the expectation of a non-negative random variable need not be non-decreasing either. This follows by the **dominated convergence theorem**, whose precise statement is the following: if $X_1, X_2, \dots$ is a sequence of non-negative random variables such that $X_n \to X$, as $n \to \infty$, pointwise a.s., and $\sup_n X_n \leq Y$ for some non-negative random variable Y with finite expectation, then $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Next we extend the definition of expectation to general random variables. For this purpose we use that every random variable $X : \Omega \to \mathbb{R}$ can be written as

$$X = X_+ - X_-,$$

where

$$X_+ = \max(0, X), \quad X_- = -\min(X, 0)$$

are respectively the positive and negative part of X. Since $X_\pm$ are non-negative random variables, then their expectation is given as in Definition 3.1.

Definition 3.2. *Let $X : \Omega \to \mathbb{R}$ be a random variable and assume that at least one of the random variables X_+, X_- has finite expectation. Then we define the expectation of X as*

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-].$$

If $X_\pm$ have both finite expectation, we say that X has finite expectation or that it is an **integrable** *random variable. The set of all integrable random variables on Ω will be denoted by $L^1(\Omega)$, or by $L^1(\Omega, \mathbb{P})$ if we want to specify the probability measure.*

Remark 3.3 (Notation)**.** Of course the expectation of a random variable depends on the probability measure. If another probability measure $\widetilde{\mathbb{P}}$ is defined on the σ-algebra of events (not necessarily equivalent to $\mathbb{P}$), we denote the expectation of X in $\widetilde{\mathbb{P}}$ by $\widetilde{\mathbb{E}}[X]$.

Remark 3.4 (Expectation=Lebesgue integral)**.** The expectation of a random variable X with respect to the probability measure $\mathbb{P}$ is also called the Lebesgue integral of X over Ω in the measure $\mathbb{P}$ and denoted by

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

We shall not use this notation.

The following theorem collects some useful properties of the expectation:

Theorem 3.1. *Let $X, Y : \Omega \to \mathbb{R}$ be integrable random variables. Then the following holds:*

(i) Linearity: For all $\alpha, \beta \in \mathbb{R}$, $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$;

(ii) If $X \leq Y$ a.s. then $\mathbb{E}[X] \leq \mathbb{E}[Y]$;

(iii) If $X \geq 0$ a.s., then $\mathbb{E}[X] = 0$ if and only if $X = 0$ a.s.;

(iv) If X, Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$;

(v) Jensen's inequality: If $f : \mathbb{R} \to \mathbb{R}$ is measurable and convex, then $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.

Sketch of the proof. The argument for the proof of (i)–(iv) is divided in three steps: STEP 1: Show that it suffices to prove the claim for non-negative random variables. STEP 2: Prove the claim for simple functions. STEP 3: Take the limit along the sequences $\{s_n^X\}_{n\in\mathbb{N}}$, $\{s_n^Y\}_{n\in\mathbb{N}}$ of simple functions converging to X, Y. Carrying out these three steps for (i), (ii) and (iii) is simpler, so let us focus on (iv). Let $X_+ = f(X)$, $X_- = g(X)$, and similarly for

Y, where $f(s) = \max(0, s)$, $g(s) = -\min(0, s)$. By Exercise 2.4, each of (X_+, Y_+), (X_-, Y_+), (X_+, Y_-) and (X_-, Y_-) is a pair of independent (non-negative) random variables. Assume that the claim is true for non-negative random variables. Then, using $X = X_+ - X_-$, $Y = Y_+ - Y_-$ and the linearity of the expectation, we find

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[(X_+ - X_-)(Y_+ - Y_-)] \\ &= \mathbb{E}[X_+Y_+] - \mathbb{E}[X_-Y_+] - \mathbb{E}[X_+Y_-] + \mathbb{E}[X_-Y_-] \\ &= \mathbb{E}[X_+]\mathbb{E}[Y_+] - \mathbb{E}[X_-]\mathbb{E}[Y_+] - \mathbb{E}[X_+]\mathbb{E}[Y_-] + \mathbb{E}[X_-]\mathbb{E}[Y_-] \\ &= (\mathbb{E}[X_+] - \mathbb{E}[X_-])(\mathbb{E}[Y_+] - \mathbb{E}[Y_-]) = \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

Hence it suffices to prove the claim for non-negative random variables. Next assume that X, Y are independent simple functions and write

$$X = \sum_{j=1}^{N} a_j \mathbb{I}_{A_j}, \quad Y = \sum_{k=1}^{M} b_k \mathbb{I}_{B_k}.$$

We have

$$XY = \sum_{j=1}^{N} \sum_{k=1}^{M} a_j b_k \mathbb{I}_{A_j} \mathbb{I}_{B_k} = \sum_{j=1}^{N} \sum_{k=1}^{M} a_j b_k \mathbb{I}_{A_j \cap B_k}.$$

Thus by linearity of the expectation, and since the events A_j, B_k are independent, for all j, k, we have

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{j=1}^{N} \sum_{k=1}^{M} a_j b_k \mathbb{E}[\mathbb{I}_{A_j \cap B_k}] = \sum_{j=1}^{N} \sum_{k=1}^{M} a_j b_k \mathbb{P}(A_j \cap B_k) \\ &= \sum_{j=1}^{N} \sum_{k=1}^{M} a_j b_k \mathbb{P}(A_j)\mathbb{P}(B_k) = \sum_{j=1}^{N} a_j \mathbb{P}(A_j) \sum_{k=1}^{M} b_k \mathbb{P}(B_k) = \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

Hence the claim holds for simple functions. It follows that

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

Letting $n \to \infty$, the right hand side converges to $\mathbb{E}[X]\mathbb{E}[Y]$. To complete the proof we have to show that the left hand side converges to $\mathbb{E}[XY]$. This follows by applying the monotone convergence theorem (see Remark 3.1) to the sequence $Z_n = s_n^X s_n^Y$. Next we prove Jensen's inequality. We assume for simplicity that f is differentiable. Then it is easy to see that for all $a \in \mathbb{R}$ the graph of $f(z)$ lies above the graph of the straight line $f'(a)(z - a) + f(a)$ tangent to $f(z)$ at $z = a$. Hence

$$f(z) \geq f'(a)(z - a) + f(a).$$

Choosing $z = X(\omega)$ and $a = \mathbb{E}[X]$, we obtain that for all sample points $\omega \in \Omega$ there holds

$$f(X(\omega)) \geq f'(\mathbb{E}[X])(X(\omega) - \mathbb{E}[X]) + f(\mathbb{E}[X]).$$

Taking the expectation of both sides and using the monotonicity and the linearity of the expectation we obtain

$$\mathbb{E}[f(X)] \geq f'(\mathbb{E}[X])(\mathbb{E}[X] - \mathbb{E}[X]) + f(\mathbb{E}[X]) = f(\mathbb{E}[X]),$$

which concludes the proof of Jensen's inequality when f is differentiable (see [7] for the general case). $\square$

As $|X| = X_+ + X_-$, a random variable X is integrable if and only if $\mathbb{E}[|X|] < \infty$. Hence we have

$$X \in L^1(\Omega) \Leftrightarrow \mathbb{E}[X] < \infty \Leftrightarrow \mathbb{E}[|X|] < \infty.$$

The set of random variables $X : \Omega \to \mathbb{R}$ such that $|X|^2$ is integrable, i.e., $\mathbb{E}[|X|^2] < \infty$, will be denoted by $L^2(\Omega)$ or $L^2(\Omega, \mathbb{P})$.

Exercise 3.3 (Sol. 15)**.** *Prove the Schwarz inequality,*

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}, \tag{3.3}$$

for all random variables $X, Y \in L^2(\Omega)$.

Letting $Y = 1$ in (3.3), we find

$$L^1(\Omega) \subset L^2(\Omega).$$

The **covariance** $\mathrm{Cov}(X, Y)$ of two random variables $X, Y \in L^2(\Omega)$ is defined as

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

Two random variables are said to be **uncorrelated** if $\mathrm{Cov}(X, Y) = 0$. By Theorem 3.1(iv), if X, Y are independent then they are uncorrelated, but the opposite is not true in general. Consider for example the simple random variables

$$X = \begin{cases} -1 & \text{with probability } 1/3 \\ 0 & \text{with probability } 1/3 \\ 1 & \text{with probability } 1/3 \end{cases}$$

and

$$Y = X^2 = \begin{cases} 0 & \text{with probability } 1/3 \\ 1 & \text{with probability } 2/3 \end{cases}$$

Then X and Y are clearly not independent, but

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X^3] - 0 = 0,$$

since $\mathbb{E}[X^3] = \mathbb{E}[X] = 0$.

Definition 3.3. *The* **variance** *of a random variable $X \in L^2(\Omega)$ is given by*

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Using the linearity of the expectation we can rewrite the definition of variance as

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - 2\mathbb{E}[\mathbb{E}[X]X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Cov}(X, X).$$

It follows that a random variable has zero variance if and only if $X = \mathbb{E}[X]$ a.s., hence we may view $\mathrm{Var}[X]$ as a measure of the "randomness of X". As a way of example, let us compute the variance of $X \in \mathcal{P}(\mu)$. We have

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k=0}^{\infty} k^2 \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k^2 \frac{\mu^k e^{-\mu}}{k!} = e^{-\mu} \sum_{k=1}^{\infty} \frac{k}{(k-1)!}\mu^k \\ &= e^{-\mu} \sum_{r=0}^{\infty} \frac{r+1}{r!}\mu^{r+1} = e^{-\mu}\mu \sum_{r=0}^{\infty} \frac{\mu^r}{r!} + \mu \sum_{r=0}^{\infty} r\mathbb{P}(X = r) = \mu + \mu\mathbb{E}[X] = \mu + \mu^2.\end{aligned}$$

Hence

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu + \mu^2 - \mu^2 = \mu.$$

Exercise 3.4. *Compute the variance of binomial random variables.*

Exercise 3.5 (Sol. 16)**.** *Prove the following:*

1. $\mathrm{Var}[\alpha X] = \alpha^2 \mathrm{Var}[X]$, *for all constants* $\alpha \in \mathbb{R}$*;*
2. $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X, Y)$*;*
3. $-\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]} \leq \mathrm{Cov}(X, Y) \leq \sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}$. *The left (resp. right) inequality becomes an equality if and only if there exists a negative (resp. positive) constant* a_0 *and a real constant* b_0 *such that* $Y = a_0 X + b_0$ *almost surely.*

By the previous exercise, $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ holds if and only if X, Y are uncorrelated. Moreover, if we define the **correlation** of X, Y as

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}},$$

then $\mathrm{Cor}(X, Y) \in [-1, 1]$ and $|\mathrm{Cor}(X, Y)| = 1$ if and only if Y is a linear function of X. The interpretation is the following: the closer is $\mathrm{Cor}(X, Y)$ to 1 (resp. -1), the more the variables X and Y have tendency to move in the same (resp. opposite) direction (for instance, ($\mathrm{Cor}(X, -2X) = -1$, $\mathrm{Cor}(X, 2X) = 1$). An important problem in quantitative finance is to find correlations between the price of different assets.

Exercise 3.6. *Let* $\{M_k\}_{k\in\mathbb{N}}$ *be a random walk (not necessarily symmetric). Compute* $\mathbb{E}[M_k]$ *and* $\mathrm{Var}[M_k]$*, for all* $k \in \mathbb{N}$*.*

Exercise 3.7. *Show that the function* $\|\cdot\|_2$ *which maps a random variable* Z *to* $\|Z\|_2 = \sqrt{\mathbb{E}[Z^2]}$ *is a norm in* $L^2(\Omega)$*.*

Remark 3.5 (L^2-norm)**.** The norm defined in the previous exercise is called L^2 **norm.** It can be shown that it is a complete norm, i.e., if $\{X_n\}_{n\in\mathbb{N}} \subset L^2(\Omega)$ is a Cauchy sequence of random variables in the norm L^2, then there exists a random variable $X \in L^2(\Omega)$ such that $\|X_n - X\|_2 \to 0$ as $n \to \infty$.

Exercise 3.8 (Sol. 17)**.** *Let $\{W_n(t)\}_{t\geq 0}$, $n \in \mathbb{N}$, be the sequence of stochastic processes defined in* (2.13)*. Compute* $\mathbb{E}[W_n(t)]$, $\mathrm{Var}[W_n(t)]$, $\mathrm{Cov}[W_n(t), W_n(s)]$*. Show that*

$$\mathrm{Var}(W_n(t)) \to t, \quad \mathrm{Cov}(W_n(t), W_n(s)) \to \min(s,t), \quad \textit{as } n \to +\infty.$$

Next we want to present a first application in finance of the theory outlined above. In particular we establish a sufficient condition which ensures that a portfolio is not an arbitrage.

Theorem 3.2. *Let a portfolio process be given with value $\{V(t)\}_{t\geq 0}$. Let $V^*(t) = D(t)V(t)$ be the discounted portfolio value. If there exists a measure $\widetilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ such that $\widetilde{\mathbb{E}}[V^*(t)]$ is constant (independent of t), then for all $T > 0$ the portfolio is not an arbitrage in the interval $[0,T]$.*

Proof. Assume that the portfolio is an arbitrage in some interval $[0,T]$. Then $V(0) = 0$ almost surely; as $V^*(0) = V(0)$, the assumption of constant expectation in the probability measure $\widetilde{\mathbb{P}}$ gives

$$\widetilde{\mathbb{E}}[V^*(t)] = 0, \quad \text{for all } t \geq 0. \tag{3.4}$$

Moreover $\mathbb{P}(V(T) \geq 0) = 1$ and $\mathbb{P}(V(T) > 0) > 0$. Since $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, we also have $\widetilde{\mathbb{P}}(V(T) \geq 0) = 1$ and $\widetilde{\mathbb{P}}(V(T) > 0) > 0$. Since the discount process is positive, we also have $\widetilde{\mathbb{P}}(V^*(T) \geq 0) = 1$ and $\widetilde{\mathbb{P}}(V^*(T) > 0) > 0$. However this contradicts (3.4), due to Theorem 3.1(iii). Hence our original hypothesis that the portfolio is an arbitrage portfolio is false. □

Radon-Nikodým theorem

Theorem 3.2 will be applied in Chapter 6. To this purpose we shall need the following characterization of equivalent probability measures.

Theorem 3.3. *Given a probability measure $\mathbb{P}$, the following are equivalent:*

(i) $\widetilde{\mathbb{P}}$ *is a probability measure equivalent to* $\mathbb{P}$*;*

(ii) There exists a random variable $Z : \Omega \to \mathbb{R}$ such that $Z > 0$ $\mathbb{P}$-almost surely, $\mathbb{E}[Z] = 1$ and $\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z\mathbb{I}_A]$, for all $A \in \mathcal{F}$.

Moreover, assuming any of these two equivalent conditions, the random variable Z is unique (up to a $\mathbb{P}$-null set) and for all random variables X such that $XZ \in L^1(\Omega, \mathbb{P})$, we have $X \in L^1(\Omega, \widetilde{\mathbb{P}})$ and

$$\widetilde{\mathbb{E}}[X] = \mathbb{E}[ZX]. \tag{3.5}$$

Proof. The implication $(i) \Rightarrow (ii)$ is the **Radon-Nikodým** theorem, whose proof can be found for instance in [7]. As to the implication $(ii) \Rightarrow (i)$, we first observe that $\widetilde{\mathbb{P}}(\Omega) = \mathbb{E}[Z\mathbb{I}_\Omega] = \mathbb{E}[Z] = 1$. Hence, to prove that $\widetilde{\mathbb{P}}$ is a probability measure, it remains to show that it satisfies the countable additivity property: for all families $\{A_k\}_{k\in\mathbb{N}}$ of disjoint events, $\widetilde{\mathbb{P}}(\cup_k A_k) = \sum_k \widetilde{\mathbb{P}}(A_k)$. To prove this let

$$B_n = \cup_{k=1}^n A_k.$$

Clearly, $Z\mathbb{I}_{B_n}$ is an increasing sequence of random variables. Hence, by the monotone convergence theorem (see Remark 3.1) we have

$$\lim_{n\to\infty} \mathbb{E}[Z\mathbb{I}_{B_n}] = \mathbb{E}[Z\mathbb{I}_{B_\infty}], \quad B_\infty = \cup_{k=1}^\infty A_k,$$

i.e.,

$$\lim_{n\to\infty} \widetilde{\mathbb{P}}(B_n) = \widetilde{\mathbb{P}}(B_\infty). \tag{3.6}$$

On the other hand, by linearity of the expectation,

$$\widetilde{\mathbb{P}}(B_n) = \mathbb{E}[ZB_n] = \mathbb{E}[Z\mathbb{I}_{\cup_{k=1}^n A_k}] = \mathbb{E}[Z(\mathbb{I}_{A_1} + \cdots + \mathbb{I}_{A_n})] = \sum_{k=1}^n \mathbb{E}[ZA_k] = \sum_{k=1}^n \widetilde{\mathbb{P}}(A_k).$$

Hence (3.6) becomes

$$\sum_{k=1}^\infty \widetilde{\mathbb{P}}(A_k) = \widetilde{\mathbb{P}}(\cup_{k=1}^\infty A_k).$$

This proves that $\widetilde{\mathbb{P}}$ is a probability measure. To show that $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, let A be such that $\widetilde{\mathbb{P}}(A) = 0$. Since $Z\mathbb{I}_A \geq 0$ almost surely, then $\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z\mathbb{I}_A] = 0$ is equivalent, by Theorem 3.1(iii), to $Z\mathbb{I}_A = 0$ almost surely. Since $Z > 0$ almost surely, then this is equivalent to $\mathbb{I}_A = 0$ a.s., i.e., $\mathbb{P}(A) = 0$. Thus $\widetilde{\mathbb{P}}(A) = 0$ if and only if $\mathbb{P}(A) = 0$, i.e., the probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent. It remains to prove the identity (3.5). If X is the simple random variable $X = \sum_k a_k \mathbb{I}_{A_k}$, then the proof is straightforward:

$$\widetilde{\mathbb{E}}[X] = \sum_k a_k \widetilde{\mathbb{P}}(A_k) = \sum_k a_k \mathbb{E}[Z\mathbb{I}_{A_k}] = \mathbb{E}[Z \sum_k a_k \mathbb{I}_{A_k}] = \mathbb{E}[ZX].$$

For a general non-negative random variable X the result follows by applying (3.5) to an increasing sequence of simple random variables converging to X and then passing to the limit (using the monotone convergence theorem). The result for a general random variable $X : \Omega \to \mathbb{R}$ follows by applying (3.5) to the positive and negative part of X and using the linearity of the expectation. □

Remark 3.6 (Radon-Nikodým derivative)**.** Using the Lebesgue integral notation (see Remark 3.4) we can write (3.5) as

$$\int_\Omega X(\omega) d\widetilde{\mathbb{P}}(\omega) = \int_\Omega X(\omega)Z(\omega) d\mathbb{P}(\omega).$$

This leads to the formal identity $d\widetilde{\mathbb{P}}(\omega) = Z(\omega)d\mathbb{P}(\omega)$, or $Z(\omega) = \frac{d\widetilde{\mathbb{P}}(\omega)}{d\mathbb{P}(\omega)}$, which explains why Z is also called the **Radon-Nikodým derivative** of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$.

An application of Theorem 3.3 is given in Exercise 3.12 below.

Computing the expectation of a random variable

Next we discuss how to compute the expectation of a random variable X. Definition 3.1 is clearly not very useful to this purpose, unless X is a simple random variable. There exist several methods to compute the value for $\mathbb{E}[X]$, some of which will be presented later in these notes. In this section we show that the expectation and the variance of a random variable can be computed easily when the random variable admits a density.

Theorem 3.4. *Let $X : \Omega \to \mathbb{R}$ be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $g(X) \in L^1(\Omega)$. Assume that X admits the density f_X. Then*

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

In particular, the expectation and the variance of X are given by

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)\, dx, \quad \mathrm{Var}[X] = \int_{\mathbb{R}} x^2 f_X(x)\, dx - \left(\int_{\mathbb{R}} x f_X(x)\, dx\right)^2.$$

Proof. We prove the theorem under the assumption that g is a simple measurable function, the proof for general functions g follows by a limit argument similar to the one used in the proof of Theorem 3.1, see Theorem 1.5.2 in [26] for the details. Hence we assume

$$g(x) = \sum_{k=1}^{N} \alpha_k \mathbb{I}_{U_k}(x),$$

for some disjoint Borel sets $U_1, \dots, U_N \subset \mathbb{R}$. Thus

$$\mathbb{E}[g(X)] = \mathbb{E}\big[\sum_k \alpha_k \mathbb{I}_{U_k}(X)\big] = \sum_k \alpha_k \mathbb{E}[\mathbb{I}_{U_k}(X)].$$

Let $Y_k = \mathbb{I}_{U_k}(X) : \Omega \to \mathbb{R}$. Then Y_k is the simple random variable that takes value 1 if $\omega \in A_k$ and 0 if $\omega \in A_k^c$, where $A_k = \{X \in U_k\}$. Thus the expectation of Y_k is given by $\mathbb{E}[Y_k] = \mathbb{P}(A_k)$ and so

$$\mathbb{E}[g(X)] = \sum_k \alpha_k \mathbb{P}(X \in U_k) = \sum_k \alpha_k \int_{U_k} f(x)\, dx = \int_{\mathbb{R}} \sum_k \alpha_k \mathbb{I}_{U_k}(x) f(x)\, dx = \int_{\mathbb{R}} g(x) f(x)\, dx,$$

as claimed. $\square$

For instance, if $X \in \mathcal{N}(m, \sigma^2)$, we have

$$\mathbb{E}[X] = \int_{\mathbb{R}} x e^{-\frac{(x-m)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} = m,$$
$$\mathrm{Var}[X] = \int_{\mathbb{R}} x^2 e^{-\frac{(x-m)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} - m^2 = \sigma^2,$$

which explains why m is called the expectation and σ^2 the variance of the normal random variable X. Note in particular that, for a Brownian motion $\{W(t)\}_{t\geq 0}$, there holds

$$\mathbb{E}[W(t) - W(s)] = 0, \quad \mathrm{Var}[W(t) - W(s)] = |t-s|, \quad \text{for all } s, t \geq 0. \tag{3.7}$$

Let us show that[1]

$$\mathrm{Cov}(W(t), W(s)) = \min(s, t). \tag{3.8}$$

For $s = t$, the claim is equivalent to $\mathrm{Var}[W(t)] = t$, which holds by definition of Brownian motion (see (3.7)). For $t > s$ we have

$$\begin{aligned}\mathrm{Cov}(W(t), W(s)) &= \mathbb{E}[W(t)W(s)] - \mathbb{E}[W(t)]\mathbb{E}[W(s)] \\ &= \mathbb{E}[W(t)W(s)] \\ &= \mathbb{E}[(W(t) - W(s))W(s)] + \mathbb{E}[W(s)^2].\end{aligned}$$

Since $W(t) - W(s)$ and $W(s)$ are independent random variables, then $\mathbb{E}[(W(t) - W(s))W(s)] = \mathbb{E}(W(t) - W(s)]\mathbb{E}[W(s)] = 0$, and so

$$\mathrm{Cov}(W(t), W(s)) = \mathbb{E}[W(s)^2] = \mathrm{Var}[W(s)] = s = \min(s, t), \quad \text{for } t > s.$$

A similar argument applies for $t < s$.

Exercise 3.9 (Sol. 18)**.** *Let $g : (0, \infty) \to \mathbb{R}$ be a differentiable function and let*

$$X(t) = g(t)W(t) - \int_0^t g'(s)W(s)\, ds.$$

Show that

$$X(t) \in \mathcal{N}(0, \Delta(t)), \quad \Delta(t) = \int_0^t g(s)^2\, ds.$$

HINT: Use the Riemann sum approximation of the integral.

Exercise 3.10. *The moment of order n of a random variable X is the quantity $\mu_n = \mathbb{E}[X^n]$, $n = 1, 2, \ldots$. Let $X \in \mathcal{N}(0, \sigma^2)$. Prove that*

$$\mu_n = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (n-1)\sigma^n & \text{if } n \text{ is even.} \end{cases}$$

[1] Compare (3.8) with the result of Exercise 3.8.

Exercise 3.11 (Sol. 19)**.** *Compute the expectation and the variance of exponential random variables.*

Exercise 3.12 (Sol. 20)**.** *Let $X \in \mathcal{E}(\lambda)$ be an exponential random variable with intensity λ. Given $\widetilde{\lambda} > 0$, let*

$$Z = \frac{\widetilde{\lambda}}{\lambda} e^{-(\widetilde{\lambda}-\lambda)X}.$$

Define $\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z\mathbb{I}_A]$, $A \in \mathcal{F}$. Show that $\widetilde{\mathbb{P}}$ is a probability measure equivalent to $\mathbb{P}$. Prove that $X \in \mathcal{E}(\widetilde{\lambda})$ in the probability measure $\widetilde{\mathbb{P}}$.

Remark 3.7. Exercise 3.12 shows that one can pick a probability measure $\widetilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ such that exponential random variables in the probability $\mathbb{P}$ remain exponentially distributed in the probability $\widetilde{\mathbb{P}}$. The extension of this result to more general random variables is Girsanov's theorem discussed in the next chapter.

Exercise 3.13. *Compute the expectation and the variance of Cauchy distributed random variables. Compute the expectation and the variance of Lévy distributed random variables.*

Exercise 3.14. *Compute the expectation and the variance of the geometric Brownian motion* (2.14)*.*

Exercise 3.15 ($\star$)**.** *Show that the paths of the Brownian motion have unbounded linear variation. Namely, given $0 = t_0 < t_1 < \cdots < t_n = t$ with $t_k - t_{k-1} = h$, for all $k = 1, \dots, n$, show that*

$$\mathbb{E}[\sum_{k=1}^{n} |W(t_k) - W(t_{k-1})|] \to \infty, \quad \text{as } n \to \infty.$$

(However, Brownian motions have finite quadratic variation, see Section 3.2).

A result similar to Theorem 3.4 can be used to compute the correlation between two random variables that admit a joint density.

Theorem 3.5. *Let $X, Y : \Omega \to \mathbb{R}$ be two random variables with joint density $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ and let $g : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function such that $g(X, Y) \in L^1(\Omega)$. Then*

$$\mathbb{E}[g(X, Y)] = \int_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y)\, dx\, dy.$$

In particular, for $X, Y \in L^2(\Omega)$,

$$\mathrm{Cov}(X, Y) = \int_{\mathbb{R}^2} xy f_{X,Y}(x, y)\, dx\, dy - \int_{\mathbb{R}} x f_X(x)\, dx \int_{\mathbb{R}} y f_Y(y)\, dy,$$

where

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy, \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dx$$

are the (marginal) densities of X and Y.

Exercise 3.16. *Show that if $X_1, X_2 : \Omega \to \mathbb{R}$ are jointly normally distributed with covariant matrix $C = (C_{ij})_{i,j=1,2}$, then $C_{ij} = \mathrm{Cov}(X_i, X_j)$.*

Combining the results of Exercises 2.11 and 3.16, we see that the parameter ρ in Equation (2.10) is precisely the correlation of the two jointly normally distributed random variables X, Y. It follows by Remark 2.3 that two jointly normally distributed random variables are independent if and only if they are uncorrelated. Recall that for general random variables, independence implies uncorrelation, but the opposite is in general not true.

Characteristic function

In this section, and occasionally in the rest of the notes, we shall need to take the expectation of a complex-valued random variable $Z : \Omega \to \mathbb{C}$. Letting $Z = \mathrm{Re}(Z) + i\mathrm{Im}(Z)$, the expectation of Z is the complex number defined by

$$\mathbb{E}[Z] = \mathbb{E}[\mathrm{Re}(Z)] + i\mathbb{E}[\mathrm{Im}(Z)].$$

Definition 3.4. *Let $X \in L^1(\Omega)$. The function $\theta_X : \mathbb{R} \to \mathbb{C}$ given by*

$$\theta_X(u) = \mathbb{E}[e^{iuX}]$$

is called the **characteristic function of X.** *The positive, real-valued function $M_X(u) = \mathbb{E}[e^{uX}]$, when it exists in some neighborhood of $u = 0$, is called the* **moment-generating function of X.**

Note that if the random variable X admits the density f_X, then

$$\theta_X(u) = \int_{\mathbb{R}} e^{iux} f_X(x)\, dx,$$

i.e., the characteristic function is the inverse Fourier transform of the density. Table 3.1 contains some examples of characteristic functions.

Remark 3.8. While θ_X is defined for all $u \in \mathbb{R}$, the moment-generating function of a random variable may be defined only in a subset of the real line, or not defined at all (see Exercise 3.17). For instance, when $X \in \mathcal{E}(\lambda)$ we have

$$M_X(u) = \mathbb{E}[e^{uX}] = \lambda \int_0^\infty e^{(u-\lambda)x}\, dx = \begin{cases} +\infty & \text{if } u \geq \lambda \\ (1 - u/\lambda)^{-1} & \text{if } u < \lambda \end{cases}$$

Hence $M_X(u)$ is defined (as a positive function) only for $u < \lambda$.

Exercise 3.17. *Show that Cauchy random variables do not have a well-defined moment-generating function.*

Density	Characteristic function
$\mathcal{N}(m,\sigma^2)$	$\exp(ium - \frac{1}{2}\sigma^2 u^2)$
$\mathcal{E}(\lambda)$	$(1 - iu/\lambda)^{-1}$
$\chi^2(\delta)$	$(1 - 2iu)^{-\delta/2}$
$\chi^2(\delta,\beta)$	$(1 - 2iu)^{-\delta/2} \exp\left(-\frac{\beta u}{2u+i}\right)$

Table 3.1: Examples of characteristic functions

The characteristic function of a random variable provides a lot of information. In particular, it determines completely the distribution function of the random variable, as shown in the following theorem (for the proof, see [7, Sec. 9.5]).

Theorem 3.6. *Let $X, Y \in L^1(\Omega)$. Then $\theta_X = \theta_Y$ if and only if $F_X = F_Y$. In particular, if $\theta_X = \theta_Y$ and one of the two variables admits a density, then the other does too and the densities are equal.*

According to the previous theorem, if we want for instance to prove that a random variable X is normally distributed, we may try to show that its characteristic function θ_X is given by $\theta_X(u) = \exp(ium - \frac{1}{2}\sigma^2 u^2)$, see Table 3.1. Another useful property of characteristic functions is proved in the following theorem.

Theorem 3.7. *Let $X_1, \dots, X_N \in L^1(\Omega)$ be independent random variables. Then*

$$\theta_{X_1+\dots+X_N} = \theta_{X_1} \cdots \theta_{X_N}.$$

Proof. We have

$$\theta_{X_1+\dots+X_N}(u) = \mathbb{E}[e^{iu(X_1+\dots+X_N)}] = \mathbb{E}[e^{iuX_1} e^{iuX_2} \cdots e^{iuX_N}].$$

Using that the variables $Y_1 = e^{iuX_1}, \dots, Y_N = e^{iuX_N}$ are independent (see Theorem 2.1) and that the expectation of the product of independent random variables is equal to the product of their expectations (see Theorem 3.1(iv)) we obtain

$$\mathbb{E}[e^{iuX_1} e^{iuX_2} \cdots e^{iuX_N}] = \mathbb{E}[e^{iuX_1}] \cdots \mathbb{E}[e^{uX_N}] = \theta_{X_1}(u) \cdots \theta_{X_N}(u),$$

which concludes the proof. □

As an application of the previous theorem, we now show that if $X_1, \dots X_N$ are independent normally distributed random variables with expectations $m_1, \dots, m_N$ and variances $\sigma_1^2, \dots, \sigma_N^2$, then the random variable

$$Y = X_1 + \dots + X_N$$

is normally distributed with mean m and variance σ^2 given by

$$m = m_1 + \cdots + m_N, \quad \sigma^2 = \sigma_1^2 + \cdots + \sigma_N^2. \tag{3.9}$$

In fact,

$$\theta_{X_1+\cdots+X_N}(u) = \theta_{X_1}(u)\cdots\theta_{X_N}(u) = e^{ium_1 - \frac{1}{2}\sigma_1^2 u^2} \cdots e^{ium_N - \frac{1}{2}\sigma_N^2 u^2} = e^{ium - \frac{1}{2}\sigma^2 u^2}.$$

The right hand side of the previous equation is the characteristic function of a normal variable with expectation m and variance σ^2 given by (3.9). Thus Theorem 3.6 implies that $X_1 + \cdots + X_N \in \mathcal{N}(m, \sigma^2)$.

Exercise 3.18 ($\star$)**.** *Let $X_1 \in \mathcal{N}(m_1, \sigma_1^2), \dots, X_N \in \mathcal{N}(m_N, \sigma_N^2)$, $N \geq 2$, be independent. Show that $Y = \sum_{k=1}^N (X_k/\sigma_k)^2 \in \chi^2(N, \beta)$ where $\beta = (m_1/\sigma_1)^2 + \cdots + (m_N/\sigma_N)^2$ (compare with Exercise 2.8).*

Exercise 3.19. *Let $X, Y \in \mathcal{N}(0,1)$ be independent and jointly normally distributed. Show that the random variable Z defined by*

$$Z = \begin{cases} Y/X & \text{for } X \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

is Cauchy distributed.

In a similar fashion, we define the characteristic function of the vector-valued random variable $X = (X_1, \dots, X_n)$ as $\theta_X : \mathbb{R}^n \to \mathbb{C}$,

$$\theta_X(u_1, \dots, u_n) = \mathbb{E}[e^{u_1 X_1 + u_2 X_2 + \cdots + u_n X_n}].$$

For instance, it can be shown that two random variables X_1, X_2 are jointly normal with mean $m = (m_1\ m_2)$ and covariance matrix $C = (C_{ij})_{i,j=1,2}$ if and only if the characteristic function of $X = (X_1, X_2)$ is given by

$$\theta_X(u_1, u_2) = e^{imu - \frac{1}{2}u^T C\, u}, \quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

Exercise 3.20 ($\star$)**.** *Let $X_1, X_2 \in \mathcal{N}(0,1)$ be independent and define*

$$Y_1 = aX_1 + bX_2, \quad Y_2 = cX_1 + dX_2,$$

for some constants $a, b, c, d \in \mathbb{R}$. Assume that the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is invertible. Show that Y_1, Y_2 are jointly normally distributed with zero mean and covariant matrix $C = AA^T$.

Exercise 3.21. *Let $\{W(t)\}_{t\geq 0}$ be a Brownian motion and $t_1 < t_2$. Show that $W(t_1)$, $W(t_2)$ are jointly normally distributed with zero mean and covariance matrix*

$$C = \begin{pmatrix} t_1 & t_1 \\ t_1 & t_2 \end{pmatrix}.$$

Exercise 3.22. *Let $X, Y \in L^1(\Omega)$ be independent random variables with densities f_X, f_Y. Show that $X + Y$ has the density*

$$f_{X+Y}(x) = \int_{\mathbb{R}} f_X(x-y) f_Y(y)\, dy.$$

Remark: The right hand side of the previous identity defines the **convolution product** *of the functions f_X, f_Y.*

The characteristic function is also very useful to establish whether two random variables are independent, as shown in the following exercise.

Exercise 3.23. *Let $X, Y \in L^1(\Omega)$ and define their joint characteristic function as*

$$\theta_{X,Y}(u,v) = \mathbb{E}[e^{iuX+ivY}], \quad u, v \in \mathbb{R}.$$

Show that X, Y are independent if and only if $\theta_{X,Y}(u,v) = \theta_X(u)\theta_Y(v)$.

3.2 Quadratic variation of stochastic processes

We continue this chapter by discussing the important concept of quadratic variation. We introduce this concept to measure how "wild" a stochastic process oscillates in time, which in financial mathematics is a measure of the volatility of an asset price.

Let $\{X(t)\}_{t\geq 0}$ be a stochastic process. A partition of the interval $[0,T]$ is a set of points $\Pi = \{t_0, t_1, \dots t_m\}$ such that

$$0 = t_0 < t_1 < t_2 < \dots < t_m = T.$$

The size of the partition is given by

$$\|\Pi\| = \max_{j=0,\dots,m-1} (t_{j+1} - t_j).$$

To measure the amount of oscillations of $\{X(t)\}_{t\geq 0}$ in the interval $[0,T]$ along the partition Π, we compute

$$Q_\Pi(\omega) = \sum_{j=0}^{m-1} (X(t_{j+1},\omega) - X(t_j,\omega))^2.$$

Q_Π is a random variable and it depends on the partition.

To define the quadratic variation of the stochastic process $\{X(t)\}_{t\geq 0}$, we compute Q_{Π_n} along a sequence $\{\Pi_n\}_{n\in\mathbb{N}}$ of partitions to the interval $[0,T]$ such that $\|\Pi_n\| \to 0$ as $n\to\infty$ and then we take the limit of Q_{Π_n} as $n\to\infty$. Since $\{Q_{\Pi_n}\}_{n\in\mathbb{N}}$ is a sequence of random variables, there are several ways to define its limit as $n\to\infty$. The precise definition that we adopt is that of L^2-quadratic variation, in which the limit is taken in the norm $\|\cdot\|_2$ defined in Exercise 3.5.

Definition 3.5. *We say that the stochastic process $\{X(t)\}_{t\geq 0}$ has* **L^2-quadratic variation** *$[X,X](T)$ in the interval $[0,T]$ along the sequence of partitions $\{\Pi_n\}_{n\in\mathbb{N}}$ if*

$$\lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{j=0}^{m(n)-1}(X(t_{j+1}^{(n)})-X(t_j^{(n)}))^2-[X,X](T)\right)^2\right]=0,$$

where $m(n)+1$ is the number of points in the partition $\Pi_n=\{t_0,t_1^{(n)},t_2^{(n)},\dots,t_{m(n)-1}^{(n)},T\}$.

Remark 3.9. If the limit in the previous definition does not exist, the quadratic variation cannot be defined as we did (an alternative definition is possible, but we shall not need it).

The quadratic variation is a random variable that depends in general on the sequence of partitions of the interval $[0,T]$ along which it is computed, although this is not reflected in our notation $[X,X](T)$. However for several important examples of stochastic processes—and in particular for all applications considered in these notes—the quadratic variation is independent of the sequence of partitions. To distinguish this important special case, we shall use the following (standard) notation:

$$dX(t)dX(t)=q(t)dt,$$

to indicate that the quadratic variation of the stochastic process $\{X(t)\}_{t\geq 0}$ in any interval $[0,T]$ is given by

$$[X,X](T)=\int_0^T q(t)\,dt$$

independently from the sequence of partitions of the interval $[0,T]$ along which it is computed. Here $\{q(t)\}_{t\geq 0}$ is a stochastic process called **rate of quadratic variation** of $\{X(t)\}_{t\geq 0}$ and which measures how fast quadratic variation accumulates in time.

Now we show that if the paths of the stochastic process $\{X(t)\}_{t\geq 0}$ are sufficiently regular, then its quadratic variation is zero along any sequence of partitions.

Theorem 3.8. *Assume that the paths of the stochastic process $\{X(t)\}_{t\geq 0}$ satisfy*

$$\mathbb{P}(|X(t)-X(s)|\leq C|t-s|^\gamma)=1, \tag{3.10}$$

for all $t,s\geq 0$ and for some positive constants $C>0$, $\gamma>1/2$. Then

$$dX(t)dX(t)=0.$$

Proof. We have

$$\mathbb{E}\left[\left(\sum_{j=0}^{m(n)-1}(X(t_{j+1}^{(n)})-X(t_j^{(n)}))^2\right)^2\right] \leq C^4\mathbb{E}\left[\left(\sum_{j=0}^{m(n)-1}(t_{j+1}^{(n)}-t_j^{(n)})^{2\gamma}\right)^2\right]$$
$$= C^4\mathbb{E}\left[\left(\sum_{j=0}^{m(n)-1}(t_{j+1}^{(n)}-t_j^{(n)})^{2\gamma-1}(t_{j+1}^{(n)}-t_j^{(n)})\right)^2\right].$$

Now we use that $t_{j+1}^{(n)}-t_j^{(n)} \leq \|\Pi_n\|$ and $\sum_j(t_{j+1}^{(n)}-t_j^{(n)}) = T$, so that

$$\mathbb{E}\left[\left(\sum_{j=0}^{m(n)-1}(X(t_{j+1}^{(n)})-X(t_j^{(n)}))^2\right)^2\right] \leq \left(C^2\|\Pi_n\|^{2\gamma-1}T\right)^2 \to 0, \quad \text{as } \|\Pi_n\| \to 0.$$

□

As a special important case we have that

$$dtdt = 0. \tag{3.11}$$

Next we compute the quadratic variation of Brownian motions.

Theorem 3.9. *For a Brownian motion $\{W(t)\}_{t\geq 0}$ there holds*

$$dW(t)dW(t) = dt. \tag{3.12}$$

Proof. Let

$$Q_{\Pi_n}(\omega) = \sum_{j=0}^{m(n)-1}(W(t_{j+1}^{(n)},\omega)-W(t_j^{(n)},\omega))^2,$$

where we recall that $m(n)+1$ is the number of points in the partition Π_n of $[0,T]$. We compute

$$\mathbb{E}[(Q_{\Pi_n}-T)^2] = \mathbb{E}[Q_{\Pi_n}^2] + T^2 - 2T\mathbb{E}[Q_{\Pi_n}].$$

But

$$\mathbb{E}[Q_{\Pi_n}] = \sum_{j=0}^{m(n)-1}\mathbb{E}[(W(t_{j+1}^{(n)})-W(t_j^{(n)}))^2] = \sum_{j=0}^{m(n)-1}\mathrm{Var}[W(t_{j+1}^{(n)})-W(t_j^{(n)})]$$
$$= \sum_{j=0}^{m(n)-1}(t_{j+1}^{(n)}-t_j^{(n)}) = T.$$

Hence we have to prove that

$$\lim_{\|\Pi_n\|\to 0}\mathbb{E}[Q_{\Pi_n}^2] - T^2 = 0,$$

or equivalently (as we have just proved that $\mathbb{E}[Q_{\Pi_n}] = T$),

$$\lim_{\|\Pi_n\|\to 0} \mathrm{Var}(Q_{\Pi_n}) = 0. \tag{3.13}$$

Since the increments of a Brownian motion are independent, we have

$$\begin{aligned}\mathrm{Var}(Q_{\Pi_n}) = \sum_{j=0}^{m(n)-1} \mathrm{Var}[(W(t_{j+1}^{(n)}) - W(t_j^{(n)}))^2] = \sum_{j=0}^{m(n)-1} \mathbb{E}[(W(t_{j+1}^{(n)}) - W(t_j^{(n)}))^4] \\ - \sum_{j=0}^{m(n)-1} \mathbb{E}[(W(t_{j+1}^{(n)}) - W(t_j^{(n)}))^2]^2\end{aligned}$$

Now we use that

$$\mathbb{E}[(W(t_{j+1}^{(n)}) - W(t_j^{(n)}))^2] = \mathrm{Var}[W(t_{j+1}^{(n)}) - W(t_j^{(n)})] = t_{j+1}^{(n)} - t_j^{(n)},$$

and, as it follows by Exercise 3.10,

$$\mathbb{E}[(W(t_{j+1}^{(n)}) - W(t_j^{(n)}))^4] = 3(t_{j+1}^{(n)} - t_j^{(n)})^2.$$

We conclude that

$$\mathrm{Var}[Q_{\Pi_n}] = 2\sum_{j=0}^{m(n)-1} (t_{j+1}^{(n)} - t_j^{(n)})^2 \leq 2\|\Pi_n\| T \to 0, \quad \text{as } \|\Pi_n\| \to 0,$$

which proves (3.13) and thus the theorem. □

Remark 3.10 (No-where differentiability of Brownian motions)**.** Combining Theorem 3.9 and Theorem 3.8, we conclude that the paths of a Brownian motion $\{W(t)\}_{t\geq 0}$ cannot satisfy the regularity condition (3.10). In fact, while the paths of a Brownian motion are a.s. continuous by definition, they turn out to be **no-where differentiable**, in the sense that the event $\{\omega \in \Omega : \gamma_W^\omega \in C^1\}$ is a null set. A proof of this can be found for instance in [9].

Finally we need to consider a slight generalization of the concept of quadratic variation.

Definition 3.6. *We say that two stochastic processes $\{X_1(t)\}_{t\geq 0}$ and $\{X_2(t)\}_{t\geq 0}$ have L^2-* **cross variation** $[X_1, X_2](T)$ *in the interval $[0, T]$ along the sequence of partitions $\{\Pi_n\}_{n\in\mathbb{N}}$, if*

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{j=0}^{m(n)-1} (X_1(t_{j+1}^{(n)}) - X_1(t_j^{(n)}))(X_2(t_{j+1}^{(n)}) - X_2(t_j^{(n)})) - [X_1, X_2](T)\right)^2\right] = 0,$$

where $m(n) + 1$ is the number of points in the partition Π_n.

As for the quadratic variation of a stochastic process, we use a special notation to express that the cross variation of two stochastic processes is independent of the sequence of partitions along which it is computed. Namely, we write

$$dX_1(t)dX_2(t) = \xi(t)\,dt,$$

to indicate that the cross variation $[X_1, X_2](T)$ equals $\int_0^T \xi(t)\,dt$ along any sequence of partitions $\{\Pi_n\}_{n\in\mathbb{N}}$ of the interval $[0, T]$. The following generalization of Theorem 3.8 is easily established.

Theorem 3.10. *Assume that the paths of the stochastic processes $\{X_1(t)\}_{t\geq 0}$, $\{X_2(t)\}_{t\geq 0}$ satisfy*

$$\mathbb{P}(|X_1(t) - X_1(s)| \leq C|t-s|^{\gamma}) = 1, \quad \mathbb{P}(|X_2(t) - X_2(s)| \leq C|t-s|^{\lambda}) = 1,$$

for all $s, t \geq 0$ and some positive constants C, γ, λ such that $\gamma+\lambda > 1/2$. Then $dX_1(t)dX_2(t) = 0$.

Exercise 3.24. *Prove the theorem.*

As a special case we find that

$$dW(t)dt = 0. \tag{3.14}$$

It is important to memorize the identities (3.11), (3.12) and (3.14), as they will be used several times in the following chapters.

Exercise 3.25 ($\star$)**.** *Let $\{W_1(t)\}_{t\geq 0}$, $\{W_2(t)\}_{t\geq 0}$ be two independent Brownian motions. Use the definition of cross variation to show that that $dW_1(t)dW_2(t) = 0$.*

3.3 Conditional expectation

Recall that the expectation value $\mathbb{E}[X]$ is an estimate on the average value of the random variable X. This estimate does not depend on the σ-algebra $\mathcal{F}$, nor on any sub-σ-algebra thereof. However, if some information is known in the form of a σ-algebra $\mathcal{G}$, then one expects to be able to improve the estimate on the value of X. To quantify this we introduce the definition of "expected value of X given $\mathcal{G}$", or conditional expectation, which we denote by $\mathbb{E}[X|\mathcal{G}]$. We want the conditional expectation to verify the following properties:

(i) If X is $\mathcal{G}$-measurable, then it should hold that $\mathbb{E}[X|\mathcal{G}] = X$, because the information provided by $\mathcal{G}$ is sufficient to determine X;

(ii) If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$, because the occurrence of the events in $\mathcal{G}$ does not effect the probability distribution of X;

Property (i) already indicates that $\mathbb{E}[X|\mathcal{G}]$ is a random variable. To begin with we define the conditional expectation of a random variable X with respect to an event $A \in \mathcal{F}$. Let's assume first that X is the simple random variable

$$X = \sum_{k=1}^{N} a_k \mathbb{I}_{A_k}.$$

Let $B \in \mathcal{F} : \mathbb{P}(B) > 0$ and $\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)}$ be the conditional probability of A_k given B, see Definition 1.6. It is natural to define the conditional expectation of X given the event B as

$$\mathbb{E}[X|B] = \sum_{k=1}^{N} a_k \mathbb{P}(A_k|B).$$

Moreover, since

$$X\mathbb{I}_B = \sum_{k=1}^{N} a_k \mathbb{I}_{A_k} \mathbb{I}_B = \sum_{k=1}^{N} a_k \mathbb{I}_{A_k \cap B},$$

we also have the identity $\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbb{I}_B]}{\mathbb{P}(B)}$. We use the latter identity to define the conditional expectation given B of general random variables.

Definition 3.7. *Let $X \in L^1(\Omega)$ and $B \in \mathcal{F}$. When $\mathbb{P}(B) > 0$ we define the* **conditional expectation of X given the event B** *as*

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbb{I}_B]}{\mathbb{P}(B)}.$$

When $\mathbb{P}(B) = 0$ we define $\mathbb{E}[X|B] = \mathbb{E}[X]$.

Remark 3.11. $\mathbb{E}[X|B]$ is a deterministic constant.. Moreover using the Lebesgue integral notation of the expectation, see Remark 3.4, we may rewrite

$$\mathbb{E}[X|B] = \frac{1}{\mathbb{P}(B)} \int_B X(\omega)\, d\mathbb{P}(\omega),$$

which shows that the conditional expectation of a random variable X with respect to an event B is just the average of X within B.

Next we discuss the concept of conditional expectation given a σ-algebra $\mathcal{G}$. We first assume that $\mathcal{G}$ is generated by a (say, finite) partition $\{A_k\}_{k=1,\dots,M}$ of Ω, see Exercise 1.4. Then it is natural to define

$$\mathbb{E}[X|\mathcal{G}] = \sum_{k=1}^{M} \mathbb{E}[X|A_k]\mathbb{I}_{A_k}, \tag{3.15}$$

which is a $\mathcal{G}$-measurable simple function. It will now be shown that (3.15) satisfies the identity

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|B] = \mathbb{E}[X|B], \quad \text{for all } B \in \mathcal{G} : \mathbb{P}(B) > 0. \tag{3.16}$$

In fact,

$$\begin{aligned}
\mathbb{P}(B)\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|B] &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{I}_B] \\
&= \mathbb{E}[\sum_{k=1}^{M} \mathbb{E}[X|A_k]\mathbb{I}_{A_k}\mathbb{I}_B] \\
&= \sum_{k=1}^{M} \mathbb{E}[\mathbb{E}[X|A_k]\mathbb{I}_{A_k\cap B}] \\
&= \sum_{k=1}^{M} \mathbb{E}[\frac{\mathbb{E}[X\mathbb{I}_{A_k}]}{\mathbb{P}(A_k)}\mathbb{I}_{A_k\cap B}] \\
&= \sum_{k=1}^{M} \frac{1}{\mathbb{P}(A_k)}\mathbb{E}[X\mathbb{I}_{A_k}]\mathbb{E}[\mathbb{I}_{A_k\cap B}].
\end{aligned}$$

Since $\{A_1, \dots, A_M\}$ is a partition of Ω, there exists $I \subset \{1, \dots, M\}$ such that $B = \cup_{k\in I} A_k$; hence the above sum may be restricted to $k \in I$. Since $\mathbb{E}[\mathbb{I}_{A_k\cap B}] = \mathbb{E}[\mathbb{I}_{A_k}] = \mathbb{P}(A_k)$, for $k \in I$, we obtain

$$\mathbb{P}(B)\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|B] = \sum_{k\in I} \mathbb{E}[X\mathbb{I}_{A_k}] = \mathbb{E}[X\mathbb{I}_{\cup_{k\in I}A_k}] = \mathbb{E}[X\mathbb{I}_B],$$

by which (3.16) follows.

Exercise 3.26. *What is the interpretation of* (3.16)*?*

The conditional expectation of a random variable with respect to a general σ-algebra can be constructed explicitly only in some special cases. However an abstract definition is still possible, which we give after the following theorem.

Theorem 3.11. *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and $X \in L^1(\Omega)$. If $Y_1, Y_2 \in L^1(\Omega)$ are $\mathcal{G}$-measurable and satisfy*

$$\mathbb{E}[Y_i|A] = \mathbb{E}[X|A], \quad \textit{for } i = 1, 2 \textit{ and all } A \in \mathcal{G} : \mathbb{P}(A) > 0, \tag{3.17}$$

then $Y_1 = Y_2$ a.s.

Proof. We want to prove that $\mathbb{P}(B) = 0$, where

$$B = \{\omega \in \Omega : Y_1(\omega) \neq Y_2(\omega)\}.$$

Let $B_+ = \{Y_1 > Y_2\}$ and assume $\mathbb{P}(B_+) > 0$. Then, by (3.17) and Definition 3.7,

$$\mathbb{E}[(Y_1 - Y_2)\mathbb{I}_{B_+}] = \mathbb{E}[Y_1\mathbb{I}_{B_+}] - \mathbb{E}[Y_2\mathbb{I}_{B_+}] = \mathbb{P}(B_+)(\mathbb{E}[Y_1|B_+] - \mathbb{E}[Y_2|B_+]) = 0.$$

By Theorem 3.1(iii), this is possible if and only if $(Y_1 - Y_2)\mathbb{I}_{B_+} = 0$ a.s., which entails $\mathbb{P}(B_+) = 0$. At the same fashion one proves that $\mathbb{P}(B_-) = 0$, where $\mathbb{P}(B_-) = \{Y_1 < Y_2\}$. Hence $\mathbb{P}(B) = \mathbb{P}(B_+) + \mathbb{P}(B_-) = 0$, as claimed. □

Theorem 3.12 (and Definition). *Let $X \in L^1(\Omega)$ and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. There exists a $\mathcal{G}$-measurable random variable $\mathbb{E}[X|\mathcal{G}] \in L^1(\Omega)$ such that*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|A] = \mathbb{E}[X|A], \quad \text{for all } A \in \mathcal{G} : \mathbb{P}(A) > 0. \tag{3.18}$$

The random variable $\mathbb{E}[X|\mathcal{G}]$, which by Theorem 3.11 is uniquely defined up to a null set, is called the **conditional expectation of X given the σ-algebra $\mathcal{G}$**. *If $\mathcal{G}$ is the σ-algebra generated by a random variable Y, i.e., $\mathcal{G} = \sigma(Y)$, we write $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|Y]$.*

Proof. See [26, Appendix B]. □

Remark 3.12. Following Remark 3.3, we denote by $\widetilde{\mathbb{E}}[X|\mathcal{G}]$ the conditional expectation of X in a new probability measure $\widetilde{\mathbb{P}}$, not necessarily equivalent to $\mathbb{P}$.

We conclude this section with a list of properties satisfied by the conditional expectation.

Theorem 3.13. *Let $X, Y \in L^1(\Omega)$ and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The following properties hold almost surely:*

(i) Linearity: $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}]$, *for all* $\alpha, \beta \in \mathbb{R}$;

(ii) Monotonicity: If $X \leq Y$ *then* $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$.

(iii) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$;

(iv) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$;

(v) Tower property: If $\mathcal{H} \subset \mathcal{G}$ is a sub-σ-algebra, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$;

(vi) If $\mathcal{G}$ consists of trivial events only, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$;

(vii) If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$;

(viii) Take it out what is known: If X is $\mathcal{G}$-measurable, then $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$;

(ix) Jensen's inequality: Given $f : \mathbb{R} \to \mathbb{R}$ measurable and convex there holds $\mathbb{E}[f(X)|\mathcal{G}] \geq f(\mathbb{E}[X|\mathcal{G}])$;

(x) Independence Lemma: If X is $\mathcal{G}$-measurable and $\mathcal{Y}$ is independent of $\mathcal{G}$, then for any measurable function $g : \mathbb{R}^2 \to [0, \infty)$, the function $f : \mathbb{R} \to [0, \infty)$ defined by

$$f(x) = \mathbb{E}[g(x, Y)]$$

is measurable and moreover

$$\mathbb{E}[g(X, Y)|\mathcal{G}] = f(X).$$

Proof. (i) Linearity follows easily by (3.18).

(ii) This follows easily by (3.18) and the fact that $X \leq Y \Rightarrow \mathbb{E}[X|A] \leq \mathbb{E}[Y|A]$.

(iii) Replace $A = \Omega$ into (3.18).

(iv) As (3.17) is satisfied by $Y_1 = \mathbb{E}[X|\mathcal{G}]$ and $Y_2 = X$, when X is $\mathcal{G}$-measurable we have $\mathbb{E}[X|\mathcal{G}] = X$ a.s., by uniqueness.

(v) Using (3.18), and since $\mathcal{H} \subset \mathcal{G}$, the random variables $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$, $\mathbb{E}[X|\mathcal{G}]$ and $\mathbb{E}[X|\mathcal{H}]$ satisfy

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]|A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|A], \\
&\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|A] = \mathbb{E}[X|A], \\
&\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|A] = \mathbb{E}[X|A],
\end{aligned}$$

for all $A \in \mathcal{H} : \mathbb{P}(A) > 0$. It follows that

$$\mathbb{E}[\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]|A] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|A]$$

and thus by uniqueness the claim follows.

(vi) This again follows at once by (3.18).

(vii) $\mathbb{E}[X|\mathcal{G}]$ is uniquely characterized by the identity (3.18). As X is independent of $\mathcal{G}$, the random variables X and $\mathbb{I}_A$ are independent, for all $A \in \mathcal{G}$, hence $\mathbb{E}[X|A] = \mathbb{E}[X\mathbb{I}_A]/\mathbb{P}(A) = \mathbb{E}[X]\mathbb{E}[\mathbb{I}_A]/\mathbb{P}(A) = \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X]|A]$. Hence (3.18) becomes $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|A] = \mathbb{E}[\mathbb{E}[X]|A]$, for all $A \in \mathcal{G}$ with positive probability, which implies the claim.

(viii)–(x) The proofs of these properties are more complicated and can be found in [7].

□

Remark 3.13. The meaning of (iv) and (viii) has been discussed at the beginning of the section. The meaning of property (v) (tower property) is that upon estimating X with the information in $\mathcal{G}$ and then with the information in $\mathcal{H} \subset \mathcal{G}$, the information contained in $\mathcal{G} \setminus \mathcal{H}$ is lost (property (iii) follows as a special case). The meaning of (vi) is that trivial events do not help to estimate random variables. The meaning of property (x) (independence lemma) is that, under the stated assumptions, we can compute the random variable $\mathbb{E}[g(X,Y)|\mathcal{G}]$ as if X were a deterministic constant.

Exercise 3.27 (Sol. 21)**.** *The purpose of this exercise is to show that the conditional expectation is the best estimator of a random variable when some information is given in the form of a sub-σ-algebra. Let $X \in L^1(\Omega)$ and $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Define* $\mathrm{Err} = X - \mathbb{E}[X|\mathcal{G}]$. *Show that* $\mathbb{E}[\mathrm{Err}] = 0$ *and*

$$\mathrm{Var}[\mathrm{Err}] = \min_Y \mathrm{Var}[Y - X],$$

where the minimum is taken with respect to all $\mathcal{G}$-measurable random variables Y.

3.4 Martingales

A martingale is a stochastic process which has no tendency to rise or fall. The precise definition is the following:

Definition 3.8. *A stochastic process $\{M(t)\}_{t\geq 0}$ is called a* **martingale** *relative to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ if it is adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$, $M(t) \in L^1(\Omega)$ for all $t \geq 0$, and*

$$\mathbb{E}[M(t)|\mathcal{F}(s)] = M(s), \quad \textit{for all } 0 \leq s \leq t, \tag{3.19}$$

for all $t \geq 0$.

Hence a stochastic process is martingale if the information available up to time s does not help to predict whether the stochastic process will raise or fall after time s.

Remark 3.14. If the condition (3.19) is replaced by $\mathbb{E}[M(t)|\mathcal{F}(s)] \geq M(s)$, for all $0 \leq s \leq t$, the stochastic process $\{M(t)\}_{t\geq 0}$ is called a **sub-martingale**. The interpretation is that $M(t)$ has no tendency to fall, but our expectation is that it will increase. If the condition (3.19) is replaced by $\mathbb{E}[M(t)|\mathcal{F}(s)] \leq M(s)$, for all $0 \leq s \leq t$, the stochastic process $\{M(t)\}_{t\geq 0}$ is called a **super-martingale**. The interpretation is that $M(t)$ has not tendency to rise, but our expectation is that it will decrease.

Remark 3.15. If we want to emphasize that the martingale property is satisfied with respect to the probability measure $\mathbb{P}$, we shall say that $\{M(t)\}_{t\geq 0}$ is a $\mathbb{P}$-martingale.

Since the conditional expectation of a random variable X is uniquely determined by (3.18), then the property (3.19) is satisfied if and only if

$$\mathbb{E}[M(s)\mathbb{I}_A] = \mathbb{E}[M(t)\mathbb{I}_A], \quad \text{for all } 0 \leq s \leq t \text{ and for all } A \in \mathcal{F}(s). \tag{3.20}$$

In particular, letting $A = \Omega$, we obtain that the expectation of a martingale is constant, i.e.,

$$\mathbb{E}[M(t)] = \mathbb{E}[M(0)], \quad \text{for all } t \geq 0. \tag{3.21}$$

Combining the latter result with Theorem 3.2, we obtain the following sufficient condition for no arbitrage.

Theorem 3.14. *Let a portfolio be given with value $\{V(t)\}_{t\geq 0}$. If there exists a measure $\widetilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ and a filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ such that the discounted value of the portfolio $\{V^*(t)\}_{t\geq 0}$ is a martingale, then for all $T > 0$ the portfolio is not an arbitrage in the interval $[0, T]$.*

Proof. The assumption is that

$$\widetilde{\mathbb{E}}[D(t)V(t)|\mathcal{F}(s)] = D(s)V(s), \quad \text{for all } 0 \leq s \leq t.$$

Hence, by (3.21), $\widetilde{\mathbb{E}}[D(t)V(t)] = \widetilde{\mathbb{E}}[D(0)V(0)] = \widetilde{\mathbb{E}}[V(0)]$. The result follows by Theorem 3.2. □

Theorem 3.15. *Let $\{\mathcal{F}(t)\}_{t\geq 0}$ be a non-anticipating filtration for the Brownian motion $\{W(t)\}_{t\geq 0}$. Then $\{W(t)\}_{t\geq 0}$ is a martingale relative to $\{\mathcal{F}(t)\}_{t\geq 0}$.*

Proof. The martingale property for $s = t$, i.e., $\mathbb{E}[W(t)|\mathcal{F}(t)] = W(t)$, follows by the fact $W(t)$ is $\mathcal{F}(t)$-measurable, and thus Theorem 3.13(iv) applies. For $0 \leq s < t$ we have

$$\begin{aligned}\mathbb{E}[W(t)|\mathcal{F}(s)] &= \mathbb{E}[(W(t) - W(s))|\mathcal{F}(s)] + \mathbb{E}[W(s)|\mathcal{F}(s)],\\ &= \mathbb{E}[W(t) - W(s)] + W(s) = W(s),\end{aligned}$$

where we used that $W(t) - W(s)$ is independent of $\mathcal{F}(s)$, and so $\mathbb{E}[(W(t) - W(s))|\mathcal{F}(s)] = \mathbb{E}[(W(t) - W(s))]$ by Theorem 3.13(vii), and the fact that $W(s)$ is $\mathcal{F}(s)$-measurable (and so $\mathbb{E}[W(s)|\mathcal{F}(s)] = W(s)$). $\square$

Exercise 3.28 ($\star$)**.** *Consider the stochastic process $\{Z(t)\}_{t\geq 0}$ given by*

$$Z(t) = \exp\big(\sigma W(t) - \frac{1}{2}\sigma^2 t\big),$$

where $\{W(t)\}_{t\geq 0}$ is a Brownian motion and $\sigma \in \mathbb{R}$ is a constant. Let $\{\mathcal{F}(t)\}_{t\geq 0}$ be a non-anticipating filtration for $\{W(t)\}_{t\geq 0}$. Use the definition of martingale to show that $\{Z(t)\}_{t\geq 0}$ is a martingale relative to $\{\mathcal{F}(t)\}_{t\geq 0}$.

Brownian motions are martingales, have a.s. continuous paths and have quadratic variation t in the interval $[0, t]$, see Theorem 3.9. The following theorem shows that these three properties characterize Brownian motions and is often used to prove that a given stochastic process is a Brownian motion. The proof can be found in [18].

Theorem 3.16 (Lévy characterization theorem). *Let $\{M(t)\}_{t\geq 0}$ be a martingale relative to a filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. Assume that (i) $M(0) = 0$ a.s., (ii) the paths $t \to M(t, \omega)$ are a.s. continuous and (iii) $dM(t)dM(t) = dt$. Then $\{M(t)\}_{t\geq 0}$ is a Brownian motion and $\{\mathcal{F}(t)\}_{t\geq 0}$ a non-anticipating filtration thereof.*

Exercise 3.29 (Sol. 22)**.** *Let $\{N(t)\}_{t\geq 0}$ be a Poisson process generating the filtration $\{\mathcal{F}_N(t)\}_{t\geq 0}$. Show that (i) $\{N(t)\}_{t\geq 0}$ is a sub-martingale relative to $\{\mathcal{F}_N(t)\}_{t\geq 0}$ and (ii) the so-called* **compound Poisson process** *$\{N(t) - \lambda t\}_{t\geq 0}$ is a martingale relative to$\{\mathcal{F}_N(t)\}_{t\geq 0}$, where λ is the rate of the Poisson process (see Definition 2.16).*

Exercise 3.30 ($\star$)**.** *Let $\{\mathcal{F}(t)\}_{t\in[0,T]}$ be a filtration and $\{M(t)\}_{t\in[0,T]}$ be a stochastic process adapted to $\{\mathcal{F}(t)\}_{t\in[0,T]}$ such that $M(t) \in L^1(\Omega)$, for all $t \in [0, T]$. Show that $\{M(t)\}_{t\in[0,T]}$ is a martingale if and only if there exists a $\mathcal{F}(T)$-measurable random variable $H \in L^1(\Omega)$ such that*

$$M(t) = \mathbb{E}[H|\mathcal{F}(t)].$$

Now assume that $\{Z(t)\}_{t\geq 0}$ is a martingale such that $Z(t) > 0$ a.s. and $\mathbb{E}[Z(0)] = 1$. Since martingales have constant expectation, then $\mathbb{E}[Z(t)] = 1$ for all $t \geq 0$. By Theorem 3.3, the map $\widetilde{\mathbb{P}} : \mathcal{F} \to [0, 1]$ given by

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z(T)\mathbb{I}_A], \quad A \in \mathcal{F} \tag{3.22}$$

is a probability measure equivalent to $\mathbb{P}$, for all $T > 0$. Note that $\widetilde{\mathbb{P}}$ depends on $T > 0$ and $\widetilde{\mathbb{P}} = \mathbb{P}$, for $T = 0$. The dependence on T is however not reflected in our notation. As usual, the (conditional) expectation in the probability measure $\widetilde{\mathbb{P}}$ will be denoted $\widetilde{\mathbb{E}}$. The relation between $\mathbb{E}$ and $\widetilde{\mathbb{E}}$ is revealed in the following theorem.

Theorem 3.17. *Let $\{Z(t)\}_{t\geq 0}$ be a $\mathbb{P}$-martingale relative to a filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ such that $Z(t) > 0$ a.s. and $\mathbb{E}[Z(0)] = 1$. Let $T > 0$ and let $\widetilde{\mathbb{P}}$ be the probability measure equivalent to $\mathbb{P}$ defined by (3.22). Let $t \in [0, T]$ and let X be a $\mathcal{F}(t)$-measurable random variable such that $Z(t)X \in L^1(\Omega, \mathbb{P})$. Then $X \in L^1(\Omega, \widetilde{\mathbb{P}})$ and*

$$\widetilde{\mathbb{E}}[X] = \mathbb{E}[Z(t)X]. \tag{3.23}$$

Moreover, for all $0 \leq s \leq t$ and for all random variables Y such that $Z(t)Y \in L^1(\Omega, \mathbb{P})$, there holds

$$\widetilde{\mathbb{E}}[Y|\mathcal{F}(s)] = \frac{1}{Z(s)}\mathbb{E}[Z(t)Y|\mathcal{F}(s)]. \tag{3.24}$$

Proof. As shown in Theorem 3.3, $\widetilde{\mathbb{E}}[X] = \mathbb{E}[Z(T)X]$. By Theorem 3.13(iii), Theorem 3.13(viii), and the martingale property of $\{Z(t)\}_{t\geq 0}$, we have

$$\mathbb{E}[Z(T)X] = \mathbb{E}[\mathbb{E}[Z(T)X|\mathcal{F}(t)]] = \mathbb{E}[X\mathbb{E}[Z(T)|\mathcal{F}(t)]] = \mathbb{E}[Z(t)X].$$

To prove (3.24), recall that the conditional expectation is uniquely defined (up to null sets) by (3.18). Hence the identity (3.24) follows if we show that

$$\widetilde{\mathbb{E}}[Z(s)^{-1}\mathbb{E}[Z(t)Y|\mathcal{F}(s)]\mathbb{I}_A] = \widetilde{\mathbb{E}}[Y\mathbb{I}_A],$$

for all $A \in \mathcal{F}(s)$. Since $\mathbb{I}_A$ is $\mathcal{F}(s)$-measurable, and using (3.23) at $t = s$ and with

$$X = Z(s)^{-1}\mathbb{E}[Z(t)Y\mathbb{I}_A|\mathcal{F}(s)]$$

we have

$$\begin{aligned}\widetilde{\mathbb{E}}[Z(s)^{-1}\mathbb{E}[Z(t)Y|\mathcal{F}(s)]\mathbb{I}_A] &= \widetilde{\mathbb{E}}[Z(s)^{-1}\mathbb{E}[Z(t)Y\mathbb{I}_A|\mathcal{F}(s)]] = \mathbb{E}[\mathbb{E}[Z(t)Y\mathbb{I}_A|\mathcal{F}(s)]] \\ &= \mathbb{E}[Z(t)Y\mathbb{I}_A] = \widetilde{\mathbb{E}}[Y\mathbb{I}_A],\end{aligned}$$

where in the last step we used again (3.23). The proof is complete. $\square$

3.5 Markov processes

In this section we introduce another class of stochastic processes, which will play a fundamental role in the following chapters.

Definition 3.9. *A stochastic process* $\{X(t)\}_{t\geq 0}$ *is called a* **Markov process** *with respect to the filtration* $\{\mathcal{F}(t)\}_{t\geq 0}$ *if it is adapted to* $\{\mathcal{F}(t)\}_{t\geq 0}$ *and if for every measurable function* $g : \mathbb{R} \to \mathbb{R}$ *such that* $g(X(t)) \in L^1(\Omega)$*, for all* $t \geq 0$*, there exists a measurable function* $f_g : [0,\infty) \times [0,\infty) \times \mathbb{R} \to \mathbb{R}$ *such that*

$$\mathbb{E}[g(X(t))|\mathcal{F}(s)] = f_g(t, s, X(s)), \quad \text{for all } 0 \leq s \leq t. \tag{3.25}$$

The function $f_g(t, s, \cdot)$ *is called the* **transition probability** *of* $\{X(t)\}_{t\geq 0}$ *from time* s *to time* t*. If* $f_g(t, s, x) = f_g(t - s, 0, x)$*, for all* $t \geq s$ *and* $x \in \mathbb{R}$*, we say that the Markov process is* **time-homogeneous**. *If there exists a measurable function* $p : [0,\infty) \times [0,\infty) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *such that* $y \to p(t, s, x, y)$ *is integrable for all* $(t, s, x) \in [0,\infty) \times [0,\infty) \times \mathbb{R}$ *and*

$$f_g(t, s, x) = \int_{\mathbb{R}} g(y)p(t, s, x, y)\, dy, \quad \text{for } 0 \leq s < t, \tag{3.26}$$

holds for all bounded measurable functions g*, then we call* p *the* **transition probability density** *of the Markov process.*

The interpretation is the following: for a Markov process, the conditional expectation of $g(X(t))$ at the future time t depends only on the random variable $X(s)$ at time s, and not on the behavior of the process before or after time s.

Remark 3.16. For a time-homogeneous Markov process the transition between any two different times is equivalent to a transition starting at $s = 0$.

Remark 3.17. We will say that a stochastic process is a $\mathbb{P}$-Markov process if we want to emphasize that the Markov property holds in the probability measure $\mathbb{P}$.

Exercise 3.31 (Sol. 23). *Show that the function* $f_g(t, s, x)$ *in the right hand side of* (3.25) *is given by*

$$f_g(t, s, x) = \mathbb{E}[g(X(t))|X(s) = x] \quad \text{for all } 0 \leq s \leq t. \tag{3.27}$$

Theorem 3.18. *Let* $\{X(t)\}_{t\geq 0}$ *be a Markov process with transition density* $p(t, s, x, y)$ *relative to the filtration* $\{\mathcal{F}(t)\}_{t\geq 0}$*. Assume* $X(s) = x \in \mathbb{R}$ *is a deterministic constant and that* $\mathcal{F}(s)$ *is the trivial* σ*-algebra. Then* $X(t)$ *has probability density* $f_{X(t)}$ *given by*

$$f_{X(t)}(y) = p(t, s, x, y), \quad \text{for all } t > s.$$

Proof. By definition of density, we have to show that

$$\mathbb{P}(X(t) \leq z) = \int_{-\infty}^{z} f_{X(t)}(y)\, dy,$$

see Definition 2.7. Letting $X(s) = x$ into (3.25)-(3.26) we obtain

$$\mathbb{E}[g(X(t))] = \int_{\mathbb{R}} g(y)p(t, s, x, y)\, dy.$$

Choosing $g = \mathbb{I}_{(-\infty,z]}$, we obtain

$$\mathbb{P}(X(t) \leq z) = \int_{-\infty}^{z} p(t, s, x, y)\, dy,$$

for all $z \in \mathbb{R}$, hence $f_{X(t)}(y) = p(t, s, x, y)$. $\square$

Theorem 3.19. *Let $\{\mathcal{F}(t)\}_{t\geq 0}$ be a non-anticipating filtration for the Brownian motion $\{W(t)\}_{t\geq 0}$. Then $\{W(t)\}_{t\geq 0}$ is a homogeneous Markov process relative to $\{\mathcal{F}(t)\}_{t\geq 0}$ with transition density $p(t, s, x, y) = p_*(t - s, x, y)$, where*

$$p_*(\tau, x, y) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(y-x)^2}{2\tau}}. \tag{3.28}$$

Proof. The statement holds for $s = t$, with $f_g(t, t, x) = g(x)$. For $s < t$ we write

$$\mathbb{E}[g(W(t))|\mathcal{F}(s)] = \mathbb{E}[g(W(t) - W(s) + W(s))|\mathcal{F}(s)] = \mathbb{E}[\widetilde{g}(W(s), W(t) - W(s))|\mathcal{F}(s)],$$

where $\widetilde{g}(x, y) = g(x + y)$. Since $W(t) - W(s)$ is independent of $\mathcal{F}(s)$ and $W(s)$ is $\mathcal{F}(s)$ measurable, then we can apply Theorem 3.13(x). Precisely, letting

$$f_g(t, s, x) = \mathbb{E}[\widetilde{g}(x, W(t) - W(s))],$$

we have

$$\mathbb{E}[g(W(t))|\mathcal{F}(s)] = f_g(t, s, W(s)),$$

which proves that the Brownian motion is a Markov process relative to $\{\mathcal{F}(t)\}_{t\geq 0}$. To derive the transition density we use that $Y = W(t) - W(s) \in \mathcal{N}(0, t - s)$, so that

$$\mathbb{E}[g(x + Y)] = \frac{1}{\sqrt{2\pi(t-s)}} \int_{\mathbb{R}} g(x + y) e^{-\frac{y^2}{2(t-s)}}\, dy = \frac{1}{\sqrt{2\pi(t-s)}} \int_{\mathbb{R}} g(y) e^{-\frac{(y-x)^2}{2(t-s)}}\, dy,$$

hence

$$\mathbb{E}[g(W(t))|\mathcal{F}(s)] = \left[\int_{\mathbb{R}} g(y) p_*(t - s, x, y)\, dy\right]_{x=W(s)},$$

where p_* is given by (3.28). This concludes the proof of the theorem. $\square$

Remark 3.18. According to Theorem 3.18, the random variable $x + W(t)$ has density $f_{x+W(t)}(y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}$, for all $x \in \mathbb{R}$ and $t > 0$, which is of course correct because $x + W(t) \in \mathcal{N}(x, t)$.

Exercise 3.32. *Show that, when p is given by* (3.28)*, the function*

$$u(t, x) = \int_{\mathbb{R}} g(y) p_*(t - s, x, y)\, dy \tag{3.29}$$

solves the **heat equation** *with initial datum g at time $t = s$, namely*

$$\partial_t u = \frac{1}{2}\partial_x^2 u, \quad u(s, x) = g(x), \quad t > s,\ x \in \mathbb{R}. \tag{3.30}$$

Exercise 3.33. *Let* $\{\mathcal{F}(t)\}_{t\geq 0}$ *be a non-anticipating filtration for the Brownian motion* $\{W(t)\}_{t\geq 0}$*. Show that the geometric Brownian motion*

$$S(t) = S(0)e^{\sigma W(t)+\alpha t}$$

is a homogeneous Markov process in the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ *with transition density* $p(t,s,x,y) = p_*(t-s,x,y)$*, where*

$$p_*(\tau,x,y) = \frac{1}{\sigma y\sqrt{2\pi\tau}} \exp\left\{-\frac{(\log(y/x)-\alpha\tau)^2}{2\sigma^2\tau}\right\} \mathbb{I}_{y>0}. \tag{3.31}$$

Show also that, when p *is given by* (3.31)*, the function* $v : (s,\infty)\times(0,\infty) \to \mathbb{R}$ *given by*

$$v(t,x) = \int_{\mathbb{R}} g(y)p_*(t-s,x,y)\,dy \tag{3.32}$$

satisfies

$$\partial_t v_s - (\alpha+\sigma^2/2)x\partial_x v_s - \frac{1}{2}\sigma^2 x^2 \partial_x^2 v_s = 0, \quad \textit{for } x>0,\ t>s, \tag{3.33a}$$

$$v(s,x) = g(x), \quad \textit{for } x>0. \tag{3.33b}$$

The correspondence between Markov processes and PDE's alluded to in the last two exercises is a general property which will be further discussed later in Chapter 5.

Chapter 4

Stochastic calculus

Throughout this chapter we assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the Brownian motion $\{W(t)\}_{t\geq 0}$ are given. Moreover we denote by $\{\mathcal{F}(t)\}_{t\geq 0}$ a non-anticipating filtration for the Brownian motion, e.g., $\mathcal{F}(t) = \mathcal{F}_W(t)$ (see Definition 2.15).

4.1 Introduction

So far we have studied in detail only one example of stochastic process, namely the Brownian motion $\{W(t)\}_{t\geq 0}$. In this chapter we define several other processes which are naturally derived from $\{W(t)\}_{t\geq 0}$ and which in particular are adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$. To begin with, if $f : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ is a measurable function, then we can introduce the stochastic processes

$$\{f(t, W(t))\}_{t\geq 0}, \quad \{\int_0^t f(s, W(s))\, ds\}_{t\geq 0}.$$

The integral in the second stochastic process is the standard Lebesgue integral on the s-variable. It is well-defined for instance when f is a continuous function.

The next class of stochastic processes that we want to consider are those obtained by integrating along the paths of a Brownian motion, i.e., we want to give sense to the integral

$$I(t) = \int_0^t X(s)dW(s), \tag{4.1}$$

where $\{X(t)\}_{t\geq 0}$ is a stochastic process adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$. For our purposes we need to give a meaning to $I(t)$ when $\{X(t)\}_{t\geq 0}$ has merely continuous paths a.s. (e.g., $X(t) = W(t)$). The problem now is that the integral $\int X(t)dg(t)$ is well-defined for continuous functions X (in the Riemann-Stieltjes sense) only when g is of bounded variation. As shown in Exercise 3.15, the paths of the Brownian motion are not of bounded variation, hence the definition of (4.1) requires some new ideas. We begin in the next section by defining (4.1) when $\{X(t)\}_{t\geq 0}$ is a step process. Then we shall extend the definition to stochastic processes $\{X(t)\}_{t\geq 0}$ such that

$$\{X(t)\}_{t\geq 0} \text{ is } \{\mathcal{F}(t)\}_{t\geq 0}\text{-adapted and } \mathbb{E}[\int_0^T X(t)^2\,dt] < \infty, \text{ for all } T > 0. \tag{4.2}$$

We denote by $\mathbb{L}^2[\mathcal{F}(t)]$ the family of stochastic processes satisfying (4.2). The integral (4.1) can be defined for more general processes than those in the class $\mathbb{L}^2[\mathcal{F}(t)]$, as we briefly discuss in Theorem 4.4. In particular the stochastic integral is well-defined for adapted processes with a.s. continuous paths.

4.2 The Itô integral of step processes

Given $0 = t_0 < t_1 < \cdots < t_k < t_{k+1} < \ldots$ and a sequence $X_1, X_2, \ldots$ of random variables such that, for all $j \in \mathbb{N}$, $X_j \in L^2(\Omega)$ and X_j is $\mathcal{F}(t_j)$-measurable, consider the $\{\mathcal{F}(t)\}_{t\geq 0}$-adapted step process

$$\Delta(t) = \sum_{j=0}^{\infty} \Delta(t_j)\mathbb{I}_{[t_j,t_{j+1})}, \quad \Delta(t_j) = X_j. \tag{4.3}$$

Note that $\{\Delta(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$, by the assumption $X_j \in L^2(\Omega)$, for all $j \in \mathbb{N}$. If we had to integrate $\Delta(t)$ along a stochastic process $\{Y(t)\}_{t\geq 0}$ with differentiable paths, we would have, assuming $t \in (t_k, t_{k+1})$,

$$\begin{aligned}\int_0^t \Delta(s)dY(s) &= \int_0^t \sum_{j=0}^{\infty} \Delta(t_j)\mathbb{I}_{[t_j,t_{j+1})}dY(t) = \sum_{j=0}^{k-1} \Delta(t_j)\int_{t_j}^{t_{j+1}} dY(t) + \Delta(t_k)\int_{t_k}^t dY(t)\\ &= \sum_{j=0}^{k-1} \Delta(t_j)(Y(t_{j+1}) - Y(t_j)) + \Delta(t_k)(Y(t) - Y(t_k)).\end{aligned}$$

The second line makes sense also for stochastic processes $\{Y(t)\}_{t\geq 0}$ whose paths are nowhere differentiable, and thus in particular for the Brownian motion. We then introduce the following definition.

Definition 4.1. *The* **Itô integral** *over the interval* $[0,t]$ *of a step process* $\{\Delta(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$ *is given by*

$$I(t) = \int_0^t \Delta(s)dW(s) = \sum_{j=0}^{k-1} \Delta(t_j)(W(t_{j+1}) - W(t_j)) + \Delta(t_k)(W(t) - W(t_k)),$$

where t_k *is such that* $t_k \leq t < t_{k+1}$.

Note that $\{I(t)\}_{t\geq 0}$ is a $\{F(t)\}_{t\geq 0}$-adapted stochastic process with a.s. continuous paths (in fact, the only dependence on the t variable is through $W(t)$). The following theorem collects some other important properties of the Itô integral of a step process.

Theorem 4.1. *The Itô integral of a step process $\{\Delta(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$ satisfies the following properties.*

(i) **Linearity:** *for every pair of $\{\mathcal{F}(t)\}_{t\geq 0}$-adapted step processes $\{\Delta_1(t)\}_{t\geq 0}, \{\Delta_2(t)\}_{t\geq 0}$ and real constants $c_1, c_2 \in \mathbb{R}$ there holds*

$$\int_0^t (c_1\Delta_1(s) + c_2\Delta_2(s))dW(s) = c_1 \int_0^t \Delta_1(s)dW(s) + c_2 \int_0^t \Delta_2(s)dW(s).$$

(ii) **Martingale property:** *the stochastic process $\{I(t)\}_{t\geq 0}$ is a martingale in the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. In particular, $\mathbb{E}[I(t)] = \mathbb{E}[I(0)] = \mathbb{E}[0] = 0$.*

(iii) **Quadratic variation:** *the quadratic variation of the stochastic process $\{I(t)\}_{t\geq 0}$ on the interval $[0,T]$ is independent of the sequence of partitions along which it is computed and it is given by*

$$[I,I](T) = \int_0^T \Delta^2(s)\,ds, \quad i.e., \quad dI(t)dI(t) = \Delta(t)^2 dt.$$

(iv) **Itô's isometry:** $\mathbb{E}[I^2(t)] = \mathbb{E}[\int_0^t \Delta^2(s)\,ds]$*, for all $t \geq 0$.*

Proof. The proof of (i) is straightforward. For the remaining claims, see the following theorems in [26]: Theorem 4.2.1 (martingale property), Theorem 4.2.2 (Itô's isometry), Theorem 4.2.3 (quadratic variation). Here we present the proof of (ii). First we remark that the condition $I(t) \in L^1(\Omega)$, for all $t \geq 0$, follows easily by the assumption that $\Delta(t_j) = X_j \in L^2(\Omega)$, for all $j \in \mathbb{N}$ and the Schwartz inequality. Hence we have to prove that

$$\mathbb{E}[I(t)|\mathcal{F}(s)] = I(s), \quad \text{for all } 0 \leq s \leq t.$$

There are two possibilities: (1) either $s,t \in [t_k, t_{k+1})$, for some $k \in \mathbb{N}$, or (2) there exists $l < k$ such that $s \in [t_l, t_{l+1})$ and $t \in [t_k, t_{k+1})$. We assume that (2) holds, the proof in the case (1) being simpler. We write

$$\begin{aligned} I(t) &= \sum_{j=0}^{l-1} \Delta(t_j)(W(t_{j+1}) - W(t_j)) + \Delta(t_l)(W(t_{l+1}) - W(t_l)) \\ &\quad + \sum_{j=l+1}^{k-1} \Delta(t_j)(W(t_{j+1}) - W(t_j)) + \Delta(t_k)(W(t) - W(t_k)) \\ &= I(t_{l+1}) + \int_{t_{l+1}}^t \Delta(u)dW(u). \end{aligned}$$

Taking the conditional expectation of $I(t_{l+1})$ we obtain

$$\mathbb{E}[I(t_{l+1})|\mathcal{F}(s)] = \sum_{j=0}^{l-1} \mathbb{E}[\Delta(t_j)(W(t_{j+1}) - W(t_j))|\mathcal{F}(s)] + \mathbb{E}[\Delta(t_l)(W(t_{l+1}) - W(t_l))|\mathcal{F}(s)].$$

As $t_{l-1} < s$, all random variables in the sum in the right hand side of the latter identity are $\mathcal{F}(s)$-measurable. Hence, by Theorem 3.13(iv),

$$\sum_{j=0}^{l-1} \mathbb{E}[\Delta(t_j)(W(t_{j+1}) - W(t_j))|\mathcal{F}(s)] = \sum_{j=0}^{l-1} \Delta(t_j)(W(t_{j+1}) - W(t_j)).$$

Similarly,

$$\begin{aligned}
\mathbb{E}[\Delta(t_l)(W(t_{l+1}) - W(t_l))|\mathcal{F}(s)] &= \mathbb{E}[\Delta(t_l)W(t_{l+1})|\mathcal{F}(s)] - \mathbb{E}[\Delta(t_l)W(t_l)|\mathcal{F}(s)] \\
&= \Delta(t_l)\mathbb{E}[W(t_{l+1})|\mathcal{F}(s)] - \Delta(t_l)W(t_l) \\
&= \Delta(t_l)W(s) - \Delta(t_l)W(t_l),
\end{aligned}$$

where for the last equality we used that $\{W(t)\}_{t\geq 0}$ is a martingale in the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. Hence

$$\mathbb{E}[I(t_{l+1})|\mathcal{F}(s)] = \sum_{j=0}^{l-1} \Delta(t_j)(W(t_{j+1}) - W(t_j)) + \Delta(t_l)(W(s) - W(t_l)) = I(s).$$

To conclude the proof we have to show that

$$\mathbb{E}[\int_{t_{l+1}}^{t} \Delta(u)dW(u)] = \sum_{j=l+1}^{k-1} \mathbb{E}[\Delta(t_j)(W(t_{j+1})-W(t_j))|\mathcal{F}(s)] + \mathbb{E}[\Delta(t_k)(W(t)-W(t_k))|\mathcal{F}(s)] = 0.$$

To prove this, we first note that, as before,

$$\mathbb{E}[\Delta(t_j)(W(t_{j+1}) - W(t_j))|\mathcal{F}(t_j)] = \Delta(t_j)W(t_j) - \Delta(t_j)W(t_j) = 0.$$

Moreover, since $\mathcal{F}(s) \subset \mathcal{F}(t_j)$, for $j = l+1, \dots, k-1$, as using Theorem 3.13(v),

$$\mathbb{E}[\Delta(t_j)(W(t_{j+1}) - W(t_j))] = \mathbb{E}[\mathbb{E}[\Delta(t_j)(W(t_{j+1}) - W(t_j))|\mathcal{F}(t_j)]|\mathcal{F}(s)] = \mathbb{E}[0|\mathcal{F}(s)] = 0.$$

At the same fashion, since $\mathcal{F}(s) \subset \mathcal{F}(t_k)$, we have

$$\mathbb{E}[\Delta(t_k)(W(t) - W(t_k))|\mathcal{F}(s)] = \mathbb{E}[\mathbb{E}[\Delta(t_k)(W(t) - W(t_k))|\mathcal{F}(t_k)]|\mathcal{F}(s)] = 0.$$

□

Next we show that any stochastic process can be approximated, in a suitable sense, by step processes.

Theorem 4.2. *Let $\{X(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$. Then for all $T > 0$ there exists a sequence of step processes $\{\{\Delta_n^T(t)\}_{t\geq 0}\}_{n\in\mathbb{N}}$ such that $\Delta_n^T(t) \in \mathbb{L}^2[\mathcal{F}(t)]$ for all $n \in \mathbb{N}$ and*

$$\lim_{n\to\infty} \mathbb{E}[\int_0^T |\Delta_n^T(t) - X(t)|^2\, dt] = 0. \tag{4.4}$$

Proof. For simplicity we argue under the stronger assumption that the stochastic process $\{X(t)\}_{t\geq 0}$ is bounded and with continuous paths, namely

$$\omega \to X(t,\omega) \text{ is bounded in } \Omega, \text{ for all } t \geq 0,$$
$$t \to X(t,\omega) \text{ is continuous for all } \omega \in \Omega \text{ and } t \geq 0.$$

Now consider the partition of $[0,T]$ given by

$$0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = T, \quad t_j^{(n)} = \frac{jT}{n}$$

and define

$$\Delta_n^T(t) = \sum_{j=0}^{n-1} X(t_k^{(n)}) \mathbb{I}_{[t_k^{(n)}, t_{k+1}^{(n)})}, \quad t \geq 0,$$

see Figure 4.1. Let us show that $\{\Delta_n^T(t)\}_{t\geq 0}$ is adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$. This is obvious for $t \geq T$

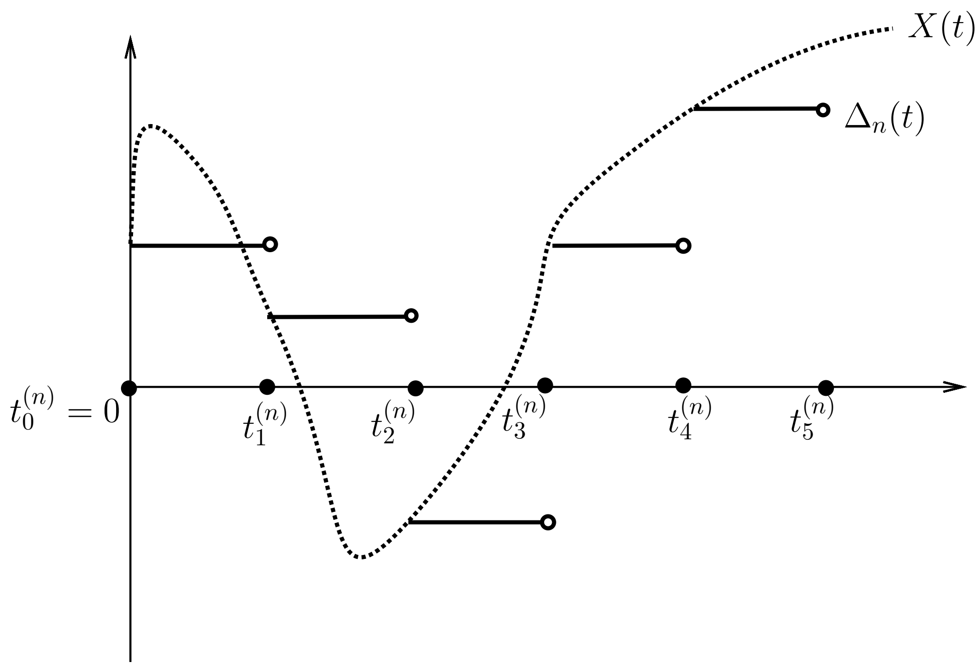


Figure 4.1: A step process approximating a general stochastic process

(since the step process is identically zero for $t \geq T$). For $t \in [0,T)$ we have $\Delta_n^T(t) = X(t_k^{(n)})$, for $t \in [t_k^{(n)}, t_{k+1}^{(n)})$, hence

$$\mathcal{F}_{\Delta_n^T(t)} = \mathcal{F}_{X(t_k^{(n)})} \underset{(*)}{\subset} \mathcal{F}(t_k^{(n)}) \underset{(**)}{\subset} \mathcal{F}(t),$$

where in (*) we used that $\{X(t)\}_{t\geq 0}$ is adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$ and in (**) the fact that $t_k^{(n)} < t$. Moreover,

$$\lim_{n\to\infty} \Delta_n^T(t) = X(t), \quad \text{for all } \omega \in \Omega,$$

by the assumed continuity of the paths of $\{X(t)\}_{t\geq 0}$. For the next step we use the dominated convergence theorem, see Remark 3.2. Since $\Delta_n^T(t)$ and $X(t)$ are bounded on $[0,T]\times\Omega$, there exists a constant C_T such that $|\Delta_n(t) - X(t)|^2 \leq C_T$. Hence we may move the limit on the left hand side of (4.4) across the expectation and integral operators and conclude that

$$\lim_{n\to\infty} \mathbb{E}[\int_0^T |\Delta_n^T(t) - X(t)|^2\, dt] = \mathbb{E}[\int_0^T \lim_{n\to\infty} |\Delta_n^T(t) - X(t)|^2\, dt] = 0,$$

as claimed. □

4.3 Itô's integral of general stochastic processes

The Itô integral of a general stochastic process is defined as the limit of the Itô integral along a sequence of approximating step processes (in the sense of Theorem 4.2). The precise definition is the following.

Theorem 4.3 (and Definition). *Let $\{X(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$, $T > 0$ and $\{\{\Delta_n^T(t)\}_{t\geq 0}\}_{n\in\mathbb{N}}$ be a sequence of $\mathbb{L}^2[\mathcal{F}(t)]$-step processes converging to $\{X(t)\}_{t\geq 0}$ in the sense of (4.4). Let*

$$I_n(T) = \int_0^T \Delta_n^T(s)dW(s).$$

Then there exists a random variable $I(T)$ such that

$$\|I_n(T) - I(T)\|_2 := \sqrt{\mathbb{E}[|I_n(T) - I(T)|^2]} \to 0, \quad \text{as } n \to \infty.$$

The random variable $I(T)$ is independent of the sequence of $\mathbb{L}^2[\mathcal{F}(t)]$-step processes converging to $\{X(t)\}_{t\geq 0}$. $I(T)$ is called **Itô's integral** *of $\{X(t)\}_{t\geq 0}$ on the interval $[0,T]$ and denoted by*

$$I(T) = \int_0^T X(s)dW(s).$$

Proof. By Itô's isometry,

$$\mathbb{E}[|I_n(T) - I_m(T)|^2] = \mathbb{E}[\int_0^T |\Delta_n^T(s) - \Delta_m^T(s)|^2 ds].$$

We have

$$\begin{aligned}\mathbb{E}[\int_0^T |\Delta_n^T(s) - \Delta_m^T(s)|^2 ds] &\leq 2\mathbb{E}[\int_0^T |\Delta_n^T(s) - X(s)|^2 ds] \\ &\quad + 2\mathbb{E}[\int_0^T |\Delta_m^T(s) - X(s)|^2 ds] \to 0 \text{ as } n,m\to\infty.\end{aligned}$$

It follows that $\{I_n(T)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in the norm $\|\cdot\|_2$. As mentioned in Remark 3.5, the norm $\|\cdot\|_2$ is complete, i.e., Cauchy sequences converge. This proves the existence of $I(T)$ such that $\|I_n(T)-I(T)\|_2\to 0$. To prove that the limit is the same along any sequence of $\mathbb{L}^2[\mathcal{F}(t)]$-step processes converging to $\{X(t)\}_{t\geq 0}$, assume that $\{\{\Delta_n(t)\}_{t\geq0}\}_{n\in\mathbb{N}}$, $\{\{\widetilde{\Delta}_n(t)\}_{t\geq0}\}_{n\in\mathbb{N}}$ are two such sequences and denote

$$I_n(T)=\int_0^T\Delta_n(s)dW(s),\quad \widetilde{I}_n(t)=\int_0^T\widetilde{\Delta}_n(s)dW(s).$$

Then, using (i), (iv) in Theorem 4.1, we compute

$$\begin{aligned}\mathbb{E}[(I_n(T)-\widetilde{I}_n(T))^2]&=\mathbb{E}[(\int_0^T(\Delta_n(s)-\widetilde{\Delta}_n(s))dW(s))^2]=\mathbb{E}[\int_0^T|\Delta_n(s)-\widetilde{\Delta}_n(s)|^2ds]\\&\leq 2\mathbb{E}[\int_0^T|\Delta_n(s)-X(s)|^2ds]+2\mathbb{E}[\int_0^T|\widetilde{\Delta}_n(s)-X(s)|^2ds]\to 0,\\&\text{as } n\to\infty,\end{aligned}$$

which proves that $I_n(T)$ and $\widetilde{I}_n(T)$ have the same limit. This completes the proof of the theorem. □

As a way of example, we compute the Itô integral of the Brownian motion. We claim that, for all $T>0$,

$$\int_0^T W(t)dW(t)=\frac{W^2(T)}{2}-\frac{T}{2}. \tag{4.5}$$

To prove the claim, we approximate the Brownian motion by the sequence of step processes introduced in the proof of Theorem 4.2. Hence we define

$$\Delta_n^T(t)=\sum_{j=0}^{n-1}W(\frac{jT}{n})\mathbb{I}_{[\frac{jT}{n},\frac{j+1}{n}T)}.$$

By definition

$$I_n(T)=\int_0^T\Delta_n^T(t)dW(t)=\sum_{j=0}^{n-1}W(\frac{jT}{n})[W(\frac{(j+1)T}{n})-W(\frac{jT}{n})].$$

To simplify the notation we let $W_j=W(jT/n)$. Hence our goal is to prove

$$\mathbb{E}[\left|\sum_{j=0}^{n-1}W_j(W_{j+1}-W_j)-\frac{W^2(T)}{2}+\frac{T}{2}\right|^2]\to 0,\quad \text{as } n\to\infty. \tag{4.6}$$

We prove below that the sum within the expectation can be rewritten as

$$\sum_{j=0}^{n-1}W_j(W_{j+1}-W_j)=\frac{1}{2}W(T)^2-\frac{1}{2}\sum_{j=0}^{n-1}(W_{j+1}-W_j)^2. \tag{4.7}$$

Hence (4.6) is equivalent to

$$\frac{1}{4}\mathbb{E}[\left|\sum_{j=0}^{n-1}(W_{j+1}-W_j)^2 - T\right|^2] \to 0, \quad \text{as } n\to\infty,$$

which holds true by the already proven fact that $[W,W](T) = T$, see Theorem 3.9. It remains to establish (4.7). Since $W(T) = W_n$, we have

$$\begin{aligned}
\frac{W(T)}{2} - \frac{1}{2}\sum_{j=0}^{n-1}(W_{j+1}-W_j)^2 &= \frac{1}{2}W_n^2 - \frac{1}{2}\sum_{j=0}^{n-1}W_{j+1}^2 - \frac{1}{2}\sum_{j=0}^{n-1}W_j^2 + \sum_{j=0}^{n-1}W_jW_{j+1}\\
&= -\frac{1}{2}\sum_{j=0}^{n-2}W_{j+1}^2 - \frac{1}{2}\sum_{j=1}^{n-1}W_j^2 + \sum_{j=1}^{n-1}W_jW_{j+1}\\
&= -\sum_{j=1}^{n-1}W_j^2 + \sum_{j=1}^{n-1}W_jW_{j+1} = \sum_{j=1}^{n-1}W_j(W_{j+1}-W_j)\\
&= \sum_{j=0}^{n-1}W_j(W_{j+1}-W_j).
\end{aligned}$$

The proof of (4.5) is complete.

Exercise 4.1 ($\star$). *Use the definition of Itô's integral to prove that*

$$TW(T) = \int_0^T W(t)dt + \int_0^T t\,dW(t). \tag{4.8}$$

The Itô integral can be defined under weaker assumptions on the integrand stochastic process than those considered so far. As this fact will be important in the following sections, it is worth to briefly discuss it. Let $\mathcal{M}^2$ denote the set of $\{\mathcal{F}(t)\}_{t\geq 0}$-adapted stochastic processes $\{X(t)\}_{t\geq 0}$ such that $\int_0^T X(t)^2\,dt$ is bounded a.s. for all $T>0$ (of course, $\mathbb{L}^2[\mathcal{F}(t)]\subset\mathcal{M}^2$).

Theorem 4.4 (and Definition). *For every process $\{X(t)\}_{t\geq 0}\in\mathcal{M}^2$ and $T>0$ there exists a sequence of step processes $\{\{\Delta_n^T(t)\}_{t\geq 0}\}_{n\in\mathbb{N}}\subset\mathcal{M}^2$ such that*

$$\lim_{n\to\infty}\int_0^T |X(s)-\Delta_n^T(s)|^2\,ds \to 0 \quad a.s.$$

and

$$\int_0^T \Delta_n(t)dW(t)$$

converges in probability as $n\to\infty$. The limit is independent of the sequence of step processes converging to $\{X(t)\}_{t\geq 0}$ and is called the Itô integral of the process $\{X(t)\}_{t\geq 0}$ in the interval $[0,T]$. If $\{X(t)\}_{t\geq 0}\in\mathbb{L}^2[\mathcal{F}(t)]$, the Itô integral just defined coincides (a.s.) with the one defined in Theorem 4.3.

For the proof of the previous theorem, see [1, Sec. 4.4]. We remark that Theorem 4.4 implies that all $\{\mathcal{F}(t)\}_{t\geq 0}$-adapted stochastic processes with a.s. continuous paths are Itô integrable. In fact, if $\{X(t)\}_{t\geq 0}$ has a.s. continuous paths, then for all $T > 0$, there exists $C_T(\omega)$ such that $\sup_{t\in[0,T]} |X(t,\omega)| \leq C_T(\omega)$ a.s. Hence

$$\int_0^T |X(s)|^2 \, ds \leq T C_T^2(\omega), \quad \text{a.s.}$$

and thus Theorem 4.4 applies. The case of stochastic processes with a.s. continuous paths covers all the applications in the following chapters, hence we shall restrict to it from now on.

Definition 4.2. *We define $\mathcal{C}^0[\mathcal{F}(t)]$ to be the space of all $\{\mathcal{F}(t)\}_{t\geq 0}$-adapted stochastic processes $\{X(t)\}_{t\geq 0}$ with a.s. continuous paths.*

In particular, if $\{X(t)\}_{t\geq 0}, \{Y(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$, then for all continuous functions f the process $\{f(t, X(t), Y(t))\}_{t\geq 0}$ belongs to $\mathcal{C}^0[\mathcal{F}(t)]$ and thus it is Itô integrable.

The properties listed in Theorem 4.1 carry over to the Itô integral of a general stochastic process. For easy reference, we rewrite these properties in the following theorem, together with an additional property, the martingale representation theorem.

Theorem 4.5. *Let $\{X(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$. Then the Itô integral*

$$I(t) = \int_0^t X(s) dW(s) \tag{4.9}$$

satisfies the following properties for all $t \geq 0$.

(0) $\{I(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$. *If* $\{X(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$, *then* $\{I(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$.

(i) **Linearity:** *For all stochastic processes $\{X_1(t)\}_{t\geq 0}, \{X_2(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$ and real constants $c_1, c_2 \in \mathbb{R}$ there holds*

$$\int_0^t (c_1 X_1(s) + c_2 X_2(s)) dW(s) = c_1 \int_0^t X_1(s) dW(s) + c_2 \int_0^t X_2(s) dW(s).$$

(ii) **Martingale property:** *If $\{X(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$, the stochastic process $\{I(t)\}_{t\geq 0}$ is a martingale in the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. In particular, $\mathbb{E}[I(t)] = \mathbb{E}[I(0)] = 0$, for all $t \geq 0$.*

(iii) **Quadratic variation:** *For all $T > 0$, the quadratic variation of the stochastic process $\{I(t)\}_{t\geq 0}$ on the interval $[0,T]$ is independent of the sequence of partitions along which it is computed and it is given by*

$$[I, I](T) = \int_0^T X^2(s) \, ds, \quad \text{i.e.,} \quad dI(t)dI(t) = X^2(t)dt \tag{4.10}$$

(iv) **Itô's isometry:** *If* $\{X(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$, *then* $\mathrm{Var}[I(t)] = \mathbb{E}[I^2(t)] = \mathbb{E}[\int_0^t X^2(s)\, ds]$, *for all* $t \geq 0$.

(v) **Martingale representation theorem:** *Let* $\{M(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}_W(t)]$, *for all* $t \geq 0$, *be a martingale stochastic process relative to the filtration* $\{\mathcal{F}_W(t)\}_{t\geq 0}$. *Then there exists a stochastic process* $\{\Gamma(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}_W(t)]$ *such that*

$$M(t) = M(0) + \int_0^t \Gamma(s) dW(s), \quad \textit{for all } t \geq 0.$$

Proof of (ii). By (iv) and the Schwartz inequality, $\mathbb{E}[I(t)] \leq \sqrt{\mathbb{E}[I^2(t)]} < \infty$. According to (3.20), it now suffices to show that

$$\mathbb{E}[I(t)\mathbb{I}_A] = \mathbb{E}[I(s)\mathbb{I}_A], \quad \text{for all } A \in \mathcal{F}(s).$$

Let $\{\{I_n(t)\}_{t\geq 0}\}_{n\in\mathbb{N}}$ be a sequence of Itô integrals of step processes which converges to $\{I(t)\}_{t\geq 0}$ in $L^2(\Omega)$, uniformly in compact intervals of time (see Theorem 4.3). Since $\{I_n(t)\}_{t\geq 0}$ is a martingale for each $n \in \mathbb{N}$, see Theorem 4.1, then

$$\mathbb{E}[I_n(t)\mathbb{I}_A] = \mathbb{E}[I_n(s)\mathbb{I}_A], \quad \text{for all } A \in \mathcal{F}(s).$$

Hence the claim follows if we show that $\mathbb{E}[I_n(t)\mathbb{I}_A] \to \mathbb{E}[I(t)\mathbb{I}_A]$, for all $t \geq 0$. Using the Schwarz inequality (3.3), we have

$$\begin{aligned} \mathbb{E}[(I_n(t) - I(t))\mathbb{I}_A] &\leq \sqrt{\mathbb{E}[(I_n(t) - I(t))^2]\mathbb{E}[\mathbb{I}_A]} \leq \|I_n(t) - I(t)\|_2 \sqrt{\mathbb{P}(A)} \\ &\leq \|I_n(t) - I(t)\|_2 \to 0, \quad \text{as } n \to \infty, \end{aligned}$$

and the claim follows. □

Remark 4.1. Note carefully that the martingale property (ii) requires $\{X(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$. A stochastic process in $\mathcal{C}^0[\mathcal{F}(t)] \setminus \mathbb{L}^2[\mathcal{F}(t)]$ is not a martingale in general (although it is a **local martingale**, see [1]).

Remark 4.2. For a proof of the martingale representation theorem see for instance Theorem 4.3.4 in [22]. This important result will be used in Chapter 6 to show the existence of hedging portfolios of European derivatives, see Theorem 6.2. Note carefully that the filtration used in the martingale representation theorem must be the one generated by the Brownian motion.

Exercise 4.2. *Prove the following generalization of Itô's isometry. Let* $\{X(t)\}_{t\geq 0}, \{Y(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)] \cap \mathbb{L}^2[\mathcal{F}(t)]$ *and denote by* $I_X(t), I_Y(t)$ *their Itô integral over the interval* $[0,t]$. *Then*

$$\mathrm{Cov}(I_X(t), I_Y(t)) = \mathbb{E}[\int_0^t X(s)Y(s)\, ds].$$

We conclude this section by introducing the "differential notation" for stochastic integrals, which means that instead of (4.9) we write

$$dI(t) = X(t)dW(t).$$

For instance, the identities (4.5), (4.8) are also expressed as

$$d(W^2(t)) = dt + 2W(t)dW(t), \quad d(tW(t)) = W(t)dt + tdW(t).$$

The differential notation is very useful to provide informal proofs in stochastic calculus. For instance, using $dI(t) = X(t)dW(t)$, and $dW(t)dW(t) = dt$, see (3.12), we obtain the following simple "proof" of Theorem 4.5(iii):

$$dI(t)dI(t) = X(t)dW(t)X(t)dW(t) = X^2(t)dW(t)dW(t) = X^2(t)dt.$$

4.4 Diffusion processes

Now that we know how to integrate along the paths of a Brownian motion, we can define a new class of stochastic processes.

Definition 4.3. *Given $\{\alpha(t)\}_{t\geq 0}, \{\sigma(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$, the stochastic process $\{X(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$ given by*

$$X(t) = X(0) + \int_0^t \sigma(s)dW(s) + \int_0^t \alpha(s)\, ds, \quad t \geq 0 \tag{4.11}$$

is called **diffusion process** *(or* **Itô process***) with* **diffusion rate** $\{\sigma(t)\}_{t\geq 0}$ *and* **drift rate** $\{\alpha(t)\}_{t\geq 0}$.

We denote diffusion processes also as

$$dX(t) = \sigma(t)dW(t) + \alpha(t)dt. \tag{4.12}$$

Note that

$$dX(t)dX(t) = \sigma^2(t)dW(t)dW(t) + \alpha^2(t)dtdt + \sigma(t)\alpha(t)dW(t)dt$$

and thus, by (3.11), (3.12) and (3.14), we obtain

$$dX(t)dX(t) = \sigma^2(t)dt,$$

which means that the quadratic variation of the diffusion process (4.12) is given by

$$[X, X](t) = \int_0^t \sigma^2(s)\, ds, \quad t \geq 0.$$

Thus the squared of the diffusion rate in a diffusion process is the rate of quadratic variation of the diffusion process. Furthermore, assuming $\{\sigma(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$, we have

$$\mathbb{E}[\int_0^t \sigma(s)dW(s)] = 0.$$

Hence the term $\int_0^t \alpha(s)ds$ is the only one contributing to the evolution of the average of $\{X(t)\}_{t\geq 0}$, which is the reason to call $\alpha(t)$ the drift rate of the diffusion process (if $\alpha = 0$ and $\{\sigma(t)\}_{t\geq 0} \in \mathbb{L}^2[\mathcal{F}(t)]$, the diffusion process is a martingale, as it follows by Theorem 4.5(ii)). Finally, the integration along the paths of the diffusion process (4.12) is defined as

$$\int_0^t Y(s)dX(s) := \int_0^t Y(s)\sigma(s)dW(s) + \int_0^t Y(s)\alpha(s)ds, \tag{4.13}$$

for all $\{Y(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$.

The product rule in stochastic calculus

Recall that if $f, g : \mathbb{R} \to \mathbb{R}$ are two differentiable functions, the product (or Leibnitz) rule of ordinary calculus states that

$$(fg)' = f'g + fg',$$

and thus

$$fg(t) = fg(0) + \int_0^t (g(s)df(s) + f(s)dg(s)).$$

Can this rule be true in stochastic calculus when f and g are general diffusion processes with continuous paths? The answer is clearly no. In fact, letting for instance $f(t) = g(t) = W(t)$, Leibnitz's rule give us the relation $d(W^2(t)) = 2W(t)dW(t)$, while we have seen before that the correct formula in Itô's calculus is $d(W^2(t)) = 2W(t)dW(t) + t$. The correct product rule in Itô's calculus is the following.

Theorem 4.6. *Let $\{X_1(t)\}_{t\geq 0}$ and $\{X_2(t)\}_{t\geq 0}$ be the diffusion processes*

$$dX_i(t) = \sigma_i(t)dW(t) + \alpha_i(t)dt.$$

Then $\{X_1(t)X_2(t)\}_{t\geq 0}$ is the diffusion process given by

$$d(X_1(t)X_2(t)) = X_2(t)dX_1(t) + X_1(t)dX_2(t) + \sigma_1(t)\sigma_2(t)dt. \tag{4.14}$$

Exercise 4.3 (Sol. 24)**.** *Prove the theorem in the case that α_i and σ_i are deterministic constants and $X_i(0) = 0$, for $i = 1, 2$.*

Recall that the correct way to interpret (4.14) is

$$X_1(t)X_2(t) = X_1(0)X_2(0) + \int_0^t X_2(s)dX_1(s) + \int_0^t X_1(s)dX_2(s) + \int_0^t \sigma_1(s)\sigma_2(s)ds, \tag{4.15}$$

where the integrals along the paths of the processes $\{X_i(t)\}_{t\geq 0}$ are defined as in (4.13). All integrals in (4.15) are well-defined, since the integrand stochastic processes have a.s. continuous paths. We also remark that, since

$$\begin{aligned}
dX_1(t)dX_2(t) &= (\sigma_1(t)dW(t) + \alpha_1(t)dt)(\sigma_2(t)dW(t) + \alpha_2(t)dt) \\
&= \sigma_1(t)\sigma_2(t)dW(t)dW(t) + (\alpha_1(t)\sigma_2(t) + \alpha_2(t)\sigma_1(t))dW(t)dt + \alpha_1(t)\alpha_2(t)dtdt \\
&= \sigma_1(t)\sigma_2(t)dt,
\end{aligned}$$

then we may rewrite (4.14) as

$$d(X_1(t)X_2(t)) = X_2(t)dX_1(t) + X_1(t)dX_2(t) + dX_1(t)dX_2(t), \tag{4.16}$$

which is somehow easier to remember. Going back to the examples considered in Section 4.3, the Itô product rule gives

$$d(W^2(t)) = W(t)dW(t) + W(t)dW(t) + dW(t)dW(t) = 2W(t)dW(t) + dt,$$

$$d(tW(t)) = tdW(t) + W(t)dt + dW(t)dt = tdW(t) + W(t)dt,$$

in agreement with our previous calculations, see (4.5) and (4.8).

The chain rule in stochastic calculus

Next we consider the generalization to Itô's calculus of the chain rule. Let us first recall how the chain rule works in ordinary calculus. Assume that $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are differentiable functions. Then

$$\frac{d}{dt}f(t, g(t)) = \partial_t f(t, g(t)) + \partial_x f(t, g(t))\frac{d}{dt}g(t),$$

by which we derive

$$f(t, g(t)) = f(0, g(0)) + \int_0^t \partial_s f(s, g(s))\, ds + \int_0^t \partial_x f(s, g(s))\, dg(s).$$

Can this formula be true in stochastic calculus when g is a general diffusion process with continuous paths? The answer is clearly no. In fact by setting $f(t, x) = x^2$, $g(t) = W(t)$ and $t = T$ in the previous formula we obtain

$$W^2(T) = 2\int_0^T W(t)dW(t), \quad \text{i.e.,} \quad \int_0^T W(t)dW(t) = \frac{W^2(T)}{2},$$

while the Itô integral of the Brownian motion is given by (4.5). The correct formula for the chain rule in stochastic calculus is given in the following theorem.

Theorem 4.7. *Let $f : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, $f = f(t, x)$, be a C^1 function such that $\partial_x^2 f$ is continuous and let $\{X(t)\}_{t\geq 0}$ be the diffusion process $dX(t) = \sigma(t)dW(t) + \alpha(t)dt$. Then* **Itô's formula** *holds:*

$$df(t, X(t)) = \partial_t f(t, X(t))\, dt + \partial_x f(t, X(t))\, dX(t) + \frac{1}{2}\partial_x^2 f(t, X(t))\, dX(t)dX(t), \tag{4.17}$$

i.e.,

$$df(t, X(t)) = \partial_t f(t, X(t))\, dt + \partial_x f(t, X(t))\, (\sigma(t)dW(t) + \alpha(t)dt) + \frac{1}{2}\partial_x^2 f(t, X(t))\sigma^2(t)\, dt. \tag{4.18}$$

For instance, letting $X(t) = W(t)$ and $f(t, x) = x^2$, we obtain $d(W^2(t)) = 2W(t)dW(t) + dt$, i.e., (4.5), while for $f(t, x) = tx$ we obtain $d(tW(t)) = W(t)dt + tdW(t)$, which is (4.8). In fact, the proof of Theorem 4.7 is similar to proof of (4.5) and (4.8). We omit the details (see [26, Theorem 4.4.1] for a sketch of the proof).

Recall that (4.18) is a shorthand for

$$f(t, X(t)) = f(0, X(0)) + \int_0^t (\partial_t f + \alpha(s)\partial_x f + \frac{1}{2}\sigma^2(s)\partial_x^2 f)(s, X(s))\, ds + \int_0^t \partial_x f(s, X(s))dW(s).$$

All integrals in the right hand side of the previous equation are well defined, as the integrand stochastic processes have continuous paths. We conclude with the generalization of Itô's formula to functions of several random variables, which again we give without proof.

Theorem 4.8. *Let $f : \mathbb{R}\times\mathbb{R}^N \to \mathbb{R}$ be a C^1 function such that $f = f(t, x)$ is twice continuously differentiable on the variable $x \in \mathbb{R}^N$. Let $\{X_1(t)\}_{t\geq 0}, \dots, \{X_N(t)\}_{t\geq 0}$ be diffusion processes and let $X(t) = (X_1(t), \dots, X_N(t))$. Then there holds:*

$$\begin{aligned} df(t, X(t)) &= \partial_t f(t, X(t))\, dt + \sum_{i=1}^N \partial_{x^i} f(t, X(t))\, dX_i(t) \\ &\quad + \frac{1}{2}\sum_{i,j=1}^N \partial_{x^i}\partial_{x^j} f(t, X(t))\, dX_i(t)dX_j(t). \end{aligned} \tag{4.19}$$

For instance, for $N = 2$ and letting $f(t, x_1, x_2) = x_1x_2$ into (4.19), we obtain the Itô product rule (4.16).

Remark 4.3. Let $\{X(t)\}_{t\geq 0}$, $\{Y(t)\}_{t\geq 0}$ be diffusion processes and define the complex-valued stochastic process $\{Z(t)\}_{t\geq 0}$ by $Z(t) = X(t) + iY(t)$. Then any stochastic process of the form $g(t, Z(t))$ can be written in the form $f(t, X(t), Y(t))$, where $f(t, x, y) = g(t, x + iy)$. Hence $dg(t, Z(t))$ can be computed using Theorem 4.8. An application to this formula is given in Exercise 4.8 below.

The following exercises help to get familiar with the rules of stochastic calculus.

Exercise 4.4 (Sol. 25)**.** *Let $\{W_1(t)\}_{t\geq 0}$, $\{W_2(t)\}_{t\geq 0}$ be Brownian motions. Assume that there exists a constant $\rho \in [-1,1]$ such that $dW_1(t)dW_2(t) = \rho dt$. Show that ρ is the correlation of the two Brownian motions at time t. Assuming that $\{W_1(t)\}_{t\geq 0}$, $\{W_2(t)\}_{t\geq 0}$ are independent, compute $\mathbb{P}(W_1(t) > W_2(s))$, for all $s,t > 0$.*

Exercise 4.5 (Sol. 26)**.** *Consider the stochastic process $\{X(t)\}_{t\geq 0}$ defined by $X(t) = W(t)^3 - 3tW(t)$. Show that $\{X(t)\}_{t\geq 0}$ is a martingale and find a process $\{\Gamma(t)\}_{t\geq 0}$ adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ such that*

$$X(t) = X(0) + \int_0^t \Gamma(s)dW(s).$$

(The existence of the process $\{\Gamma(t)\}_{t\geq 0}$ is ensured by Theorem 4.5(v).)

Exercise 4.6 (Sol. 27)**.** *Let $\{\theta(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$ and define the stochastic process $\{Z(t)\}_{t\geq 0}$ by*

$$Z(t) = \exp\left(-\int_0^t \theta(s)dW(s) - \frac{1}{2}\int_0^t \theta^2(s)ds\right).$$

Show that

$$Z(t) = 1 - \int_0^t \theta(s)Z(s)\, dW(s).$$

Processes of the form considered in Exercise 4.6 are fundamental in mathematical finance. In particular, it is important to know whether $\{Z(t)\}_{t\geq 0}$ is a martingale. By Exercise 4.6 and Theorem 4.5(ii), $\{Z(t)\}_{t\geq 0}$ is a martingale if $\theta(t)Z(t) \in \mathbb{L}^2[\mathcal{F}(t)]$, which is however difficult in general to verify directly. The following condition, known as **Novikov's condition**, is more useful in the applications, as it involves only the given process $\{\theta(t)\}_{t\geq 0}$. The proof can be found in [17].

Theorem 4.9. *Let $\{\theta(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$ satisfy*

$$\mathbb{E}[\exp(\frac{1}{2}\int_0^T \theta(t)^2\, dt)] < \infty, \quad \textit{for all } T > 0. \tag{4.20}$$

Then the stochastic process $\{Z(t)\}_{t\geq 0}$ given by

$$Z(t) = \exp\left(-\int_0^t \theta(s)dW(s) - \frac{1}{2}\int_0^t \theta^2(s)ds\right).$$

is a martingale relative the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$.

In particular, the stochastic process $\{Z(t)\}_{t\geq 0}$ is a martingale when $\theta(t) = \mathit{const}$, hence we recover the result of Exercise 3.28. The following exercise extends the result of Exercise 4.6 to the case of several independent Brownian motions.

Exercise 4.7. *Let $\{W_1(t)\}_{t\geq 0}, \dots, \{W_N(t)\}_{t\geq 0}$ be independent Brownian motions and let*

$\{\mathcal{F}(t)\}_{t\geq 0}$ be a non-anticipating filtration for all of them. Let $\{\theta_1(t)\}_{t\geq 0}, \dots \{\theta_N(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}(t)]$ be adapted to $\{\mathcal{F}(t)\}_{t\geq 0}$ and set $\theta(t) = (\theta_1(t), \dots, \theta_N(t))$. Compute $dZ(t)$, where

$$Z(t) = \exp\left(-\sum_{j=1}^{N}\int_0^t \theta_j(s)dW_j(s) - \frac{1}{2}\int_0^t \|\theta(s)\|^2 ds\right),$$

where $\|\theta(t)\| = \sqrt{\theta_1(t)^2 + \cdots + \theta_N(t)^2}$ is the Euclidean norm of $\theta(t)$.

Exercise 4.8 ($\star$)**.** *Let $A : \mathbb{R} \to \mathbb{R}$ be continuous deterministic function of time. Show that the random variable*

$$I(t) = \int_0^t A(s)\, dW(s)$$

is normally distributed with zero expectation and variance $\int_0^t A(s)^2 ds$.

Exercise 4.9. *Show that the process $\{W^2(t) - t\}_{t\geq 0}$ is a martingale relative to $\{\mathcal{F}(t)\}_{t\geq 0}$, where $\{W(t)\}_{t\geq 0}$ is a Brownian motion and $\{\mathcal{F}(t)\}_{t\geq 0}$ a non-anticipating filtration thereof. Prove also the following logically opposite statement: assume that $\{X(t)\}_{t\geq 0}$ and $\{X^2(t) - t\}_{t\geq 0}$ are martingales relative to $\{\mathcal{F}(t)\}_{t\geq 0}$, $\{X(t)\}_{t\geq 0}$ has a.s. continuous paths and $X(0) = 0$ a.s.. Then $\{X(t)\}_{t\geq 0}$ is a Brownian motion with non-anticipating filtration $\{\mathcal{F}(t)\}_{t\geq 0}$.*

4.5 Girsanov's theorem

In this section we assume that the non-anticipating filtration of the Brownian motion coincides with $\{\mathcal{F}_W(t)\}_{t\geq 0}$. Let $\{\theta(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ satisfy the Novikov condition (4.20). It follows by Theorem 4.9 that the positive stochastic process $\{Z(t)\}_{t\geq 0}$ given by

$$Z(t) = \exp\left(-\int_0^t \theta(s)dW(s) - \frac{1}{2}\int_0^t \theta^2(s)ds\right) \tag{4.21}$$

is a martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. As $Z(0) = 1$, then $\mathbb{E}[Z(t)] = 1$ for all $t \geq 0$. Thus we can use the stochastic process $\{Z(t)\}_{t\geq 0}$ to generate a measure $\widetilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ as we did at the end of Section 3.4, namely $\widetilde{\mathbb{P}} : \mathcal{F} \to [0, 1]$ is given by

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z(T)\mathbb{I}_A], \quad A \in \mathcal{F}, \tag{4.22}$$

for some given $T > 0$. The relation between $\mathbb{E}$ and $\widetilde{\mathbb{E}}$ has been determined in Theorem 3.17, where we showed that

$$\widetilde{\mathbb{E}}[X] = \mathbb{E}[Z(t)X], \tag{4.23}$$

for all $t \geq 0$ and $\mathcal{F}_W(t)$-measurable random variables X, and

$$\widetilde{\mathbb{E}}[Y|\mathcal{F}_W(s)] = \frac{1}{Z(s)}\mathbb{E}[Z(t)Y|\mathcal{F}_W(s)] \tag{4.24}$$

for all $0 \leq s \leq t$ and random variables Y. We can now state and sketch the proof of **Girsanov**'s theorem, which is a fundamental result with deep applications in mathematical finance.

Theorem 4.10. *Define the stochastic process* $\{\widetilde{W}(t)\}_{t\geq 0}$ *by*

$$\widetilde{W}(t) = W(t) + \int_0^t \theta(s)ds, \tag{4.25}$$

i.e., $d\widetilde{W}(t) = dW(t) + \theta(t)dt$. *Then* $\{\widetilde{W}(t)\}_{t\geq 0}$ *is a* $\widetilde{\mathbb{P}}$*-Brownian motion with non-anticipating filtration* $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

Sketch of the proof. We prove only that $\{\widetilde{W}(t)\}_{t\geq 0}$ is a $\widetilde{\mathbb{P}}$-Brownian motion using the Lévy characterization of Brownian motions, see Theorem 3.16. Clearly, $\{\widetilde{W}(t)\}_{t\geq 0}$ starts from zero and has continuous paths a.s. Moreover we (formally) have $d\widetilde{W}(t)d\widetilde{W}(t) = dW(t)dW(t) = dt$. Hence it remains to show that the Brownian motion $\{\widetilde{W}(t)\}_{t\geq 0}$ is $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. By Itô's product rule we have

$$\begin{aligned} d(\widetilde{W}(t)Z(t)) &= \widetilde{W}(t)dZ(t) + Z(t)d\widetilde{W}(s) + d\widetilde{W}(t)dZ(t) \\ &= (1 - \theta(t)\widetilde{W}(t))Z(t)dW(t), \end{aligned}$$

that is to say,

$$\widetilde{W}(t)Z(t) = \int_0^t (1 - \widetilde{W}(u)\theta(u))Z(u)dW(u).$$

It follows by Theorem 4.5(ii) that the stochastic process $\{Z(t)\widetilde{W}(t)\}_{t\geq 0}$ is a $\mathbb{P}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$, i.e.,

$$\mathbb{E}[Z(t)\widetilde{W}(t)|\mathcal{F}_W(s)] = Z(s)\widetilde{W}(s).$$

But according to (4.24),

$$\mathbb{E}[Z(t)\widetilde{W}(t)|\mathcal{F}_W(s)] = Z(s)\widetilde{\mathbb{E}}[\widetilde{W}(t)|\mathcal{F}_W(s)].$$

Hence $\widetilde{\mathbb{E}}[\widetilde{W}(t)|\mathcal{F}_W(s)] = \widetilde{W}(s)$, as claimed. $\square$

Later we shall need also the multi-dimensional version of Girsanov's theorem. Let

$$\{W_1(t)\}_{t\geq 0}, \dots, \{W_N(t)\}_{t\geq 0}$$

be independent Brownian motions and let $\{\mathcal{F}_W(t)\}_{t\geq 0}$ be their own generated filtration. Let $\{\theta_1(t)\}_{t\geq 0}, \dots, \{\theta_N(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ and set $\theta(t) = (\theta_1(t), \dots, \theta_N(t))$. We assume that the Novikov condition (4.20) is satisfied (with $\theta(t)^2 = \|\theta(t)\|^2 = \theta_1(t)^2 + \cdots + \theta_N(t)^2$). Then, as shown in Exercise 4.7, the stochastic process $\{Z(t)\}_{t\geq 0}$ given by

$$Z(t) = \exp\left(-\sum_{j=1}^N \int_0^t \theta_j(s)dW_j(s) - \frac{1}{2}\int_0^t \|\theta(s)\|^2 ds \right)$$

is a martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. It follows as before that the map $\widetilde{\mathbb{P}} : \mathcal{F} \to [0,1]$ given by

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z(T)\mathbb{I}_A], \quad A \in \mathcal{F} \tag{4.26}$$

is a new probability measure equivalent to $\mathbb{P}$ and the following N-dimensional generalization of Girsanov's theorem holds.

Theorem 4.11. *Define the stochastic processes $\{\widetilde{W}_1(t)\}_{t\geq 0}, \dots, \{\widetilde{W}_N(t)\}_{t\geq 0}$ by*

$$\widetilde{W}_k(t) = W_k(t) + \int_0^t \theta_k(s)ds, \quad k = 1, \dots, N. \tag{4.27}$$

Then $\{\widetilde{W}_1(t)\}_{t\geq 0}, \dots, \{\widetilde{W}_N(t)\}_{t\geq 0}$ are independent Brownian motions in the probability measure $\widetilde{\mathbb{P}}$. Moreover the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$ generated by $\{W_1(t)\}_{t\geq 0}, \dots, \{W_N(t)\}_{t\geq 0}$ is a non-anticipating filtration for $\{\widetilde{W}_1(t)\}_{t\geq 0}, \dots, \{\widetilde{W}_N(t)\}_{t\geq 0}$.

4.6 Diffusion processes in financial mathematics

The purpose of this final section is to introduce some important examples of diffusion processes in financial mathematics. The analysis of the properties of such processes is the subject of Chapter 6.

Generalized geometric Brownian motion

Given two stochastic processes $\{\alpha(t)\}_{t\geq 0}, \{\sigma(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$, the stochastic process $\{S(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ given by

$$S(t) = S(0) \exp\left(\int_0^t \alpha(s)ds + \int_0^t \sigma(s)dW(s)\right) \tag{4.28}$$

is called **generalized geometric Brownian motion** with **mean of log-return** (or **log-drift**) $\{\alpha(t)\}_{t\geq 0}$ and **volatility** $\{\sigma(t)\}_{t\geq 0}$. Since

$$d(\log S(t)) = \alpha(t)\, dt + \sigma(t)\, dW(t),$$

then $\alpha(t)$ is the drift rate of the log-price while $\sigma(t)$ is the diffusion rate of the log-price (i.e., $\sigma(t)^2$ is the rate of quadratic variation of $\log S(t)$). When $\alpha(t) = \alpha \in \mathbb{R}$ and $\sigma(t) = \sigma > 0$ are deterministic constant, the process above reduces to the geometric Brownian motion, see (2.14). The generalized geometric Brownian motion provides a quite more general and realistic model for the dynamics of stock prices than the simple geometric Brownian motion. In the rest of these notes we assume that stock prices are modeled by (4.28).

Since

$$S(t) = S(0)e^{X(t)}, \quad dX(t) = \alpha(t)dt + \sigma(t)dW(t),$$

then Itô's formula gives

$$\begin{aligned}
dS(t) &= S(0)e^{X(t)}dX(t) + \frac{1}{2}S(0)e^{X(t)}dX(t)dX(t) \\
&= S(t)\alpha(t)dt + S(t)\sigma(t)dW(s) + \frac{1}{2}\sigma^2(t)S(t)dt \\
&= \mu(t)S(t)dt + \sigma(t)S(t)dW(t), \quad \text{where } \mu(t) = \alpha(t) + \frac{1}{2}\sigma^2(t),
\end{aligned}$$

hence a generalized geometric Brownian motion is a diffusion process in which the diffusion and the drift depend on the process itself.

In the presence of several stocks, it is reasonable to assume that each of them introduced a new source of randomness in the market. Thus, when dealing with N stocks, we assume the existence of N independent Brownian motions $\{W_1(t)\}_{t\geq 0}, \dots, \{W_N(t)\}_{t\geq 0}$ and model the evolution of the stocks prices $\{S_1(t)\}_{t\geq 0}, \dots, \{S_N(t)\}_{t\geq 0}$ by the following **N-dimensional generalized geometric Brownian motion**:

$$dS_k(t) = \left(\mu_k(t) + \sum_{j=1}^{N} \sigma_{kj}(t)dW_j(t)\right) S_k(t) \tag{4.29}$$

for some stochastic processes $\{\mu_k(t)\}_{t\geq 0}, \{\sigma_{kj}(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$, $j, k = 1, \dots, N$.

Self-financing portfolios

Consider a portfolio $\{h_S(t), h_B(t)\}_{t\geq 0}$ invested in a 1+1-dimensional market. We assume that the price of the stock follows the generalized geometric Brownian motion

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW(t), \tag{4.30}$$

while the value of the risk-free asset is given by (2.15), i.e.,

$$dB(t) = B(t)r(t)dt, \tag{4.31}$$

where $\{r(t)\}_{t\geq 0}$ is the risk-free rate of the money market. Moreover we assume that the **market parameters** $\{\mu(t)\}_{t\geq 0}$, $\{\sigma(t)\}_{t\geq 0}$, $\{r(t)\}_{t\geq 0}$ have continuous paths a.s. and are adapted to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. The value of the portfolio is given by

$$V(t) = h_S(t)S(t) + h_B(t)B(t). \tag{4.32}$$

We say that the portfolio is self-financing if purchasing more shares of one asset is possible only by selling shares of the other asset for an equivalent value (and not by infusing new cash into the portfolio), and, conversely, if any cash obtained by selling one asset is immediately re-invested to buy shares of the other asset (and not withdrawn from the portfolio). To translate this condition into a mathematical formula, assume that (h_S, h_B) is the investor position on the stock and the risk-free asset during the "infinitesimal" time interval $[t, t+\delta t)$. Let $V^-(t+\delta t)$ be the value of this portfolio immediately before the time $t+\delta t$ at which the position is changed, i.e.,

$$V^-(t+\delta t) = \lim_{u\to t+\delta t} h_S S(u) + h_B B(u) = h_S S(t+\delta t) + h_B B(t+\delta t),$$

where we used the continuity in time of the assets price. At the time $t+\delta t$, the investor sells/buys shares of the assets. Let (h_S', h_B') be the new position on the stock and the risk-free asset. Then the value of the portfolio at time $t+\delta t$ is given by

$$V(t+\delta t) = h_S' S(t+\delta t) + h_B' B(t+\delta t).$$

The difference $V(t+\delta t) - V^-(t+\delta t)$, if not zero, corresponds to cash withdrawn or added to the portfolio as a result of the change in the position on the assets. In a self-financing portfolio, however, this difference must be zero. We obtain

$$V(t+\delta t) - V^-(t+\delta t) = 0 \Leftrightarrow (h_S - h_S')S(t+\delta t) + (h_B - h_B')B(t+\delta t) = 0.$$

Hence, the change of the portfolio value in the interval $[t, t+\delta t]$ is given by

$$\delta V = V(t+\delta t) - V(t) = h_S' S(t+\delta t) + h_B' B(t+\delta t) - (h_S S(t) + h_B B(t)) = h_S \delta S + h_B \delta B,$$

where $\delta S = S(t+\delta t) - S(t)$, and $\delta B = B(t+\delta t) - B(t)$ are the changes of the assets value in the interval $[t, t+\delta t]$. This discussion leads to the following definition.

Definition 4.4. *A portfolio process $\{h_S(t), h_B(t)\}_{t\geq 0}$ invested in the $1+1$-dimensional market* (4.30)-(4.31) *is said to be* **self-financing** *if it is adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ and if its value process $\{V(t)\}_{t\geq 0}$ satisfies*

$$dV(t) = h_S(t)dS(t) + h_B(t)dB(t). \tag{4.33}$$

Exercise 4.10 (Sol. 28)**.** *Show that given a diffusion process $\{h_S(t)\}_{t\geq 0}$, it is always possible to find a diffusion process $\{h_B(t)\}_{t\geq 0}$ such that the portfolio process $\{h_S(t), h_B(t)\}_{t\geq 0}$ is self-financing.*

We conclude with the important definition of hedging portfolio. Suppose that at time t a European derivative with pay-off Y at the time of maturity $T > t$ is sold for the price $\Pi_Y(t)$. An important problem in financial mathematics is to find a strategy for how the seller should invest the premium $\Pi_Y(t)$ of the derivative in order to **hedge** the derivative, i.e., in order to ensure that the portfolio value of the seller at time T is enough to pay-off the buyer of the derivative. We assume that the seller invests the premium of the derivative only on the 1+1 dimensional market consisting of the underlying stock and the risk-free asset (Δ-hedging).

Definition 4.5. *Consider the European derivative with pay-off Y and time of maturity T, where we assume that Y is $\mathcal{F}_W(T)$-measurable. A portfolio process $\{h_S(t), h_B(t)\}_{t\geq 0}$ invested in the underlying stock and the risk-free asset is said to be an* **hedging portfolio** *if* *(i)* $\{h_S(t), h_B(t)\}_{t\geq 0}$ *is adapted to* $\{\mathcal{F}_W(t)\}_{t\geq 0}$*;*

(ii) The value of the portfolio satisfies $V(T) = Y$*.*

In Chapter 6 we shall answer the following questions:

1) What is a reasonable "fair" price for the European derivative at time $t \in [0, T]$?

2) What investment strategy (on the underlying stock and the risk-free asset) should the seller undertake in order to hedge the derivative?

Chapter 5

Stochastic differential equations and partial differential equations

Throughout this chapter, the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given and $\{\mathcal{F}(t)\}_{t\geq 0}$ denotes a non-anticipating filtration for the given Brownian motion $\{W(t)\}_{t\geq 0}$ (e.g., $\mathcal{F}(t) \equiv \mathcal{F}_W(t)$). Given $T > 0$, we denote by $\mathcal{D}_T$ the open region in the (t, x)-plane given by

$$\mathcal{D}_T = \{t \in (0, T), x \in \mathbb{R}\} = (0, T) \times \mathbb{R}.$$

The closure and the boundary of $\mathcal{D}_T$ are given respectively by

$$\overline{\mathcal{D}_T} = [0, T] \times \mathbb{R}, \quad \partial \mathcal{D}_T = \{t = 0, x \in \mathbb{R}\} \cup \{t = T, x \in \mathbb{R}\}.$$

Similarly we denote $\mathcal{D}_T^+$ the open region

$$\mathcal{D}_T^+ = \{t \in (0, T), x > 0\} = (0, T) \times (0, \infty),$$

whose closure and boundary are given by

$$\overline{\mathcal{D}_T^+} = [0, T] \times [0, \infty), \quad \partial \mathcal{D}_T^+ = \{t = 0, x \geq 0\} \cup \{t = T, x \geq 0\} \cup \{t \in [0, T], x = 0\}.$$

Moreover we shall employ the following notation for functions spaces: For $\mathcal{D} = \mathcal{D}_T$ or $\mathcal{D}_T^+$,

- $C^k(\mathcal{D})$ is the space of k-times continuously differentiable functions $u : \mathcal{D} \to \mathbb{R}$;
- $C^{1,2}(\mathcal{D})$ is the space of functions $u \in C^1(\mathcal{D})$ such that $\partial_x^2 u \in C(\mathcal{D})$;
- $C^k(\overline{\mathcal{D}})$ is the space of functions $u : \overline{\mathcal{D}} \to \mathbb{R}$ such that $u \in C^k(\mathcal{D})$ and the partial derivatives of u up to order k extend continuously on $\overline{\mathcal{D}}$.
- $C_c^k(\mathbb{R}^n)$ is the space of k-times continuously differentiable functions $u : \mathbb{R}^n \to \mathbb{R}$ with compact support. We also let $C_c^\infty(\mathbb{R}^n) = \cap_{k\in\mathbb{N}} C_c^k(\mathbb{R}^n)$

A function $u : \mathcal{D} \to \mathbb{R}$ is **uniformly bounded** if there exists $C_T > 0$ such that $|u(t, x)| \leq C_T$, for all $(t, x) \in \mathcal{D}$. Unless otherwise stated, all functions are real-valued.

5.1 Stochastic differential equations

Definition 5.1. *Given $s \geq 0$, $\alpha, \beta \in C^0([s,\infty) \times \mathbb{R})$, and a deterministic constant $x \in \mathbb{R}$, we say that a stochastic process $\{X(t)\}_{t\geq s}$ is a* **global (strong) solution** *to the* **stochastic differential equation (SDE)**

$$dX(t) = \alpha(t, X(t))\, dt + \beta(t, X(t))\, dW(t) \tag{5.1}$$

with initial value $X(s,\omega) = x$ at time $t = s$, if $\{X(t)\}_{t\geq s} \in \mathcal{C}^0[\mathcal{F}(t)]$ and

$$X(t) = x + \int_s^t \alpha(\tau, X(\tau))\, d\tau + \int_s^t \beta(\tau, X(\tau))\, dW(\tau), \quad t \geq s. \tag{5.2}$$

The initial value of a SDE can be a random variable instead of a deterministic constant, but we shall not need this more general case. Note also that the integrals in the right hand side of (5.2) are well-defined, as the integrand functions have continuous paths a.s. Of course one needs suitable assumptions on the functions α, β to ensure that there is a (unique) process $\{X(t)\}_{t\geq s}$ satisfying (5.2). The precise statement is contained in the following global existence and uniqueness theorem for SDE's, which is reminiscent of the analogous result for ordinary differential equations (Picard's theorem).

Theorem 5.1. *Assume that for each $T > s$ there exist constants $C_T, D_T > 0$ such that α, β satisfy*

$$|\alpha(t,x)| + |\beta(t,x)| \leq C_T(1 + |x|), \tag{5.3}$$
$$|\alpha(t,x) - \alpha(t,y)| + |\beta(t,x) - \beta(t,y)| \leq D_T|x - y|, \tag{5.4}$$

for all $t \in [s,T]$, $x, y \in \mathbb{R}$. Then there exists a unique global solution $\{X(t)\}_{t\geq s}$ of the SDE (5.1) *with initial value $X(s) = x$. Moreover $\{X(t)\}_{t\geq s} \in \mathbb{L}^2[\mathcal{F}(t)]$.*

A proof of Theorem 5.1 can be found in [22, Theorem 5.2.1]. Note that the result proved in [22] is a bit more general than the one stated above, as it covers the case of systems of SDE's with random initial value.

The solution of (5.1) with initial initial value x at time $t = s$ will be also denoted by $\{X(t; s, x)\}_{t\geq s}$. It can be shown that, under the assumptions of Theorem 5.1, the random variable $X(t; s, x)$ depends (a.s.) continuously on the initial conditions (s, x), see [1, Sec. 7.3].

Remark 5.1. The uniqueness statement in Theorem 5.1 is to be understood "up to null sets". Precisely, if $\{X_i(t)\}_{t\geq s}$, $i = 1, 2$ are two solutions with the same initial value x, then

$$\mathbb{P}(\sup_{t\in[s,T]} |X_1(t) - X_2(t)| > 0) = 0, \quad \text{for all } T > s.$$

Remark 5.2. If the assumptions of Theorem 5.1 are satisfied only up to a *fixed* time $T > 0$, then the solution of (5.1) could *explode* at some finite time in the future of T. For example, the stochastic process given by $X(t) = \log(W(t) + e^x)$ solves (5.1) with $\alpha = -\exp(-2x)/2$ and $\beta = \exp(-x)$, but only up to the time $T_* = \inf\{t : W(t) = -e^x\} > 0$. Note that T_* is a random variable in this example[1]. In these notes we are only interested in global solutions of SDE's, hence we require (5.3)-(5.4) to hold for all $T > 0$.

Remark 5.3. The growth condition (5.3) alone is sufficient to prove the existence of a global solution to (5.1). The Lipschitz condition (5.4) is used to ensure uniqueness. By using a more general notion of solution (**weak solution**[2]) and uniqueness (**pathwise uniqueness**), one can extend Theorem 5.1 to a larger class of SDE's, which include in particular the CIR process considered in Section 5.3; see [24] for details.

Exercise 5.1 (Sol. 29)**.** *Within many applications in finance, the drift term $\alpha(t, x)$ is linear, an so it can be written in the form*

$$\alpha(t, x) = a(b - x), \quad a, b \text{ constant.} \tag{5.5}$$

A stochastic process $\{X(t)\}_{t\geq 0}$ is called **mean reverting** *if there exists a constant c such that $\mathbb{E}[X(t)] \to c$ as $t \to +\infty$. Most financial variables are required to satisfy the mean reversion property. Assume that β satisfies the assumptions in Theorem 5.1. Prove that the solution $\{X(t; s, x)\}_{t\geq 0}$ of (5.1) with linear drift (5.5) satisfies*

$$\mathbb{E}[X(t; s, x)] = xe^{-a(t-s)} + b(1 - e^{-a(t-s)}). \tag{5.6}$$

Hence the process $\{X(t; s, x)\}_{t\geq 0}$ is mean reverting if and only if $a > 0$ and in this case the long time mean is given by $c = b$.

Linear SDE's

A SDE of the form

$$dX(t) = (a(t) + b(t)X(t))\,dt + (\gamma(t) + \sigma(t)X(t))\,dW(t), \quad X(s) = x, \tag{5.7}$$

where a, b, γ, σ are deterministic functions of time, is called a linear stochastic differential equation. We assume that for all $T > 0$ there exists a constant C_T such that

$$\sup_{t\in[s,T]} (|a(t)| + |b(t)| + |\gamma(t)| + |\sigma(t)|) < C_T,$$

and so by Theorem 5.1 there exists a unique global solution of (5.7). For example, the geometric Brownian motion (2.14) solves the linear SDE $dS(t) = \mu S(t)\,dt + \sigma S(t)\,dW(t)$, where $\mu = \alpha + \sigma^2/2$. Another example of linear SDE in finance is the Vasicek interest rate model, see Exercise 6.36. Linear SDE's can be solved explicitly, as shown in the following theorem.

[1] More precisely, a stopping time, see Definition 6.10.

[2] A weak solution of (5.1) is a stochastic process $\{X(t)\}_{t\geq s}$ that satisfy (5.2) for *some* Brownian motion $\{W(t)\}_{t\geq 0}$ (not necessarily equal to the given one). See also Remark 5.9.

Theorem 5.2. *The solution* $\{X(t)\}_{t\geq s}$ *of* (5.7) *is given by* $X(t) = Y(t)Z(t)$*, where*

$$Z(t) = \exp\left(\int_s^t \sigma(\tau)dW(\tau) + \int_s^t (b(\tau) - \frac{\sigma(\tau)^2}{2})d\tau\right),$$

$$Y(t) = x + \int_s^t \frac{a(\tau) - \sigma(\tau)\gamma(\tau)}{Z(\tau)}\,d\tau + \int_s^t \frac{\gamma(\tau)}{Z(\tau)}\,dW(\tau).$$

Exercise 5.2 ($\star$)**.** *Proof Theorem 5.2.*

For example, in the special case in which the functions a, b, γ, σ are constant (independent of time), the solution of (5.7) with initial value $X(0) = x$ at time $t = 0$ is

$$X(t) = e^{\sigma W(t)+(b-\frac{\sigma^2}{2})t}\Big(x + (a-\gamma\sigma)\int_0^t e^{-\sigma W(\tau)-(b-\frac{\sigma^2}{2})\tau}d\tau + \gamma\int_0^t e^{-\sigma W(\tau)-(b-\frac{\sigma^2}{2})\tau}dW(\tau)\Big).$$

Exercise 5.3 (Sol. 30)**.** *Consider the linear SDE* (5.7) *with constant coefficients* a, b, γ *and* $\sigma = 0$*, namely*

$$dX(t) = (a + bX(t))\,dt + \gamma dW(t), \quad t \geq s, \quad X(s) = x. \tag{5.8}$$

Find the solution and show that $X(t; s, x) \in \mathcal{N}(m(t-s, x), \Delta(t-s)^2)$*, where*

$$m(\tau, x) = xe^{b\tau} + \frac{a}{b}(e^{b\tau} - 1), \quad \Delta(\tau)^2 = \frac{\gamma^2}{2b}(e^{2b\tau} - 1). \tag{5.9}$$

Exercise 5.4 (Sol. 31)**.** *Find the solution* $\{X(t)\}_{t\geq 0}$ *of the linear SDE*

$$dX(t) = tX(t)\,dt + dW(t), \quad t \geq 0$$

with initial value $X(0) = 1$*. Find* $\mathrm{Cov}(X(s), X(t))$*.*

Exercise 5.5. *Compute* $\mathrm{Cov}(W(t), X(t))$ *and* $\mathrm{Cov}(W^2(t), X(t))$*, where* $X(t) = X(t; s, x)$ *is the stochastic process in Exercise 5.3.*

Markov property

It can be shown that, under the assumptions of Theorem 5.1, the solution $\{X(t; s, x)\}_{t\geq s}$ of (5.1) is a Markov process, see for instance [1, Th. 9.2.3]. Moreover when α, β in (5.1) are time-independent, $\{X(t; s, x)\}_{t\geq s}$ is a homogeneous Markov process. The fact that solutions of SDE's should satisfy the Markov property is quite intuitive, for, as shown in Theorem 5.1, the solution at time t is uniquely characterized by the initial value at time $s < t$. Consider for example the linear SDE (5.8). As shown in Exercise 5.3, the solution satisfies $X(t; s, x) \in \mathcal{N}(m(t-s, x), \Delta(t-s)^2)$, where $m(\tau, x)$ and $\Delta(\tau)$ are given by (5.9). By Theorem 3.18, the transition density of the Markov process $\{X(t; s, x)\}_{t\geq 0}$ is given by the pdf of the random variable $X(t; s, x)$, namely $p(t, s, x, y) = p_*(t-s, x, y)$, where

$$p_*(\tau, x, y) = e^{\frac{(y-m(\tau,x))^2}{2\Delta(\tau)^2}} \frac{1}{\sqrt{2\pi\Delta(\tau)^2}}. \tag{5.10}$$

This example rises the question of how one can find the transition density of the solution to a SDE (assuming that such density exists). This problem is one of the subjects of Section 5.2.

Systems of SDE's

Occasionally in the next chapter we need to consider systems of several SDE's. All the results presented in this section extend *mutatis mutandis* to systems of SDE's, the difference being merely notational. For example, given two Brownian motions $\{W_1(t)\}_{t\geq 0}$, $\{W_2(t)\}_{t\geq 0}$ and continuous functions $\alpha_1, \alpha_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22} : [s, \infty) \times \mathbb{R}^2 \to \mathbb{R}$, the relations

$$dX_i(t) = \alpha_i(t, X_1(t), X_2(t))\, dt + \sum_{j=1,2} \beta_{ij}(t, X_1(t), X_2(t)) dW_j(t), \tag{5.11a}$$

$$X_i(s) = x_i, \quad i = 1, 2 \tag{5.11b}$$

define a system of two SDE's on the stochastic processes $\{X_1(t)\}_{t\geq 0}$, $\{X_2(t)\}_{t\geq 0}$ with initial values $X_1(s) = x_1, X_2(s) = x_2$ at time s. As usual, the correct way to interpret the relations above is in the integral form:

$$X_i(t) = x_i + \int_s^t \alpha_i(\tau, X_1(\tau), X_2(\tau))\, d\tau + \sum_{j=1,2} \int_s^t \beta_{ij}(\tau, X_1(\tau), X_2(\tau)) dW_j(\tau) \quad i = 1, 2.$$

Upon defining the vector and matrix valued functions $\alpha = (\alpha_1, \alpha_2)^T$, $\beta = (\beta_{ij})_{i,j=1,2}$, and letting $X(t) = (X_1(t), X_2(t))$, $x = (x_1, x_2)$, $W(t) = (W_1(t), W_2(t))$, we can rewrite (5.11) as

$$dX(t) = \alpha(t, X(t))\, dt + \beta(t, X(t)) \cdot dW(t), \quad X(s) = x, \tag{5.12}$$

where $\cdot$ denotes the row by column matrix product. In fact, every system of any arbitrary number of SDE's can be written in the form (5.12). Theorem 5.1 continues to be valid for systems of SDE's, the only difference being that $|\alpha|$, $|\beta|$ in (5.3)-(5.4) stand now for the vector norm of α and for the matrix norm of β.

5.2 Kolmogorov equations

Most financial variables are represented by stochastic processes solving (systems of) SDE's. In this context, a problem which recurs often is to find a function f such that the process $\{Y(t)\}_{t\geq 0}$ given by $Y(t) = f(t, X(t))$ is a martingale, where $\{X(t)\}_{t\geq 0}$ is the global solution of (5.1) with initial value $X(0) = x$. To this regard we have the following result.

Theorem 5.3. *Let $T > 0$ and $u : \overline{\mathcal{D}_T} \to \mathbb{R}$ such that $u \in C^{1,2}(\mathcal{D}_T)$ and $\partial_x u$ is uniformly bounded. Assume that u satisfies the partial differential equation*

$$\partial_t u + \alpha(t, x)\partial_x u + \frac{1}{2}\beta(t, x)^2 \partial_x^2 u = 0 \tag{5.13}$$

in the region $\mathcal{D}_T$. Assume also that α, β satisfy the conditions in Theorem 5.1 (with $s = 0$) and let $\{X(t)\}_{t\geq 0}$ be the unique global solution of (5.1) with initial value $X(0) = x$. The stochastic process $\{u(t, X(t))\}_{t\in[0,T]}$ is a martingale and satisfies

$$u(t, X(t)) = u(0, x) + \int_0^t \beta(\tau, X(\tau))\partial_x u(\tau, X(\tau))\, dW(\tau), \quad t \in [0, T]. \tag{5.14}$$

Proof. By Itô's formula we find

$$du(t, X(t)) = (\partial_t u + \alpha \partial_x u + \frac{\beta^2}{2}\partial_x^2 u)(t, X(t))\, dt + (\beta \partial_x u)(t, X(t))\, dW(t).$$

As u solves (5.13), then $du(t, X(t)) = (\beta \partial_x u)(t, X(t))\, dW(t)$, which is equivalent to (5.14) (because $u(0, X(0)) = u(0, x)$). As $\partial_x u$ is uniformly bounded, there exists a constant $C_T > 0$ such that $|\partial_x u(t, x)| \leq C_T$ and so, due also to (5.3), the process $Y(t) = \beta \partial_x u(t, X(t))$ satisfies $|Y(t)| \leq C_T(1 + |X(t)|)$. Since $\{X(t)\}_{t\geq 0} \in \mathbb{L}^2(\mathcal{F}(t))$, then $\{Y(t)\}_{t\geq 0} \in \mathbb{L}^2(\mathcal{F}(t))$ as well and so the Itô integral in the right hand side of (5.14) is a martingale. This concludes the proof of the theorem. □

Definition 5.2. *The partial differential equation (PDE)* (5.13) *is called the* **backward Kolmogorov equation** *associated to the SDE* (5.1). *We say that* $u : \overline{\mathcal{D}_T} \to \mathbb{R}$ *is a* **strong solution** *of* (5.13) *in the region* $\mathcal{D}_T$ *if* $u \in C^{1,2}(\mathcal{D}_T)$, $\partial_x u$ *is uniformly bounded and* u *solves* (5.13) *for all* $(t, x) \in \mathcal{D}_T$. *Similarly, replacing* $\mathcal{D}_T$ *with* $\mathcal{D}_T^+$, *one defines strong solutions of* (5.13) *in the region* $\mathcal{D}_T^+$

Exercise 5.6. *Derive the backward Kolmogorov PDE associated to the system of SDE's* (5.11) *when the Brownian motions* $\{W_1(t)\}_{t\geq 0}$, $\{W_2(t)\}_{t\geq 0}$ *have constant correlation* $\rho \in [-1, 1]$. *HINT: Remember that* $dW_1(t)dW_2(t) = \rho\, dt$, *see Exercise 4.4.*

The statement of Theorem 5.3 rises the question of whether the backward Kolmogorov PDE admits in general strong solutions. This problem is discussed, with different degrees of generality, in any textbook on PDE's, see [12] for example. Here we are particularly interested in which conditions ensure the *uniqueness* of the strong solution. To this regard we have the following theorem.

Theorem 5.4. *Assume that* $\alpha, \beta \in C^0([0, \infty) \times \mathbb{R})$ *satisfy* (5.3)-(5.4) *for all* $T > 0$ *and* $(t, x) \in [0, T] \times \mathbb{R}$ *and let* $\{X(t; x, s)\}_{t\geq s}$ *be the unique global solution of* (5.1) *with initial value* $X(s) = x$. *Let* $g \in C^2(\mathbb{R})$, *resp.* $g \in C^2([0, \infty))$ *such that* g' *is uniformly bounded. The backward Kolmogorov PDE*

$$\partial_t u + \alpha(t, x)\partial_x u + \frac{1}{2}\beta(t, x)^2 \partial_x^2 u = 0, \quad (t, x) \in \mathcal{D}_T, \quad \textit{resp. } (t, x) \in \mathcal{D}_T^+, \tag{5.15}$$

with the **terminal condition**

$$\lim_{t\to T} u(t, x) = g(x), \quad \textit{for all } x \in \mathbb{R}, \textit{ resp. } x > 0, \tag{5.16}$$

admits at most one strong solution. Moreover, when it exists, the strong solution is given by the **Feynman-Kac** *formula:*

$$u_T(t, x) = \mathbb{E}[g(X(T; t, x))], \quad 0 \leq t \leq T. \tag{5.17}$$

Proof. Let v be a strong solution and set $Y(\tau) = v(\tau, X(\tau;t,x))$, for $t \leq \tau \leq T$. By Itô's formula and using that v solves (5.15) we find $dY(\tau) = \beta\partial_x v(\tau, X(\tau;t,x))dW(\tau)$. Hence

$$v(T, X(T;t,x)) - v(t, X(t;t,x)) = \int_t^T \beta\partial_x v(\tau, X(\tau;t,x))dW(\tau). \tag{5.18}$$

Moreover $v(T, X(T;t,x)) = g(X(T;t,x))$, $v(t, X(t;t,x)) = v(t,x)$ and in addition, by (5.3) and the fact that of $\partial_x v$ is uniformly bounded, the Itô integral in the right hand side of (5.18) is a martingale. Hence taking the expectation we find $v(t,x) = \mathbb{E}[g(T, X(T;t,x))] = u(t,x)$. □

Remark 5.4. As shown in [22, Theorem 8.1.1], the function (5.17) is indeed the strong solution of the Kolmogorov PDE in the whole space $x \in \mathbb{R}$ under quite general conditions on the terminal value g and the coefficients α, β. The case when the problem is posed on the half-space $x > 0$ is however more subtle, see the discussion at the end of Section 5.3 for the Kolmogorov PDE associated to the CIR process.

Remark 5.5. The conditions on the function g in Theorem 5.4 can be considerably weakened. In particular the theorem still holds if one chooses g to be the pay-off function of call (or put) options, i.e., $g(x) = (x-K)_+$, although of course in this case the solution does not have a smooth extension on the terminal time boundary $t = T$.

Remark 5.6. It is often convenient to study the backward Kolmogorov PDE with an initial, rather than terminal, condition. To this purpose it suffices to make the change of variable $t \to T - t$ in (5.15). Letting $\bar{u}(t,x) = u(T-t,x)$, we now see that $\bar{u}$ satisfies the PDE

$$-\partial_t\bar{u} + \alpha(T-t,x)\partial_x\bar{u} + \frac{1}{2}\beta(T-t,x)^2\partial_x^2\bar{u} = 0, \tag{5.19}$$

with initial condition $\bar{u}(0,x) = g(x)$. Note that this is the equation considered in [22, Theorem 8.1.1]

Remark 5.7. It is possible to define other concepts of solution to the backward Kolmogorov PDE than the strong one, e.g., weak solution, entropy solution, etc. In general these solutions are not uniquely characterized by their terminal value. In these notes we only consider strong solutions, which, as proved in Theorem 5.4, are uniquely determined by (5.16).

Exercise 5.7. *Find the strong solution in the region $\mathcal{D}_T^+$ of the Kolmogorov PDE associated to the linear SDE* (5.8) *and with terminal condition $u(T,x) = e^{-x}$. HINT: Use the ansatz $u(t,x) = e^{-xA(t)+B(t)}$.*

The study of the backward Kolmogorov equation is also important to establish whether the solution of a SDE admits a transition density. In fact, it can be shown that when $\{X(t)\}_{t\geq s}$ admits a smooth transition density, then the latter coincides with the **fundamental solution** of the backward Kolmogorov equation. To state the result, let us denote by $\delta(x-y)$ the **δ-distribution** centered in $y \in \mathbb{R}$, i.e., the distribution satisfying

$$\int_{\mathbb{R}} \psi(x)\delta(x-y)\,dx = \psi(y), \quad \text{for all } \psi \in C_c^\infty(\mathbb{R}).$$

A sequence of measurable functions $(g_n)_{n\in\mathbb{N}}$ is said to converge to $\delta(x-y)$ in the sense of distributions if

$$\lim_{n\to\infty}\int_{\mathbb{R}} g_n(x)\psi(x)\,dx \to \psi(y), \quad \text{as } n\to\infty, \text{ for all } \psi\in C_c^\infty(\mathbb{R}).$$

Theorem 5.5. *Assume the conditions in Theorem 5.1 are satisfied. Let $\{X(t;s,x)\}_{t\geq s}$ be the global solution of* (5.1) *with initial value $X(s)=x$; recall that this solution is a Markov stochastic process.*

(i) If $\{X(t;s,x)\}_{t\geq s}$ admits a transition density $p(t,s,x,y)$ which is C^1 in the variable s and C^2 in the variable x, then $p(t,s,x,y)$ solves the backward Kolmogorov PDE

$$\partial_s p + \alpha(s,x)\partial_x p + \frac{1}{2}\beta(s,x)^2\partial_x^2 p = 0, \quad 0<s<t,\ x\in\mathbb{R}, \tag{5.20}$$

with terminal value

$$\lim_{s\to t} p(t,s,x,y) = \delta(x-y). \tag{5.21}$$

(ii) If $\{X(t;s,x)\}_{t\geq s}$ admits a transition density $p(t,s,x,y)$ which is C^1 in the variable t and C^2 in the variable y then $p(t,s,x,y)$ solves the **forward Kolmogorov** *PDE*[3]

$$\partial_t p + \partial_y(\alpha(t,y)p) - \frac{1}{2}\partial_y^2(\beta(t,y)^2 p) = 0, \quad t>s,\ x\in\mathbb{R}, \tag{5.22}$$

with initial value

$$\lim_{t\to s} p(t,s,x,y) = \delta(x-y). \tag{5.23}$$

Exercise 5.8. *Prove Theorem 5.5. HINT: See Exercises 6.8 and 6.9 in [26].*

Remark 5.8. The solution p of the problem (5.20)-(5.21) is called the **fundamental solution** for the backward Kolmogorov PDE, since any other solution can be reconstructed from it. For example, for all sufficiently regular functions g, the solution of (5.15) with the terminal condition (5.16) is given by

$$u_T(t,x) = \int_{\mathbb{R}} p(T,t,x,y)g(y)\,dy.$$

This can be verified either by a direct calculation or by using the interpretation of the fundamental solution as transition density. Similarly, p is the fundamental solution of the forward Kolmogorov equation.

Let us discuss a simple application of Theorem 5.5. First recall that when the functions α,β in (5.1) are time-independent, then the Markovian stochastic process $\{X(t;s,x)\}_{t\geq s}$ is homogeneous and therefore the transition density, when it exists, has the form $p(t,s,x,y) =$

[3] Also known as **Fokker-Planck** PDE.

$p_*(t-s,x,y)$. By the change of variable $s \to t-s=\tau$ in (5.20), and by (5.22), we find that $p_*(\tau,x,y)$ satisfies

$$-\partial_\tau p_* + \alpha(x)\partial_x p_* + \frac{1}{2}\sigma(x)^2\partial_x^2 p_* = 0, \tag{5.24}$$

as well as

$$\partial_\tau p_* + \partial_y(\alpha(y)p_*) - \frac{1}{2}\partial_y^2(\sigma(y)^2 p_*) = 0, \tag{5.25}$$

with the initial condition $p_*(0,x,y) = \delta(x-y)$. For example the Brownian motion is a Markov process with transition density (3.28). In this case, (5.24) and (5.25) both reduce to the heat equation $-\partial_\tau p_* + \frac{1}{2}\partial_x^2 p_* = 0$. It is straightforward to verify that (3.28) satisfies the heat equation for $(\tau,x) \in (0,\infty)\times\mathbb{R}$. Now we show that, as claimed in Theorem 5.5, the initial condition $p_*(0,x,y) = \delta(x-y)$ is also verified, that is

$$\lim_{\tau\to 0}\int_{\mathbb{R}} p_*(\tau,x,y)\psi(y)\,dy = \psi(x), \quad \text{for all } \psi \in C_c^\infty(\mathbb{R}) \text{ and } x \in \mathbb{R}.$$

Indeed with the change of variable $y = x + \sqrt{\tau}z$, we have

$$\int_{\mathbb{R}} p_*(\tau,x,y)\psi(y)\,dy = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{z^2}{2}}\psi(x+\sqrt{\tau}z)\,dz \to \psi(0)\int_{\mathbb{R}} e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}} = \psi(0),$$

as claimed. Moreover, as $W(0)=0$ a.s., Theorem 5.5 entails that the density of the Brownian motion is $f_{W(t)}(y) = p_*(t,0,y) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{y^2}{2t}}$, which is of course correct (because $W(t) \in \mathcal{N}(0,t)$).

Exercise 5.9. *Show that the transition density* (5.10) *is the fundamental solution of the Kolmogorov equation for the linear SDE* (5.8)*.*

5.3 The CIR process

A **CIR process** is a stochastic process $\{X(t)\}_{t\geq s}$ satisfying the SDE

$$dX(t) = a(b - X(t))\,dt + c\sqrt{X(t)}\,dW(t), \quad X(s) = x > 0, \tag{5.26}$$

where a,b,c are constant ($c \neq 0$). CIR processes are used in finance to model the stock volatility in the Heston model (see Section 6.6) and the spot interest rate of bonds in the CIR model (see Section 6.7). Note that the SDE (5.26) is not of the form considered so far, as the function $\beta(t,x) = c\sqrt{x}$ is defined only for $x \geq 0$ and, more importantly, it is not Lipschitz continuous in a neighborhood of $x=0$ as required in Theorem 5.1. Nevertheless, as already mentioned in Remark 5.3, it can be shown that (5.26) admits a unique global solution for all $x > 0$. Clearly the solution satisfies $X(t) \geq 0$ a.s., for all $t \geq 0$, otherwise the Itô integral in the right hand side of (5.26) would not even be defined. For future applications, it is important to know whether the solution can hit zero in finite time with positive probability. This question is answered in the following theorem, whose proof is outlined for instance in Exercise 37 of [20, Sec. 6.3].

Theorem 5.6. *Let $\{X(t)\}_{t\geq 0}$ be the CIR process with initial value $X(0) = x > 0$ at time $t = 0$. Define the (stopping*[4]*) time*

$$\tau_0^x = \inf\{t \geq 0 : X(t) = 0\}.$$

Then $\mathbb{P}(\tau_0^x < \infty) = 0$ if and only if $ab \geq c^2/2$, which is called **Feller's condition**.

The following theorem shows how to build a CIR process from a family of linear SDE's.

Theorem 5.7. *Let $\{W_1(t)\}_{t\geq 0}, \dots \{W_N(t)\}_{t\geq 0}$ be $N \geq 2$ independent Brownian motions and assume that $\{X_1(t)\}_{t\geq 0}, \dots, \{X_N(t)\}_{t\geq 0}$ solve*

$$dX_j(t) = -\frac{\theta}{2} X_j(t)\, dt + \frac{\sigma}{2}\, dW_j(t), \quad j = 1, \dots, N, \quad X_j(0) = x_j \in \mathbb{R}, \tag{5.27}$$

where θ, σ are deterministic constant. There exists a Brownian motion $\{W(t)\}_{t\geq 0}$ such that the stochastic process $\{X(t)\}_{t\geq 0}$ given by

$$X(t) = \sum_{j=1}^{N} X_j(t)^2$$

solves (5.26) *with $a = \theta$, $c = \sigma$ and $b = \frac{N\sigma^2}{4\theta}$.*

Proof. Let $X(t) = \sum_{j=1}^{N} X_j(t)^2$. Applying Itô's formula we find, after straightforward calculations,

$$dX(t) = (\frac{N\sigma^2}{4} - \theta X(t))\, dt + \sigma \sum_{j=1}^{N} X_j(t)\, dW_j(t).$$

Letting $a = \theta$, $c = \sigma$, $b = \frac{N\sigma^2}{4\theta}$ and

$$dW(t) = \sum_{j=1}^{N} \frac{X_j(t)}{\sqrt{X(t)}}\, dW_j(t),$$

we obtain that $X(t)$ satisfies

$$dX(t) = a(b - X(t))\, dt + c\sqrt{X(t)}\, dW(t).$$

Thus $\{X(t)\}_{t\geq 0}$ is a CIR process, provided we prove that $\{W(t)\}_{t\geq 0}$ is a Brownian motion. Clearly, $W(0) = 0$ a.s. and the paths $t \to W(t, \omega)$ are a.s. continuous. Moreover $\{W(t)\}_{t\geq 0}$ is a martingale, as it is the sum of matingale Itô integrals. Hence to conclude that $\{W(t)\}_{t\geq 0}$

[4]See Definition 6.10 for the general definition of stopping time.

is a Brownian motion we must show that $dW(t)dW(t) = dt$, see Theorem 3.16. We have

$$\begin{aligned}dW(t)dW(t) &= \frac{1}{X(t)}\sum_{i,j=1}^{N} X_i(t)X_j(t)dW_i(t)dW_j(t) = \frac{1}{X(t)}\sum_{i,j=1}^{N} X_i(t)X_j(t)\delta_{ij}dt\\ &= \frac{1}{X(t)}\sum_{j=1}^{N} X_j^2(t)dt = dt,\end{aligned}$$

where we used that $dW_i(t)dW_j(t) = \delta_{ij}dt$, since the Brownian motions are independent. $\square$

Remark 5.9. The process $\{X(t)\}_{t\geq 0}$ in Theorem 5.7 is said to be a **weak solution** of (5.26), because the Brownian motion $\{W(t)\}_{t\geq 0}$ in the SDE is not given advance, but rather it depends on the solution itself.

Since $N \geq 2$ in the previous theorem implies the Feller condition $ab \geq c^2/2$, then (provided $x_j \neq 0$ for some j, so that $X(0) > 0$) the CIR process constructed in Theorem 5.7 does not hit zero, see Theorem 5.6. Moreover, since the solution of (5.27) is

$$X_j(t) = e^{-\frac{1}{2}\theta t}\left(x_j + \frac{\sigma}{2}\int_0^t e^{\frac{1}{2}\theta\tau}dW_j(\tau)\right),$$

then it follows by Exercise 4.8 that the random variables $X_1(t), \dots, X_N(t)$ are normally distributed with

$$\mathbb{E}[X_j(t)] = e^{-\frac{1}{2}\theta t}x_j, \quad \mathrm{Var}[X_j(t)] = \frac{\sigma^2}{4\theta}(e^{\frac{1}{2}\theta t} - 1).$$

It follows by Exercise 3.18 that the CIR process constructed Theorem 5.7 is non-central χ^2 distributed. The following theorem shows that this is a general property of CIR processes.

Theorem 5.8. *Assume $ab > 0$. The CIR process starting at $x > 0$ at time $t = s$ satisfies*

$$X(t; s, x) = \frac{1}{2k}Y, \quad Y(t; s, x) \in \chi^2(\delta, \beta),$$

where

$$k = \frac{2a}{(1 - e^{-a(t-s)})c^2}, \quad \delta = \frac{4ab}{c^2}, \quad \beta = 2kxe^{-a(t-s)}.$$

Proof. As the CIR process is a time-homogeneous Markov process, it is enough to prove the claim for $s = 0$. Let $X(t) = X(t; 0, x)$ for short. The characteristic function of $X(t)$ is given by

$$\theta_{X(t)}(u) = \mathbb{E}[e^{iuX(t)}] = \mathbb{E}[e^{iu\frac{Y(t)}{2k}}] = \theta_{Y(t)}(\frac{u}{2k})$$

where $Y(t) = Y(t, 0, x)$. Thus the statement of the theorem is equivalent to

$$h(t,u) := \theta_{X(t)}(u) = \frac{\exp\left(-\frac{\beta u}{2(u+ik)}\right)}{(1 - iu/k)^{\delta/2}}, \tag{5.28}$$

where $k = \frac{2a}{(1-e^{-at})c^2}$, $\delta = \frac{4ab}{c^2}$, $\beta = 2kxe^{-at}$, see Table 3.1. To prove this denote $p(t,0,x,y) = p_*(t,x,y)$ the transition density of $X(t)$. Then

$$h(t,u) = \int_{\mathbb{R}} e^{iuy} p_*(t,x,y)\, dy. \tag{5.29}$$

By Theorem 5.5, p_* solves the Fokker-Planck equation

$$\partial_t p_* + \partial_y(a(b-y)p_*) - \frac{1}{2}\partial_y^2(c^2 y\, p_*) = 0, \tag{5.30}$$

with initial datum $p_*(0,x,y) = \delta(x-y)$. After straightforward calculations we derive the following equation on h

$$\partial_t h - iabuh + (au - \frac{c^2}{2}iu^2)\partial_u h = 0. \tag{5.31a}$$

The initial condition for equation (5.31a) is

$$h(0,u) = e^{iux}, \tag{5.31b}$$

which is equivalent to $p_*(0,x,y) = \delta(x-y)$. Now the proof can be completed by showing that (5.28) satisfies the initial value problem (5.31a)-(5.31b). □

Exercise 5.10. *Derive* (5.31a) *from* (5.30) *using* (5.29). *Show that* (5.28) *satisfies the initial value problem* (5.31a)-(5.31b).

Exercise 5.11. *Use the result of Theorem 5.32 to show that, for $ab > 0$, the density of the CIR process starting at x is $f_{\mathrm{CIR}}(y; t-s, x)$, where*

$$f_{\mathrm{CIR}}(y;\tau,x) = ke^{aq\tau/2}\exp(-k(y+xe^{-a\tau}))\left(\frac{y}{x}\right)^{q/2} I_q(2ke^{-a\tau/2}\sqrt{xy}), \quad q = \frac{\delta}{2} - 1. \tag{5.32}$$

Finally we discuss briefly the question of existence of strong solutions to the Kolmogorov equation for the CIR process, which is

$$\partial_t u + a(b-x)\partial_x u + \frac{c^2}{2}x\partial_x^2 u = 0, \quad (t,x) \in \mathcal{D}_T^+, \quad u(T,x) = g(x). \tag{5.33}$$

Note carefully that the Kolmogorov PDE is now defined only for $x > 0$, as the initial value x in (5.26) must be positive. Now, if a strong solution of (5.33) exists, then it must be given by $u(t,x) = \mathbb{E}[g(X(T;t,x))]$ (see Theorem 5.4). Supposing $ab > 0$, then

$$u(t,x) = \mathbb{E}[g(X(T;t,x))] = \int_0^\infty f_{\mathrm{CIR}}(y; T-t, x) g(y)\, dy,$$

where $f_{\mathrm{CIR}}(y;\tau,x)$ is given by (5.32). Using the asymptotic behavior of $f_{\mathrm{CIR}}(y;\tau,x)$ as $x \to 0^+$, it can be shown $\partial_x u(t,x)$ is bounded near the axis $x = 0$ only if the Feller condition $ab \geq c^2/2$ is satisfied. Hence u is the (unique) strong solution of (5.33) if and only if $ab \geq c^2/2$.

5.4 Finite different solutions of PDE's

The finite difference methods are techniques to find (numerically) approximate solutions to ordinary differential equations (ODE's), stochastic differential equations (SDE's) and partial differential equations (PDE's). They are based on the idea to replace the ordinary/partial derivatives with a finite difference quotient, e.g., $y'(x) \approx (y(x+h) - y(x))/h$. The various methods differ by the choice of the finite difference used in the approximation. We shall present a number of methods by examples.

ODE's

Consider the first order ODE

$$\frac{dy}{dt} = ay + bt, \quad y(0) = y_0, \quad t \in [0, T], \tag{5.34}$$

for some constants $a, b \in \mathbb{R}$ and $T > 0$. The solution is given by

$$y(t) = y_0 e^{at} + \frac{b}{a^2}(e^{at} - at - 1). \tag{5.35}$$

We shall apply three different finite difference methods to approximate the solution of (5.34). In all cases we divide the time interval $[0, T]$ into a uniform partition,

$$0 = t_0 < t_1 < \cdots < t_n = T, \quad t_j = j\frac{T}{n}, \; \Delta t = t_{j+1} - t_j = \frac{T}{n}$$

and define

$$y(t_j) = y_j, \quad j = 0, \dots, n.$$

Forward Euler method

In this method we introduce the following approximation of dy/dt at time t:

$$\frac{dy}{dt}(t) = \frac{y(t + \Delta t) - y(t)}{\Delta t} + O(\Delta t),$$

i.e.,

$$y(t + \Delta t) = y(t) + \frac{dy}{dt}(t)\Delta t + O(\Delta t^2). \tag{5.36}$$

For Equation (5.34) this becomes

$$y(t + \Delta t) = y(t) + (ay(t) + bt)\Delta t + O(\Delta t^2).$$

Setting $t = t_j$, $\Delta t = T/n$, $t + \Delta t = t_j + T/n = t_{j+1}$ and neglecting second order terms we obtain

$$y_{j+1} = y_j + (ay_j + bt_j)\frac{T}{n}, \quad j = 0, \dots, n-1. \tag{5.37}$$

As y_0 is known, the previous iterative equation can be solved at any step j. This method is called *explicit*, because the solution at the step $j+1$ is given explicitly in terms of the solution at the step j. It is a simple matter to implement this method numerically, for instance using the following Matlab function:[5]

```
function [time,sol]=exampleODEexp(T,y0,n)
dt=T/n;
sol=zeros(1,n+1);
time=zeros(1,n+1);
a=1; b=1;
sol(1)=y0;
for j=2:n+1
sol(j)=sol(j-1)+(a*sol(j-1)+b*time(j-1))*dt;
time(j)=time(j-1)+dt;
end
```

Exercise 5.12. *Compare the approximate solution with the exact solution for increasing values of n. Compile a table showing the difference between the approximate solution and the exact solution at time T for increasing value of n.*

Backward Euler method

This method consists in approximating dy/dt at time t as

$$\frac{dy}{dt}(t) = \frac{y(t) - y(t - \Delta t)}{\Delta t} + O(\Delta t),$$

hence

$$y(t + \Delta t) = y(t) + \frac{dy}{dt}(t + \Delta t)\Delta t + O(\Delta t^2). \tag{5.38}$$

The iterative equation for (5.34) now is

$$y_{j+1} = y_j + (ay_{j+1} + bt_{j+1})\frac{T}{n}, \quad j = 0, \dots, n-1. \tag{5.39}$$

This method is called implicit, because the solution at the step $j+1$ depends on the solution at both the step j *and* the step $j+1$ itself. Therefore implicit methods involve an extra computation, which is to find y_{j+1} in terms of y_j only. For the present example this is a trivial step, as we have

$$y_{j+1} = \big(1 - \frac{aT}{n}\big)^{-1}\big(y_j + bt_{j+1}\frac{T}{n}\big), \tag{5.40}$$

[5]The Matlab codes presented in this text are not optimized. Moreover the powerful vectorization tools of Matlab are not employed, so as to make the codes easily adaptable to other computer softwares and languages.

provided $n \neq aT$. Here is a Matlab function implementing the backward Euler method for the ODE (5.34):

```
function [time,sol]=exampleODEimp(T,y0,n)
dt=T/n;
sol=zeros(1,n+1);
time=zeros(1,n+1);
a=1; b=1;
sol(1)=y0;
for j=2:n+1
time(j)=time(j-1)+dt;
sol(j)=1/(1-a*dt)*(sol(j-1)+b*time(j)*dt);
end
```

Exercise 5.13. *Compare the approximate solution obtained with the backward Euler method with the exact solution and the approximate one obtained via the forward Euler method. Compile a table for increasing values of n as in Exercise 1.*

Central difference method

By a Taylor expansion,

$$y(t+\Delta t) = y(t) + \frac{dy}{dt}(t)\Delta t + \frac{1}{2}\frac{d^2y}{dt^2}(t)\Delta t^2 + O(\Delta t^3), \tag{5.41}$$

and replacing Δt with $-\Delta t$,

$$y(t-\Delta t) = y(t) - \frac{dy}{dt}(t)\Delta t + \frac{1}{2}\frac{d^2y}{dt^2}(t)\Delta t^2 + O(\Delta t^3). \tag{5.42}$$

Subtracting the two equations we obtain the following approximation for dy/dt at time t:

$$\frac{dy}{dt}(t) = \frac{y(t+\Delta t) - y(t-\Delta t)}{2\Delta t} + O(\Delta t^2),$$

which is called central difference approximation. Hence

$$y(t+\Delta t) = y(t-\Delta t) + 2\frac{dy}{dt}(t)\Delta t + O(\Delta t^3). \tag{5.43}$$

Note that, compared to (5.36) and (5.38), we have gained one order in accuracy. The iterative equation for (5.34) becomes

$$y_{j+1} = y_{j-1} - 2(ay_j + bt_j)\frac{T}{n}, \quad j = 0, \ldots, n-1. \tag{5.44}$$

The first step $j = 0$ requires y_{-1}. This is fixed by the backward method

$$y_{-1} = y_0 - \frac{T}{n} a y_0, \tag{5.45}$$

which is (5.39) for $j = -1$.

Exercise 5.14. *Write a Matlab function that implements the central difference method for* (5.34)*. Compile a table comparing the exact solution with the approximate solutions at time T obtained by the three methods presented above for increasing value of n.*

A second order ODE

Consider the second order ODE for the harmonic oscillator:

$$\frac{d^2y}{dt^2} = -\omega^2 y, \quad y(0) = y_0, \quad \dot{y}(0) = \widetilde{y}_0. \tag{5.46}$$

The solution to this problem is given by

$$y(t) = y_0 \cos(\omega t) + \frac{\widetilde{y}_0}{\omega} \sin(\omega t). \tag{5.47}$$

One can define forward/backward/central difference approximations for second derivatives in a way similar as for first derivatives. For instance, adding (5.41) and (5.42) we obtain the following central difference approximation for d^2y/dt^2 at time t:

$$\frac{d^2y}{dt^2}(t) = \frac{y(t+\Delta t) - 2y(t) + y(t-\Delta t)}{\Delta t^2} + O(\Delta t),$$

which leads to the following iterative equation for (5.46):

$$y_{j+1} = 2y_j - y_{j-1} - \left(\frac{T}{n}\right)^2 \omega^2 y_j, \quad j = 1, \ldots, n-1, \tag{5.48}$$

$$y_1 = y_0 + \widetilde{y}_0 \frac{T}{n}. \tag{5.49}$$

The approximate solution y_1 at the first node is computed using the forward method and the initial datum $\dot{y}(0) = \widetilde{y}_0$. The Matlab function solving this iteration is the following.

```
function [time,sol]=harmonic(w,T,y0,n)
dt=T/n;
sol=zeros(1,n+1);
time=zeros(1,n+1);
sol(1)=y0(1);
sol(2)=sol(1)+y0(2)*dt;
for j=3:n+1
```

```
sol(j)=2*sol(j-1)-sol(j-2)-dt^2*w^2*sol(j-1);
time(j)=time(j-1)+dt;
end
```

Exercise 5.15. *Compare the exact and approximate solutions at time T for increasing values of n.*

SDE's

The Euler method can be straightforwardly generalized to SDE's, see [15]. In this section we present briefly the so-called Euler-Maruyama method, which is the generalization to SDE's of the forward Euler method for ODE's.

Consider the SDE

$$dX(t) = \alpha(t, X(t))\, dt + \beta(t, X(t))\, dW(t), \tag{5.50}$$

where we assumed that the assumptions in Theorem 5.1 are satisfied. Given the uniform partition

$$0 = t_0 < t_1 < \cdots < t_n = T, \quad t_j = j\frac{T}{n},\ \Delta t = t_{j+1} - t_j = \frac{T}{n}$$

of the interval $[0, T]$, we define

$$X(t_j) = X_j, \quad j = 0, \ldots, n, \quad W_j = W(t_j).$$

Note that X_j, W_j are random variables and that

$$G_j = \frac{W_j - W_{j-1}}{\sqrt{\Delta t}}$$

are independent standard normal random variables for $j = 1, \ldots, n$. The (explicit) finite difference approximation of (5.50) is

$$X_j = X_{j-1} + \alpha(t_{j-1}, X_{j-1})\frac{T}{n} + \beta(t_{j-1}, X_{j-1})\sqrt{\frac{T}{n}}\, G_j. \tag{5.51}$$

The following Matlab function applies the iterative equation (5.51) to the linear SDE (5.8), for which (5.51) becomes

$$X_j = X_{j-1} + (a + bX_{j-1})(T/n) + \gamma\sqrt{T/n}\, G_j, \quad X_0 = x_0,$$

where x_0 is the (constant) initial datum. The output `sol` contains one path of the stochastic process $X(t)$ along the time partition $\{t_0 = 0, t_1, \ldots, t_n = T\}$, that is, a path $(X_0 = x_0, X(t_1), \ldots, X(t_n) = X(T))$.

```
function [time,sol] = linearSDEexample(T,x0,n,a,b,gamma)
dt=T/n;
sol=zeros(1,n+1);
time=zeros(1,n+1);
G=randn(1,n);
sol(1)=x0;
for j=2:n+1
sol(j)=sol(j-1)+(a+b*sol(j-1))*dt+gamma*sqrt(dt)*G(j-1);
time(j)=time(j-1)+dt;
end
```

PDE's

In this section we present three finite difference methods to find approximate solutions to the one-dimensional heat equation

$$\partial_t u = \partial_x^2 u, \quad u(0,x) = u_0(x), \tag{5.52}$$

where u_0 is continuous. We refer to t as the time variable and to x as the spatial variable, since this is what they typically represent in the applications of the heat equation. As before, we let $t \in [0,T]$. As to the domain of the spatial variable x, we distinguish two cases

(i) x runs over the whole real line, i.e., $x \in (-\infty, \infty)$, and we are interested in finding an approximation to the solution $u \in C^{1,2}(\overline{\mathcal{D}_T})$.

(ii) x runs over a finite interval, say $x \in (x_{\min}, x_{\max})$, and we want to find an approximation of the solution $u \in C^{1,2}(\overline{\mathcal{D}})$, where $\mathcal{D} = (0,T) \times (x_{\min}, x_{\max})$, which satisfies the boundary conditions[6]

$$u(t, x_{\min}) = u_L(t), \quad u(t, x_{\max}) = u_r(t), \quad t \in [0,T],$$

for some given continuous functions u_L, u_R. We also require $u_L(0) = u_0(x_{\min})$, $u_R(0) = u_0(x_{\max})$, so that the solution is continuous on the boundary.

In fact, for numerical purposes, problem (i) is a special case of problem (ii), for the domain $(-\infty, \infty)$ must be approximated by $(-A, A)$ for $A >> 1$ when we solve problem (i) in a computer. Note however that in the finite domain approximation of problem (i), *the boundary conditions at $x = \pm A$ cannot be prescribed freely!* Rather they have to be given by suitable approximations of the limit values at $x = \pm\infty$ of the solution to the heat equation on the real line.

[6] These are called Dirichlet type boundary conditions. Other types of boundary conditions can be imposed, but the Dirichlet type is sufficient for our forthcoming applications to financial problems.

By what we have just said we can focus on problem (ii). To simplify the discussion we assume that the domain of the x variable is given by $x \in (0, X)$ and we assign zero boundary conditions, i.e., $u_L = u_R = 0$. Hence we want to study the problem

$$\partial_t u = \partial_x^2 u, \quad (t, x) \in (0, T) \times (0, X), \tag{5.53a}$$

$$u(0, x) = u_0(x), \quad u(t, 0) = u(t, X) = 0, \quad x \in [0, X],\ t \in [0, T]; \quad u_0(0) = u_0(X) = 0. \tag{5.53b}$$

We introduce the partition of the interval $(0, X)$ given by

$$0 = x_0 < x_1 < \cdots < x_m = X, \quad \Delta_j x = x_{j+1} - x_j, \quad j = 0, \ldots, m-1,$$

and the partition of the time interval $[0, T]$ given by

$$0 = t_0 < t_1 < \cdots < t_n = T, \quad t_i = i\frac{T}{n}, \quad \Delta t = t_{i+1} - t_i = \frac{T}{n}.$$

Note that we use a uniform partition for the time interval while the partition for the spatial domain is in general not uniform. Of course

$$\Delta_0 x + \Delta_1 x + \cdots + \Delta_{m-1} x = X$$

and the spatial partition is uniform if and only if

$$\Delta_0 x = \Delta_1 x = \cdots = \Delta_{m-1} x = \Delta x = \frac{X}{m}, \quad x_j = j\Delta x. \tag{5.54}$$

Using non-uniform partitions is important when one needs more accuracy in some region. For instance, when computing the price of options with the finite difference methods, a more refine partition is recommended in the "nearly at the money" region. In the rest of this section we assume that the spatial partition is uniform and leave the generalization to non-uniform partitions as an exercise (Exercise 5.17). We denote

$$u_{i,j} = u(t_i, x_j), \quad i = 0, \ldots, n, \quad j = 0, \ldots, m.$$

Hence $u_{i,j}$ is a $n+1 \times m+1$ matrix. The i^{th} row contains the value of the approximate solution at each point of the spatial mesh at the fixed time t_i. For instance, the zeroth row is the initial datum: $u_{0,j} = u_0(x_j)$, $i = 0, \ldots m$. The columns of the matrix $u_{i,j}$ contain the values of the approximate solution at one spatial point for different times. For instance, the column $u_{i,0}$ are the values of the approximate solution at $x_0 = 0$ for different times t_i, while $u_{i,m}$ contains the values at $x_m = X$. By the given boundary conditions we then have

$$u_{i,0} = u_{i,m} = 0, \quad i = 0, \ldots, n.$$

We define (for a uniform spatial partition)

$$d = \frac{\Delta t}{\Delta x^2} = \frac{T}{X^2}\frac{m^2}{n}. \tag{5.55}$$

Method 1: Forward in time, centered in space

In this method we use a forward difference approximation for the time derivative and a centered difference approximation for the second spatial derivative:

$$\partial_t u(t,x) = \frac{u(t+\Delta t,x) - u(t,x)}{\Delta t} + O(\Delta t),$$

$$\partial_x^2 u(t,x) = \frac{u(t,x+\Delta x) - 2u(t,x) + u(t,x-\Delta x)}{\Delta x^2} + O(\Delta x).$$

We find

$$u(t+\Delta t,x) = u(t,x) + d(u(t,x+\Delta x) - 2u(t,x) + u(t,x-\Delta x)).$$

Hence we obtain the following iterative equation

$$u_{i+1,j} = u_{i,j} + d(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}), \quad i = 0,\dots,n-1,\ j = 1,\dots,m-1, \tag{5.56}$$

where we recall that $u_{0,j} = u_0(x_j)$, $u_{i,0} = u_{i,m+1} = 0$, $i = 0,\dots,n$, $j = 0,\dots m$. Let

$$\mathbf{u_i} = (u_{i,1}\ u_{i,1}\ \dots\ u_{i,m-1})^T$$

be the *column* vector containing the approximate solution at time t_i; note that we do not need to include $u_{i,0}$, $u_{i,m}$ in the vector $\mathbf{u}_i$, as these components are fixed equal to zero by the boundary conditions. We can rewrite (5.56) in matrix form as follows:

$$\mathbf{u_{i+1}} = A(d)\mathbf{u_i}, \tag{5.57}$$

where $A(z)$ is the $(m-1)\times(m-1)$ matrix with non-zero entries given by

$$A_{k,k}(z) = 1 - 2z, \quad k = 1,\dots,m-1, \qquad A_{q,q+1}(z) = A_{q+1,q}(z) = z, \quad q = 1,\dots,m-2. \tag{5.58}$$

This method is completely explicit, as the solution at the time step $i+1$ is explicitly given in terms of the solution at the time step i. A Matlab function solving the iteration (5.57) with the initial datum $u_0(x) = \exp(X^2/4) - \exp((x - X/2)^2)$ is the following.

```
function [time,space,sol]=heatexp(T,X,n,m)
dt=T/n; dx=X/m;
d=dt/dx^2;
sol=zeros(n+1,m+1);
time=zeros(1,n+1);
space=zeros(1,m+1);
for i=2:n+1
time(i)=time(i-1)+dt;
end
for j=2:m+1
space(j)=space(j-1)+dx;
```

```
end
for j=1:m+1
sol(1,j)=exp(X^2/4)-exp((space(j)-X/2)^2);
end
sol(:,1)=0; sol(:,m+1)=0;
A=zeros(m-1,m-1);
for k=1:m-1
A(k,k)=1-2*d;
end
for q=1:m-2
A(q,q+1)=d;
A(q+1,q)=d;
end
for i=2:n+1
sol(i,2:m)=sol(i-1,2:m)*transpose(A);
end
```

To visualize the result it is convenient to employ an animation which plots the approximate solution at each point on the spatial mesh for some increasing sequence of times in the partition $\{t_0, t_1, \dots, t_n\}$. This visualization can be achieved with the following simple Matlab function:

```
function anim(r,F)
N=length(F(:,1));
Max=max(max(F));
for i=1:N
plot(r,F(i,:));
axis([0 1 0 Max]);
drawnow;
pause(0.01);
end
```

Upon running the command `anim(space,sol)`, the previous function will plot the approximate solutions at different increasing times.

Let us try the following: `[time,space,sol]=heatexp(1,1,2500,50)`. Hence we solve the problem on the unit square $(t, x) \in (0, 1)^2$ on a mesh of $(n, m) = 2500 \times 50$ points. The value of the parameter (5.55) is

$$d = 1.$$

If we now try to visualize the solution by running `anim(space,sol,0.1)`, we find that the approximate solution behaves very strangely (it produces just random oscillations). However by increasing the number of time steps with `[time,space,sol]=heatexp(1,1,5000,50)`, so that

$$d = 0.5,$$

and visualize the solution, we shall find that the approximate solution converges quickly and smoothly to $u \equiv 0$, which is the equilibrium of our problem (i.e., the time independent solution of (5.53)). In fact, this is not a coincidence, as it can be shown that *the forward-centered method for the heat equation is unstable if $d > 0.5$ and stable for $d \leq 0.5$.* The term unstable here refers to the fact that numerical errors, due for instance to the truncation and round-off of the initial datum on the spatial grid, will increase in time. On the other hand, stability of a finite difference method means that the error will remain small at all times. The stability condition $d \leq 0.5$ for the forward-centered method applied to the heat equation is very restrictive: it forces us to choose a very high number of points on the time partition. To avoid such a restriction, which could be very costly in terms of computation time, implicit methods are preferred, such as the one we present next.

Method 2: Backward in time, centered in space

In this method we employ the backward finite difference approximation for the time derivative and the central difference for the second spatial derivative (same as before). This results in the following iterative equation:

$$u_{i+1,j} = u_{i,j} + d(u_{i+1,j+1} - 2u_{i+1,j} + u_{i+1,j-1}), \quad i = 0, \ldots, n-1, \; j = 1, \ldots, m-1, \tag{5.59}$$

where we recall that $u_{0,j} = u_0(x_j)$, $u_{i,0} = u_{i,m+1} = 0$, $i = 0, \ldots, n$, $j = 0, \ldots m$. This method is implicit and we need therefore to solve for the solution at time $i+1$ in terms of the solution at time i. To this purpose we let, as before,

$$\mathbf{u_i} = (u_{i,1} \; u_{i,1} \; \ldots \; u_{i,m-1})^T$$

and rewrite (5.59) in matrix form as follows:

$$A(-d)\mathbf{u_{i+1}} = \mathbf{u_i}, \tag{5.60}$$

where $A(z)$ is the matrix with non-zero entries (5.58). The matrix A is invertible, hence we can express $\mathbf{u_{i+1}}$ in terms of $\mathbf{u_i}$ as

$$\mathbf{u_{i+1}} = A(-d)^{-1}\mathbf{u_i}. \tag{5.61}$$

This method is unconditionally stable, i.e., it is stable for all values of the parameter d. We can test this property by using the following Matlab function, which solves the iterative equation (5.61):

```
function [time,space,sol]=heatimp(T,X,n,m)
dt=T/n; dx=X/m;
d=dt/dx^2;
```

```
sol=zeros(n+1,m+1);
time=zeros(1,n+1);
space=zeros(1,m+1);
for i=2:n+1
time(i)=time(i-1)+dt;
end
for j=2:m+1
space(j)=space(j-1)+dx;
end
for j=1:m+1
sol(1,j)=exp(X^2/4)-exp((space(j)-X/2)^2);
end
sol(:,1)=0; sol(:,m+1)=0;
A=zeros(m-1,m-1);
for k=1:m-1
A(k,k)=1+2*d;
end
for q=1:m-2
A(q,q+1)=-d;
A(q+1,q)=-d;
end
for i=2:n+1
sol(i,2:m)=sol(i-1,2:m)*transpose(inv(A));
end
```

If we now run `[time,space,sol]=heatexp(1,1,500,50)`, for which $d = 5$, and visualize the solution we shall obtain that the approximate solution behaves smoothly as expected, indicating that the instability problem of the forward-centered method has been solved.

Method 3: The θ-method

This is an implicit method with higher order of accuracy than the backward-centered method. It is obtained by simply averaging between methods 1 and 2 above, as follows

$$u_{i+1,j} = \theta u_{i+1,j}^{\text{back}} + (1-\theta) u_{i+1,j}^{\text{forw}}, \quad \theta \in (0,1),$$

where the first term in the right hand side is computed with method 1 and the second term with method 2. Thus we obtain the following iterative equation

$$u_{i+1,j} = u_{i,j} + d[(1-\theta)(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}) + \theta(u_{i+1,j+1} - 2u_{i+1,j} + u_{i+1,j-1})], \quad (5.62)$$

or, in matrix form

$$A(-d\theta)\mathbf{u}_{i+1} = A(d(1-\theta))\mathbf{u}_i$$

Remark 5.10. Note carefully that the solution obtained with the θ-method is *not* the average of the solutions obtained with methods 1 and 2, but rather the solution obtained by averaging the two methods.

Remark 5.11. *For $\theta = 1/2$, the θ-method is also called* **Crank-Nicolson method**.

Exercise 5.16. *Write a Matlab function that implements the θ-method for the heat equation.*

Exercise 5.17. *Generalize the methods 1-3 to the case of non-uniform spatial partition. Write the iterative equations in matrix form. Write a Matlab function that implements the θ-method with non-uniform spatial partition to the heat equation.*

Chapter 6

Applications to finance

Throughout this chapter we assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the Brownian motion $\{W(t)\}_{t\geq 0}$ are given. Moreover, in order to avoid the need of repeatedly specifying technical assumptions, we stipulate the following conventions:

- All stochastic processes in this chapter are assumed to belong to the space $\mathcal{C}^0[\mathcal{F}_W(t)]$, i.e., they are adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ and have a.s. continuous paths. This assumption may be relaxed, but for our applications it is general enough.

- All Itô integrals in this chapter are assumed to be martingales, which holds for instance when the integrand stochastic process is in the space $\mathbb{L}^2[\mathcal{F}_W(t)]$.

6.1 Arbitrage-free markets

The ultimate purpose of this section is to prove that any self-financing portfolio invested in a 1+1 dimensional market is not an arbitrage. We shall prove the result by using Theorem 3.14, i.e., by showing that there exists a probability measure $\widetilde{\mathbb{P}}$, equivalent to $\mathbb{P}$, with respect to which the discounted value of the portfolio is a martingale. We first define such a measure. We have seen in Theorem 4.9 that, given a stochastic process $\{\theta(t)\}_{t\geq 0}$ satisfying the Novikov condition (4.20), the stochastic process $\{Z(t)\}_{t\geq 0}$ defined by

$$Z(t) = \exp\left(-\int_0^t \theta(s)\, dW(s) - \frac{1}{2}\int_0^t \theta(s)^2\, ds\right) \tag{6.1}$$

is a $\mathbb{P}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$ and that the map $\widetilde{\mathbb{P}} : \mathcal{F} \to [0, 1]$ given by

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z(T)\mathbb{I}_A] \tag{6.2}$$

is a probability measure equivalent to $\mathbb{P}$, for all given $T > 0$.

Definition 6.1. *Consider the 1+1 dimensional market*

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW(t), \quad dB(t) = B(t)r(t)dt.$$

Assume that $\sigma(t) > 0$ almost surely for all times. Let $\{\theta(t)\}_{t\geq 0}$ be the stochastic process given by

$$\theta(t) = \frac{\mu(t) - r(t)}{\sigma(t)}, \tag{6.3}$$

and define $\{Z(t)\}_{t\geq 0}$ by (6.1). Assume that $\{Z(t)\}_{t\geq 0}$ is a martingale (e.g., $\{\theta(t)\}_{t\geq 0}$ satisfies the Novikov condition (4.20)). The probability measure $\widetilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ given by (6.2) is called the **risk-neutral probability measure** *of the market at time T, while the process $\{\theta(t)\}_{t\geq 0}$ is called the* **market price of risk**.

By the definition (6.3) of the stochastic process $\{\theta(t)\}_{t\geq 0}$, we can rewrite $dS(t)$ as

$$dS(t) = r(t)S(t)dt + \sigma(t)S(t)d\widetilde{W}(t), \tag{6.4}$$

where

$$d\widetilde{W}(t) = dW(t) + \theta(t)dt. \tag{6.5}$$

By Girsanov theorem, Theorem 4.10, the stochastic process $\{\widetilde{W}(t)\}_{t\geq 0}$ is a $\widetilde{\mathbb{P}}$-Brownian motion. Moreover, $\{\mathcal{F}_W(t)\}_{t\geq 0}$ is a non-anticipating filtration for $\{\widetilde{W}(t)\}_{t\geq 0}$. We also recall that a portfolio $\{h_S(t), h_B(t)\}_{t\geq 0}$ is self-financing if its value $\{V(t)\}_{t\geq 0}$ satisfies

$$dV(t) = h_S(t)dS(t) + h_B(t)dB(t), \tag{6.6}$$

see Definition 4.4. Moreover $S^*(t) = D(t)S(t)$ is the discounted price (at time $t = 0$) of the stock, where $D(t) = \exp(-\int_0^t r(s)\,ds)$ is the discount process.

Theorem 6.1. *Consider the 1+1 dimensional market*

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW(t), \quad dB(t) = B(t)r(t)dt, \tag{6.7}$$

where $\sigma(t) > 0$ almost surely for all times.

(i) The discounted stock price $\{S^*(t)\}_{t\geq 0}$ is a $\widetilde{\mathbb{P}}$-martingale in the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

(ii) A portfolio process $\{h_S(t), h_B(t)\}_{t\geq 0}$ is self-financing if and only if its discounted value satisfies

$$V^*(t) = V(0) + \int_0^t D(s)h_S(s)\sigma(s)S(s)d\widetilde{W}(s). \tag{6.8}$$

In particular the discounted value of self-financing portfolios is a $\widetilde{\mathbb{P}}$-martingale in the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

(iii) If $\{h_S(t), h_B(t)\}_{t\geq 0}$ *is a self-financing portfolio, then* $\{h_S(t), h_B(t)\}_{t\geq 0}$ *is not an arbitrage.*

Proof. (i) By (6.4) and $dD(t) = -D(t)r(t)dt$ we have

$$\begin{aligned} dS^*(t) &= S(t)dD(t) + D(t)dS(t) + dD(t)dS(t) \\ &= -S(t)r(t)D(t)\,dt + D(t)(r(t)S(t)\,dt + \sigma(t)S(t)\,d\widetilde{W}(t)) \\ &= D(t)\sigma(t)S(t)d\widetilde{W}(t), \end{aligned}$$

and so the discounted price $\{S^*(t)\}_{t\geq 0}$ of the stock is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

(ii) By (6.4) and $h_S(t)S(t) + h_B(t)B(t) = V(t)$, the definition (6.6) of self-financing portfolio is equivalent to

$$dV(t) = h_S(t)S(t)[(\mu(t) - r(t))dt + \sigma(t)dW(t)] + V(t)r(t)dt. \tag{6.9}$$

Hence

$$dV(t) = h_S(t)S(t)\sigma(t)d\widetilde{W}(t) + V(t)r(t)dt.$$

In terms of the discounted portfolio value $V^*(t) = D(t)V(t)$ the previous equation reads

$$\begin{aligned} dV^*(t) &= V(t)dD(t) + D(t)dV(t) + dD(t)dV(t) \\ &= -D(t)V(t)r(t)\,dt + D(t)h_S(t)S(t)\sigma(t)d\widetilde{W}(t) + D(t)V(t)r(t)dt \\ &= D(t)h_S(t)S(t)\sigma(t)d\widetilde{W}(t), \end{aligned}$$

which proves (6.8).

(iii) By (6.8), the discounted value of self-financing portfolios is a $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. As $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent, (iii) follows by Theorem 3.14. □

Remark 6.1 (Arbitrage-free principle)**.** The absence of self-financing arbitrage portfolios in the 1+1 dimensional market (6.7) is consistent with the observations. In fact, even though arbitrage opportunities do exist in real markets, they typically last only for very short times, as they are quickly exploited by investors. In general, when a stochastic model for the price of assets is introduced, we require that it should satisfy the **arbitrage-free principle**, namely that any self-financing portfolio invested in these assets and in the money market should be no arbitrage. Theorem 6.1 shows that a stock market consisting of one single stock with price modeled by the generalized geometric Brownian motion satisfies the arbitrage-free principle, provided $\sigma(t) > 0$ a.s. for all times. The generalization of this result to stock markets with several stocks is discussed in Section 6.9.

6.2 The risk-neutral pricing formula

Consider the European derivative with pay-off Y and time of maturity $T > 0$. We assume that Y is $\mathcal{F}_W(T)$-measurable. Suppose that the derivative is sold at time $t < T$ for the price

$\Pi_Y(t)$. The first concern of the seller is to hedge the derivative, that is to say, to invest the amount $\Pi_Y(t)$ in such a way that the value of the seller portfolio at time T is enough to pay-off the buyer of the derivative. The purpose of this section is to define a theoretical price for the derivative which makes it possible for the seller to set-up an hedging portfolio. We argue under the following assumptions:

1. the seller is only allowed to invest the amount $\Pi_Y(t)$ in the 1+1 dimensional market consisting of the underlying stock and the risk-free asset (**Δ-hedging**);

2. the investment strategy of the seller is self-financing.

It follows by Theorem 6.1 that the sought hedging portfolio is not an arbitrage. We may interpret this fact as a "fairness" condition on the price of the derivative $\Pi_Y(t)$. In fact, if the seller can hedge the derivative and still be able to make a risk-less profit on the underlying stock, this may be considered unfair for the buyer.

We thus consider the 1+1 dimensional market

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW(t), \quad dB(t) = B(t)r(t)dt,$$

where $\sigma(t) > 0$ almost surely for all times. Let $\{h_S(t), h_B(t)\}_{t\geq 0}$ be a self-financing portfolio invested in this market and let $\{V(t)\}_{t\geq 0}$ be its value. By Theorem 6.1, the discounted value $\{V^*(t)\}_{t\geq 0}$ of the portfolio is a $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$, hence

$$D(t)V(t) = \widetilde{\mathbb{E}}[D(T)V(T)|\mathcal{F}_W(t)].$$

Requiring the hedging condition $V(T) = Y$ gives

$$V(t) = \frac{1}{D(t)}\widetilde{\mathbb{E}}[D(T)Y|\mathcal{F}_W(t)].$$

Since $D(t)$ is $\mathcal{F}_W(t)$-measurable, we can move it inside the conditional expectation and write the latter equation as

$$V(t) = \widetilde{\mathbb{E}}[Y\frac{D(T)}{D(t)}|\mathcal{F}_W(t)] = \widetilde{\mathbb{E}}[Y\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)],$$

where we used the definition $D(t) = \exp(-\int_0^t r(s)\,ds)$ of the discount process. Assuming that the derivative is sold at time t for the price $\Pi_Y(t)$, then the value of the seller portfolio at this time is precisely equal to the premium $\Pi_Y(t)$, which leads to the following definition.

Definition 6.2. *Let Y be a $\mathcal{F}_W(T)$-measurable random variable with finite expectation. The* **risk-neutral price** *(or* **fair price***, or* **arbitrage-free price***) at time $t \in [0,T]$ of the European derivative with pay-off Y and time of maturity $T > 0$ is given by*

$$\Pi_Y(t) = \widetilde{\mathbb{E}}[Y\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)], \tag{6.10}$$

i.e., it is equal to the value at time t of any self-financing hedging portfolio invested in the underlying stock and the risk-free asset.

Since the risk-neutral price of the European derivative equals the value of self-financing hedging portfolios invested in a 1+1 dimensional market, then, by Theorem 6.1, the discounted risk-neutral price $\{\Pi_Y^*(t)\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. In fact, this property follows directly also by Definition 6.10, as shown in the first part of the following theorem.

Theorem 6.2. *Consider the 1+1 dimensional market*

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW(t), \quad dB(t) = B(t)r(t)dt,$$

where $\sigma(t) > 0$ almost surely for all times. Assume that the European derivative on the stock with pay-off Y and time of maturity $T > 0$ is priced by (6.10) and let $\Pi_Y^*(t) = D(t)\Pi_Y(t)$ be the discounted price of the derivative. Then the following holds.

(i) The process $\{\Pi_Y^*(t)\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

(ii) There exists a stochastic process $\{\Delta(t)\}_{t\in[0,T]}$, adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$, such that

$$\Pi_Y^*(t) = \Pi_Y(0) + \int_0^t \Delta(s)d\widetilde{W}(s), \quad t \in [0,T]. \tag{6.11}$$

(iii) The portfolio $\{h_S(t), h_B(t)\}_{t\in[0,T]}$ given by

$$h_S(t) = \frac{\Delta(t)}{D(t)\sigma(t)S(t)}, \quad h_B(t) = (\Pi_Y(t) - h_S(t)S(t))/B(t) \tag{6.12}$$

is self-financing and replicates the derivative at any time, i.e., its value $V(t)$ is equal to $\Pi_Y(t)$ for all $t \in [0,T]$. In particular, $V(T) = \Pi_Y(T) = Y$, i.e., the portfolio is hedging the derivative.

Proof. (i) We have

$$\Pi_Y^*(t) = D(t)\Pi_Y(t) = \widetilde{\mathbb{E}}[\Pi_Y(T)D(T)|\mathcal{F}_W(t)] = \widetilde{\mathbb{E}}[\Pi_Y^*(T)|\mathcal{F}_W(t)],$$

where we used that $\Pi_Y(T) = Y$. Hence, for $s \leq t$, and using Theorem 3.13(v),

$$\widetilde{\mathbb{E}}[\Pi_Y^*(t)|\mathcal{F}_W(s)] = \widetilde{\mathbb{E}}[\widetilde{\mathbb{E}}[\Pi_Y^*(T)|\mathcal{F}_W(t)]|\mathcal{F}_W(s)] = \widetilde{\mathbb{E}}[\Pi_Y^*(T)|\mathcal{F}_W(s)] = \Pi_Y^*(s).$$

This shows that the discounted price of the derivative is a $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

(ii) By (i) and (3.24) we have

$$Z(s)\Pi_Y^*(s) = Z(s)\widetilde{\mathbb{E}}[\Pi_Y^*(t)|\mathcal{F}_W(s)] = \mathbb{E}[Z(t)\Pi_Y^*(t)|\mathcal{F}_W(s)], \tag{6.13}$$

i.e., the stochastic process $\{Z(t)\Pi_Y^*(t)\}_{t\in[0,T]}$ is a $\mathbb{P}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. Hence, by the martingale representation theorem, Theorem 4.5(v), there exists a stochastic process $\{\Gamma(t)\}_{t\in[0,T]}$ adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ such that

$$Z(t)\Pi_Y^*(t) = \Pi_Y(0) + \int_0^t \Gamma(s)dW(s), \quad t\in[0,T],$$

i.e.,

$$d(Z(t)\Pi_Y^*(t)) = \Gamma(t)dW(t). \tag{6.14a}$$

On the other hand, by Itô's product rule,

$$\begin{aligned} d\Pi_Y^*(t) =& d(Z(t)\Pi_Y^*(t)/Z(t)) = d(1/Z(t))Z(t)\Pi_Y^*(t) + 1/Z(t)d(Z(t)\Pi_Y^*(t)) \\ &+ d(1/Z(t))d(Z(t)\Pi_Y^*(t)). \end{aligned} \tag{6.14b}$$

By Itô's formula and $dZ(t) = -\theta(t)Z(t)dW(t)$, we obtain

$$d(1/Z(t)) = -\frac{1}{Z(t)^2}dZ(t) + \frac{1}{Z(t)^3}dZ(t)dZ(t) = \frac{\theta(t)}{Z(t)}d\widetilde{W}(t). \tag{6.14c}$$

Hence

$$d(1/Z(t))d(Z(t)\Pi_Y^*(t)) = \frac{\theta(t)\Gamma(t)}{Z(t)}dt. \tag{6.14d}$$

Combining Equations (6.14) we have

$$d\Pi_Y^*(t) = \Delta(t)d\widetilde{W}(t), \quad \text{where} \;\; \Delta(t) = \theta(t)\Pi_Y^*(t) + \frac{\Gamma(t)}{Z(t)},$$

which proves (6.11).

(iii) It is clear that the portfolio $\{h_S(t), h_B(t)\}_{t\in[0,T]}$ given by (6.12) is adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. By the definition of $h_B(t)$ we have $V(t) = h_S(t)S(t)+h_B(t)B(t) = \Pi_Y(t)$, hence the portfolio replicates the derivative. Furthermore (6.11) entails that $V^*(t) = \Pi_Y^*(t)$ satisfies (6.8), hence, by Theorem 6.1(ii), $\{h_S(t), h_B(t)\}_{t\in[0,T]}$ is a self-financing portfolio, and the proof is completed. $\square$

Consider now the 2+1 dimensional market consisting of a stock, a European derivative on the stock and the risk-free asset. The value of a self-financing portfolio invested in this market satisfies

$$dV(t) = h_S(t)dS(t)+h_Y(t)d\Pi_Y(t)+h_B(t)dB(t), \quad V(t) = h_S(t)S(t)+h_Y(t)\Pi_Y(t)+h_B(t)B(t),$$

where $h_Y(t)$ is the number of shares of the derivative in the portfolio. It follows by (6.11) that the discounted value of this portfolio satisfies

$$\begin{aligned} d(V^*(t)) &= -r(t)D(t)V(t)dt + D(t)h_S(t)dS(t) + D(t)h_Y(t)d\Pi_Y(t) + D(t)h_B(t)B(t)r(t)dt \\ &= -r(t)D(t)h_S(t)S(t)dt + D(t)h_S(t)dS(t) \\ &\quad - r(t)D(t)h_Y(t)\Pi_Y(t)dt + D(t)h_Y(t)d\Pi_Y(t) \\ &= -h_S(t)d(D(t)S(t)) - h_Y(t)d(D(t)\Pi_Y(t)) \\ &= -h_S(t)D(t)\sigma(t)S(t)d\widetilde{W}(t) - h_Y(t)\Delta(t)d\widetilde{W}(t). \end{aligned}$$

We infer that the discounted value process $\{V^*(t)\}_{t\geq 0}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. Hence, by Theorem 3.14, the portfolio is not an arbitrage and therefore the risk-neutral price model for European derivatives satisfies the arbitrage-free principle, see Remark 6.1.

Put-call parity

Being defined as a conditional expectation, the risk-neutral price (6.10) can be computed explicitly only for simple models on the market parameters, see Sections 6.3, 6.6 and 6.7. However the formula (6.10) can be used to derive a number of general qualitative properties on the fair price of options. The most important is the put-call parity relation.

Theorem 6.3. *Let $\Pi_{\mathrm{call}}(t)$ be the fair price at time t of the European call option on the stock with maturity $T > t$ and strike $K > 0$. Let $\Pi_{\mathrm{put}}(t)$ be the price of the European put option with the same strike and maturity. Then the* **put-call parity identity** *holds:*

$$\Pi_{\mathrm{call}}(t) - \Pi_{\mathrm{put}}(t) = S(t) - KB(t,T), \tag{6.15}$$

where $B(t,T) = \widetilde{\mathbb{E}}[D(T)/D(t)|\mathcal{F}_W(t)]$ is the fair value at time t of the contract[1] *with constant pay-off=1 at time T.*

Proof. The pay-off of the call/put option is

$$Y_{\mathrm{call}} = (S(T)-K)_+, \quad Y_{\mathrm{put}} = (K - S(T))_+.$$

Using $(x-K)_+ - (K-x)_+ = (x-K)$, for all $x \in \mathbb{R}$, we obtain

$$\begin{aligned}
\Pi_{\mathrm{call}}(t) - \Pi_{\mathrm{put}}(t) &= \widetilde{\mathbb{E}}[D(t)^{-1}D(T)(S(T)-K)_+|\mathcal{F}_W(t)] - \widetilde{\mathbb{E}}[D(t)^{-1}D(T)(K-S(T))_+|\mathcal{F}_W(t)]\\
&= \widetilde{\mathbb{E}}[D(t)^{-1}D(T)(S(T)-K)|\mathcal{F}_W(t)]\\
&= D(t)^{-1}\widetilde{\mathbb{E}}[D(T)S(T)|\mathcal{F}_W(t)] - K\widetilde{\mathbb{E}}[D(t)^{-1}D(T)|\mathcal{F}_W(t)]\\
&= S(t) - KB(t,T),
\end{aligned}$$

where in the last step we use that the discounted stock price process is a martingale in the risk-neutral probability measure. □

Exercise 6.1. *Prove that the risk-neutral price of European call/put options satisfy the following properties:*

(i) $\Pi_{\mathrm{call}}(t) \leq S(t)$, $\Pi_{\mathrm{put}}(t) \leq KB(t,T)$

(ii) $\Pi_{\mathrm{call}}(t)$ *is non-increasing and convex in the strike price. Similarly* $\Pi_{\mathrm{put}}(t)$ *is non-decreasing and convex in the strike price.*

[1] This contract is called zero-coupon bond, see Section 6.7

(iii) If $B(t,T) \leq 1$, then $\Pi_{\mathrm{call}}(t) \geq (S(t)-K)_+$ and $\Pi_{\mathrm{call}}(t)$ is non-decreasing with respect to T.

Give an intuitive explaination of why these properties are reasonable.

Remark 6.2. Since $B(t,T) = \widetilde{\mathbb{E}}[\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)]$, the condition $B(t,T) \leq 1$ in the property (iii) is satisfied in particular when the interest rate process is non-negative.

Exercise 6.2 (Sol. 32)**.** *Consider the European derivative with pay-off Y at maturity T and the derivative with pay-off $Z = \Pi_Y(t_*)$ at maturity $t_* < T$. Show that $\Pi_Z(t) = \Pi_Y(t)$, $t \in [0, t_*]$.*

6.3 Black-Scholes price of standard European derivatives

In the particular case of a standard European derivative, i.e., when $Y = g(S(T))$, for some measurable function g, the risk-neutral price formula (6.10) becomes

$$\Pi_Y(t) = \widetilde{\mathbb{E}}[g(S(T))\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)].$$

By (6.4) we have

$$S(T) = S(t)\exp\left(\int_t^T (r(s) - \frac{1}{2}\sigma^2(s))ds + \int_t^T \sigma(s)d\widetilde{W}(s)\right),$$

hence the risk-neutral price of standard European derivatives takes the form

$$\Pi_Y(t) = \widetilde{\mathbb{E}}[g(S(t)e^{\int_t^T (r(s)-\frac{1}{2}\sigma^2(s))ds + \int_t^T \sigma(s)d\widetilde{W}(s)})\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)]. \tag{6.16}$$

In this section we compute (6.16) in a **Black-Scholes market**, i.e., in the case when the market parameters $\{\mu(t)\}_{t\geq 0}$, $\{\sigma(t)\}_{t\geq 0}$, $\{r(t)\}_{t\geq 0}$ are deterministic constants. Letting $\mu(t) = \mu$, $r(t) = r$, $\sigma(t) = \sigma > 0$ into (6.4) we obtain that the stock price satisfies

$$dS(t) = rS(t)\,dt + \sigma S(t)\,d\widetilde{W}(t), \tag{6.17}$$

where $\widetilde{W}(t) = W(t) + \frac{\mu - r}{\sigma}t$ is a Brownian motion in the risk-neutral probability measure. The market price of risk $\theta = (\mu - r)/\sigma$ is constant. Integrating (6.17) we obtain that, in the risk-neutral probability, $S(t)$ is given by the geometric Brownian motion

$$S(t) = S(0)e^{(r-\frac{\sigma^2}{2})t + \sigma\widetilde{W}(t)}. \tag{6.18}$$

By (6.16), the risk-neutral price of the standard European derivative is

$$\Pi_Y(t) = e^{-r\tau}\widetilde{\mathbb{E}}[g(S(t)e^{(r-\frac{1}{2}\sigma^2)\tau}e^{\sigma(\widetilde{W}(T)-\widetilde{W}(t))})|\mathcal{F}_W(t)],$$

where $\tau = T - t$ is the time left to maturity. As the increment $\widetilde{W}(T) - \widetilde{W}(t)$ is independent of $\mathcal{F}_W(t)$ and $S(t)$ is $\mathcal{F}_W(t)$-measurable, the conditional expectation above can be computed using Theorem 3.13(x), namely

$$\Pi_Y(t) = v_g(t, S(t)), \tag{6.19a}$$

where the **Black-Scholes price function** $v_g : \overline{\mathcal{D}_T^+} \to \mathbb{R}$ is given by

$$v_g(t,x) = e^{-r\tau}\widetilde{\mathbb{E}}[g(xe^{(r-\frac{1}{2}\sigma^2)\tau}e^{\sigma(\widetilde{W}(T)-\widetilde{W}(t))})] = \frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{\mathbb{R}} g(xe^{(r-\frac{1}{2}\sigma^2)\tau}e^{\sigma\sqrt{\tau}y})e^{-\frac{y^2}{2}}\,dy. \tag{6.19b}$$

Of course we need some conditions on the function g in order for the integral in the right-hand side of (6.19b) to converge to a smooth function. For our purposes it suffices to require that $g \in \mathcal{G}$, where

$$\mathcal{G} = \{g : [0,\infty) \to \mathbb{R} : \begin{array}{l} (i)\ g \text{ is almost everywhere twice differentiable} \\ (ii)\ |g(z)| \leq A + B|z| \text{ for some constants } A, B > 0 \\ (iii)\ g', g'' \text{ are uniformly bounded}\} \end{array} \tag{6.20}$$

Conditions (i)-(iii) are satisfied by the typical pay-off functions used in the applications. It is an easy exercise to show that $g \in \mathcal{G} \Rightarrow v_g \in C^{1,2}(\mathcal{D}_T^+)$.

Definition 6.3. *Let $g \in \mathcal{G}$. The stochastic process $\{\Pi_Y(t)\}_{t\in[0,T]}$ given by* (6.19)*, is called the* **Black-Scholes price** *of the standard European derivative with pay-off $Y = g(S(T))$ and time of maturity $T > 0$.*

Remark 6.3. The fact that the Black-Scholes price of the derivative at time t is a deterministic function of $S(t)$, that is, $\Pi_Y(t) = v_g(t, S(t))$, is an important property for the applications. In fact, thanks to this property, at time t we may look at the price $S(t)$ of the stock in the market and compute explicitly the theoretical price $\Pi_Y(t)$ of the derivative. This theoretical value is, in general, different from the real market price. The difference between the Black-Scholes price and the market price is expressed in terms of the implied volatility of the derivative, as we discuss below.

Next we show that the formula (6.19) is equivalent to the Markov property of the geometric Brownian motion (6.18) in the risk-neutral probability measure $\widetilde{\mathbb{P}}$. To this purpose we rewrite the Black-Scholes price function as $v_g(t,x) = h(T-t, x)$, where, by a change of variable in the integral on the right hand side of (6.19b),

$$h(\tau, x) = \int_{\mathbb{R}} g(y)q(\tau, x, y)\,dy,$$

where

$$q(\tau,x,y)=\frac{e^{-r\tau}\mathbb{I}_{y>0}}{y\sqrt{2\pi\sigma^2\tau}}\exp\left[-\frac{1}{2\sigma^2\tau}\left(\log\frac{y}{x}-(r-\frac{1}{2}\sigma^2)\tau\right)^2\right].$$

Comparing this expression with (3.31), we see that we can write the function q as

$$q(\tau,x,y)=e^{-r\tau}p_*(\tau,x,y),$$

where p_* is the transition density of the geometric Brownian motion (6.18). In particular, the risk-neutral pricing formula of standard European derivatives in a market with constant parameters is equivalent to the identity

$$\widetilde{\mathbb{E}}[g(S(T))|\mathcal{F}_{\widetilde{W}}(t)]=\int_{\mathbb{R}}p_*(T-t,S(t),y)g(y)\,dy,$$

and thus, since $0\leq t\leq T$ are arbitrary, it is equivalent to the Markov property of the geometric Brownian motion (6.18) in the risk-neutral probability measure $\widetilde{\mathbb{P}}$, see again Exercise 3.33. We shall generalize this discussion to markets with non-deterministic parameters in Section 6.6. Moreover replacing $s=0$, $t=\tau$, $\alpha=r-\sigma^2/2$ into (3.33), and letting $u(\tau,x)=e^{r\tau}h(\tau,x)$, we obtain that u satisfies

$$-\partial_\tau u+rx\partial_x u+\frac{1}{2}\sigma^2x^2\partial_x^2u=0,\quad u(0,x)=h(0,x)=v_g(T,x)=g(x).$$

Hence the function $h(\tau,x)$ satisfies

$$-\partial_\tau h+rx\partial_x h+\frac{1}{2}\sigma^2x^2\partial_x^2h=rh,\quad h(0,x)=g(x).$$

As $v_g(t,x)=h(T-t,x)$, we obtain the following result.

Theorem 6.4. *Given $g\in\mathcal{G}$, the Black-Scholes price function v_g is the unique strong solution of the* **Black-Scholes** *PDE*

$$\partial_t v_g+rx\partial_x v_g+\frac{1}{2}\sigma^2x^2\partial_x^2v_g=rv_g,\quad (t,x)\in\mathcal{D}_T^+ \tag{6.21a}$$

with the **terminal condition**

$$v_g(T,x)=g(x). \tag{6.21b}$$

Exercise 6.3. *Write a Matlab code that computes the finite difference solution of the problem* (6.21)*. Use the θ-method presented in Section 5.4.*

Remark 6.4. For the previous exercise one needs to fix the boundary condition at $x=0$ for the strong solution of (6.21a). To derive this boundary condition we just let $x=0$ into (6.21a) and use that $\partial_x v_g,\partial_x^2v_g$ are bounded to obtain that $v(t)=v_g(t,0)$ satisfies $dv/dt=rv$, hence $v(t)=v(T)e^{r(t-T)}$. Moreover $v(T)=v_g(T,0)=g(0)$ and thus the boundary condition at $x=0$ of the Black-Scholes pricing function v_g is

$$v_g(t,0)=g(0)e^{r(t-T)},\quad \text{for all } t\in[0,T]. \tag{6.22}$$

For instance, in the case of a call, i.e., when $g(z) = (z - K)_+$, we obtain $v_g(t, 0) = 0$, for all $t \in [0, T]$, hence the risk-neutral price of a call option is zero when the price of the underlying stock tends to zero. That this should be the case is clear, for the call will never expire in the money if the price of the stock is arbitrarily small. For a put option, i.e., when $g(z) = (K - z)_+$, we have $v_g(t, 0) = Ke^{-r\tau}$, hence the risk-neutral price of a put option is given by the discounted value of the strike price when the price of the underlying stock tends to zero. This is also clear, since in this case the put option will certainly expire in the money, i.e., its value at maturity is K with probability one, and so the value at any earlier time is given by discounting its terminal value.

Next we consider the problem of constructing a replicating (and thus hedging) portfolio of the derivative.

Theorem 6.5. *Consider a standard European derivative priced according to Definition 6.3. The portfolio* $\{h_S(t), h_B(t)\}$ *given by*

$$h_S(t) = \partial_x v_g(t, S(t)), \quad h_B(t) = (\Pi_Y(t) - h_S(t)S(t))/B(t)$$

is a self-financing replicating portfolio for the derivative.

Proof. According to Theorem 6.2, we have to show that the discounted value of the Black-Scholes price satisfies

$$d\Pi_Y^*(t) = D(t)S(t)\sigma\partial_x v_g(t, S(t))d\widetilde{W}(t).$$

A straightforward calculation, using $\Pi_Y(t) = v_g(t, S(t))$, Itô's formula and Itô's product rule, gives

$$\begin{aligned} d(D(t)\Pi_Y(t)) =& D(t)[\partial_t v_g(t, x) + rx\partial_x v_g(t, x) + \frac{1}{2}\sigma^2 x^2 \partial_x^2 v_g(t, x) - rv_g(t, x)]_{x=S(t)} \\ &+ D(t)\sigma S(t)\partial_x v_g(t, S(t))d\widetilde{W}(t). \end{aligned} \tag{6.23}$$

Since v_g solves the Black-Scholes PDE (6.21a), the result follows. □

Exercise 6.4. *Work out the details of the computation leading to* (6.23).

Exercise 6.5. *Derive the price function of standard European derivatives assuming that the market parameters are deterministic functions of time.*

Black-Scholes price of European vanilla options

In this section we focus the discussion on call/put options, which are also called **vanilla options**. We thereby assume that the pay-off of the derivative is given by

$$Y = (S(T) - K)_+, \text{ i.e., } Y = g(S(T)),\ g(x) = (x - K)_+, \quad \text{for a call option,}$$

$$Y = (K - S(T))_+, \text{ i.e., } Y = g(S(T)),\ g(x) = (K - x)_+, \quad \text{for a put option.}$$

The function v_g given by (6.19b) will be denoted by C, for a call option, and by P, for a put option.

Theorem 6.6. *The Black-Scholes price at time t of a European call option with strike price $K > 0$ and maturity $T > 0$ is given by $C(t, S(t))$, where*

$$C(t,x) = x\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \tag{6.24a}$$

$$d_2 = \frac{\log\left(\frac{x}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}, \quad d_1 = d_2 + \sigma\sqrt{\tau}, \tag{6.24b}$$

and where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\, dy$ is the standard normal distribution. The Black-Scholes price of the corresponding put option is given by $P(t, S(t))$, where

$$P(t,x) = \Phi(-d_2)Ke^{-r\tau} - \Phi(-d_1)x. \tag{6.24c}$$

Moreover the put-call parity identity holds:

$$C(t, S(t)) - P(t, S(t)) = S(t) - Ke^{-r\tau}. \tag{6.25}$$

Proof. We derive the Black-Scholes price of call options only, the argument for put options being similar (see Exercise 6.6). We substitute $g(z) = (z - K)_+$ into the right hand side of (6.19b) and obtain

$$C(t,x) = \frac{e^{-r\tau}}{\sqrt{2\pi}} \int_{\mathbb{R}} \left(xe^{(r-\frac{1}{2}\sigma^2)\tau} e^{\sigma\sqrt{\tau}y} - K \right)_+ e^{-\frac{y^2}{2}}\, dy.$$

Now we use that $xe^{(r-\frac{1}{2}\sigma^2)\tau}e^{\sigma\sqrt{\tau}y} > K$ if and only if $y > -d_2$. Hence

$$C(t,x) = \frac{e^{-r\tau}}{\sqrt{2\pi}} \left[\int_{-d_2}^{\infty} xe^{(r-\frac{1}{2}\sigma^2)\tau} e^{\sigma\sqrt{\tau}y} e^{-\frac{y^2}{2}} - K \int_{-d_2}^{\infty} e^{-\frac{y^2}{2}}\, dy \right].$$

Using $-\frac{1}{2}y^2 + \sigma\sqrt{\tau}y = -\frac{1}{2}(y - \sigma\sqrt{\tau})^2 + \frac{\sigma^2}{2}\tau$ and changing variable in the integrals we obtain

$$\begin{aligned}
C(t,x) &= \frac{e^{-r\tau}}{\sqrt{2\pi}} \left[xe^{r\tau} \int_{-d_2}^{\infty} e^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}\, dy - K \int_{-d_2}^{\infty} e^{-\frac{y^2}{2}}\, dy \right] \\
&= \frac{e^{-r\tau}}{\sqrt{2\pi}} \left[xe^{r\tau} \int_{-\infty}^{d_2+\sigma\sqrt{\tau}} e^{-\frac{1}{2}y^2}\, dy - K \int_{-\infty}^{d_2} e^{-\frac{y^2}{2}}\, dy \right] \\
&= s\Phi(d_1) - Ke^{-r\tau}\Phi(d_2).
\end{aligned}$$

The put-call parity (6.25) follows by replacing $r(t) = r$ in (6.15), or directly by (6.24):

$$\begin{aligned}
C(t,x) - P(t,x) &= s\Phi(d_1) - Ke^{-r\tau}\Phi(d_2) - \Phi(-d_2)Ke^{-r\tau} + s\Phi(-d_1) \\
&= x(\Phi(d_1) + \Phi(-d_1)) - Ke^{-r\tau}(\Phi(d_2) + \Phi(-d_2)).
\end{aligned}$$

As $\Phi(z) + \Phi(-z) = 1$, the claim follows. $\square$

Exercise 6.6. *Derive the Black-Scholes price $P(t, S(t))$ of European put options claimed in Theorem 6.6.*

Remark 6.5. The formulas (6.24)-(6.24c) appeared for the first time in the seminal paper [2].

As to the self-financing replicating portfolio for the call/put option, we have $h_S(t) = \partial_x C(t, S(t))$ for call options and $h_S(t) = \partial_x P(t, S(t))$ for put options, see Theorem 6.5, while the number of shares of the risk-free asset in the hedging portfolio is given by

$$h_B(t) = (C(t, S(t)) - S(t)\partial_x C(t, S(t)))/B(t), \quad \text{for call options,}$$

and

$$h_B(t) = (P(t, S(t)) - S(t)\partial_x P(t, S(t)))/B(t), \quad \text{for put options.}$$

Let us compute $\partial_x C$:

$$\partial_x C = \Phi(d_1) + x\Phi'(d_1)\partial_x d_1 - Ke^{-r\tau}\Phi'(d_2)\partial_x d_2.$$

As $\partial_x d_1 = \partial_x d_2 = \frac{1}{\sigma\sqrt{\tau}x}$, and $\Phi'(x) = e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, we obtain

$$\partial_x C = \Phi(d_1) + \frac{1}{\sigma\sqrt{2\pi\tau}}\left(e^{-\frac{1}{2}d_1^2} - \frac{K}{x}e^{-r\tau}e^{-\frac{1}{2}d_2^2}\right).$$

Replacing $d_1 = d_2 + \sigma\sqrt{\tau}$ in the second term we obtain

$$\partial_x C = \Phi(d_1) + \frac{e^{-\frac{1}{2}d_2^2}}{\sigma\sqrt{2\pi\tau}}\left(e^{-\frac{1}{2}\sigma^2\tau - d_2\sigma\sqrt{\tau}} - \frac{K}{x}e^{-r\tau}\right).$$

Using the definition of d_2, the term within round brackets in the previous expression is easily found to be zero, hence

$$\partial_x C = \Phi(d_1).$$

By the put-call parity we find also

$$\partial_x P = \Phi(d_1) - 1 = -\Phi(-d_1).$$

Note that $\partial_x C > 0$, while $\partial_x P < 0$. This agrees with the fact that call options are bought to protect a short position on the underlying stock, while put options are bought to protect a long position on the underlying stock.

Exercise 6.7 (Sol. 33)**.** *Consider a European derivative with maturity T and pay-off Y given by*

$$Y = k + S(T)\log S(T),$$

where $k > 0$ is a constant. Find the Black-Scholes price of the derivative at time $t < T$ and the replicating self-financing portfolio. Find the probability that the derivative expires in the money.

Exercise 6.8 (Sol. 34). *A* **binary** *(or* **digital***) call option with strike K and maturity T pays-off the buyer if and only if $S(T) > K$. If the pay-off is a fixed amount of cash L, then the binary call option is said to be "cash-settled", while if the pay-off is the stock itself then the option is said to be "physically settled". Compute the Black-Scholes price and the replicating portfolio of the physically settled and of the cash-settled binary call option. Repeat the exercise for put options. Is there a put-call parity?*

Exercise 6.9 (Sol. 35). *Given $T_2 > T_1$, a chooser option with maturity T_1 is a contract which gives to the buyer the right to choose at time T_1 whether the derivative becomes a call or a put option with strike K and maturity T_2. Show that the Black-Scholes price of a chooser option is given by*

$$\Pi_Y(t) = C(t, S(t), K, T_2) + P(t, S(t), Ke^{-r(T_2-T_1)}, T_1).$$

where $C(t, S(t), K, T)$ (resp. $P(t, S(t), K, T)$) is the Black-Scholes price of a European call (resp. put) with strike K and maturity T. HINT: You need the result of Exercise 6.2 and the identity $\max(a, b) = a + \max(0, b - a)$.

Exercise 6.10 (Sol. 36). *Let $0 < s < T$. Find the Black-Schole price $\Pi_Y(t)$, $t \in [0, T]$, of the European derivative with pay-off $Y = (S(T) - S(s))_+$ and maturity T. Find also the put-call parity relation satisfied by this derivative and the derivative with pay-off $Z = (S(s) - S(T))_+$.*

The greeks. Implied volatility and volatility curve

The Black-Scholes price of a call (or put) option derived in Theorem 6.6 depends on the price of the underlying stock, the time to maturity, the strike price, as well as on the (constant) market parameters r, σ (it does not depend on α). The partial derivatives of the price function C with respect to these variables are called **greeks**. We collect the most important ones (for call options) in the following theorem.

Theorem 6.7. *The price function C of call options satisfies the following:*

$$\Delta := \partial_x C = \Phi(d_1), \tag{6.26}$$

$$\Gamma := \partial_x^2 C = \frac{\phi(d_1)}{x\sigma\sqrt{\tau}}, \tag{6.27}$$

$$\rho := \partial_r C = K\tau e^{-r\tau}\Phi(d_2), \tag{6.28}$$

$$\Theta := \partial_t C = -\frac{x\phi(d_1)\sigma}{2\sqrt{\tau}} - rKe^{-r\tau}\Phi(d_2), \tag{6.29}$$

$$\nu := \partial_\sigma C = x\phi(d_1)\sqrt{\tau} \quad \textit{(called "vega")}. \tag{6.30}$$

In particular:

- $\Delta > 0$, i.e., the price of a call is increasing on the price of the underlying stock;

- $\Gamma > 0$, i.e., the price of a call is convex on the price of the underlying stock;
- $\rho > 0$, i.e., the price of the call is increasing on the risk-free interest rate;
- $\Theta < 0$, i.e., the price of the call is decreasing in time;
- $\nu > 0$, i.e., the price of the call is increasing on the volatility of the stock.

Exercise 6.11. *Use the put-call parity to derive the greeks of put options.*

The greeks measure the sensitivity of options prices with respect to the market conditions. This information can be used to draw some important conclusions. Let us comment for instance on the fact that vega is positive. It implies that the wish of an investor with a long position on a call option is that the volatility of the underlying stock increased. As usual, since this might not happen, the investor portfolio is exposed to possible losses due to the decrease of the stock volatility (which makes the call option in the portfolio loose value). This exposure can be secured by taking a short position on a swap on the variance of the underlying stock[2]; see Section 6.6.

Exercise 6.12. *Prove that*

$$\lim_{\sigma\to 0^+} C(t,x) = (x - Ke^{-r\tau})_+, \quad \lim_{\sigma\to\infty} C(t,x) = x.$$

Exercise 6.13 (Sol. 37)**.** *Derive the probability density function $f_{C(t)}$ of the Black-Scholes price $C(t) = C(t, S(t))$ of the call option.*

Implied volatility

Let us temporarily re-denote the Black-Scholes price of the call as

$$C(t, S(t), K, T, \sigma),$$

which reflects the dependence of the price on the parameters K, T, σ (we disregard the dependence on r). As shown in Theorem 6.7,

$$\partial_\sigma C(t, S(t), K, T, \sigma) = \text{vega} = \frac{S(t)}{\sqrt{2\pi}} e^{-\frac{d_1^2}{2}} \sqrt{\tau} > 0.$$

Hence the Black-Scholes price of the option is an increasing function of the volatility. Furthermore, by Exercise 6.12,

$$\lim_{\sigma\to 0^+} C(t, S(t), K, T) = (S(t) - Ke^{-r\tau})_+, \quad \lim_{\sigma\to+\infty} C(t, S(t), K, T) = S(t).$$

[2] At least theoretically, since variance swaps on the underlying stock might not be available in the market.

Therefore the function $C(t, S(t), K, T, \cdot)$ is a one-to-one map from $(0, \infty)$ into the interval $I = ((S(t) - Ke^{-r\tau})_+, S(t))$, see Figure 6.1. Now suppose that at some given *fixed* time t the real market price of the call is $\widetilde{C}(t) \in I$. Then there exists a unique value of σ, which depends on the fixed parameters T, K and which we denote by $\sigma_{\text{imp}}(T, K)$, such that

$$C(t, S(t), K, T, \sigma_{\text{imp}}(T, K)) = \widetilde{C}(t).$$

$\sigma_{\text{imp}}(T, K)$ is called the **implied volatility** of the option. The implied volatility must be computed numerically (for instance using Newton's method), since there is no close formula for it.

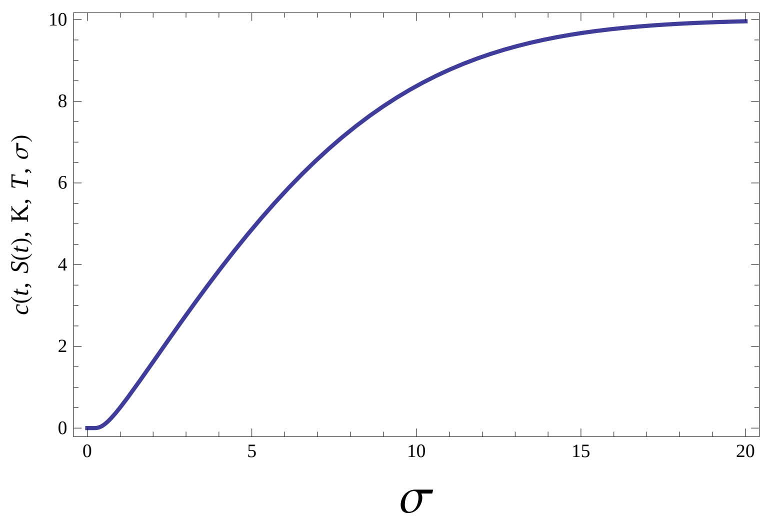


Figure 6.1: We fix $S(t) = 10$, $K = 12$, $r = 0.01$, $\tau = 1/12$ and depict the Black-Scholes price of the call as a function of the volatility. Note that in practice only the very left part of this picture is of interest, because typically $0 < \sigma < 1$.

The implied volatility of an option (in this example of a call option) is a very important parameter and it is often quoted together with the price of the option. If the market followed exactly the assumptions in the Black-Scholes theory, then the implied volatility would be a constant, independent of T, K and equal to the volatility of the underlying asset. In this respect, $\sigma_{\text{imp}}(T, K)$ may be viewed as a quantitative measure of how real markets deviate from ideal Black-Scholes markets.

As a way of example, in Figure 6.2 the implied volatility is determined (graphically) for various Apple call options on May 12, 2014, when the stock was quoted at 585.54 dollars (closing price of the previous market day). All options expire on June 13, 2014 ($\tau = 1$ month $= 1/12$). The value $r = 0.01$ has been used, but the results do no change significantly even assuming $r = 0.05$. In the pictures, K denotes the strike price and $\widetilde{C}(t)$ the call price. We observe that the implied volatility is 20 % in three cases, while for the call with strike $K = 565$ dollars the implied volatility is a little smaller ($\approx 16\%$), which means that the latter call is slightly underpriced compare to the others.

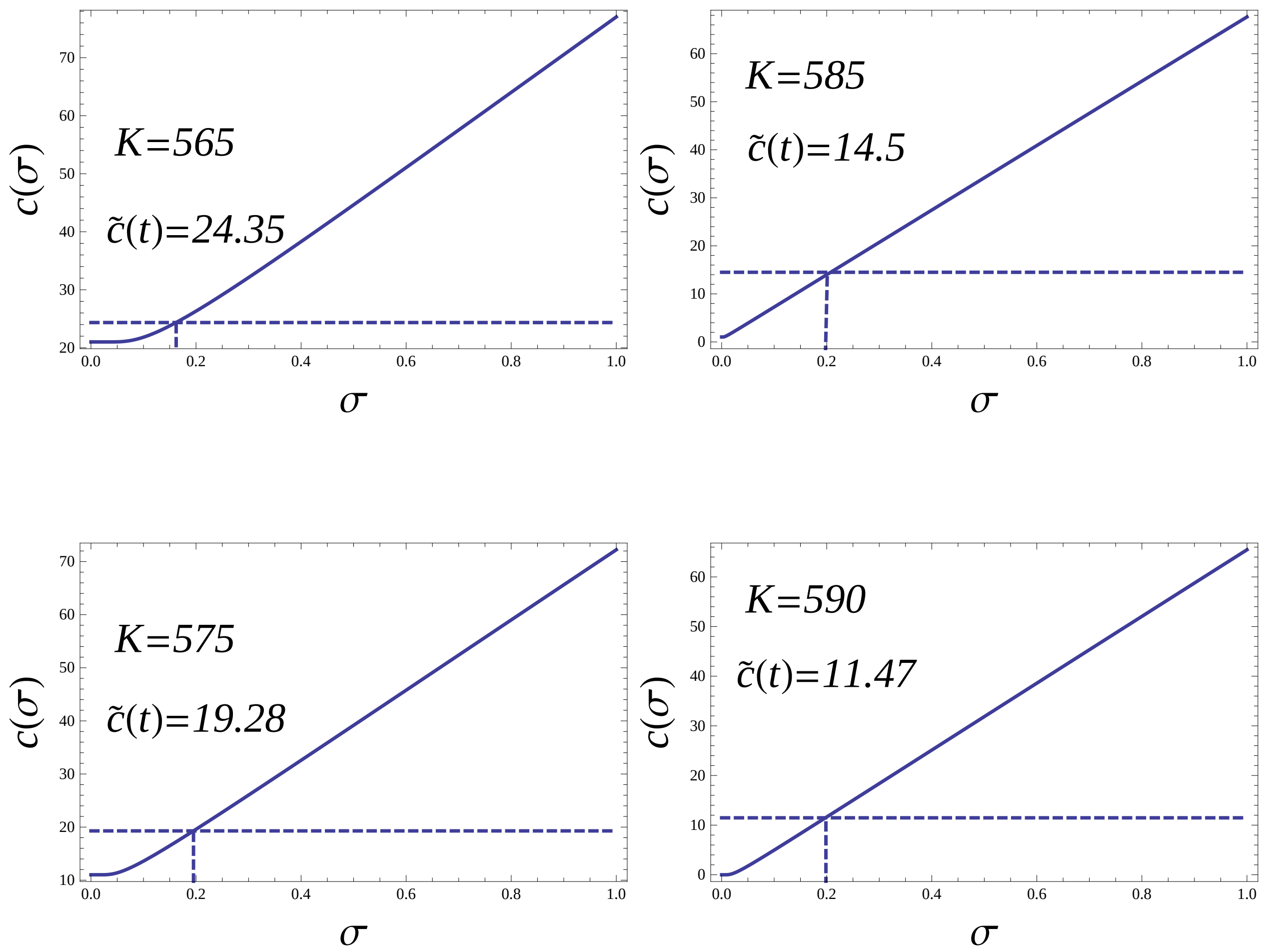
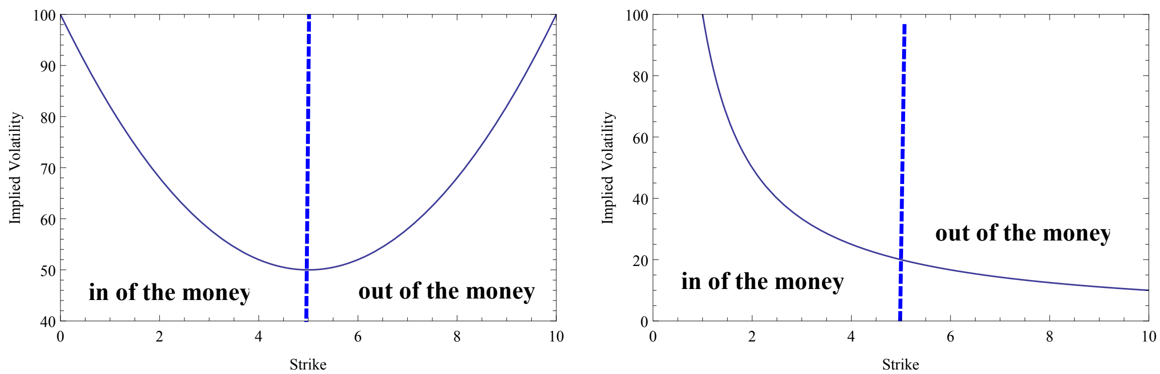


Figure 6.2: Implied volatility of various call options on the Apple stock

Volatility curve

As mentioned before, the implied volatility depends on the parameters T, K. Here we are particularly interested in the dependence on the strike price, hence we re-denote the implied volatility as $\sigma_{\rm imp}(K)$. If the market behaved exactly as in the Black-Scholes theory, then $\sigma_{\rm imp}(K) = \sigma$ for all values of K, hence the graph of $K \to \sigma_{\rm imp}(K)$ would be just a straight horizontal line. Given that real markets do not satisfy exactly the assumptions in the Black-Scholes theory, what can we say about the graph of the **volatility curve** $K \to \sigma_{\rm imp}(K)$? Remarkably, it has been found that there exists recurrent convex shapes for the graph of volatility curves, which are known as **volatility smile** and **volatility skew**, see Figure 6.3. In the case of a volatility smile, the minimum of the graph is reached at the strike price $K \approx S(t)$, i.e., when the call is at the money. This behavior indicates that the more the call is far from being at the money, the more it will be overpriced. Volatility smiles have been common in the market since after the crash in October 19$^{\rm th}$, 1987 (Black Monday), indicating that this event led investors to be more cautious when trading on options that are in or out of the money. Devising mathematical models of volatility and asset prices able to reproduce volatility curves is an active research topic in financial mathematics. We discuss the most popular volatility models in Section 6.6.



(a) Smile

(b) Skew

Figure 6.3: Volatility smile and skew of a call option (not from real data!)

European derivatives on a dividend-paying stock

In this section we consider Black-Scholes markets with a dividend-paying stock. For modeling purposes we assume that, at the dividend payment day, the price of the stock decreases of exactly the amount paid by the dividend. Let $t_0 \in (0, T)$ be the payment day and $a \in (0, 1)$ be the fraction of the stock price payed by the dividend. Letting $S(t_0^-) = \lim_{t \to t_0^-} S(t)$, we then have

$$S(t_0) = S(t_0^-) - aS(t_0^-) = (1 - a)S(t_0^-). \tag{6.31}$$

We assume that on each of the intervals $[0, t_0)$, $[t_0, T]$, the stock price follows a geometric Brownian motion, namely,

$$S(s) = S(t)e^{\alpha(s-t)+\sigma(W(s)-W(t))}, \quad t \in [0, t_0), \ s \in [t, t_0) \tag{6.32}$$

$$S(s) = S(u)e^{\alpha(s-u)+\sigma(W(s)-W(u))}, \quad u \in [t_0, T], \ s \in [u, T]. \tag{6.33}$$

Theorem 6.8. *Consider the standard European derivative with pay-off $Y = g(S(T))$ and maturity T. Let $\Pi_Y^{(a,t_0)}(t)$ be the Black-Scholes price of the derivative at time $t \in [0, T]$ assuming that the underlying stock pays the dividend $aS(t_0^-)$ at time $t_0 \in (0, T)$. Then*

$$\Pi_Y^{(a,t_0)}(t) = \begin{cases} v_g(t, (1 - a)S(t)), & \text{for } t < t_0, \\ v_g(t, S(t)), & \text{for } t \geq t_0, \end{cases}$$

where $v_g(t, x)$ is the Black-Scholes pricing function in the absence of dividends, which is given by (6.19b).

Proof. Using $\frac{S(T)}{S(t)} = e^{\alpha\tau+\sigma(W(T)-W(t))}$, we can rewrite the Black-Scholes price in the form

$$\Pi_Y(t) = e^{-r\tau}\mathbb{E}[g(S(T)e^{(r-\frac{\sigma^2}{2}-\alpha)\tau})|\mathcal{F}_W(t)]. \tag{6.34}$$

Taking the limit $s \to t_0^-$ in (6.32) and using the continuity of the paths of the Brownian motion we find

$$S(t_0^-) = S(t)e^{\alpha(t_0-t)+\sigma(W(t_0)-W(t))}, \quad t \in [0, t_0).$$

Replacing in (6.31) we obtain

$$S(t_0) = (1-a)S(t)e^{\alpha(t_0-t)+\sigma(W(t_0)-W(t))}, \quad t \in [0, t_0).$$

Hence, letting $(s, u) = (T, t_0)$ and $(s, u) = (T, t)$ into (6.33), we find

$$S(T) = \begin{cases} (1-a)S(t)e^{\alpha\tau+\sigma(W(T)-W(t))} & \text{for } t \in [0, t_0), \\ S(t)e^{\alpha\tau+\sigma(W(T)-W(t))} & \text{for } t \in [t_0, T]. \end{cases} \tag{6.35}$$

By the definition of Black-Scholes price in the form (6.34) and denoting $G = (W(T) - W(t))/\sqrt{\tau}$, we obtain

$$\Pi_Y^{(a,t_0)}(t) = e^{-r\tau}\mathbb{E}[g((1-a)S(t)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}G}|\mathcal{F}_W(t)]], \quad \text{for } t \in [0, t_0),$$

$$\Pi_Y^{(a,t_0)}(t) = e^{-r\tau}\mathbb{E}[g(S(t)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}G}|\mathcal{F}_W(t)]], \quad \text{for } t \in [t_0, T].$$

As $G \in \mathcal{N}(0, 1)$ and $S(t)$ is independent of G, the conditional expectation can be computed using Theorem 3.13(x), and the result follows. □

We conclude that for $t \geq t_0$, i.e., after the dividend has been paid, the Black-Scholes price function of the derivative is again given by (6.19b), while for $t < t_0$ it is obtained by replacing x with $(1-a)x$ in (6.19b). To see the effect of this change, suppose that the derivative is a call option; let $C(t, x)$ be the Black-Scholes price function in the absence of dividends and $C_a(t, x)$ be the price function in the case that a dividend is paid at time t_0. Then, according to Theorem 6.8,

$$C_a(t, x) = \begin{cases} C(t, (1-a)x), & \text{for } t < t_0, \\ C(t, x), & \text{for } t \geq t_0. \end{cases}$$

Since $\partial_x C > 0$ (see Theorem 6.7), it follows that $C_a(t, x) < C(t, x)$, for $t < t_0$, that is to say, the payment of a dividend makes the call option on the stock less valuable (i.e., cheaper) than in the absence of dividends until the dividend is paid.

Exercise 6.14 (?). *Give an intuitive explanation for the property just proved for call options on a dividend paying stock.*

Exercise 6.15 (Sol. 38). *A standard European derivative pays the amount* $Y = (S(T) - S(0))_+$ *at time of maturity* T. *Find the Black-Scholes price* $\Pi_Y(0)$ *of this derivative at time* $t = 0$ *assuming that the underlying stock pays the dividend* $(1 - e^{-rT})S(\frac{T}{2}-)$ *at time* $t = \frac{T}{2}$.

Exercise 6.16. *Derive the Black-Scholes price of the derivative with pay-off* $Y = g(S(T))$, *assuming that the underlying pays a dividend at each time* $t_1 < t_2 < \cdots < t_M \in [0, T]$. *Denote by* a_i *the dividend paid at time* t_i, $i = 1, \ldots, M$.

6.4 The Asian option

The Asian call option with **arithmetic average**, strike $K > 0$ and maturity $T > 0$ is the non-standard European style derivative with pay-off

$$Y_{\mathrm{AC}} = \left(\frac{1}{T}\int_0^T S(t)\,dt - K\right)_+,$$

while for the Asian put the pay-off is

$$Y_{\mathrm{AP}} = \left(K - \frac{1}{T}\int_0^T S(t)\,dt\right)_+.$$

We study the Asian option in a Black-Scholes market, i.e., assuming that the market parameters are deterministic constants. The risk-neutral price at time $t \leq T$ of the Asian call is therefore given by

$$\Pi_{\mathrm{AC}}(t) = e^{-r(T-t)}\widetilde{\mathbb{E}}[(Q(T)/T - K)_+|\mathcal{F}_W(t)],$$

where

$$Q(t) = \int_0^t S(\tau)\,d\tau, \quad S(\tau) = S(0)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\widetilde{W}(\tau)}.$$

Exercise 6.17 (Sol. 39)**.** *Prove the following put-call parity identity for Asian options:*

$$\Pi_{\mathrm{AC}}(t) - \Pi_{\mathrm{AP}}(t) = \frac{Q(t)}{T}e^{-r(T-t)} + \frac{S(t)}{rT}(1 - e^{-r(T-t)}) - Ke^{-r(T-t)}. \tag{6.36}$$

Exercise 6.18. *The Asian call with* **geometric average** *is the European style derivative with pay-off*

$$Z = \left(\exp\Big(\frac{1}{T}\int_0^T \log S(t)\,dt\Big) - K\right)_+,$$

where $T > 0$ and $K > 0$ are respectively the maturity and strike of the call. Derive an exact formula for the Black-Scholes price of this option and for the corresponding put option. Derive also the put-call parity. Prove that the Asian call with geometric average is cheaper than the corresponding Asian call with arithmetic average.

Exercise 6.19 (Sol. 40)**.** *Let $r \geq 0$. Prove the following inequalities between the Black -Scholes price of the Asian call and the European call options:*

$$\Pi_{\mathrm{AC}}(0) < \frac{1-e^{-rT}}{rT}C(0, S_0, K, T).$$

Conclude from this that for $r \geq 0$ the Asian call is less valuable than the European call. HINT: You need the Jensen inequality for integrals: $f(\frac{1}{b-a}\int_a^b g(x)\,dx) \leq \frac{1}{b-a}\int_a^b f(g(x))\,dx$, for all $b > a$ and for all functions f, g such that f is convex.

A simple closed formula for the price of the Asian option with arithmetic average is not available. In this section we discuss two numerical methods to price the Asian option with arithmetic average, namely the finite difference method applied to the pricing PDE and the Monte Carlo method applied to the risk-neutral pricing formula (at time $t = 0$).

The PDE method

To price the Asian option with arithmetic average using the PDE method we first observe that the stochastic processes $\{Q(t)\}_{t\in[0,T]}$, $\{S(t)\}_{t\in[0,T]}$ satisfy the system of SDE's

$$dQ(t) = S(t)\,dt, \quad dS(t) = rS(t)\,dt + \sigma S(t)\,d\widetilde{W}(t).$$

By the Markov property of SDE's it follows that the risk-neutral price of the Asian call satisfies

$$\Pi_{\mathrm{AC}}(t) = c(t, S(t), Q(t)). \tag{6.37}$$

for some measurable function c.

Theorem 6.9. *Let $c : [0,T]\times(0,\infty)^2 \to (0,\infty)$ be the strong solution to the terminal value problem*

$$\partial_t c + rx\partial_x c + x\partial_y c + \frac{\sigma^2}{2}x^2\partial_x^2 c = rc, \quad t\in(0,T),\ x,y>0 \tag{6.38a}$$

$$c(T,x,y) = (y/T-K)_+, \quad x,y>0. \tag{6.38b}$$

Then (6.37) holds. Moreover the number of shares of the stock in the self-financing hedging portfolio is given by $h_S(t) = \partial_x c(t,S(t),Q(t))$.

Proof. We have

$$d(e^{-rt}c(t,S(t),Q(t)) = e^{-rt}[\partial_t c + rx\partial_x c + x\partial_y c + \frac{\sigma^2}{2}x^2\partial_x^2 c - rc](t,S(t),Q(t))\,dt$$
$$+ e^{-rt}\partial_x c(t,S(t))\sigma S(t)d\widetilde{W}(t).$$

As c satisfies (6.38), then

$$d(e^{-rt}c(t,S(t),Q(t)) = e^{-rt}\partial_x c(t,S(t),Q(t))\sigma S(t)d\widetilde{W}(t). \tag{6.39}$$

It follows that the process $\{e^{-rt}c(t,S(t),Q(t))\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\in[0,T]}$. In particular

$$e^{-rT}\widetilde{\mathbb{E}}[c(T,S(T),Q(T))|\mathcal{F}_W(t)] = e^{-rt}c(t,S(t),Q(t)), \quad T\geq t.$$

Using $c(T,S(T),Q(T)) = (Q(T)/T-K)_+$ proves (6.37). Moreover by (6.39) the discounted value of the Asian call satisfies

$$\Pi^*_{\mathrm{AC}}(t) = \Pi_{\mathrm{AC}}(0) + \int_0^t e^{-r\tau}\partial_x c(\tau,S(\tau),Q(\tau))\sigma S(t)\,d\widetilde{W}(\tau),$$

hence by Theorem 6.2 the number of shares of the stock in the hedging portfolio is $h_S(t) = \partial_x c(t,S(t),Q(t))$. □

Remark 6.6. It can be shown that, as stated in the theorem, the problem (6.38) admits a unique strong solution, although the proof requires a highly non-trivial generalization of the results presented in Chapter 5. In fact, while at any time $t \in [0, T]$ the solution is a function of the two variables x, y, the **diffusion operator** (i.e., the second-order differential operator) acts only on the x variable. This type of PDE's are called **hypoelliptic** and have been studied systematically by the Swedish mathematician Hörmander (see [16]).

Exercise 6.20. *Use Theorem 6.9 to give an alternative proof of the put-call parity* (6.36).

A simple closed formula solution for the problem (6.38) is not available, hence one needs to rely on numerical methods to find approximate solutions. One such method is the finite difference method described in Chapter 5. To apply this method one needs to specify the boundary conditions for (6.38) at infinity and for $\{x = 0, y > 0\}$, $\{y = 0, x > 0\}$. Concerning the boundary condition at $x = 0$, let $\bar{c}(t, y) = c(t, 0, y)$. Letting $x = 0$ in (6.38) we obtain that $\bar{c}$ satisfies $\partial_t \bar{c} = r\bar{c}$ and $\bar{c}(T, y) = (y/T - K)_+$, from which we derive the boundary condition

$$c(t, 0, y) = e^{-r(T-t)} \left(\frac{y}{T} - K\right)_+ . \tag{6.40}$$

As to the boundary condition when $y \to \infty$, we first observe that the Asian put becomes clearly worthless if $Q(t)$ reaches arbitrarily large values. Hence the put-call parity (6.36) leads us to impose

$$c(t, x, y) \sim \frac{y}{T} e^{-r(T-t)}, \quad \text{as } y \to \infty, \text{ for all } x > 0 \tag{6.41}$$

The boundary conditions at $y = 0$ and $x \to \infty$ are not so obvious. The next theorem shows that one can avoid giving these boundary conditions by a suitable variable transformation.

Theorem 6.10. *Let* $u : [0, \infty) \times \mathbb{R} \to [0, \infty)$ *be the strong solution to the problem*

$$\partial_t u + \frac{\sigma^2}{2}(\gamma(t) - z)^2 \partial_z^2 u = 0, \quad t \in (0, T), z \in \mathbb{R} \tag{6.42a}$$

$$u(T, z) = (z)_+, \quad \lim_{z \to -\infty} u(t, z) = 0, \ \lim_{z \to \infty} (u(t, z) - z) = 0, \quad t \in [0, T), \tag{6.42b}$$

where $\gamma(t) = \frac{1 - e^{-r(T-t)}}{rT}$. *Then the function*

$$c(t, x, y) = xu\left(t, \frac{1}{rT}(1 - e^{-r(T-t)}) + \frac{e^{-r(T-t)}}{x}\left(\frac{y}{T} - K\right)\right) \tag{6.43}$$

solves (6.38) *as well as* (6.40)-(6.41)

Proof. Proving that c solves the PDE in (6.38) is a straightforward calculation which is left as an exercise. We now show that u satisfies the stated boundary conditions. First we observe that $c(T, x, y) = xu(T, x^{-1}(y/T - K)) = x[x^{-1}(y/T - K)]_+ = (y/T - K)_+$, thus c verifies the terminal condition in (6.38). Now, for all x given, $y \to \infty$ is equivalent to the second argument in the function (6.43) going to infinity, hence $u \sim z$ as $z \to \infty$ implies that, as $y \to \infty$,

$$c \sim xz = \frac{x}{rT}(1 - e^{-r(T-t)}) + e^{-r(T-t)}(y/T - K) \sim e^{-r(T-t)}y/T,$$

which proves (6.41). Similarly, the second argument in the function (6.43) tends to $-\infty$ if and only if $x \to 0^+$ and $y/T - K < 0$. Hence $u \to 0$ as $z \to -\infty$ implies (6.40) for $y/T - K < 0$. When $y > 0$ is given such that $y/T > K$, we have that $x \to 0$ is equivalent to the second argument in the function (6.43) diverging to ∞. Hence using $u \sim z$ as $z \to \infty$ as before we have

$$c \sim xz = \frac{x}{rT}(1 - e^{-r(T-t)}) + e^{-r(T-t)}(y/T - K) \sim e^{-r(T-t)}(y/T - K), \quad \text{as } x \to 0^+,$$

which is (6.40) for $y/T > K$. Finally for $y/T = K$ we have $c(t, x, y) = xu(t, (1 - e^{-r(T-t)})/rT) \to 0$ as $x \to 0^+$. This concludes the proof of the theorem. □

Exercise 6.21. *Prove that the function* (6.43) *solves the PDE in* (6.38)*.*

A project to analyze the Asian option by solving numerically the problem (6.42) is proposed in Appendix A.1. An alternative method to compute the price of Asian options is the Monte Carlo method which we discuss next.

The Monte Carlo method

A very popular method to compute numerically the price of non-standard derivatives is the Monte Carlo method. In this section we describe briefly how the method works for Asian options with arithmetic average and leave the generalization to other derivatives as an exercise.

The crude Monte Carlo method

The Monte Carlo method is, in its simplest form, a numerical method to compute the expectation of a random variable. Its mathematical validation is based on the **Law of Large Numbers**, which states the following: Suppose $\{Y_i\}_{i\geq 1}$ is a sequence of i.i.d. random variables with expectation $\mathbb{E}[Y_i] = \mu$. Then the sample average of the first n components of the sequence, i.e.,

$$\overline{Y} = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n),$$

converges (in probability) to μ as $n \to \infty$.

The law of large numbers can be used to justify the fact that if we are given a large number of independent trials $Y_1, \ldots, Y_n$ of the random variable Y, then

$$\mathbb{E}[Y] \approx \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n).$$

To measure how reliable is the approximation of $\mathbb{E}[Y]$ given by the sample average, consider the standard deviation of the trials $Y_1, \dots, Y_n$:

$$s_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\overline{Y} - Y_i)^2}.$$

A simple application of the Central Limit Theorem proves that the random variable

$$\frac{\mu - \overline{Y}}{s_Y/\sqrt{n}}$$

converges in distribution to a standard normal random variable. We use this result to show that the true value μ of $\mathbb{E}[Y]$ has about 95% probability to be in the interval

$$[\overline{Y} - 1.96\frac{s_Y}{\sqrt{n}}, \overline{Y} + 1.96\frac{s_Y}{\sqrt{n}}]. \tag{6.44}$$

Indeed, for n large,

$$\mathbb{P}\left(-1.96 \leq \frac{\mu - \overline{Y}}{s_Y/\sqrt{n}} \leq 1.96\right) \approx \int_{-1.96}^{1.96} e^{-x^2/2}\,\frac{dx}{\sqrt{2\pi}} \approx 0.95.$$

In the applications to options pricing, the random variable Y is the pay-off of a European derivative. Using the Monte Carlo method and the risk-neutral pricing formula, we can approximate the Black-Scholes price at time $t = 0$ of the European derivative with pay-off Y and maturity $T > 0$ with the sample average

$$\Pi_Y(0) = e^{-rT}\frac{Y_1 + \dots Y_n}{n}, \tag{6.45}$$

where $Y_1, \dots, Y_n$ is a large number of independent trials of the pay-off. Each trial Y_i is determined by a path of the stock price. Letting $0 = t_0 < t_1 < \dots < t_N = T$ be a partition of the interval $[0, T]$ with size $t_i - t_{i-1} = h$, we may construct a sample of n paths of the geometric Brownian motion on the given partition with the following simple Matlab function:

```
function Path=StockPath(s,sigma,r,T,N,n)
h=T/N;
W=randn(n,N);
q=ones(n,N);
Path=s*exp((r-sigma^2/2)*h.*cumsum(q')+sigma*sqrt(h)*cumsum(W'));
Path=[s*ones(1,n);Path];
```

Note carefully that the stock price is modeled as a geometric Brownian motion with mean of log return $\alpha = r - \sigma^2/2$, which means that the geometric Brownian motion is risk-neutral. This is of course correct, since the expectation in (6.45) that we want to compute is in the risk-neutral probability measure. In the case of the Asian call option with arithmetic average, strike K and maturity T the pay-off is given by

$$Y = \left(\frac{1}{T}\int_0^T S(t)\,dt - K\right)_+ \approx \left(\frac{1}{N}\sum_{i=1}^N S(t_i) - K\right)_+ .$$

The following function computes the approximate price of the Asian option using the Monte Carlo method:

```
function [price, conf95]=MonteCarlo_AC(s,sigma,r,K,T,N,n)
tic
stockPath=StockPath(s,sigma,r,T,N,n);
payOff=max(0,mean(stockPath)-K);
price=exp(-r*T)*mean(payOff);
conf95=1.96*std(payOff)/sqrt(n);
toc
```

The function also return the 95% confidence interval of the result. For example, by running the command

```
[price, conf95]=MonteCarlo_AC(100,0.5,0.05,100,1/2,100,1000000)
```

we get `price`=8.5799, `conf95`=0.0283, which means that the Black-Scholes price of the Asian option with the given parameters has 95% probability to be in the interval 8.5799 ± 0.0283. The calculation took about 4 seconds. Note that the 95% confidence is $0.0565/8.5799 * 100 \approx 0.66\%$ of the price.

Exercise 6.22 (Matlab). *Look for the definition of Barrier options and Lookback options. Write a Matlab code that computes the price of these derivatives using the Monte Carlo method. Study numerically the dependence of the price on the market parameters.*

Control variate Monte Carlo method

The crude Monte Carlo method just described can be improved in a number of ways. For instance, it follows by (6.44) that in order to shrink the confidence interval of the Monte Carlo price one can try to reduce the standard derivation s. There exist several methods to decrease the standard deviation of a Monte Carlo computation, which are collectively called **variance reduction techniques**. Here we describe the **control variate** method.

Suppose we want to compute $\mathbb{E}[Y]$. The idea of the control variate method is to introduce a second random variable Z for which $\mathbb{E}[Z]$ can be computed *exactly* and then write

$$\mathbb{E}[Y] = \mathbb{E}[X] + \mathbb{E}[Z], \quad \text{where } X = Y - Z.$$

Hence the Monte Carlo approximation of $\mathbb{E}[Y]$ can now be written as

$$\mathbb{E}[Y] \approx \frac{X_1 + \cdots + X_n}{n} + \mathbb{E}[Z],$$

where $X_1, \dots, X_n$ are independent trials of the random variable X. This approximation improves the crude Monte Carlo estimate (without control variate) if the sample average estimator of $\mathbb{E}[X]$ is better than the sample average estimator of $\mathbb{E}[Y]$. Because of (6.44), this will be the case if $(s_X)^2 < (s_Y)^2$. It will now be shown that the latter inequality holds if Y, Z *have a positive large correlation.* Letting $Y_1, \dots, Y_n$ be independent trials of Y and $Z_1, \dots, Z_n$ be independent trials of Z, we compute

$$\begin{aligned}(s_X)^2 &= \frac{1}{n-1}\sum_{i=1}^n (\overline{X} - X_i)^2 = \frac{1}{n-1}\sum_{i=1}^n ((\overline{Y} - \overline{Z}) - (Y_i - Z_i))^2 \\ &= (s_Y)^2 + (s_Z)^2 - 2C(Y, Z),\end{aligned}$$

where $C(Y, Z)$ is the sample covariance of the trials $(Y_1, \dots, Y_n)$, $(Z_1, \dots, Z_n)$, namely

$$C(Y, Z) = \sum_{i=1}^n (\overline{Y} - Y_i)(\overline{Z} - Z_i).$$

Hence $(s_X)^2 < (s_Y)^2$ holds provided $C(Y, Z)$ is sufficiently large and positive (precisely, $C(Y, Z) > s_Z/\sqrt{2}$). As $C(Y, Z)$ is an unbiased estimator of $\mathrm{Cov}(Y, Z)$, then the use of the control variate Z will improve the performance of the crude Monte Carlo method if Y, Z have a positive large correlation.

Exercise 6.23 (Matlab)**.** *Write a Matlab code that computes the Black-Scholes price at time $t = 0$ and the confidence interval of the Asian option with arithmetic average using the control variate Monte Carlo method and the pay-off of the Asian option with geometric mean as control variate. Compare the new method with the crude Monte Carlo method and show that the control variate technique improves the performance of the computation. Finally use the control variate Monte Carlo method to study numerically how the price of the Asian call with arithmetic average depends on the parameters of the option. In particular:*

(a) Verify numerically the put-call parity (6.36)

(b) Show that the Asian call is less sensitive to volatility than the standard call. Do you have an intuitive explanation for this?

(c) Show that for large volatility the Monte Carlo method becomes unstable (the confidence interval grows very fast)

6.5 The lookback option

Lookback options are non-standard European style derivatives whose pay-off depends on the minimum or maximum of the stock price within a given time period until maturity. There exists four main types of lookback options.

- A lookback **call** option with **floating strike** and maturity $T > 0$ gives to the owner the right to buy the underlying stock at maturity for the minimum price of the stock in the interval $[0, T]$. Thus the pay-off for this lookback option is

$$Y_{\mathrm{LC}}^{\mathrm{float}} = S(T) - \min\{S(t), t \in [0, T]\}.$$

- A lookback **put** option with **floating strike** and maturity $T > 0$ gives to the owner the right to sell the underlying stock at maturity for the maximum price of the stock in the interval $[0, T]$. Thus the pay-off for this lookback option is

$$Y_{\mathrm{LP}}^{\mathrm{float}} = \max\{S(t), t \in [0, T]\} - S(T).$$

- A lookback **call** option with **fixed strike** $K > 0$ and maturity $T > 0$ pays the buyer the difference between the maximum of the stock price in the interval $[0, T]$ and the strike K, provided this difference is positive. Hence the pay-off for this lookback option is

$$Y_{\mathrm{LC}}^{\mathrm{fixed}} = (\max\{S(t), t \in [0, T]\} - K)_+.$$

- A lookback **put** option with **fixed strike** $K > 0$ and maturity $T > 0$ pays the buyer the difference between the strike price K and the minimum of the stock price in the interval $[0, T]$, provided this difference is positive. Hence the pay-off for this lookback option is

$$Y_{\mathrm{LP}}^{\mathrm{fixed}} = (K - \min\{S(t), t \in [0, T]\})_+.$$

In the following section we focus on the lookback call option with floating strike in a Black-Scholes market. In particular the stock price is given by

$$S(t) = S(0)e^{(r - \frac{1}{2}\sigma^2)t + \sigma \widetilde{W}(t)}$$

and the risk-neutral price for the floating strike lookback call option with maturity $T > 0$ is

$$\Pi_{\mathrm{LC}}^{\mathrm{float}}(t) = e^{-r(T-t)}\widetilde{\mathbb{E}}[S(T) - \min\{S(\tau), \tau \in [0, T]\}|\mathcal{F}_W(t)].$$

Exercise 6.24. *Show that the price at time $t = 0$ of the lookback call and put options with floating strike is a linear increasing function of the stock price at time zero.*

Pricing PDE for the lookback call option with floating strike

The main purpose of this section is to derive the PDE satisfied by the pricing function of lookback call options with floating strike.

Theorem 6.11. *Let* $v:(0,T)\times(0,\infty)\times(0,\infty)$, $v=v(t,x,y)$, *satisfy*

$$\partial_t v + rx\partial_x v + \frac{1}{2}\sigma^2x^2\partial_x^2 v = rv, \quad t\in(0,T),\ x>0,\ 0<y<x, \tag{6.46a}$$
$$\partial_y v(t,x,x) = 0, \quad t\in[0,T], x>0, \tag{6.46b}$$
$$v(T,x,y) = x-y, \quad 0\leq y\leq x. \tag{6.46c}$$

Then $\Pi_{\mathrm{LC}}^{\mathrm{float}}(t) = v(t,S(t),\min_{0\leq\tau\leq t}S(\tau))$.

Proof. Let $Y(t)=\min_{0\leq\tau\leq t}S(\tau)$; note that $\{Y(t)\}_{t\geq0}$ is a non-increasing process in the space $\mathcal{C}^0[\mathcal{F}_W(t)]$. However $\{Y(t)\}_{t\geq0}$ is *not* a diffusion process. We now show that

$$dY(t)dY(t)=0.$$

Recall that this means that the quadratic variation of $\{Y(t)\}_{t\geq0}$ is zero in any interval $[0,T]$ along any sequence of partitions $\{\Pi_n\}_{n\in\mathbb{N}}$ of this interval such that $\|\Pi_n\|\to0$, as $n\to\infty$. Letting $\Pi_n=\{t_0=0,t_1^{(n)},\dots,t^n_{m(n)}=T\}$, we have to prove that

$$\lim_{n\to\infty}\sum_{j=1}^{m(n)}(Y(t_j^{(n)})-Y(t_{j-1}^{(n)}))^2=0 \quad \text{in } L^2(\Omega).$$

But

$$\begin{aligned}\sum_{j=1}^{m(n)}(Y(t_j^{(n)})-Y(t_{j-1}^{(n)}))^2 &\leq \max_j|Y(t_j^{(n)})-Y(t_{j-1}^{(n)})|\sum_j|Y(t_j^{(n)})-Y(t_{j-1}^{(n)})|\\ &=\max_j|Y(t_{j-1}^{(n)})-Y(t_j^{(n)})|\sum_j(Y(t_{j-1}^{(n)})-Y(t_j^{(n)}))\\ &=\max_j|Y(t_{j-1}^{(n)})-Y(t_j^{(n)})|(Y(0)-Y(T)),\end{aligned}$$

where in the sum we used that $Y(t)$ is non-increasing to write $|Y(t)-Y(s)|=Y(s)-Y(t)$, for $t\geq s$. As Y is continous in time, then $\max_j|Y(t_{j-1}^{(n)}))-Y(t_j^{(n)})|\to0$, pointwise in $\omega\in\Omega$. As $Y(t_j^{(n})\leq Y(0)$, then, by the dominated convergence theorem, the limit is zero also in L^2, which completes the proof of $dY(t)dY(t)=0$. Similarly one can prove that $dS(t)dY(t)=0$ (see Exercise 6.25). Hence applying Itô's formula we obtain

$$\begin{aligned}d(e^{-rt}v(t,S(t),Y(t))) &= e^{-rt}(\partial_t v+rx\partial_x v+\frac{1}{2}\sigma^2x^2\partial_x^2v-rv)(t,S(t),Y(t))\,dt\\ &\quad+e^{-rt}\sigma S(t)\partial_x v(t,S(t),Y(t))d\widetilde{W}(t)+e^{-rt}\partial_y v(t,S(t),Y(t))dY(t).\end{aligned}$$

The drift term $(\dots)\,dt$ is zero by the PDE (6.46a). We now show that the term $(\dots)dY(t)$ is also zero, thereby concluding that $\{e^{-rt}v(t,S(t)Y(t)\}_{t\geq0}$ is a $\widetilde{\mathbb{P}}$-martingale relative to

$\{\mathcal{F}_W(t)\}_{t\geq 0}$. Since $Y(t)$ is non-increasing, then it has bounded first variation and therefore the integral

$$\int_0^t \partial_y v(\tau, S(\tau), Y(\tau))dY(\tau)$$

can be understood in the Riemann–Stieltjes sense. We divide this integral as

$$\begin{aligned}\int_0^t \partial_y v(\tau, S(\tau), Y(\tau))dY(\tau) = &\int_{S(\tau)>Y(\tau)} \partial_y v(\tau, S(\tau), Y(\tau))dY(\tau)\\ &+ \int_{S(\tau)=Y(\tau)} \partial_y v(\tau, S(\tau), Y(\tau))dY(\tau).\end{aligned}$$

The second piece is zero by the boundary condition (6.46b). The second piece is also zero, because $\{S(\tau) > Y(\tau)\}$ is an open set (as S, Y are time-continuous) and $Y(\tau)$ is constant in this set (that is "$dY(\tau) = 0$"). It follows that the process $\{e^{-rt}v(t, S(t)Y(t)\}_{t\geq 0}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$; in particular

$$\widetilde{\mathbb{E}}[e^{-rT}v(T, S(T), Y(T)]|\mathcal{F}_W(t)] = e^{-rt}v(t, S(t), Y(t)).$$

Hence, by the terminal condition (6.46c),

$$v(t, S(t), Y(t)) = e^{-r(T-t)}\widetilde{\mathbb{E}}[S(T) - Y(T)]|\mathcal{F}_W(t)],$$

which is the claim. □

Exercise 6.25. *Prove the property $dS(t)dY(t) = 0$ used in the previous theorem.*

To study the problem (6.46) one needs a complete set of boundary conditions for strong solutions. Assume first that $y \to 0$. This means that the stock price has reached the value zero at some time $0 \leq \tau \leq t$, in which case of course the minimum stock price in the interval $[0, T]$ will be zero with probability 1. Hence for $y \to 0^+$, the lookback call price converges to its highest possible value, that is

$$v(t, x, 0) = x. \quad t \in [0, T]. \tag{6.47}$$

The boundary value as $x \to \infty$ is not so obvious. In the next theorem we show that the 1+2 dimensional problem (6.46) can be reduced to a 1+1 dimensional problem with explicit boundary conditions.

Theorem 6.12. *Let $u : (1, \infty) \to \mathbb{R}$ satisfy*

$$\partial_t u + rz\partial_z u + \frac{1}{2}\sigma^2 z^2 \partial_{zz} u = ru, \ t \in (0, T), \ z > 1 \tag{6.48a}$$

with terminal condition

$$u(T, z) = z - 1 \tag{6.48b}$$

and boundary conditions

$$\lim_{z\to\infty}(u(t,z)-z)=0, \quad u(t,1)-\partial_z u(t,1)=0. \tag{6.48c}$$

Then

$$v(t,x,y)=yu\left(t,\frac{x}{y}\right)$$

solves (6.46) *as well as* (6.47).

Exercise 6.26. *Prove the previous theorem.*

6.6 Local and Stochastic volatility models

In this and the next section we present a method to compute the risk-neutral price of European derivatives when the market parameters are not deterministic functions. We first assume in this section that the interest rate of the money market is constant, i.e., $r(t)=r$, which is quite reasonable for derivatives with short maturity such as options; stochastic interest rate models are important for pricing derivatives with long time of maturity, e.g. coupon bonds, which are discussed in Section 6.7. Assuming that the derivative is the standard European derivative with pay-off function g and maturity T, the risk-neutral price formula (6.1) becomes

$$\Pi_Y(t)=e^{-r\tau}\widetilde{\mathbb{E}}[g(S(T))|\mathcal{F}_W(t)], \quad \tau=T-t. \tag{6.49}$$

Motivated by our earlier results on the Black-Scholes price, and Remark 6.3, we attempt to re-write the risk-neutral price formula in the form

$$\Pi_Y(t)=v_g(t,S(t)) \quad \text{for all } t\in[0,T], \text{ for all } T>0, \tag{6.50}$$

for some function $v_g:\overline{\mathcal{D}_T^+}\to(0,\infty)$, which we call the pricing function of the derivative. By (6.49), this is equivalent to

$$\widetilde{\mathbb{E}}[g(S(T))|\mathcal{F}_W(t)]=e^{r\tau}v_g(t,S(t)) \tag{6.51}$$

i.e., to the property that $\{S(t)\}_{t\geq 0}$ is a Markov process in the risk-neutral probability measure $\widetilde{\mathbb{P}}$, relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. At this point it remains to understand for which stochastic processes $\{\sigma(t)\}_{t\geq 0}$ does the generalized geometric Brownian motion (6.4) satisfies this Markov property. We have seen in Section 5.1 that this holds in particular when $\{S(t)\}_{t\geq 0}$ satisfies a (system of) stochastic differential equation(s), see Section 5.1. Next we discuss two examples which encompass most of the volatility models used in the applications: Local volatility models and Stochastic volatility models.

Local volatility models

A **local volatility model** is a special case of the generalized geometric Brownian motion in which the instantaneous volatility of the stock $\{\sigma(t)\}_{t\geq 0}$ is assumed to be a deterministic function of the stock price $S(t)$. Given a continuous function $\beta : [0,\infty)\times[0,\infty)\to(0,\infty)$, we then let

$$\sigma(t)S(t) = \beta(t, S(t)), \tag{6.52}$$

into (6.4), so that the stock price process $\{S(t)\}_{t\geq 0}$ satisfies the SDE

$$dS(t) = rS(t)\,dt + \beta(t, S(t))d\widetilde{W}(t), \quad S(0) = S_0 > 0. \tag{6.53}$$

We assume that this SDE admits a unique global solution, which is true in particular under the assumptions of Theorem 5.1. To this regard we observe that the drift term $\alpha(t,x) = rx$ in (6.53) satisfies both (5.3) and (5.4), hence these conditions restrict only the form of the function $\beta(t,x)$. In the following we shall also assume that the solution $\{S(t)\}_{t\geq 0}$ of (6.53) is non-negative a.s. for all $t > 0$. Note however that the stochastic process solution of (6.53) will in general hit zero with positive probability at any finite time. For example, letting $\beta(t,x) = \sqrt{x}$, the stock price (6.53) is a CIR process (5.26) with $b = 0$ and so, according to Theorem 5.6, $S(t) = 0$ with positive probability for all $t > 0$.

Theorem 6.13. *Let $g \in \mathcal{G}$ and assume that the Kolmogorov PDE*

$$\partial_t u + rx\partial_x u + \frac{1}{2}\beta(t,x)^2\partial_x^2 u = 0 \quad (t,x)\in\mathcal{D}_T^+, \tag{6.54}$$

associated to (6.53) admits a (necessarily unique) strong solution in the region $\mathcal{D}_T^+$ satisfying $u(T,x) = g(x)$. Let also

$$v_g(t,x) = e^{-r\tau}u(t,x).$$

Then we have the following.

(i) v_g satisfies

$$\partial_t v_g + rx\partial_x v_g + \frac{1}{2}\beta(t,x)^2\partial_x^2 v_g = rv_g \quad (t,x)\in\mathcal{D}_T^+, \tag{6.55}$$

and the terminal condition

$$v_g(T,x) = g(x). \tag{6.56}$$

(ii) The price of the European derivative with pay-off $Y = g(S(T))$ and maturity $T > 0$ is given by (6.50).

(iii) The portfolio given by

$$h_S(t) = \partial_x v_g(t, S(t)), \quad h_B(t) = (\Pi_Y(t) - h_S(t)S(t))/B(t)$$

is a self-financing hedging portfolio.

Proof. (i) It is straightforward to verify that v_g satisfies (6.55).

(ii) Let $X(t) = v_g(t, S(t))$. By Itô's formula we find

$$dX(t) = (\partial_t v_g(t, S(t)) + rS(t)\partial_x v_g(t, S(t)) + \frac{1}{2}\beta(t, S(t))^2\partial_x^2 v_g(t, S(t)))dt \\ + \beta(t, S(t))\partial_x v_g(t, S(t))d\widetilde{W}(t).$$

Hence

$$d(e^{-rt}X(t)) = e^{-rt}(\partial_t v_g + rx\partial_x v_g + \frac{1}{2}\beta(t, x)^2\partial_x^2 v_g - rv_g)(t, S(t))dt \\ + e^{-rt}\beta(t, S(t))\partial_x v_g(t, S(t))d\widetilde{W}(t).$$

As $v_g(t, x)$ satisfies (6.55), the drift term in the right hand side of the previous equation is zero. Hence

$$e^{-rt}v_g(t, S(t)) = v_g(t, S_0) + \int_0^t e^{-ru}\beta(u, S(u))\partial_x v_g(u, S(u))d\widetilde{W}(u). \tag{6.57}$$

It follows that[3] the stochastic process $\{e^{-rt}v_g(t, S(t))\}_{t\geq 0}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. Hence

$$\widetilde{\mathbb{E}}[e^{-rT}v_g(T, S(T))|\mathcal{F}_W(t)] = e^{-rt}v_g(t, S(t)), \quad \text{for all } 0 \leq t \leq T.$$

Using the boundary condition (6.56), we find

$$v_g(t, S(t)) = e^{-r\tau}\widetilde{\mathbb{E}}[g(S(T))|\mathcal{F}_W(t)],$$

which proves (6.50).

(iii) Replacing $\Pi_Y(t) = v_g(t, S(t))$ into (6.57), we find

$$\Pi_Y^*(t) = \Pi_Y(0) + \int_0^t e^{-ru}\beta(u, S(u))\partial_x v_g(u, S(u))d\widetilde{W}(u).$$

Hence the claim on the hedging portfolio follows by Theorem 6.2.

□

A closed formula for the solution of (6.54) is rarely available, hence to compute the price of the derivative one needs to rely on numerical methods, such as those discussed in Section 5.4.

[3]Recall that we assume that Itô's integrals are martingales!

Example: The CEV model

For the **constant elasticity variance (CEV)** model, we have $\beta(t, S(t)) = \sigma S(t)^\delta$, where $\sigma > 0$, $\delta > 0$ are constants. The SDE for the stock price becomes

$$dS(t) = rS(t)dt + \sigma S(t)^\delta d\widetilde{W}(t), \quad S(0) = S_0 > 0. \tag{6.58}$$

For $\delta = 1$ we recover the Black-Scholes model. For $\delta \neq 1$, we can construct the solution of (6.58) using a CIR process, as shown in the following exercise.

Exercise 6.27. *Given σ, r and $\delta \neq 1$, define*

$$a = 2r(\delta - 1),\ c = -2\sigma(\delta - 1),\ b = \frac{\sigma^2}{2r}(2\delta - 1),\ \theta = -\frac{1}{2(\delta - 1)}.$$

Let $\{X(t)\}_{t\geq 0}$ be the CIR process

$$dX(t) = a(b - X(t))\,dt + c\sqrt{X(t)}d\widetilde{W}(t), \quad X(0) = x > 0.$$

Show that $S(t) = X(t)^\theta$ solves (6.58) with $S_0 = x^\theta$.

It follows by Exercise 6.27, and by Feller's condition $ab \geq c^2/2$ for the positivity of the CIR process, that the solution of (6.58) remains strictly positive a.s. if $\delta \geq 1$, while for $0 < \delta < 1$, the stock price hits zero in finite time with positive probability.

The Kolmogorov PDE (6.54) associated to the CEV model is

$$\partial_t u + rx\partial_x u + \frac{\sigma^2}{2}x^{2\delta}\partial_x^2 u = 0,\ (t, x) \in \mathcal{D}_T^+.$$

Given a terminal value g at time T as in Theorem 6.13, the previous equation admits a unique solution. However a fundamental solution, in the sense of Theorem 5.5, exists only for $\delta > 1$, as otherwise the stochastic process $\{S(t)\}_{t\geq 0}$ hits zero at any finite time with positive probability and therefore the density of the random variable $S(t)$ has a discrete part. The precise form of the (generalized) density $f_{S(t)}(x)$ in the CEV model is known for all values of δ and are given for instance in [21]. An exact formula for call options can be found in [25]. A project that aims to analyze the CEV model by the finite difference method is proposed in Appendix A.2.

Stochastic volatility models

For local volatility models, the stock price and the instantaneous volatility are both stochastic processes. However there is only one source of randomness which drives both these processes, namely a single Brownian motion $\{W(t)\}_{t\geq 0}$. The next level of generalization consists in assuming that the stock price and the volatility are driven by two different sources of randomness.

Definition 6.4. *Let $\{W_1(t)\}_{t\geq 0}$, $\{W_2(t)\}_{t\geq 0}$ be two independent Brownian motions and $\{\mathcal{F}_W(t)\}_{t\geq 0}$ be their own generated filtration. Let $\rho \in [-1,1]$ be a deterministic constant and $\mu, \eta, \beta : [0,\infty)^3 \to \mathbb{R}$ be continuous functions. A* ***stochastic volatility model*** *is a pair of (non-negative) stochastic diffusion processes $\{S(t)\}_{t\geq 0}$, $\{v(t)\}_{t\geq 0}$ satisfying the following system of SDE's:*

$$dS(t) = \mu(t, S(t), v(t))S(t)\,dt + \sqrt{v(t)}S(t)\,dW_1(t), \tag{6.59}$$

$$dv(t) = \eta(t, S(t), v(t))\,dt + \beta(t, S(t), v(t))\sqrt{v(t)}(\rho\,dW_1(t) + \sqrt{1-\rho^2}\,dW_2(t)). \tag{6.60}$$

We see from (6.59) that $\{v(t)\}_{t\geq 0}$ is the instantaneous variance of the stock price $\{S(t)\}_{t\geq 0}$. Moreover the process $\{W^{(\rho)}(t)\}_{t\geq 0}$ given by

$$W^{(\rho)}(t) = \rho\,W_1(t) + \sqrt{1-\rho^2}\,W_2(t)$$

is a Brownian motion satisfying

$$dW_1(t)dW^{(\rho)}(t) = \rho\,dt;$$

in particular the two Brownian motions $\{W_1(t)\}_{t\geq 0}$, $\{W^{(\rho)}\}_{t\geq 0}$ are not independent, as their cross variation is not zero; in fact, by Exercise 4.4, ρ is the correlation of the two Brownian motions. Hence in a stochastic volatility model the stock price and the volatility are both stochastic processes driven by two correlated Brownian motions. We assume that $\{S(t)\}_{t\geq 0}$ is non-negative and $\{v(t)\}_{t\geq 0}$ is positive a.s. for all times, although we refrain from discussing under which general conditions this is verified (we will present an example below).

Our next purpose is to introduce a risk-neutral probability measure such that the discounted price of the stock is a martingale. As we have two Brownian motions in this model, we shall apply the two-dimensional Girsanov Theorem 4.11 to construct such a probability measure. Precisely, let $r > 0$ be the constant interest rate of the money market and $\gamma : [0,\infty)^3 \to \mathbb{R}$ be a continuous function We define

$$\theta_1(t) = \frac{\mu(t, S(t), v(t)) - r}{\sqrt{v(t)}}, \quad \theta_2(t) = \gamma(t, S(t), v(t)), \quad \theta(t) = (\theta_1(t), \theta_2(t)).$$

Given $T > 0$, we introduce the new probability measure $\widetilde{\mathbb{P}}^{(\gamma)}$ equivalent to $\mathbb{P}$ by $\widetilde{\mathbb{P}}^{(\gamma)}(A) = \mathbb{E}[Z(T)\mathbb{I}_A]$, for all $A \in \mathcal{F}$, where

$$Z(t) = \exp\left(-\int_0^t \theta_1(s)\,dW_1(s) - \int_0^t \theta_2(s)\,dW_2(s) - \frac{1}{2}\int_0^t \|\theta(s)\|^2\,ds\right).$$

Then by Theorem 4.11, the stochastic processes

$$\widetilde{W}_1(t) = W_1(t) + \int_0^t \theta_1(s)\,ds, \quad \widetilde{W}_2^{(\gamma)}(t) = W_2(t) + \int_0^t \gamma(s)\,ds$$

are two $\widetilde{\mathbb{P}}^{(\gamma)}$-independent Brownian motions. Moreover (6.59)-(6.60) can be rewritten as

$$dS(t) = rS(t)dt + \sqrt{v(t)}S(t)d\widetilde{W}_1(t), \tag{6.61a}$$
$$dv(t) = [\eta(t, S(t), v(t)) - \sqrt{v(t)}\psi(t, S(t), v(t))\beta(t, S(t), v(t))]dt$$
$$+ \beta(t, S(t), v(t))\sqrt{v(t)}d\widetilde{W}^{(\rho,\gamma)}, \tag{6.61b}$$

where $\{\psi(t, S(t), v(t))\}_{t\geq 0}$ is the $\{\mathcal{F}_W(t)\}_{t\geq 0}$-adapted stochastic process given by

$$\psi(t, S(t), v(t)) = \frac{\mu(t, S(t), v(t)) - r}{\sqrt{v(t)}}\rho + \gamma(t, S(t), v(t))\sqrt{1-\rho^2} \tag{6.62}$$

and where

$$\widetilde{W}^{(\rho,\gamma)}(t) = \rho\widetilde{W}_1(t) + \sqrt{1-\rho^2}\widetilde{W}_2^{(\gamma)}(t).$$

Note that the $\widetilde{\mathbb{P}}^{(\gamma)}$-Brownian motions $\{\widetilde{W}_1(t)\}_{t\geq 0}$, $\{\widetilde{W}^{(\rho,\gamma)}(t)\}_{t\geq 0}$ satisfy

$$d\widetilde{W}_1(t)d\widetilde{W}^{(\rho,\gamma)}(t) = \rho dt, \quad \text{for } \rho \in [-1, 1]. \tag{6.63}$$

It follows immediately that the discounted price $\{e^{-rt}S(t)\}_{t\geq 0}$ is a $\widetilde{\mathbb{P}}^{(\gamma)}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. Hence *all* probability measures $\widetilde{\mathbb{P}}^{(\gamma)}$ are equivalent risk-neutral probability measures.

Remark 6.7 (Incomplete markets)**.** As the risk-neutral probability measure is not uniquely defined, the market under discussion is said to be **incomplete**. Within incomplete markets there is no unique value for the price of derivatives (it depends on which specific risk-neutral probability measure is used to price the derivative). The stochastic process $\{\psi(t)\}_{t\geq 0}$ is called the **market price of volatility risk** and reduces to (6.3) for $\gamma \equiv 0$ (or $\rho = 1$).

Consider now the standard European derivative with pay-off $Y = g(S(T))$ at time of maturity T. For stochastic volatility models it is reasonable to assume that the risk-neutral price $\Pi_Y(t)$ of the derivative is a local function of the stock price *and* of the instantaneous variance, i.e., we make the following *ansatz* which generalizes (6.50):

$$\Pi_Y(t) = e^{-r(T-t)}\widetilde{\mathbb{E}}[g(S(T))|\mathcal{F}_W(t)] = v_g(t, S(t), v(t)) \tag{6.64}$$

for all $t \in [0, T]$, for all $T > 0$ and for some measurable pricing v_g. Of course, as in the case of local volatility models, (6.64) is motivated by the Markov property of solutions to systems of SDE's. In fact, it is useful to consider a more general European derivative with pay-off Y given by

$$Y = h(S(T), v(T)),$$

for some function $h : [0, \infty)^2 \to \mathbb{R}$, i.e., the pay-off of the derivative depends on the stock value *and* on the instantaneous variance of the stock at the time of maturity. We have the following analogue of Theorem 6.13.

Theorem 6.14. *Assume that the functions $\eta(t,x,y)$, $\beta(t,x,y)$, $\psi(t,x,y)$ in* (6.61) *are such that the PDE*

$$\partial_t u + rx\partial_x u + A\partial_y u + \frac{1}{2}yx^2\partial_x^2 u + \frac{1}{2}\beta^2 y\partial_y^2 u + \rho\beta xy\partial_{xy}^2 u = 0, \tag{6.65a}$$

$$A = \eta - \sqrt{y}\beta\psi, \quad (t,x,y) \in (0,T)\times(0,\infty)^2 \tag{6.65b}$$

admits a unique strong solution u satisfying $u(T,x,y) = h(x,y)$. Then the risk-neutral price of the derivative with pay-off $Y = h(S(T), v(T))$ and maturity T is given by

$$\Pi_Y(t) = f_h(t, S(t), v(t))$$

where the pricing function f_h is given by $f_h(t,x,y) = e^{-r\tau}u(t,x,y)$, $\tau = T - t$.

Exercise 6.28. *Prove the theorem. Hint: use Itô's formula in two dimensions, see Theorem 4.8, and the argument in the proof of Theorem 6.13.*

As for the local volatility models, a closed formula solution of (6.65) is rarely available and the use numerical methods to price the derivative becomes essential.

Heston model

A popular stochastic volatility model is the **Heston** model, which is obtained by the following substitutions in (6.59)-(6.60):

$$\mu(t, S(t), v(t)) = \mu_0, \quad \beta(t,x,y) = c, \quad \eta(t,x,y) = a(b-y),$$

where μ_0, a, b, c are constant. Hence the stock price and the volatility dynamics in the Heston model are given by the following stochastic differential equations:

$$dS(t) = \mu_0 S(t)\,dt + \sqrt{v(t)}S(t)dW_1(t), \tag{6.66a}$$

$$dv(t) = a(b - v(t))dt + c\sqrt{v(t)}dW^{(\rho)}(t). \tag{6.66b}$$

Note in particular that the variance in the Heston model is a CIR process in the physical probability $\mathbb{P}$, see (5.26). The condition $2ab \geq c^2$ ensures that $v(t)$ is strictly positive almost surely. To pass to the risk neutral world we need to fix a risk-neutral probability measure, that is, we need to fix the market price of volatility risk function ψ in (6.62). In the Heston model it is assumed that

$$\psi(t,x,y) = \lambda\sqrt{y}, \tag{6.67}$$

for some constant $\lambda \in \mathbb{R}$, which leads to the following form of the pricing PDE (6.65):

$$\partial_t u + rx\partial_x u + (k - my)\partial_y u + \frac{1}{2}yx^2\partial_x^2 u + \frac{c^2}{2}y\partial_y^2 u + \rho cxy\partial_{xy}^2 u = 0, \tag{6.68}$$

where the constant k, m are given by $k = ab$, $m = (a + c\lambda)$. Note that the choice (6.67) implies that the variance of the stock remains a CIR process in the risk-neutral probability measure.

The general solution of (6.68) with terminal datum $u(T, x, y) = h(x, y)$ is not known. However in the case of a call option (i.e., $h(x, y) = g(x) = (x - K)_+$) an explicit formula for the Fourier transform of the solution is available, see [14]. With this formula at hand one can compute the price of call options by very efficient numerical methods, which is one of the main reasons for the popularity of the Heston model.

Variance swaps

Variance swaps are financial derivatives[4] on the realized annual variance of an asset (or index). We first describe how the realized annual variance is computed from the historical data of the asset price. Let $T > 0$ be measured in days and consider the partition

$$0 = t_0 < t_1 < \cdots < t_n = T, \quad t_{j+1} - t_j = h > 0,$$

of the interval $[0, T]$. Assume for instance that the asset is a stock and let $S(t_j) = S_j$ be the stock price at time t_j. Here $S_1, \dots S_n$ are historical data for the stock price and *not* random variables (i.e., the interval $[0, T]$ lies in the past of the present time). The **realized annual variance** of the stock in the interval $[0, T]$ along this partition is defined as

$$\sigma^2_{\text{1year}}(n, T) = \frac{\kappa}{T} \sum_{j=0}^{n-1} \left(\log \frac{S_{j+1}}{S_j} - \frac{1}{n} \log \frac{S(T)}{S(0)} \right)^2, \tag{6.69}$$

where κ is the number of trading days in one year (typically, $\kappa = 252$). Using $T = nh$ we see that, up to a normalization factor, (6.69) coincides with the sample variance of the log-returns of the stock in the intervals $[t_j, t_{j+1}]$, $j = 0, \dots, n-1$. A **variance swap** stipulated at time $t = 0$, with maturity T and strike variance K is a contract between two parties which, at the expiration date, entails the exchange of cash given by $N(\sigma^2_{\text{1year}} - K)$, where N (called **variance notional**) is a conversion factor from units of variance to units of currency. In particular, the holder of the long position on the swap is the party who receives the cash in the case that the realized annual variance at the expiration date is larger than the strike variance. Variance swaps are traded over the counter and they are used by investors to protect their exposure to the volatility of the asset. For instance, suppose that an investor has a position on an asset which is profitable if the volatility of the stock price increases (e.g., the investor owns call options on the stock). Then it is clearly important for the investor to secure such position against a possible decrease of the volatility. To this purpose the investor opens a short position on a variance swap with another investor who is exposed to the opposite risk.

[4] More precisely, forward contracts, see Section 6.8.

Let us now discuss variance swaps from a mathematical modeling point of view. We assume that the stock price follows the generalized geometric Brownian motion

$$S(t) = S(0) \exp\left(\int_0^t \alpha(s)ds + \int_0^t \sigma(s)dW(s) \right).$$

Next we show that, as $n \to \infty$, the realized annual variance in the future time interval $[0, T]$ converges in L^2 to the random variable

$$Q_T = \frac{\kappa}{T}[\log S, \log S](T) = \frac{\kappa}{T} \int_0^T \sigma^2(t)\, dt.$$

To see this we first rewrite the definition of realized annual variance as

$$\sigma^2_{\text{1year}}(n, T) = \frac{\kappa}{T} \sum_{j=0}^{n-1} \left(\log \frac{S_{j+1}}{S_j} \right)^2 - \frac{\kappa}{nT} \left(\log \frac{S(T)}{S(0)} \right)^2. \tag{6.70}$$

Hence

$$\lim_{n\to\infty} \sigma^2_{\text{1year}}(n, T) = \lim_{n\to\infty} \frac{\kappa}{T} \sum_{j=0}^{n-1} \left(\log \frac{S_{j+1}}{S_j} \right)^2 \quad \text{in } L^2.$$

Moreover, by the definition of quadratic variation, it follows that

$$\mathbb{E}\left[\left(\frac{\kappa}{T} \sum_{j=0}^{n-1} \left(\log \frac{S_{j+1}}{S_j} \right)^2 - Q_T \right)^2 \right] \to 0, \quad \text{as } n \to \infty.$$

A variance swap can thus be defined as the (non-standard) European derivative with pay-off $Y = Q_T - K$. Assuming that the interest rate of the money market is the constant $r \in \mathbb{R}$, the risk-neutral value of a variance swap is given by

$$\Pi_Y(t) = e^{-r\tau} \widetilde{\mathbb{E}}[Q_T - K | \mathcal{F}_W(t)]. \tag{6.71}$$

In particular, at time $t = 0$, i.e., when the contract is stipulated, we have

$$\Pi_Y(0) = e^{-rT} \widetilde{\mathbb{E}}[Q_T - K], \tag{6.72}$$

where we used that $\mathcal{F}_W(0)$ is a trivial σ-algebra, and therefore the conditional expectation with respect to $\mathcal{F}_W(0)$ is a pure expectation. As none of the two parties in a variance swap has a privileged position on the contract, there is no premium associated to variance swaps, that is to say, the fair value of a variance swap is zero[5]. The value K_* of the variance strike which makes the risk-neutral price of a variance swap equal to zero at time $t = 0$, i.e., $\Pi_Y(0) = 0$, is called the **fair variance strike**. By (6.72) we find

$$K_* = \frac{\kappa}{T} \int_0^T \widetilde{\mathbb{E}}[\sigma^2(t)]\, dt. \tag{6.73}$$

[5] This is a general property of forward contracts, see Section 6.8.

To compute K_* explicitly, we need to fix a stochastic model for the variance process $\{\sigma^2(t)\}_{t\geq 0}$. Let us consider the Heston model

$$d\sigma^2(t) = a(b - \sigma^2(t))dt + c\sigma(t)d\widetilde{W}(t), \tag{6.74}$$

where a, b, c are positive constants satisfying $2ab \geq c^2$ and where $\{\widetilde{W}(t)\}_{t\geq 0}$ is a Brownian motion in the risk-neutral probability measure. To compute the fair variance strike of the swap using the Heston model we use that

$$\widetilde{\mathbb{E}}[\sigma^2(t)] = abt - a\int_0^t \widetilde{\mathbb{E}}[\sigma^2(s)]\,ds,$$

which implies $\frac{d}{dt}\widetilde{\mathbb{E}}[\sigma^2(t)] = ab - a\widetilde{\mathbb{E}}[\sigma^2(t)]$ and so

$$\widetilde{\mathbb{E}}[\sigma^2(t)] = b + (\sigma_0^2 - b)e^{-at}, \quad \sigma_0^2 = \widetilde{\mathbb{E}}[\sigma^2(0)] = \sigma^2(0). \tag{6.75}$$

Replacing into (6.73) we obtain

$$K_* = \kappa\left[b + \frac{\sigma_0^2 - b}{aT}(1 - e^{-aT})\right].$$

Exercise 6.29 (Sol. 41)**.** *Given $\sigma_0 > 0$, let $\sigma(t) = \sigma_0\sqrt{S(t)}$, which is an example of CEV model. Compute the fair strike of the variance swap.*

Exercise 6.30 (Sol. 42)**.** *Assume that the price $S(t)$ of a stock follows a generalized geometric Brownian motion with instantaneous volatility $\{\sigma(t)\}_{t\geq 0}$ given by the Heston model $d\sigma^2(t) = a(b - \sigma^2(t))\,dt + c\sigma(t)\,d\widetilde{W}(t)$, where $\{\widetilde{W}(t)\}_{t\geq 0}$ is a Brownian motion in the risk-neutral probability measure and a, b, c are constants such that $2ab \geq c^2 > 0$. The volatility call option with strike K and maturity T is the financial derivative with pay-off*

$$Y = N\left(\sqrt{\frac{\kappa}{T}\int_0^T \sigma^2(t)\,dt} - K\right)_+,$$

where κ is the number of trading days in one year and N is a dimensional constant that converts units of volatility into units of currency. Assuming that the interest rate of the money market is constant, find the partial differential equation and the terminal value satisfied by the pricing function of the volatility option.

6.7 Interest rate contracts

Zero-coupon bonds

A **zero-coupon bond** (ZCB) with **face** (or **nominal**) value K and maturity $T > 0$ is a contract that promises to pay to its owner the amount K at time T in the future. Zero-coupon bonds, and related contracts described below, are issued by national governments and private companies as a way to borrow money and fund their activities. In the following we assume that all ZCB's are issued by one given institution, so that all bonds differ merely by their face values and maturities. Moreover without loss of generality we assume from now on that $K = 1$, as owning a ZCB with face value K is clearly equivalent to own K shares of a ZCB with face value 1.

Once a debt is issued in the so-called **primary market**, it becomes a tradable asset in the **secondary** bond market. It is therefore natural to model the value at time t of the ZCB maturing at time $T > t$ (and face value 1) as a random variable, which we denote by $B(t,T)$. We assume throughout the discussion that the institution issuing the bond bears no risk of default, i.e., $B(t,T) > 0$, for all $t \in [0,T]$. Clearly $B(T,T) = 1$ and, under normal market conditions, $B(t,T) < 1$, for $t < T$, although exceptions are not rare[6]. A **zero-coupon bond market** (ZCB market) is a market in which the objects of trading are ZCB's with different maturities. Our main goal is to introduce models for the prices of ZCB's observed in the market. For modeling purposes we assume that zero-coupon bonds are available with a continuum range of maturities $T \in [0,S]$, where $S > 0$ is sufficiently large so that all ZCB's in the market mature before the time S (e.g., $S \approx 50$ years). Mathematically this means that we model the prices of ZCB's in the market as a stochastic process depending on 2 parameters, namely

$$\{B(t,T), t \in [0,T], T \in [0,S]\}.$$

All processes $\{X(t,T), t \in [0,T], T \in [0,S]\}$ introduced in this section are assumed to have a.s. continuous paths in both variables t, T and to be adapted to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$ generated by the given Brownian motion $\{W(t)\}_{t\geq 0}$. This means that for all *given* $T \in [0,S]$, the stochastic process $\{X(t,T)\}_{t\in[0,T]}$ is adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. By abuse of notation we continue to denote by $\mathcal{C}^0[\mathcal{F}_W(t)]$ this class of processes.

Remark 6.8. The term "bond" is more specifically used for long term loans with maturity $T > 1$ year. Short term loans have different names (e.g., bills, repo, etc.) and they constitute the component of the loan market called **money market**. A bond which has less than one year left to maturity is also considered a money market asset.

Forward and spot rate

The difference in value of zero-coupon bonds with different maturities is expressed through the implied forward rate of the bond. To define this concept, suppose first that at the present time t we open a portfolio that consists of -1 share of a zero-coupon bond with maturity $T > t$ and $B(t,T)/B(t,T+\delta)$ shares of a zero-coupon bond expiring at time $T+\delta$. Note that the value of this portfolio is $V(t) = 0$. This investment entails that we pay 1 at time T and receive $B(t,T)/B(t,T+\delta)$ at time $T+\delta$. Hence our investment at time t is equivalent

[6]For instance, zero-coupon (and other) national bonds in Sweden with maturity shorter than 5 years yield currently (2017) a negative return.

to an investment in the future time interval $[T, T+\delta]$ with (annualized) return given by

$$F_\delta(t,T) = \frac{1}{\delta}\left(B(t,T)/B(t,T+\delta) - 1\right) = \frac{B(t,T) - B(t,T+\delta)}{\delta B(t,T+\delta)}. \tag{6.76}$$

The quantity $F_\delta(t,T)$ is also called the **simply compounded forward rate** in the interval $[T, T+\delta]$ **locked at time** t (or forward LIBOR, as it is commonly applied to LIBOR interest rate contracts). The name is intended to emphasize that the return in the future interval $[T, T+\delta]$ is locked at the time $t \leq T$, that is to say, we know today which interest rate has to be charged to borrow in the future time interval $[T, T+\delta]$ (if a different rate were locked today, then an arbitrage opportunity would arise). In the limit $\delta \to 0^+$ we obtain the so called **continuously compounded T-forward rate** of the bond locked at time t:

$$F(t,T) = \lim_{\delta\to 0^+} F_\delta(t,T) = -\frac{1}{B(t,T)}\partial_T B(t,T) = -\partial_T \log B(t,T), \tag{6.77}$$

where $0 \leq t \leq T$ and $0 \leq T \leq S$. Inverting (6.77) we obtain

$$B(t,T) = \exp\left(-\int_t^T F(t,v)\,dv\right), \quad 0 \leq t \leq T \leq S. \tag{6.78}$$

By (6.78), a model for the price $B(t,T)$ of the ZCB's in the market can be obtained by a model on the **forward rate curve** $T \to F(t,T)$. This approach to the problem of ZCB pricing is known as HJM approach, from Heath, Jarrow, Morton, who introduced this method in the late 1980s.

Letting $T \to t^+$ in (6.76) we obtain the **simply compounded spot rate**,

$$R_\delta(t) = \lim_{T\to t^+} F_\delta(t,T), \tag{6.79}$$

that is to say, the interest rate locked "on the spot", i.e., at the present time t, to borrow in the interval $[t, t+\delta]$. Letting $\delta \to 0^+$ we obtain the **instantaneous** (or **continuously compounded**) **spot rate** $\{r(t)\}_{t\in[0,S]}$ of the ZCB market:

$$r(t) = \lim_{\delta\to 0^+} R_\delta(t) = \lim_{T\to t^+} F(t,T), \quad t \in [0,S]. \tag{6.80}$$

Note that $r(t)$ is the interest rate locked at time t to borrow in the "infinitesimal interval" of time $[t, t+dt]$. Hence $r(t)$ coincides with the risk-free rate of the money market used in the previous sections. For the options pricing problem studied in Sections 6.3–6.6 we assumed that $r(t)$ was equal to a constant r, which is reasonable for short maturity contracts ($T \lesssim 1$ year). However when large maturity assets such as ZCB's are considered, we have to relax this assumption and promote $\{r(t)\}_{t\in[0,S]}$ to a stochastic process. In the so-called **classical approach** to the problem of ZCB's pricing, the fair value of $B(t,T)$ is derived from a model on the spot rate process.

Yield to maturity of ZCB's

If a ZCB is bought at time t and kept until its maturity $T > t$, the annualized log-return of the investment is

$$Y(t,T) = -\frac{1}{T-t}\log B(t,T) = \frac{1}{T-t}\int_t^T F(t,v)\,dv, \tag{6.81}$$

which is called the **(continuously compounded) yield to maturity** of the zero-coupon bond, while $T \to Y(t,T)$ is called the **yield curve**; see Figure 6.4 for an example of yield curve (on Swedish bonds).

Inverting (6.81) we find

$$B(t,T) = e^{-Y(t,T)(T-t)} = \frac{D_Y(T)}{D_Y(t)}, \quad \text{where } D_Y(s) = e^{-Y(t,T)s}. \tag{6.82}$$

Hence we may interpret the yield also as the *constant interest rate which entails that the value of the ZCB at time t equals the discounted value of the future payment 1 at time T.*

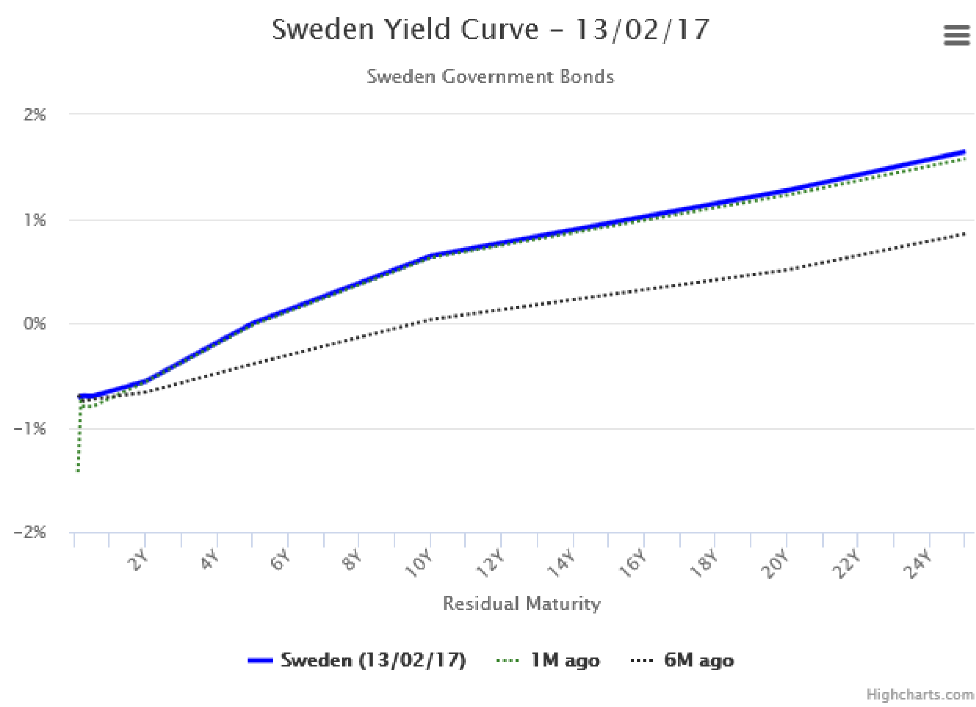


Figure 6.4: Yield curve for Swedish bonds. Note that the yield is negative for maturities shorter than 5 years.

Exercise 6.31. *Yield curves observed in the market are classified based on their shape (e.g., steep, flat, inverted, etc.). Find out on the Internet what the different shapes mean from an economical point of view.*

Coupon bonds

Let $0 < t_1 < t_2 < \cdots < t_M = T$ be a partition of the interval $[0,T]$. A **coupon bond** with maturity T, face value 1 and coupons $c_1, c_2, \dots, c_M \in [0,1)$ is a contract that promises to pay the amount c_k at time t_k and the amount $1+c_M$ at maturity $T = t_M$. Note that some c_k may be zero, which means that no coupon is actually paid at that time. We set $c = (c_1, \dots, c_M)$ and denote by $B_c(t,T)$ the value at time t of the bond paying the coupons $c_1, \dots, c_M$ and maturing at time T. Now, let $t \in [0,T]$ and $k(t) \in \{1, \dots, M\}$ be the smallest index such that $t_{k(t)} > t$, that is to say, $t_{k(t)}$ is the first time after t at which a coupon is paid. Holding the coupon bond at time t is clearly equivalent to holding a portfolio containing $c_{k(t)}$ shares of the ZCB expiring at time $t_{k(t)}$, $c_{k(t)+1}$ shares of the ZCB expiring at time $t_{k(t)+1}$, and so on, hence

$$B_c(t,T) = \sum_{j=k(t)}^{M-1} c_j B(t,t_j) + (1+c_M)B(t,T), \tag{6.83}$$

the sum being zero when $k(t) = M$.

Remark 6.9. The value of the coupon bond at time t does not depend on the coupons paid at or before t. This of course makes sense, because purchasing a bond at time t does not give the buyer any right concerning previous coupon payments.

The yield to maturity of a coupon bond is the quantity $Y_c(t,T)$ defined implicitly by the equation

$$B_c(t,T) = \sum_{j=k(t)}^{M-1} c_j e^{-Y_c(t,T)(t_j-t)} + (1+c_M)e^{-Y_c(t,T)(T-t)}. \tag{6.84}$$

It follows that *the yield of the coupon bond is the constant interest rate used to discount the total future payments of the coupon bond.*

Example. Consider a 3 year maturity coupon bond with face value 1 which pays 2% coupon semiannually. Suppose that the bond is listed with an yield of 1%. What is the value of the bond at time zero? The coupon dates are

$$(t_1, t_2, t_3, t_4, t_5, t_6) = (1/2, 1, 3/2, 2, 5/2, 3),$$

and $c_1 = c_2 = \cdots = c_6 = 0.02$. Hence

$$\begin{aligned} B_c(0,T) &= 0.02e^{-0.01*\frac{1}{2}} + 0.02e^{-0.01*1} + 0.02e^{-0.01*\frac{3}{2}} + 0.02e^{-0.01*2} + 0.02e^{-0.01*\frac{5}{2}} \\ &\quad + (1+0.02)e^{-0.01*3} = 1.08837. \end{aligned}$$

Remark 6.10. As in the previous example, the coupons of a coupon bond are typically all equal, i.e., $c_1 = c_2 = \cdots = c_M = c \in (0,1)$.

In the example above, the yield was given and $B_c(0,T)$ was computed. However one is most commonly faced with the opposite problem, i.e., computing the yield of the coupon bond with given initial value $B_c(0,T)$. We can easily solve this problem numerically inverting (6.84). For instance, assume that the bond is issued at time $t=0$ with maturity $T=M$ years (M integer) and that the coupons are paid annually, that is $t_1=1$, $t_2=2,\dots,t_M=M$. Then $x=\exp(-Y_c(0,T))$ solves $p(x)=0$, where p is the M-order polynomial given by

$$p(x)=c_1x+c_2x^2+\cdots+(1+c_M)x^M-B_c(0,T). \tag{6.85}$$

The roots of this polynomial can easily be computed numerically, e.g., with the command `roots[p]` in matlab, see Exercise 6.32 below.

Exercise 6.32 (Matlab)**.** *Write a matlab function*

```
yield(B, Coupon, FirstCoupDate, CoupFreq, T)
```

that computes the yield of a coupon bond. Here, `B` *is the current (i.e., at time $t=0$) price of the coupon,* `Coupon` $\in[0,1)$ *is the (constant) coupon,* `FirstCoupDate` *is the first future date at which the coupon is paid,* `CoupFreq` *is the frequency of coupon payments and* `T` *is the maturity of the coupon bond. Express all time in fraction of years, using 1 day = 1/252 years. For example*

`yield(1.01, 0.02, 46/252, 1, 4)`

computes the yield of a 2% coupon bond which today is valued 1.01, pays the first coupon in 46 days and expires in 4 years.

Classical approach to ZCB pricing

In the so-called classical approach to the problem of pricing ZCB's we interpret the ZCB as a derivative on the spot rate. Assume that a model for $\{r(t)\}_{t\in[0,S]}$ is given as a stochastic process in the space $\mathcal{C}^0[\mathcal{F}_W(t)]$. As the pay-off of the ZCB equals one, the risk-neutral price of the ZCB is given by

$$B(t,T)=\widetilde{\mathbb{E}}[D(t)^{-1}D(T)|\mathcal{F}_W(t)]=\widetilde{\mathbb{E}}\Big[\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)\Big] \tag{6.86}$$

and by Theorem 6.2(i) (with $Y=1$), the discounted price of the ZCB, $B^*(t,T)=D(t)B(t,T)$, is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$; in particular, self-financing portfolios invested in the ZCB market are not arbitrage portfolios. Note however that *the risk-neutral probability measure in* (6.86) *cannot be determined solely by the spot rate, and therefore models for the process $\{r(t)\}_{t\in[0,S]}$ must be given a priori in terms of a risk-neutral probability measure.* As the real world is not risk-neutral, the foundation of the classical approach is questionable. There are two ways to get around this problem. One is the HJM approach described below; the other is by adding a risky asset (e.g. a stock) to the ZCB market, which would then be used to determine the risk-neutral probability measure. The last procedure is referred to as "completing the ZCB market" and will be discussed in Section 6.8.

As an application of the classical approach, assume that the spot rate is given by the Cox-Ingersoll-Ross (CIR) model,

$$dr(t) = a(b - r(t))dt + c\sqrt{r(t)}d\widetilde{W}(t), \quad r(0) = r_0 > 0, \tag{6.87}$$

where $\{\widetilde{W}(t)\}_{t\in[0,T]}$ is a Brownian motion *in the risk-neutral probability measure* and R_0, a, b, c are positive constants. To compute $B(t,T)$ under a CIR interest rate model, we make the *ansatz*

$$B(t,T) = v(t, r(t)), \tag{6.88}$$

for some smooth function $v : [0,T] \times \mathbb{R} \to \mathbb{R}$, which we want to find. Note that we do not require the Feller condition $ab \geq c^2/2$, hence we allow the spot rate to become zero with positive probability, although negative values are excluded in the CIR model.

Theorem 6.15. *When the interest rate* $\{r(t)\}_{t>0}$ *follows the CIR model* (6.87)*, the value* $B(t,T)$ *of the zero-coupon bond is given by*

$$v(t,x) = e^{-xC(\tau)-A(\tau)}, \quad \tau = T - t, \tag{6.89}$$

where $C(\tau)$*,* $A(\tau)$ *satisfy the Cauchy problem*

$$C'(\tau) = 1 - aC(\tau) - \frac{c^2}{2}C(\tau)^2, \quad A'(\tau) = abC(\tau) \tag{6.90a}$$

$$C(0) = 0, \quad A(0) = 0. \tag{6.90b}$$

Moreover the solution of the Cauchy problem (6.90) *is given by*

$$C(\tau) = \frac{\sinh(\gamma\tau)}{\gamma\cosh(\gamma\tau) + \frac{1}{2}a\sinh(\gamma\tau)} \tag{6.91a}$$

$$A(\tau) = -\frac{2ab}{c^2}\log\left[\frac{\gamma e^{\frac{1}{2}a\tau}}{\gamma\cosh(\gamma\tau) + \frac{1}{2}a\sinh(\gamma\tau)}\right] \tag{6.91b}$$

and

$$\gamma = \frac{1}{2}\sqrt{a^2 + 2c^2}. \tag{6.91c}$$

Proof. Using Itô's formula and the product rule, together with (6.87), we obtain

$$\begin{aligned} d(D(t)v(t,r(t))) = D(t)[&\partial_t v(t,r(t)) + a(b - r(t))\partial_x v(t,r(t)) \\ &+ \frac{c^2}{2}r(t)\partial_x^2 v(t,r(t)) - r(t)v(t,r(t))]dt \\ &+ D(t)\partial_x v(t,r(t))c\sqrt{r(t)}d\widetilde{W}(t). \end{aligned}$$

Hence, imposing that v be a solution of the PDE

$$\partial_t v + a(b - x)\partial_x v + \frac{c^2}{2}x\partial_x^2 v = xv, \quad (t,x) \in \mathcal{D}_T^+, \tag{6.92a}$$

we obtain that the stochastic process $\{D(t)v(t,r(t))\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\in[0,T]}$. Imposing additionally the terminal condition

$$v(T,x)=1,\quad \text{for all } x>0, \tag{6.92b}$$

we obtain

$$D(t)v(t,r(t))=\widetilde{\mathbb{E}}[D(T)v(T,R(T))|\mathcal{F}_W(t)]=\widetilde{\mathbb{E}}[D(T)|\mathcal{F}_W(t)],$$

hence

$$v(t,r(t))=\widetilde{\mathbb{E}}[D(T)/D(t)|\mathcal{F}_W(t)],$$

and thus (6.88) is verified. Replacing the *ansatz* (6.89) into (6.92) we find the following first order polynomial equation

$$x(C'(\tau)+aC(\tau)+\frac{c^2}{2}C(\tau)^2-1)+A'(\tau)-abC(\tau)=0.$$

The previous equation holds for all x if and only if (6.90a) hold, while the initial conditions (6.90b) are equivalent to the terminal condition $v(T,x)=1$. The proof of the claim that (6.91) is the solution of the Cauchy problem (6.90) is left as an exercise. □

Exercise 6.33. *Show that* (6.91) *is the solution of the Cauchy problem* (6.90).

Exercise 6.34. *Use the matlab code in Exercise 6.32 to compute the yield $Y_c(0,T)$ of coupon bonds in the CIR model for a given value of r_0 and different values of a,b,c. Assume that the coupons are paid at times $t=1,2,\dots,T$. Plot the yield curve and try to reproduce all the typical profiles found in Exercise 6.31. Remark: each bond in the yield curve might pay a different coupon, but the constants r_0,a,b,c in the CIR model must be chosen to be the same for all bonds in the yield curve.*

The CIR model is an example of **affine model**, i.e., a model for the interest rate which entails a price function for the ZCB of the form $B(t,T)=\exp(-r(t)C(t)-A(t))$ (or equivalently, an yield which is a linear function of the spot rate). The most general affine model has the form

$$dr(t)=a(t)(b(t)-r(t))\,dt+c(t)\sqrt{r(t)+\delta(t)}\,d\widetilde{W}(t), \tag{6.93}$$

where a,b,c,δ are deterministic functions of time.

Exercise 6.35. *Let $B(t,T)=v(t,r(t))$ the price of the ZCB with face value 1 entailed by the general affine model* (6.93). *Set $v(t,x)=\exp(-xC(T-t)-A(T-t))$ and derive the ODE's verified by the functions A,C.*

An example of non-affine model is the CKLS model (Chan, Karolyi, Longstaff, Sanders [6]), which is given by the SDE

$$dr(t)=a(b-r(t))\,dt+cr(t)^{\gamma}\,d\widetilde{W}, \tag{6.94}$$

where $a, b, c, \gamma > 0$ and $\gamma \neq 1/2$ (for $\gamma = 1/2$ we recover the CIR model, while for $\gamma = 0$ the CKLS model reduces to the Vasicek model, see Exercise 6.36). Repeating the argument in the proof of Theorem 6.15 we find that the pricing function of ZCB's in the CKLS model satisfies the PDE

$$\partial_t v + a(b-x)\partial_x v + \frac{c^2}{2}x^{2\gamma}\partial_x^2 v = xv, \quad (t,x) \in \mathcal{D}_T^+, \tag{6.95}$$

with the terminal condition $v(T,x) = 1$. The solution to this terminal value problem is not known, hence the use of numerical methods becomes essential to price ZCB's in the CKLS model.

Exercise 6.36 (Sol. 43)**.** *Assume that the interest rate of a zero-coupon bond is given by the Vasicek model*

$$dr(t) = a(b - r(t))dt + c\, d\widetilde{W}(t), \quad r(0) = r_0 \in \mathbb{R},$$

where a, b, c are positive constants and $\{\widetilde{W}(t)\}_{t\geq 0}$ is a Brownian motion in the risk-neutral probability measure $\widetilde{\mathbb{P}}$. Show that $r(t)$ is $\widetilde{\mathbb{P}}$-normally distributed and compute its expectation and variance in the risk-neutral probability measure. Derive the PDE for the pricing function v of the ZCB with face value 1 and maturity $T > 0$. Find v using the ansatz (6.89)*.*

Interest rate swap

An **interest rate swap** can be seen as a coupon bond with variable (random) coupons, which can be positive or negative. More precisely, consider a partition $0 = T_0 < T_1 < \cdots < T_n = T$ with $T_i - T_{i-1} = \delta$, for all $i = 1, \dots, n$. Let $R_\delta(T_i) = F_\delta(T_i, T_i)$ be the simply compounded spot rate in the interval $[T_i, T_{i+1}]$. Recall that this quantity is known at time T_i (but not at time $t = 0$). An interest rate swap is a contract between two parties which at each time T_{i+1}, $i = 1, \dots, n-1$, entails the exchange of cash $N(R_\delta(T_i)\delta - r\delta)$, where r is a fixed interest rate and $N > 0$ is the notional amount converting units of interest rates into units of currency. Without loss of generality, we assume $N = 1$ in the following. The party that receives this cash flow when it is positive is called the receiver, while the opposite party is called the payer. Hence the receiver has a long position on the spot rate, while the payer has a short position on the spot rate. The risk-neutral value at time $t = 0$ of the interest rate swap is the expectation, in the risk-neutral probability measure, of the discounted cash-flow entailed by the contract, that is

$$\Pi_{\mathrm{irs}}(0) = \delta \sum_{i=1}^{n-1} \widetilde{\mathbb{E}}[(R_\delta(T_i) - r)D(T_{i+1})], \tag{6.96}$$

where $D(t)$ is the discount process. Being a forward type contract (see Section 6.8), the fair value of the interest rate swap is zero: neither the receiver nor the payer has a privileged position in the contract and thus none of them needs to pay a premium. The value of r for which $\Pi_{\mathrm{irs}}(0) = 0$ is called the **(fair) swap rate** of the interest rate swap.

Theorem 6.16. *The swap rate of an interest rate swap stipulated at time $t = 0$ and with maturity T is given by*

$$r_{\text{swap}} = \frac{\sum_{i=1}^{n-1} B(0, T_{i+1}) F_\delta(0, T_i)}{\sum_{i=1}^{n-1} B(0, T_{i+1})}. \tag{6.97}$$

Proof. We show below that, for all $T > 0$ and $\delta > 0$, the following identity holds:

$$\widetilde{\mathbb{E}}[D(T+\delta)F_\delta(T,T)] = B(0, T+\delta)F_\delta(0,T). \tag{6.98}$$

Using (6.98) in (6.96) we obtain

$$\begin{aligned}
\Pi_{\text{irs}}(0) &= \delta \sum_{i=1}^{n-1} \widetilde{\mathbb{E}}[F_\delta(T_i, T_i)D(T_i+\delta)] - \delta r \sum_{i=1}^{n-1} \widetilde{\mathbb{E}}[D(T_{i+1})] \\
&= \delta \left(\sum_{i=1}^{n-1} B(0, T_{i+1}) F_\delta(0, T_i) - r \sum_{i=1}^{n-1} B(0, T_{i+1}) \right),
\end{aligned} \tag{6.99}$$

hence $\Pi_{\text{irs}}(0) = 0$ if and only if $r = r_{\text{swap}}$. It remains to prove (6.98). As $B(t,T) = \widetilde{\mathbb{E}}[D(T)/D(t)|\mathcal{F}_W(t)]$, we have

$$\begin{aligned}
\widetilde{\mathbb{E}}[D(T+\delta)F_\delta(T,T)] &= \widetilde{\mathbb{E}}[D(T+\delta)(\frac{1 - B(T,T+\delta)}{\delta B(T,T+\delta)})] \\
&= \frac{1}{\delta}\widetilde{\mathbb{E}}[D(T+\delta)B(T,T+\delta)^{-1}] - \frac{1}{\delta}\widetilde{\mathbb{E}}[D(T+\delta)] \\
&= \frac{1}{\delta}\widetilde{\mathbb{E}}[\widetilde{\mathbb{E}}[\frac{D(T+\delta)}{D(T)}\frac{D(T)}{B(T,T+\delta)}|\mathcal{F}_W(T)]] - \frac{1}{\delta}B(0,T+\delta) \\
&= \frac{1}{\delta}\widetilde{\mathbb{E}}[\frac{D(T)}{B(T,T+\delta)}B(T,T+\delta)] - \frac{1}{\delta}B(0,T+\delta) \\
&= \frac{1}{\delta}\frac{B(0,T) - B(0,T+\delta)}{B(0,T+\delta)}B(0,T+\delta) = B(0,T+\delta)F_\delta(0,T).
\end{aligned}$$

□

Remark 6.11. Note carefully that all quantities in the right hand side of (6.97) are known at time $t = 0$, hence the swap rate is fixed by information available at the time when the interest rate swap is stipulated.

Caps and Floors

An **interest rate cap** is a contract that caps (i.e., put a maximum limit on) the spot rate. More precisely, consider, as before, a uniform partition $0 = T_0 < T_1 < \cdots < T_n = T$ of the interval $[0, T]$ with $T_i - T_{i-1} = \delta$, for all $i = 1, \ldots, n$. Let $R_\delta(T_i) = F_\delta(T_i, T_i)$ be the simply compounded spot rate in the interval $[T_i, T_{i+1}]$. An interest rate cap with strike rate

r and notional amount $N = 1$ pays to its owner the amount $(R_\delta(T_i)\delta - r\delta)_+$ at time T_{i+1}, $i = 1, \dots, n-1$. Hence the spot rate for the owner of the interest rate cap is no higher than r: any excess to the strike rate is paid by the seller of the interest rate cap. Similarly, an **interest rate floor** put a minimum on the spot rate and pays to its owner the amount $(r\delta - R_\delta(T_i)\delta)_+$ at every time T_{i+1}, $i = 1, \dots, n-1$. The risk-neutral price of the interest rate cap/floor at time $t = 0$ is given by

$$\Pi_{\text{cap}}(0) = \delta \sum_{i=1}^{n-1} \widetilde{\mathbb{E}}[(R_\delta(T_i) - r)_+ D(T_{i+1})], \tag{6.100}$$

$$\Pi_{\text{floor}}(0) = \delta \sum_{i=1}^{n-1} \widetilde{\mathbb{E}}[(r - R_\delta(T_i))_+ D(T_{i+1})]. \tag{6.101}$$

As $(R_\delta(T_i) - r)_+ - (r - R_\delta(T_i))_+ = (R_\delta(T_i) - r)$, the **cap-floor parity** identity holds:

$$\Pi_{\text{cap}}(0) - \Pi_{\text{floor}}(0) = \Pi_{\text{irs}}(0).$$

In particular if the strike rate coincides with the swap rate then the cap and the floor have the same initial price. An interest rate cap (resp. floor) on one time period (i.e., $n = 1$) is called a **caplet** (resp. **floorlet**).

The HJM approach to ZCB pricing

Next we present a different approach for the evaluation of ZCB's due to Heath, Jarrow and Morton (HJM model). The cornerstone of this approach is to use the forward rate of the market, instead of the spot rate, as the fundamental parameter to express the price of the bond.

The starting point in the HJM approach is to assume that $\{F(t, T), t \in [0, T], T \in [0, S]\}$ is given by the diffusion process

$$dF(t, T) = \alpha(t, T)\, dt + \sigma(t, T) dW(t), \quad t \in [0, T],\ T \in [0, S]. \tag{6.102}$$

where $\{\alpha(t, T), t \in [0, T], T \in [0, S]\}$, $\{\sigma(t, T), t \in [0, T], T \in [0, S]\} \in \mathcal{C}^0[\mathcal{F}_W(t)]$; in the applications they are often assumed to be deterministic functions of (t, T). Note carefully that, as opposed to the spot rate dynamics in the classical approach, the forward rate dynamics in the HJM approach is given in terms of the *physical* (real-world) probability and not *a priori* in the risk-neutral probability. Specifying the dynamics (6.102) for the forward rate in the HJM approach corresponds to the assumption made in options pricing theory that the underlying stock follows a (generalized) geometric Brownian motion.

Theorem 6.17. *When the forward rate of the ZCB market is given by* (6.102)*, the value* (6.78) *of the zero-coupon bond is the diffusion process* $\{B(t, T), t \in [0, T], T \in [0, S]\}$ *given by*

$$dB(t, T) = B(t, T)[r(t) - \overline{\alpha}(t, T) + \frac{1}{2}\overline{\sigma}(t, T)^2]dt - \overline{\sigma}(t, T)B(t, T)\, dW(t), \tag{6.103}$$

where $r(t)$ *is the instantaneous spot rate* (6.80) *and*

$$\overline{\alpha}(t,T) = \int_t^T \alpha(t,v)\,dv, \quad \overline{\sigma}(t,T) = \int_t^T \sigma(t,v)\,dv. \tag{6.104}$$

Proof. Let $X(t) = -\int_t^T F(t,v)\,dv$. By Itô's formula,

$$dB(t,T) = B(t,T)(dX(t) + \frac{1}{2}dX(t)dX(t)). \tag{6.105}$$

Moreover

$$\begin{aligned} dX(t) &= F(t,t)\,dt - \int_t^T dF(t,v)\,dv \\ &= r(t)\,dt - \int_t^T (\alpha(t,v)\,dt + \sigma(t,v)\,dW(t))\,dv \\ &= r(t)\,dt - \overline{\alpha}(t,T)\,dt - \overline{\sigma}(t,T)\,dW(t). \end{aligned}$$

Replacing in (6.105) the claim follows. □

We now establish a condition which ensures that investing on ZCB's entails no arbitrage. This can be achieved by the same argument used in Section 6 for 1+1 dimensional stock markets, namely by showing that there exists a probability measure $\widetilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ such that the discounted value of the ZCB is a martingale for *all* maturities $T \in [0,S]$. Recall that $\{\overline{\alpha}(t,T), t \in [0,T], T \in [0,S]\}$ and $\overline{\sigma}(t,T)t \in [0,T], T \in [0,S]\}$ are given by (6.104).

Theorem 6.18. *Let the forward rate process* $\{F(t,T), t \in [0,T], T \in [0,S]\}$ *be given by* (6.102) *such that* $\sigma(t,T) > 0$ *a.s. for all* $0 \leq t \leq T \leq S$. *Assume that there exists a stochastic process* $\{\theta(t)\}_{t\in[0,S]} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ *(the* **market price of risk***), independent of* $T \in [0,S]$*, such that*

$$\alpha(t,T) = \theta(t)\sigma(t,T) + \sigma(t,T)\overline{\sigma}(t,T), \text{ for all } 0 \leq t \leq T \leq S. \tag{6.106}$$

and such that the stochastic process $\{Z(t)\}_{t\in[0,S]}$ *given by*

$$Z(t) = \exp\left(-\int_0^t \theta(s)\,dW(s) - \frac{1}{2}\int_0^t \theta(s)^2\,ds\right)$$

is a $\mathbb{P}$*-martingale (e.g.,* $\{\theta(t)\}_{t\in[0,S]}$ *satisfies the Novikov condition). Let* $\widetilde{\mathbb{P}}$ *be the probability measure equivalent to* $\mathbb{P}$ *given by* $\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z(S)\mathbb{I}_A]$*, for all* $A \in \mathcal{F}$*, and denote by* $\{\widetilde{W}(t)\}_{t\in[0,S]}$ *the* $\widetilde{\mathbb{P}}$*-Brownian motion given by* $d\widetilde{W}(t) = dW(t) + \theta(t)\,dt$*. Then the following holds:*

(i) The forward rate satisfies

$$dF(t,T) = \sigma(t,T)\overline{\sigma}(t,T)\,dt + \sigma(t,T)d\widetilde{W}(t). \tag{6.107}$$

(ii) The discounted price (6.78) *of the ZCB with maturity* $T \in [0, S]$ *satisfies*

$$dB^*(t,T) = -\overline{\sigma}(t,T)B^*(t,T)d\widetilde{W}(t). \tag{6.108}$$

In particular for any given $T \in [0, S]$*, the discounted value process* $\{B^*(t,T)\}_{t\in[0,T]}$ *is a* $\widetilde{\mathbb{P}}$*-martingale relative to* $\{\mathcal{F}_W(t)\}_{t\geq 0}$*, and so any self-financing portfolio that consists of ZCB's with different maturities is not an arbitrage.*

(iii) The value of the ZCB satisfies the risk-neutral pricing formula (6.86)*.*

The probability measure $\widetilde{\mathbb{P}}$ *is called the* **risk neutral probability measure** *of the ZCB market.*

Proof. (6.107) follows at once by replacing (6.106) into (6.102) and using $d\widetilde{W}(t) = dW(t) + \theta(t)\,dt$. In order to prove (6.108) we first show that $\theta(t)$ can be rewritten as

$$\theta(t) = \frac{\overline{\alpha}(t,T)}{\overline{\sigma}(t,T)} - \frac{1}{2}\overline{\sigma}(t,T). \tag{6.109}$$

Indeed, integrating (6.106) with respect to T and using $\sigma(t,T) = \partial_T\overline{\sigma}(t,T)$ we obtain

$$\begin{aligned}
\overline{\alpha}(t,T) &= \theta(t)\overline{\sigma}(t,T) + \int_t^T \sigma(t,v)\overline{\sigma}(t,v)\,dv \\
&= \theta(t)\overline{\sigma}(t,T) + \frac{1}{2}\int_t^T \partial_v(\overline{\sigma}(t,v)^2)\,dv \\
&= \theta(t)\overline{\sigma}(t,T) + \frac{1}{2}\overline{\sigma}(t,T)^2,
\end{aligned}$$

which gives (6.109). Next, by (6.103) and Itô's product rule,

$$\begin{aligned}
d(D(t)B(t,T)) &= D(t)B(t,T)(r(t) - \overline{\alpha}(t,T) + \frac{1}{2}\overline{\sigma}(t,T)^2)\,dt \\
&\quad - D(t)B(t,T)\overline{\sigma}(t,T)\,dW(t) - B(t,T)D(t)r(t)\,dt \\
&= B^*(t,T)[(\frac{1}{2}\overline{\sigma}(t,T)^2 - \overline{\alpha}(t,T))\,dt - \overline{\sigma}(t,T)\,dW(t)] \\
&= -B^*(t,T)\overline{\sigma}(t,T)\,d\widetilde{W}(t),
\end{aligned}$$

where in the last step we use (6.109). Thus $\{B^*(t,T)\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. In particular

$$\widetilde{\mathbb{E}}[D(T)B(T,T)|\mathcal{F}_W(t)] = D(t)B(t,T).$$

As $B(T,T) = 1$, the previous equation is equivalent to (6.86). $\square$

Example: The CIR model revisited

As a way of example we show how to re-formulate the CIR model in the HJM approach. Recall that in the CIR model the spot rate is given by

$$dr(t) = a(b - r(t))\,dt + c\sqrt{r(t)}\,d\widetilde{W}(t). \tag{6.110}$$

We have seen before that the price of the zero-coupon bond with face value 1 and maturity $T > 0$ in the CIR model is given by $B(t,T) = v(t, r(t))$, where $v(t,x) = \exp(-xC(T-t) - A(T-t))$ and where the deterministic functions $A(\tau), C(\tau)$ satisfy (6.90). The instantaneous forward rate satisfies

$$F(t,T) = -\partial_T \log B(t,T) = r(t)C'(T-t) + A'(T-t).$$

Hence, using (6.110),

$$\begin{aligned}
dF(t,T) &= C'(T-t)dr(t) - r(t)C''(T-t)dt - A''(T-t)dt \\
&= [a(b - r(t))\,dt + c\sqrt{r(t)}\,d\widetilde{W}(t)] - r(t)C''(T-t)\,dt - A''(T-t)\,dt \\
&= [a(b - r(t))C'(T-t) - r(t)C''(T-t) - A''(T-t)]\,dt + C'(T-t)c\sqrt{r(t)}\,d\widetilde{W}(t).
\end{aligned}$$

Comparing this result with (6.107) we are led to set

$$\sigma(t,T) = C'(T-t)c\sqrt{r(t)}, \tag{6.111}$$
$$\sigma(t,T)\overline{\sigma}(t,T) = [a(b - r(t))C'(T-t) - r(t)C''(T-t) - A''(T-t)]. \tag{6.112}$$

As $\overline{\sigma}(t,T) = \int_t^T \sigma(t,v)\,dv$, we obtain

$$\begin{aligned}
a(b - r(t))&C'(T-t) - r(t)C''(T-t) - A''(T-t) \\
&= C'(T-t)c\sqrt{r(t)} \int_t^T C'(v-t)c\sqrt{r(t)}\,dv \\
&= c^2 r(t)C'(T-t)C(T-t).
\end{aligned}$$

The latter identity holds for all possible value of $r(t)$ if and only if the following two equations are satisfied:

$$abC' - A'' = 0, \quad -aC' - C'' = -c^2 C'C. \tag{6.113}$$

Differentiating (6.90a) we find that (6.113) are indeed satisfied. Note that the HJM approach gives the same result (i.e., the same pricing function for ZCB's) as the classical approach with the CIR model if we assume that the forward rate is given by (6.102) with $\alpha(t,T)$ given by (6.106) (where $\theta(t)$ is arbitrary, typically chosen to be constant) and $\sigma(t,T)$ is given by (6.111). The advantage of the HJM approach is that the model for the forward rate is expressed in the physical probability (as opposed to (6.110)) and thus the parameters of this model can be calibrated using real market data.

Exercise 6.37 (Sol. 44)**.** *Use the HJM method to derive the dynamics of the instantaneous forward rate $F(t,T)$ for the Vacisek model (see Exercise 6.36).*

Exercise 6.38 (Sol. 45)**.** *Assume that the spot interest rate in the risk-neutral probability is given by the Ho-Lee model:*

$$dr(t) = a(t)\,dt + c d\widetilde{W}(t),$$

where $c > 0$ is a constant and $a(t)$ is a deterministic function of time. Derive the risk-neutral price $B(t,T)$ of the ZCB with face value 1 and maturity T. Use the HJM method to derive the dynamics of the instantaneous forward rate $F(t,T)$ in the physical probability.

6.8 Forwards and Futures

Forwards

A **forward contract** with **delivery price** K and maturity (or delivery) time T on an asset $\mathcal{U}$ is a type of financial derivative stipulated by two parties in which one of the parties promises to sell and deliver to the other party the asset $\mathcal{U}$ at time T in exchange for the cash K. The party who agrees to buy, resp. sell, the asset has the long, resp. short, position on the forward contract. Note that, as opposed to option contracts, both parties in a forward contract are *obliged* to fulfill their part of the agreement. Forward contracts are traded over the counter and most commonly on commodities or currencies. Let us give two examples.

Example of forward contract on a commodity. Consider a farmer who grows wheat and a miller who needs wheat to produce flour. Clearly, the farmer interest is to sell the wheat for the highest possible price, while the miller interest is to pay the least possible for the wheat. The price of the wheat depends on many economical and non-economical factors (such as whether conditions, which affect the quality and quantity of harvests) and it is therefore quite volatile. The farmer and the miller then stipulate a forward contract on the wheat in the winter (before the plantation, which occurs in the spring) with expiration date in the end of the summer (when the wheat is harvested), in order to lock its future trading price beforehand.

Example of forward contract on a currency. Suppose that a car company in Sweden promises to deliver a stock of 100 cars to another company in the United States in exactly one month. Suppose that the price of each car is fixed in Swedish crowns, say 100.000 crowns. Clearly the American company will benefit by an increase of the exchange rate crown/dollars and will be damaged in the opposite case. To avoid possible high losses, the American company by a forward contract on 100×100.000=ten millions Swedish crowns expiring in one month which gives the company the right *and* the obligation to buy ten millions crowns for a price in dollars agreed upon today.

The delivery price K in a forward contract is the result of a pondered estimation for the price of the asset at the time T in the future. In this respect, K is also called the **forward**

price of $\mathcal{U}$. More precisely, the T-forward price of the asset $\mathcal{U}$ at time t is the strike price of a forward contract on $\mathcal{U}$ with maturity T stipulated at time t; the current, actual price $\Pi(t)$ of the asset is also called the **spot** price.

Let us apply the risk-neutral pricing theory introduced in Section 6.2 to derive a mathematical model for the forward price of financial assets. Let $f(t, \Pi(t), K, T)$ be the value at time t of the forward contract on the asset $\mathcal{U}$ with maturity T and delivery price K. Here $\{\Pi(t)\}_{t\in[0,T]}$ is the price process of the underlying asset. The pay-off for the long position on the forward is given by

$$Y_{\text{long}} = (\Pi(T) - K),$$

while the pay-off for the short position is $Y_{\text{short}} = (K - \Pi(T))$. As both positions entail the same rights/obligations, and thus no privileged position exists, none of the two parties has to pay a premium to stipulate the forward contract. Hence $f(t, \Pi(t), K, T) = 0$. Assuming that the price $\{\Pi(t)\}_{t\geq 0}$ of the underlying asset follows a generalized geometric Brownian motion with strictly positive volatility, the risk-neutral value of the forward contract for the two parties is

$$\begin{aligned} f(t, \Pi(t), K, T) &= \pm\widetilde{\mathbb{E}}[(\Pi(T) - K)D(T)/D(t)|\mathcal{F}_W(t)] \\ &= \pm\left(\frac{1}{D(t)}\widetilde{\mathbb{E}}(\Pi(T)D(T)|\mathcal{F}_W(t)) - K\widetilde{\mathbb{E}}[\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)]\right). \end{aligned}$$

As the discounted price $\{\Pi^*(t)\}_{t\geq 0}$ of the underlying asset is a $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$, then $\widetilde{\mathbb{E}}(\Pi(T)D(T)|\mathcal{F}_W(t)) = D(t)\Pi(t)$. Hence

$$f(t, \Pi(t), K, T) = \pm(\Pi(t) - KB(t, T)),$$

where

$$B(t, T) = \widetilde{\mathbb{E}}[\exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)], \tag{6.114}$$

is the value of the ZCB computed using the risk-neutral probability. This leads to the following definition.

Definition 6.5. *Assume that the price $\{\Pi(t)\}_{t\geq 0}$ of the asset satisfies*

$$d\Pi(t) = \alpha(t)\Pi(t)dt + \sigma(t)\Pi(t)dW(t),$$

where $\{\alpha(t)\}_{t\geq 0}, \{\sigma(t)\}_{t\geq 0}, \{r(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ and $\sigma(t) > 0$ almost surely for all times. The **risk-neutral T-forward price** *at time t of the asset is the $\{\mathcal{F}_W(t)\}_{t\geq 0}$-adapted stochastic process $\{\mathrm{For}_T(t)\}_{t\in[0,T]}$ given by*

$$\mathrm{For}_T(t) = \frac{\Pi(t)}{B(t, T)}, \quad t \in [0, T],$$

where $B(t, T)$ is given by (6.114).

Note that, as long as the spot rate is positive, the forward price increases with respect to the time left to delivery, i.e., the longer we delay the delivery of the asset, the more we have to pay for it. This is intuitive, as the seller of the asset is loosing money by not selling the asset on the spot (due to its devaluation compared to the bond value).

Remark 6.12. For a constant interest rate $r(t) = r$ the forward price becomes

$$\mathrm{For}_T(t) = e^{r\tau}\Pi(t), \quad \tau = T - t,$$

in which case we find that the spot price of an asset is the discounted value of the forward price. When the asset is a commodity (e.g., corn), the forward price is also inflated by the cost of storage. Letting $c > 0$ be the cost to storage one share of the asset for one year, then the forward price of the asset, for delivery in τ years in the future, is $e^{c\tau}e^{r\tau}\Pi(t)$.

Forward probability measure

Let $t \in [0, T]$ and define

$$Z^{(T)}(t) = \frac{D(t)B(t,T)}{B(0,T)} = \frac{B^*(t,T)}{B(0,T)}. \tag{6.115}$$

As the discounted value of the ZCB is a martingale in the risk-neutral probability measure, then the process $\{Z^{(T)}(t)\}_{t\in[0,T]}$ is also a $\widetilde{\mathbb{P}}$-martingale. Moreover $\widetilde{\mathbb{E}}[Z^{(T)}(t)] = 1$, hence, by Theorem (3.17), for all $t \in [0, T]$, the function $\mathbb{P}^{(T)}(A) = \widetilde{\mathbb{E}}[Z^{(T)}(t)\mathbb{I}_A]$ defines a probability measure equivalent to $\widetilde{\mathbb{P}}$, which is called the T**-forward probability measure**.

Theorem 6.19. *Assume that the price $\{\Pi(t)\}_{t\geq 0}$ of the asset satisfies*

$$d\Pi(t) = \alpha(t)\Pi(t)dt + \sigma(t)\Pi(t)dW(t),$$

where $\{\alpha(t)\}_{t\geq 0}, \{\sigma(t)\}_{t\geq 0}, \{r(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ and $\sigma(t) > 0$ almost surely for all times. The risk-neutral T-forward price process $\{\mathrm{For}_T(t)\}_{t\in[0,T]}$ of the asset is a martingale in the T-forward probability measure relative to the filtration $\{\mathcal{F}_W(t)\}$ and

$$\mathrm{For}_T(t) = \mathbb{E}^{(T)}[\Pi(T)|\mathcal{F}_W(t)]. \tag{6.116}$$

Moreover risk-neutral pricing formula (6.10) for the European derivative with pay-off Y at maturity T can be written in terms of the forward probability measure as follows:

$$\Pi_Y(t) = B(t,T)\mathbb{E}^{(T)}[Y|\mathcal{F}_W(t)] \tag{6.117}$$

Proof. By (3.24) we have

$$\begin{aligned}\mathbb{E}^{(T)}[\Pi(T)|\mathcal{F}_W(t)] &= \frac{1}{Z^{(T)}(t)}\widetilde{\mathbb{E}}[Z^{(T)}(T)\Pi(T)|\mathcal{F}_W(t)] \\ &= \frac{B(0,T)}{D(t)B(t,T)}\widetilde{\mathbb{E}}[D(T)\Pi(T)/B(0,T)|\mathcal{F}_W(t)].\end{aligned}$$

As the discounted value of the asset is a martingale in the risk-neutral probability, we have

$$\mathbb{E}^{(T)}[\Pi(T)|\mathcal{F}_W(t)] = \frac{\Pi(t)}{B(t,T)} = \mathrm{For}_T(t).$$

Moreover, by Exercise 3.30, equation (6.116) implies that the forward price is a $\mathbb{P}^{(T)}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$. Similarly,

$$\mathbb{E}^{(T)}[Y|\mathcal{F}_W(t)] = \frac{1}{Z^{(T)}(t)}\widetilde{\mathbb{E}}[Z^{(T)}(T)Y|\mathcal{F}_W(t)] = \frac{1}{B(t,T)}\widetilde{\mathbb{E}}[D^{-1}(t)D(T)Y|\mathcal{F}_W(t)]$$

and so (6.117) follows. □

An advantage of writing the risk-neutral pricing formula in the form (6.117) is that the interest rate process does not appear within the conditional expectation. Thus, while in the risk-neutral probability measure one needs the joint distribution of the discount factor $D(T)$ and the pay-off Y in order to compute the price of the derivative, in the T-forward probability measure only the distribution of the pay-off is required. This suggests a method to compute the risk-neutral price of a derivative when the market parameters are stochastic. Before we see an example of application of this argument, we remark that, by Theorem 6.2(ii) with $\Pi_Y(t) = B(t,T)$, the process $\{Z_T(t)\}_{t\in[0,T]}$ satisfies $dZ^{(T)}(t) = \Delta(t)Z^{(T)}d\widetilde{W}(t)$, for some process $\{\Delta(t)\}_{t\in[0,T]}$ adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. Hence,

$$Z^{(T)}(t) = \exp\left(-\int_0^t \theta(s)d\widetilde{W}(s) - \frac{1}{2}\int_0^t \theta^2(s)\,ds\right), \quad \theta(t) = -\Delta(t).$$

It follows by Girsanov's theorem 4.10 that the process $\{W^{(T)}(t)\}_{t\in[0,T]}$ given by

$$W^{(T)}(t) = \widetilde{W}(t) + \int_0^t \theta(s)\,ds \tag{6.118}$$

is a Brownian motion in the T-forward measure.

Theorem 6.20. *Assume that the price $\{\Pi(t)\}_{t\geq 0}$ of the asset $\mathcal{U}$ satisfies*

$$d\Pi(t) = \alpha(t)\Pi(t)dt + \sigma(t)\Pi(t)dW(t), \quad \Pi(0) = \Pi_0 > 0, \quad \sigma(t) > 0,\ t\in[0,T],$$

and that there exists a constant $\widehat{\sigma} > 0$ such that

$$d\mathrm{For}_T(t) = \widehat{\sigma}\mathrm{For}_T(t)dW^{(T)}(t), \quad \mathrm{For}_T(0) = F_0 := \Pi_0/B(0,T), \tag{6.119}$$

where $\{W^{(T)}(t)\}_{t\in[0,T]}$ is the $\mathbb{P}^{(T)}$-Brownian motion defined above. Then the value at time $t = 0$ of the European call on the asset $\mathcal{U}$ with strike K and maturity T can be written as

$$\Pi_{\mathrm{call}}(0) = \Pi_0\Phi(d_+) - KB(0,T)\Phi(d_-), \tag{6.120}$$

where

$$d_\pm = \frac{\log\left(\frac{\Pi_0}{KB(0,T)}\right) \pm \frac{1}{2}\widehat{\sigma}^2 T}{\widehat{\sigma}\sqrt{T}}. \tag{6.121}$$

Proof. According to (6.117) we have

$$\Pi_{\text{call}}(0) = B(0,T)\mathbb{E}^{(T)}[\Pi(T) - K)_+].$$

In the right hand side we replace

$$\Pi(T) = \text{For}_T(T) = F_0 e^{-\frac{1}{2}\widehat{\sigma}^2 T - \widehat{\sigma} W^{(T)}(T)} = F_0 e^{-\frac{1}{2}\widehat{\sigma}^2 T - \widehat{\sigma}\sqrt{T} G}$$

and compute the expectation using that $G \in \mathcal{N}(0,1)$ in the forward measure, that is

$$\Pi_{\text{call}}(0) = B(0,T)\int_{\mathbb{R}} (F_0 e^{-\frac{1}{2}\widehat{\sigma}^2 T - \widehat{\sigma}\sqrt{T} y} - K)_+ e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}}.$$

The result now follows by computing the integral as in the proof of Theorem 6.6. □

Futures

Futures contracts are standardized forward contracts, i.e., rather than being traded over the counter, they are negotiated in regularized markets. Perhaps the most interesting role of futures contracts is that they make trading on commodities possible for anyone. To this regard we remark that commodities, e.g. crude oil, wheat, etc, are most often sold through long term contracts, such as forward and futures contracts, and therefore they do not usually have an "official spot price", but only a future delivery price (commodities "spot markets" exist, but their role is marginal for the discussion in this section).

Futures markets are markets in which the objects of trading are futures contracts. Unlike forward contracts, all futures contracts in a futures market are subject to the same regulation, and so in particular all contracts on the same asset with the same delivery time T have the same delivery price, which is called the T-**future price** of the asset and which we denote by $\text{Fut}_T(t)$. Thus $\text{Fut}_T(t)$ is the delivery price in a futures contract on the asset with time of delivery T and which is stipulated at time $t < T$. Futures markets have been existing for more than 300 years and nowadays the most important ones are the Chicago Mercantile Exchange (CME), the New York Mercantile Exchange (NYMEX), the Chicago Board of Trade (CBOT) and the International Exchange Group (ICE).

In a futures market, anyone (after a proper authorization) can stipulate a futures contract. More precisely, holding a position in a futures contract in the futures market consists in the agreement to receive as a cash flow the change in the future price of the underlying asset during the time in which the position is held. Note that the cash flow may be positive or negative. In a long position the cash flow is positive when the future price goes up and it is negative when the future price goes down , while a short position on the same contract receives the opposite cash flow. Moreover, in order to eliminate the risk of insolvency, the cash flow is distributed in time through the mechanism of the **margin account**. More precisely, assume that at $t = 0$ we open a long position in a futures contract expiring at time T. At the same time, we need to open a margin account which contains a certain amount

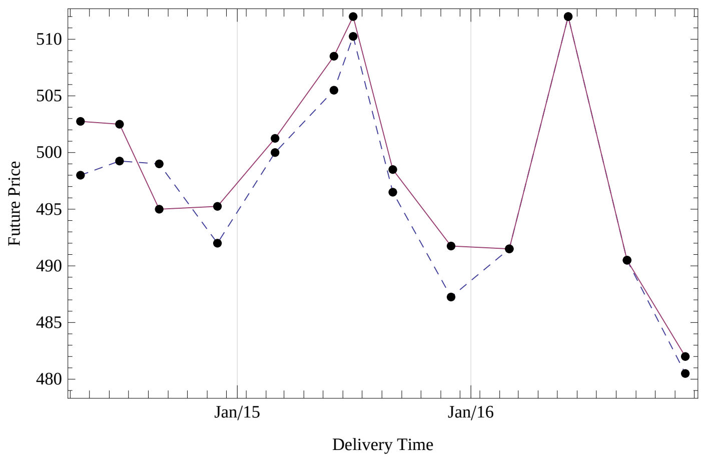


Figure 6.5: Futures price of corn on May 12, 2014 (dashed line) and on May 13, 2014 (continuous line) for different delivery times

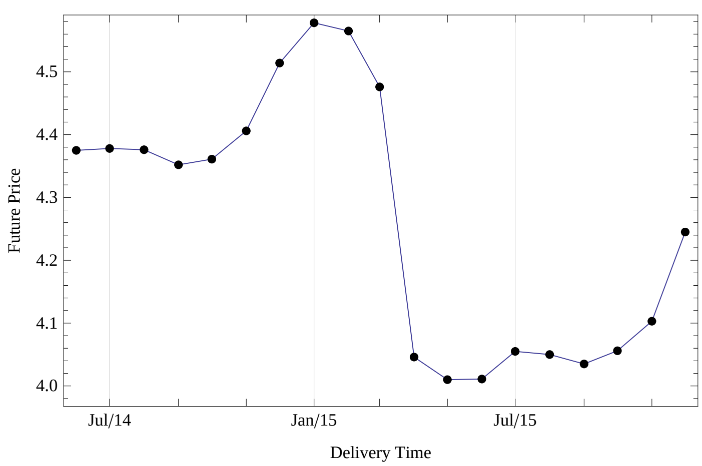


Figure 6.6: Futures price of natural gas on May 13, 2014 for different delivery times

of cash (usually, 10 % of the current value of the T-future price for each contract opened). At $t = 1$ day, the amount $\mathrm{Fut}_T(1) - \mathrm{Fut}_T(0)$ will be added to the account, if it positive, or withdrawn, if it is negative. The position can be closed at any time $t < T$ (multiple of days), in which case the total amount of cash flown in the margin account is

$$\begin{aligned}(\mathrm{Fut}_T(t) - \mathrm{Fut}_T(t-1)) + (\mathrm{Fut}_T(t-1) - \mathrm{Fut}_T(t-2))+\\ \dots + (\mathrm{Fut}_T(1) - \mathrm{Fut}_T(0)) = (\mathrm{Fut}_T(t) - \mathrm{Fut}_T(0)).\end{aligned}$$

If a long position is held up to the time of maturity, then the holder of the long position should buy the underlying asset.

Remark 6.13. Since a futures contract can be closed at any time prior to expiration, future contracts are *not* European style derivatives.

Our next purpose is to derive a mathematical model for the future price of an asset. Our guiding principle is that the **1+1 dimensional futures market** consisting of a futures contract and a risk-free asset should not admit self-financing arbitrage portfolios. Consider a portfolio invested in $h(t)$ shares of the futures contract and $h_B(t)$ shares of the risk-free asset at time t. We assume that $\{h(t), h_B(t)\}_{t\in[0,T]}$ is adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ and suppose that $\{\mathrm{Fut}_T(t)\}_{t\in[0,T]}$ is a diffusion process. Since futures contracts have zero-value, the value of the portfolio at time t is $V(t) = h_B(t)B(t)$. For a self-financing portfolio we require that any positive cash-flow generated by the futures contract in the interval $[t, t+dt]$ should be invested to buy shares of the risk-free asset and that, conversely, any negative cash flow should be settled by issuing shares of the risk-free asset (i.e., by borrowing money). Since the cash-flow generated in the interval $[t, t + dt]$ is given by $dC(t) = h(t)d\mathrm{Fut}_T(t)$, the value of a self-financing portfolio invested in the 1+1 dimensional futures market must satisfy

$$dV(t) = h_B(t)dB(t) + h(t)d\mathrm{Fut}_T(t) = r(t)V(t)dt + h(t)d\mathrm{Fut}_T(t),$$

or equivalently

$$dV^*(t) = h(t)D(t)d\mathrm{Fut}_T(t). \tag{6.122}$$

Now, we have seen that a simple condition ensuring that a portfolio is not an arbitrage is that its discounted value be a martingale in the risk-neutral measure relative to the filtration generated by the Brownian motion. By (6.122), the latter condition is achieved by requiring that $d\mathrm{Fut}_T(t) = \Delta(t)d\widetilde{W}(t)$, for some stochastic process $\{\Delta(t)\}_{t\in[0,T]}$ adapted to $\{\mathcal{F}_W(t)\}_{t\in[0,T]}$. In particular, it is reasonable to impose that

(i) $\{\mathrm{Fut}_T(t)\}_{t\in[0,T]}$ should be a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

Furthermore, it is clear that the future price of an asset at the expiration date T should be equal to its spot price at time T, and so we impose that

(ii) $\mathrm{Fut}_T(T) = \Pi(T)$.

It follows by Exercise 3.30 that the conditions **(i)-(ii)** determine a unique stochastic process $\{\mathrm{Fut}_T(t)\}_{t\in[0,T]}$, which is given in the following definition.

Definition 6.6. *Assume that the price $\{\Pi(t)\}_{t\geq 0}$ of the asset satisfies*

$$d\Pi(t) = \alpha(t)\Pi(t)dt + \sigma(t)\Pi(t)dW(t),$$

*where $\{\alpha(t)\}_{t\geq 0}, \{\sigma(t)\}_{t\geq 0}, \{r(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ and $\sigma(t) > 0$ almost surely for all times. The T-**Future price** at time t of the asset is the $\{\mathcal{F}_W(t)\}_{t\geq 0}$-adapted stochastic process $\{\mathrm{Fut}_T(t)\}_{t\in[0,T]}$ given by*

$$\mathrm{Fut}_T(t) = \widetilde{\mathbb{E}}[\Pi(T)|\mathcal{F}_W(t)], \quad t \in [0,T].$$

We now show that our goal to make the futures market arbitrage-free has been achieved.

Theorem 6.21. *There exists a stochastic process $\{\Delta(t)\}_{t\in[0,T]}$ adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ such that*

$$\mathrm{Fut}_T(t) = \mathrm{Fut}_T(0) + \int_0^t \Delta(s)d\widetilde{W}(s). \tag{6.123}$$

Moreover, any $\{\mathcal{F}_W(t)\}_{t\geq 0}$-adapted self-financing portfolio $\{h(t), h_B(t)\}_{t\in[0,T]}$ invested in the 1+1 dimensional futures market is not an arbitrage.

Proof. The second statement follows immediately by the first one, since (6.122) and (6.123) imply that the value of a self-financing portfolio invested in the 1+1 dimensional futures market is a $\widetilde{\mathbb{P}}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\in[0,T]}$. To prove (6.123), we first notice that, by (3.24),

$$Z(s)\widetilde{\mathbb{E}}[\mathrm{Fut}_T(t)|\mathcal{F}_W(s)] = \mathbb{E}[Z(t)\mathrm{Fut}_T(t)|\mathcal{F}_W(s)].$$

By the martingale property of the future price, the left hand side is $Z(s)\mathrm{Fut}_T(s)$. Hence

$$Z(s)\mathrm{Fut}_T(s) = \mathbb{E}[Z(t)\mathrm{Fut}_T(t)|\mathcal{F}_W(s)],$$

that is to say, the process $\{Z(t)\mathrm{Fut}_T(t)\}_{t\in[0,T]}$ is a $\mathbb{P}$-martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\in[0,T]}$. By the martingale representation theorem, Theorem 4.5(v), there exists a stochastic process $\{\Gamma(t)\}_{t\in[0,T]}$ adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ such that

$$Z(t)\mathrm{Fut}_T(t) = \mathrm{Fut}_T(0) + \int_0^t \Gamma(s)dW(s).$$

We now proceed as in the proof of Theorem 6.2, namely we write

$$d\mathrm{Fut}_T(t) = d(Z(t)\mathrm{Fut}_T(t)Z(t)^{-1})$$

and apply Itô's product rule and Itô's formula to derive that (6.123) holds with

$$\Delta(t) = \theta(t)\mathrm{Fut}_T(t) + \frac{\Gamma(t)}{Z(t)}.$$

□

Exercise 6.39 (Sol. 46). *Show that the* **Forward-Future spread** *of an asset, i.e., the difference between its forward and future price, satisfies*

$$\mathrm{For}_T(t)-\mathrm{Fut}_T(t)=\frac{1}{\widetilde{\mathbb{E}}[D(T)|\mathcal{F}_W(t)]}\Big\{\widetilde{\mathbb{E}}[D(T)\Pi(T)|\mathcal{F}_W(t)]-\widetilde{\mathbb{E}}[D(T)|\mathcal{F}_W(t)]\widetilde{\mathbb{E}}[\Pi(T)|\mathcal{F}_W(t)]\Big\}. \tag{6.124}$$

Moreover, show that when the interest rate $\{r(t)\}_{t\in[0,T]}$ is a deterministic function of time (e.g., a deterministic constant), then $\mathrm{For}_T(t)=\mathrm{Fut}_T(t)$, for all $t\in[0,T]$.

6.9 Multi-dimensional markets

In this section we consider $N+1$ dimensional stock markets. We denote the stocks prices by

$$\{S_1(t)\}_{t\geq 0},\dots,\{S_N(t)\}_{t\geq 0}$$

and assume the following dynamics

$$dS_k(t)=\left(\mu_k(t)\,dt+\sum_{j=1}^N\sigma_{kj}(t)dW_j(t)\right)S_k(t), \tag{6.125}$$

for some stochastic processes $\{\mu_k(t)\}_{t\geq 0}$, $\{\sigma_{kj}(t)\}_{t\geq 0}$, $j,k=1,\dots,N$ in the class $\mathcal{C}^0[\mathcal{F}_W(t)]$, where in this section $\{\mathcal{F}_W(t)\}_{t\geq 0}$ denotes the filtration generated by the Brownian motions $\{W_1(t)\}_{t\geq 0},\dots\{W_N(t)\}_{t\geq 0}$. Moreover we assume that the Brownian motions are independent, in particular

$$dW_j(t)dW_k(t)=0,\quad \text{for all } j\neq k, \tag{6.126}$$

see Exercise 3.25. Finally $\{r(t)\}_{t\geq 0}\in\mathcal{C}^0[\mathcal{F}_W(t)]$ is the interest rate of the money market.

Now, given stochastic processes $\{\theta_k(t)\}_{t\geq 0}\in\mathcal{C}^0[\mathcal{F}_W(t)]$, $k=1,\dots,N$, satisfying the Novikov condition (4.20), the stochastic process $\{Z(t)\}_{t\geq 0}$ given by

$$Z(t)=\exp\left(-\sum_{k=1}^N\left(\int_0^t\frac{1}{2}\theta_k^2(s)\,ds+\int_0^t\theta_k(s)\,dW_k(s)\right)\right) \tag{6.127}$$

is a martingale relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$ (see Exercise 4.7). Since $\mathbb{E}[Z(t)]=\mathbb{E}[Z(0)]=1$, for all $t\geq 0$, we can use the stochastic process $\{Z(t)\}_{t\geq 0}$ to define a risk-neutral probability measure associated to the $N+1$ dimensional stock market, as we did in the one dimensional case, see Definition 6.1.

Definition 6.7. *Let $T>0$ and assume that the* **market price of risk equations**

$$\mu_j(t)-r(t)=\sum_{k=1}^N\sigma_{jk}(t)\theta_k(t),\quad j=1,\dots,N, \tag{6.128}$$

admit a solution $(\theta_1(t), \dots, \theta_N(t))$*, for all* $t \geq 0$*. Define the stochastic process* $\{Z(t)\}_{t\geq 0}$ *as in* (6.127)*. Then the measure* $\widetilde{\mathbb{P}}$ *equivalent to* $\mathbb{P}$ *given by*

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}[Z(T)\mathbb{I}_A]$$

is called the risk-neutral probability measure of the market at time T*.*

Note that, as opposed to the one dimensional case, the risk-neutral measure just defined need not be unique, as the market price of risk equations may admit more than one solution. For each risk-neutral probability measure $\widetilde{\mathbb{P}}$ we can apply the multidimensional Girsanov theorem 4.11 and conclude that the stochastic processes $\{\widetilde{W}_1(t)\}_{t\geq 0}, \dots \{\widetilde{W}_N(t)\}_{t\geq 0}$ given by

$$\widetilde{W}_k(t) = W_k(t) + \int_0^t \theta_k(s)\, ds$$

are $\widetilde{\mathbb{P}}$-independent Brownian motions. Moreover these Brownian motions are $\widetilde{\mathbb{P}}$-martingales relative to the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$.

Now let $\{h_{S_1}(t)\}_{t\geq 0}, \dots, \{h_{S_N}(t)\}_{t\geq 0} \in \mathcal{C}^0[\mathcal{F}_W(t)]$ be stochastic processes representing the number of shares on the stocks in a portfolio invested in the $N+1$ dimensional stock market. Let $\{h_B(t)\}_{t\geq 0}$ be the number of shares on the risk-free asset. The portfolio value is

$$V(t) = \sum_{k=1}^N h_{S_k}(t)S_k(t) + h_B(t)B(t)$$

and the portfolio process is self-financing if its value satisfies

$$dV(t) = \sum_{k=1}^N h_{S_k}(t)dS_k(t) + h_B(t)dB(t),$$

that is

$$dV(t) = \sum_{k=1}^N h_{S_k}(t)dS_k(t) + r(t)\left(V(t) - \sum_{k=1}^N h_{S_k}(t)S_k(t)\right)dt.$$

Theorem 6.22. *Assume that a risk-neutral probability* $\widetilde{\mathbb{P}}$ *exists, i.e., the equations* (6.128) *admit a solution. Then the discounted value of any self-financing portfolio invested in the* $N+1$ *dimensional market is a* $\widetilde{\mathbb{P}}$*-martingale relative to the filtration* $\{\mathcal{F}_W(t)\}_{t\geq 0}$*. In particular (by Theorem 3.14) there exists no self-financing arbitrage portfolio invested in the* $N+1$ *dimensional stock market.*

Proof. The discounted value of the portfolio satisfies

$$\begin{aligned}
dV^*(t) &= D(t)\left(\sum_{j=1}^N h_{S_j}(t)S_j(t)(\alpha_j(t) - r(t))\,dt + \sum_{j,k=1}^N h_{S_j}(t)S_j(t)\sigma_{jk}(t)dW_k(t)\right)\\
&= D(t)\left(\sum_{j=1}^N h_{S_j}(t)S_j(t)\sum_{k=1}^N \sigma_{jk}(t)\theta_k(t)dt + \sum_{j,k=1}^N h_{S_j}(t)S_j(t)\sigma_{jk}(t)dW_k(t)\right)\\
&= D(t)\sum_{j=1}^N h_{S_j}(t)S_j(t)\sum_{k=1}^N \sigma_{jk}(t)d\widetilde{W}_k(t).
\end{aligned}$$

All Itô's integrals in the last line are $\widetilde{\mathbb{P}}$-martingales relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. The result follows. $\square$

Exercise 6.40. *Work out the details of the computations omitted in the proof of the previous theorem.*

Next we show that the existence of a risk-neutral probability measure is necessary for the absence of self-financing arbitrage portfolios in $N+1$ dimensional stock markets.

Let $N = 3$ and assume that the market parameters are constant. Let $r(t) = r > 0$, $(\mu_1, \mu_2, \mu_3) = (2, 3, 2)$ and let the volatility matrix be given by

$$\sigma_{ij} = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 1 & 2 & 0 \end{pmatrix}.$$

Thus the stocks prices satisfy

$$\begin{aligned}
dS_1(t) &= (2dt + dW_1(t) + 2dW_2(t))S_1(t),\\
dS_2(t) &= (3dt + 2dW_1(t) + 4dW_2(t))S_2(t),\\
dS_3(t) &= (2dt + dW_1(t) + 2dW_2(t))S_3(t).
\end{aligned}$$

The market price of risk equations are

$$\begin{aligned}
\theta_1 + 2\theta_2 &= 2 - r\\
2\theta_1 + 4\theta_2 &= 3 - r\\
\theta_1 + 2\theta_2 &= 2 - r.
\end{aligned}$$

This system is solvable if and only if $r = 1$, in which case there exist infinitely many solutions given by

$$\theta_1 \in \mathbb{R}, \quad \theta_2 = \frac{1}{2}(1 - \theta_1).$$

Hence for $r = 1$ there exists at least one (in fact, infinitely many) risk-neutral probability measures, and thus the market is free of arbitrage. To construct an arbitrage portfolio when $0 < r < 1$, let

$$h_{S_1}(t) = \frac{1}{S_1(t)}, \quad h_{S_2}(t) = -\frac{1}{S_2(t)}, \quad h_{S_3}(t) = \frac{1}{S_3(t)}$$

and choose $h_B(t)$ such that the portfolio process is self-financing (see Exercise 4.10). The value $\{V(t)\}_{t\geq 0}$ of this portfolio satisfies

$$\begin{aligned}
dV(t) &= h_{S_1}(t)dS_1(t) + h_{S_2}(t)dS_2(t) + h_{S_3}(t)dS_3(t) \\
&\quad + r(V(t) - h_{S_1}(t)S_1(t) - h_{S_2}(t)S_2(t) - h_{S_3}(t)S_3(t))dt \\
&= rV(t)dt + (1-r)dt.
\end{aligned}$$

Hence

$$V(t) = V(0)e^{rt} + \frac{1}{r}(1-r)(e^{rt} - 1)$$

and this portfolio is an arbitrage, because for $V(0) = 0$ we have $V(t) > 0$, for all $t > 0$. Similarly one can find an arbitrage portfolio for $r > 1$.

Next we address the question of **completeness** of $N + 1$ dimensional stock markets, i.e., the question of whether any European derivative can be hedged in this market. Consider a European derivative on the stocks with pay-off Y and time of maturity T. For instance, for a standard European derivative, $Y = g(S_1(T), \dots S_N(T))$, for some measurable function g. The risk-neutral price of the derivative is

$$\Pi_Y(t) = \widetilde{\mathbb{E}}[Y \exp(-\int_t^T r(s)\,ds)|\mathcal{F}_W(t)],$$

and coincides with the value at time t of any self-financing portfolio invested in the $N + 1$ dimensional market. The question of existence of an hedging portfolio is answered by the following theorem.

Theorem 6.23. *Assume that the volatility matrix $(\sigma_{jk}(t))_{j,k=1,\dots N}$ is invertible, for all $t \geq 0$. There exist stochastic processes $\{\Delta_1(t)\}_{t\in[0,T]}, \dots \{\Delta_N(t)\}_{t\in[0,T]}$, adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$, such that*

$$D(t)\Pi_Y(t) = \Pi_Y(0) + \sum_{k=1}^{N} \int_0^t \Delta_k(s) d\widetilde{W}_k(s), \quad t \in [0,T]. \tag{6.129}$$

Let $(Y_1(t), \dots, Y_N(t))$ be the solution of

$$\sum_{k=1}^{N} \sigma_{jk}(t) Y_j(t) = \frac{\Delta_k(t)}{D(t)}. \tag{6.130}$$

Then the portfolio $\{h_{S_1}(t), \dots, h_{S_N}(t), h_B(t)\}_{t\in[0,T]}$ given by

$$h_{S_j}(t) = \frac{Y_j(t)}{S_j(t)}, \quad h_B(t) = (\Pi_Y(t) - \sum_{j=1}^{N} h_{S_j}(t)S_j(t))/B(t) \tag{6.131}$$

is self-financing and replicates the derivative at any time, i.e., its value $V(t)$ is equal to $\Pi_Y(t)$ for all $t \in [0,T]$. In particular, $V(T) = \Pi_Y(T) = Y$, i.e., the portfolio is hedging the derivative.

The proof of this theorem is conceptually very similar to that of Theorem 6.2 and is therefore omitted (it makes use of the multidimensional version of the martingale representation theorem). Notice that, having assumed that the volatility matrix is invertible, *the risk-neutral probability measure of the market is unique.* We now show that the uniqueness of the risk-neutral probability measure is necessary to guarantee completeness. In fact, let $r = 1$ in the example considered before and pick the following solutions of the market price of risk equations:

$$(\theta_1, \theta_2) = (0, 1/2), \quad \text{and} \quad (\theta_1, \theta_2) = (1, 0)$$

(any other pair of solutions would work). The two corresponding risk-neutral probability measures, denoted respectively by $\widetilde{\mathbb{P}}$ and $\widehat{\mathbb{P}}$, are given by

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}[\widetilde{Z}\mathbb{I}_A] \quad \widehat{\mathbb{P}}(A) = \mathbb{E}[\widehat{Z}\mathbb{I}_A], \quad \text{for all } A \in \mathcal{F},$$

where

$$\widetilde{Z} = e^{-\frac{1}{8}T - \frac{1}{2}W_2(T)}, \quad \widehat{Z} = e^{-\frac{1}{2}T - W_1(T)}.$$

Let $A = \{\omega : \frac{1}{2}W_2(T,\omega) - W_1(T,\omega) < \frac{3}{8}T\}$. Hence

$$\widehat{Z}(\omega) < \widetilde{Z}(\omega), \quad \text{for } \omega \in A$$

and thus $\widehat{\mathbb{P}}(A) < \widetilde{\mathbb{P}}(A)$. Consider a financial derivative with pay-off $Q = \mathbb{I}_A/D(T)$. If there existed an hedging, self-financing portfolio for such derivative, then, since the discounted value of such portfolio is a martingale in both risk-neutral probability measures, we would have

$$V(0) = \widetilde{\mathbb{E}}[QD(T)], \quad \text{and} \quad V(0) = \widehat{\mathbb{E}}[QD(T)]. \tag{6.132}$$

But

$$\widehat{\mathbb{E}}[QD(T)] = \widehat{\mathbb{E}}(\mathbb{I}_A) = \widehat{\mathbb{P}}(A) < \widetilde{\mathbb{P}}(A) = \widetilde{\mathbb{E}}(\mathbb{I}_A) = \widetilde{\mathbb{E}}[QD(T)]$$

and thus (6.132) cannot be verified.

Multi-assets options

Multi-asset options are options on several underlying assets. Notable examples include rainbow options, basket options and quanto options.

In the following we discuss options on two stocks in a **2+1 dimensional Black-Scholes market**, i.e., a market with constant parameters. It follows that

$$dS_1(t) = \mu_1 S_1(t)dt + \sigma_{11}S_1(t)dW_1(t) + \sigma_{12}S_1(t)dW_2(t) \tag{6.133a}$$
$$dS_2(t) = \mu_2 S_2(t)dt + \sigma_{21}S_2(t)dW_1(t) + \sigma_{22}S_2(t)dW_2(t), \tag{6.133b}$$

where the volatility matrix

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

is invertible (so that the market is complete). Integrating (6.133) we obtain that $(S_1(t), S_2(t))$ is given by the **2-dimensional geometric Brownian motion**:

$$S_1(t) = S_1(0)e^{(\mu_1 - \frac{1}{2}(\sigma_{11}^2+\sigma_{12}^2))t + \sigma_{11}W_1(t) + \sigma_{12}W_2(t)}, \tag{6.134a}$$

$$S_2(t) = S_2(0)e^{(\mu_2 - \frac{1}{2}(\sigma_{21}^2+\sigma_{22}^2))t + \sigma_{21}W_1(t) + \sigma_{22}W_2(t)}, \tag{6.134b}$$

or, more concisely,

$$S_j(t) = S_j(0)e^{(\mu_j - \frac{|\sigma_j|^2}{2})t + \sigma_j \cdot W(t)},$$

where $\sigma_j = (\sigma_{j1}, \sigma_{j2})$, $j = 1, 2$, $W(t) = (W_1(t), W_2(t))$ and $\cdot$ denotes the standard scalar product of vectors.

Theorem 6.24. *The random variables $S_1(t), S_2(t)$ have the joint density*

$$f_{S_1(t),S_2(t)}(x,y) = \frac{e^{-\frac{1}{2t}\left(\log\frac{x}{S(0)} - \alpha_1 t \quad \log\frac{y}{S(0)} - \alpha_2 t\right)(\sigma\sigma^T)^{-1}\begin{pmatrix}\log\frac{x}{S(0)} - \alpha_1 t \\ \log\frac{y}{S(0)} - \alpha_2 t\end{pmatrix}}}{txy\sqrt{(2\pi)^2\det(\sigma\sigma^T)}}, \tag{6.135}$$

where $\alpha_j = \mu_j - \frac{|\sigma_j|^2}{2}$, $j = 1, 2$. Moreover $\log S_1(t), \log S_2(t)$ are jointly normally distributed with mean $m = (\log S_1(0) + \alpha_1 t, \log S_2(0) + \alpha_2 t)$ and covariant matrix $C = t\sigma\sigma^T$.

Proof. Letting $X_i = W_i(t)/\sqrt{t} \in \mathcal{N}(0,1)$, we write the stock prices as

$$S_1(t) = S_1(0)e^{\alpha_1 t + Y_1}, \quad S_2(t) = S_2(0)e^{\alpha_2 t + Y_2},$$

where

$$Y_1 = \sigma_{11}\sqrt{t}X_1 + \sigma_{12}\sqrt{t}X_2, \quad Y_2 = \sigma_{21}\sqrt{t}X_1 + \sigma_{22}\sqrt{t}X_2.$$

It follows by Exercise 3.20 that Y_1, Y_2 are jointly normally distributed with zero mean and covariant matrix $C = t\sigma\sigma^T$, which proves the second statement in the theorem. To compute the joint density of the stock prices, we notice that

$$S_1(t) \leq x \Leftrightarrow Y_1 \leq \log\left(\frac{x}{S_1(0)}\right) - \alpha_1 t, \quad S_2(t) \leq y \Leftrightarrow Y_2 \leq \log\left(\frac{y}{S_2(0)}\right) - \alpha_2 t,$$

hence

$$F_{S_1(t),S_2(t)}(x,y) = F_{Y_1,Y_2}(\log\frac{x}{S_1(0)} - \alpha_1 t, \log\frac{y}{S_2(0)} - \alpha_2 t).$$

Hence

$$f_{S_1(t),S_2(t)}(x,y) = \partial_{xy}^2 F_{S_1(t),S_2(t)}(x,y) = \frac{1}{xy}f_{Y_1,Y_2}(\log\frac{x}{S_1(0)} - \alpha_1 t, \log\frac{y}{S_2(0)} - \alpha_2 t).$$

Using the joint normal density of Y_1, Y_2 completes the proof. $\square$

Exercise 6.41. *Show that the process* (6.133) *is equivalent, in distribution, to the process*

$$dS_i(t) = \mu_i S_i(t)\,dt + \overline{\sigma}_i S_i(t) dW_i^{(\rho)}(t), \quad i = 1, 2, \tag{6.136}$$

where

$$\overline{\sigma}_i = \sqrt{\sigma_{i1}^2 + \sigma_{i2}^2}, \quad \rho = \frac{\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}}{\sqrt{(\sigma_{11}^2 + \sigma_{12}^2)(\sigma_{21}^2 + \sigma_{22}^2)}} \in [-1, 1] \tag{6.137}$$

and where $W_1^{(\rho)}(t)$, $W_2^{(\rho)}(t)$ *are correlated Brownian motions with correlation* ρ, *i.e.*,

$$dW_1^{(\rho)}(t) dW_2^{(\rho)}(t) = \rho\,dt.$$

TIP: Use Exercise 2.11.

Now let $r(t) = r$ be the constant interest rate of the money market. The solution of the market price of risk equations (6.128) can be written as

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \sigma^{-1} \begin{pmatrix} \mu_1 - r \\ \mu_2 - r \end{pmatrix} = \frac{1}{\det \sigma} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{pmatrix} \begin{pmatrix} \mu_1 - r \\ \mu_2 - r \end{pmatrix},$$

that is

$$\theta_1 = \frac{1}{\det \sigma}[\sigma_{22}(\mu_1 - r) - \sigma_{12}(\mu_2 - r)], \quad \theta_2 = \frac{1}{\det \sigma}[-\sigma_{21}(\mu_1 - r) + \sigma_{11}(\mu_2 - r)].$$

Replacing $dW_i(t) = d\widetilde{W}_i(t) - \theta_i\,dt$ into (6.133) we find

$$dS_1(t) = r\,S_1(t)dt + \sigma_{11} S_1(t) d\widetilde{W}_1(t) + \sigma_{12} S_1(t) d\widetilde{W}_2(t), \tag{6.138a}$$

$$dS_2(t) = r\,S_2(t)dt + \sigma_{21} S_2(t) d\widetilde{W}_1(t) + \sigma_{22} S_2(t) d\widetilde{W}_2(t). \tag{6.138b}$$

Note that the discounted price of both stocks is a martingale in the risk-neutral probability measure, as expected. Moreover the system (6.138) can be integrated to give

$$S_j(t) = S_j(0) e^{(r - \frac{|\sigma_j|^2}{2})t + \sigma_j \cdot \widetilde{W}(t)}, \tag{6.139}$$

where $\widetilde{W}(t) = (\widetilde{W}_1(t), \widetilde{W}_2(t))$. As $\widetilde{W}_1(t), \widetilde{W}_2(t)$ are independent $\widetilde{\mathbb{P}}$-Brownian motions, the joint distribution of the stock prices in the risk-neutral probability measure is given by (6.135) where now

$$\alpha_i = r - \frac{|\sigma_j|^2}{2}, \quad i = 1, 2.$$

Next consider a standard European style derivative on the two stocks with pay-off $Y = g(S_1(T), S_2(T))$. The risk-neutral price of the derivative is

$$\Pi_Y(t) = e^{-r(T-t)} \widetilde{\mathbb{E}}[g(S_1(T), S_2(T)) | \mathcal{F}_W(t)]. \tag{6.140}$$

By the Markov property for systems of stochastic differential equations, there exists a function $v_g : [0, T] \times (0, \infty)^2 \to (0, \infty)$ such that

$$\Pi_Y(t) = v_g(t, S_1(t), S_2(t)). \tag{6.141}$$

As in the case of options on one single stock, the pricing function can be computed in two ways: using the joint probability density of the stocks or by solving a PDE.

Black-Scholes price for options on two stocks

We show first how to compute the function v_g in (6.141) using the joint probability density of $S_1(t), S_2(t)$ derived in Theorem 6.24. We argue as in Section 6.3. By (6.139) we have

$$S_i(T) = S_j(t)e^{(r-\frac{|\sigma_j|^2}{2})\tau+\sigma_j\cdot(\widetilde{W}(T)-\widetilde{W}(t))}, \quad \tau = T-t.$$

Replacing into (6.140) we obtain

$$\Pi_Y(t) = e^{-r\tau}\widetilde{\mathbb{E}}[g(S_1(t)e^{(r-\frac{|\sigma_1|^2}{2})\tau+\sigma_1\cdot(\widetilde{W}(T)-\widetilde{W}(t))}, S_2(t)e^{(r-\frac{|\sigma_2|^2}{2})\tau+\sigma_2\cdot(\widetilde{W}(T)-\widetilde{W}(t))})|\mathcal{F}_W(t)].$$

As $(S_1(t), S_2(t))$ is measurable with respect to $\mathcal{F}_W(t)$ and $\widetilde{W}(T)-\widetilde{W}(t)$ is independent of $\mathcal{F}_W(t)$, Theorem 3.13(x) gives

$$\Pi_Y(t) = v_g(t, S_1(t), S_2(t)),$$

where

$$v_g(t,x,y) = e^{-r\tau}\widetilde{\mathbb{E}}[g(xe^{(r-\frac{|\sigma_1|^2}{2})\tau+\sigma_1\cdot(\widetilde{W}(T)-\widetilde{W}(t))}, ye^{(r-\frac{|\sigma_2|^2}{2})\tau+\sigma_2\cdot(\widetilde{W}(T)-\widetilde{W}(t))})].$$

To compute the expectation in the right hand side of the latter equation we use that the random variables

$$Y_1 = \sigma_1\cdot(\widetilde{W}(T)-\widetilde{W}(t)), \quad Y_2 = \sigma_2\cdot(\widetilde{W}(T)-\widetilde{W}(t))$$

are jointly normally distributed with zero mean and covariance matrix $C = \tau\sigma\sigma^T$. Hence

$$v_g(t,x,y) = e^{-r\tau}\int_{\mathbb{R}}\int_{\mathbb{R}} g(xe^{(r-\frac{|\sigma_1|^2}{2})\tau+\sqrt{\tau}\xi}, ye^{(r-\frac{|\sigma_2|^2}{2})\tau+\sqrt{\tau}\eta})\frac{\exp\left(-\frac{1}{2}\begin{pmatrix}\xi & \eta\end{pmatrix}(\sigma\sigma^T)^{-1}\begin{pmatrix}\xi \\ \eta\end{pmatrix}\right)}{2\pi\sqrt{\det(\sigma\sigma^T)}}d\xi\, d\eta. \tag{6.142}$$

Definition 6.8. *The stochastic process* $\{\Pi_Y(t)\}_{t\in[0,T]}$ *given by* (6.141)-(6.142), *is called the* **Black-Scholes price** *of the standard 2-stocks European derivative with pay-off* $Y = g(S_1(T), S_2(T))$ *and time of maturity* $T > 0$.

Black-Scholes PDE for options on two stocks

Next we show how to derive the pricing function v_g by solving a PDE.

Theorem 6.25. *Let* v_g *be the (unique) strong solution to the terminal value problem*

$$\partial_t v_g + r(x\partial_x v_g + y\partial_y v_g) + \frac{1}{2}(\sigma_{11}^2+\sigma_{12}^2)x^2\partial_x^2 v_g + \frac{1}{2}(\sigma_{21}^2+\sigma_{22}^2)y^2\partial_y^2 v_g$$
$$+ (\sigma_{11}\sigma_{21}+\sigma_{12}\sigma_{22})xy\partial_{xy}v_g = r\, v_g, \quad t\in(0,T), \quad x,y>0, \tag{6.143a}$$
$$v_g(T,x,y) = g(x,y), \quad x,y>0. \tag{6.143b}$$

Then (6.141) *holds. The PDE in* (6.143) *is called the* **2-dimensional Black-Scholes PDE**.

Proof. By Itô's formula in two dimensions,

$$
\begin{aligned}
d(e^{-rt}v_g) =& e^{-rt}\Big(- rv_g\,dt + \partial_t v_g\,dt + \partial_x v_g\,dS_1(t) + \partial_y v_g\,dS_2(t) \\
&+ \partial^2_{xy} v_g\,dS_1(t)dS_2(t) + \frac{1}{2}\partial^2_x v_g\,dS_1(t)\,dS_1(t) + \frac{1}{2}\partial^2_y v_g\,dS_2(t)\,dS_2(t)\Big).
\end{aligned}
$$

Moreover, using (6.138),

$$
\begin{aligned}
dS_1(t)dS_1(t) &= (\sigma_{11}^2 + \sigma_{12}^2)S_1(t)^2\,dt \\
dS_2(t)dS_2(t) &= (\sigma_{21}^2 + \sigma_{22}^2)S_2(t)^2\,dt \\
dS_1(t)dS_2(t) &= (\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22})S_1(t)S_2(t)\,dt.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
d(e^{-rt}v_g(t,S_1(t),S_2(t))) = \alpha(t)\,dt &+ e^{-rt}S_1(t)\partial_x v_g(t,S_1(t),S_2(t))\,(\sigma_{11}\,d\widetilde{W}_1(t) + \sigma_{12}d\widetilde{W}_2(t)) \\
&+ e^{-rt}S_2(t)\partial_y v_g(t,S_1(t),S_2(t))\,(\sigma_{21}\,d\widetilde{W}_1(t) + \sigma_{22}d\widetilde{W}_2(t))
\end{aligned}
$$

where the drift term is

$$
\begin{aligned}
\alpha(t) =& e^{-rt}\big(- rv_g + \partial_t v_g + r(x\partial_x v_g + y\partial_y v_g) \\
&+ \frac{1}{2}(\sigma_{11}^2 + \sigma_{12}^2)x^2\partial_x^2 v_g + \frac{1}{2}(\sigma_{21}^2 + \sigma_{22}^2)y^2\partial_y^2 v_g \\
&+ (\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22})xy\partial_{xy} v_g\big)(t,S_1(t),S_2(t)) = 0,
\end{aligned}
$$

due to v_g solving (6.143). It follows that the stochastic process $\{e^{-rt}v_g(t,S_1(t),S_2(t))\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$, hence, using the terminal condition $v_g(T) = g$, we have

$$
e^{-rT}\widetilde{\mathbb{E}}[g(S_1(T),S_2(T))|\mathcal{F}_W(t)] = e^{-rt}v_g(t,S_1(t),S_2(t)), \quad t\in[0,T],
$$

which proves (6.141). □

Hedging portfolio

Finally we derive the formulas for the hedging portfolio for standard 2-stocks European derivatives in Black-Scholes markets.

Theorem 6.26. *The numbers of shares $h_{S_1}(t), h_{S_2}(t)$ in the self-financing hedging portfolio for the European derivative with pay-off $Y = g(S_1(T),S_2(T))$ and maturity T are given by*

$$
h_{S_1}(t) = \partial_x v_g(t,S_1(t),S_2(t)), \quad h_{S_2}(t) = \partial_y v_g(t,S_1(t),S_2(t)).
$$

Proof. The discounted value of the derivative satisfies $d\Pi_Y^*(t) = \Delta_1(t)d\widetilde{W}_1(t) + \Delta_2(t)d\widetilde{W}_2(t)$,

where

$$\Delta_1(t) = e^{-rt}(S_1(t)\sigma_{11}\partial_x v_g + S_2(t)\sigma_{21}\partial_y v_g)(t, S_1(t), S_2(t))$$
$$\Delta_2(t) = e^{-rt}(S_1(t)\sigma_{12}\partial_x v_g + S_2(t)\sigma_{22}\partial_y v_g)(t, S_1(t), S_2(t))$$

Letting $\Delta = (\Delta_1\ \Delta_2)^T$, we have $\Delta/e^{-rt} = \sigma^T Y$, where

$$Y = \begin{pmatrix} S_1(t)\partial_x v_g(t, S_1(t), S_2(t)) \\ S_2(t)\partial_y v_g(t, S_1(t), S_2(t)) \end{pmatrix}.$$

Hence by Theorem 6.23, the number of stock shares in the hedging portfolio is $h_{S_1}(t) = Y_1/S_1(t) = \partial_x v_g(t, S_1(t), S_2(t))$, $h_{S_2}(t) = Y_2/S_2(t) = \partial_y v_g(t, S_1(t), S_2(t))$, which concludes the proof of the theorem. □

An example of option on two stocks (outperformance option)

Let $K, T > 0$ and consider a standard European derivative with pay-off

$$Y = \left(\frac{S_1(T)}{S_2(T)} - K\right)_+$$

at time of maturity T. This is an example of outperformance option, i.e., an option that allows investors to benefit from the relative performance of two underslying assets. Using (6.139), we can write the risk-neutral price of the derivative as

$$\begin{aligned}\Pi_Y(t) &= e^{-r\tau}\widetilde{\mathbb{E}}\left[\left(\frac{S_1(T)}{S_2(T)} - K\right)_+ |\mathcal{F}_W(t)\right] \\ &= e^{-r\tau}\widetilde{\mathbb{E}}\left[\left(\frac{S_1(t)}{S_2(t)} e^{(\frac{|\sigma_2|^2}{2} - \frac{|\sigma_1|^2}{2})\tau + (\sigma_1 - \sigma_2)\cdot(\widetilde{W}(T) - \widetilde{W}(t))} - K\right)_+ |\mathcal{F}_W(t)\right].\end{aligned}$$

Now we write

$$(\sigma_1 - \sigma_2)\cdot(\widetilde{W}(T) - \widetilde{W}(t)) = \sqrt{\tau}[(\sigma_{11} - \sigma_{21})G_1 + (\sigma_{12} - \sigma_{22})G_2] = \sqrt{\tau}(X_1 + X_2),$$

where $G_j = (\widetilde{W}_j(T) - \widetilde{W}_j(t))/\sqrt{\tau} \in \mathcal{N}(0,1)$, $j = 1,2$, hence $X_j \in \mathcal{N}(0, (\sigma_{1j} - \sigma_{2j})^2)$, $j = 1,2$. In addition, X_1, X_2 are independent random variables, hence, as shown in Section 3.1, $X_1 + X_2$ is normally distributed with zero mean and variance $(\sigma_{11} - \sigma_{21})^2 + (\sigma_{12} - \sigma_{22})^2 = |\sigma_1 - \sigma_2|^2$. It follows that

$$\Pi_Y(t) = e^{-r\tau}\widetilde{\mathbb{E}}\left[\left(\frac{S_1(t)}{S_2(t)} e^{(\frac{|\sigma_2|^2}{2} - \frac{|\sigma_1|^2}{2})\tau + \sqrt{\tau}|\sigma_1 - \sigma_2|G)} - K\right)_+\right],$$

where $G \in \mathcal{N}(0,1)$. Hence, letting

$$\hat{r} = \frac{|\sigma_1 - \sigma_2|^2}{2} + \left(\frac{|\sigma_2|^2}{2} - \frac{|\sigma_1|^2}{2}\right)$$

and $a = e^{(\hat{r}-r)\tau}$, we have

$$\Pi_Y(t) = ae^{-\hat{r}\tau}\mathbb{E}\left[\left(\frac{S_1(t)}{S_2(t)}e^{(\hat{r}-\frac{|\sigma_1-\sigma_2|^2}{2})\tau+\sqrt{\tau}|\sigma_1-\sigma_2|G} - K\right)_+\right]$$

Up to the multiplicative parameter a, this is the Black-Scholes price of a call on a stock with price $S_1(t)/S_2(t)$, volatility $|\sigma_1 - \sigma_2|$ and for an interest rate of the money market given by $\hat{r}$. Hence, Theorem 6.6 gives

$$\Pi_Y(t) = a\left(\frac{S_1(t)}{S_2(t)}\Phi(d_+) - Ke^{-\hat{r}\tau}\Phi(d_-)\right) := v(t, S_1(t), S_2(t)) = u(t, \frac{S_1(t)}{S_2(t)}) \qquad (6.144)$$

where

$$d_\pm = \frac{\log\frac{S_1(t)}{KS_2(t)} + (\hat{r} \pm \frac{|\sigma_1-\sigma_2|^2}{2})\tau}{|\sigma_1 - \sigma_2|\sqrt{\tau}}.$$

As to the self-financing hedging portfolio, we have $h_{S_1}(t) = \partial_x v(t, S_1(t), S_2(t))$, $h_{S_2}(t) = \partial_y v(t, S_1(t), S_2(t))$, $j = 1, 2$. Therefore, recalling the delta function of the standard European call (see Theorem 6.7), we obtain

$$h_{S_1}(t) = \frac{a}{S_2(t)}\Phi(d_+), \quad h_{S_2}(t) = -\frac{aS_1(t)}{S_2(t)^2}\Phi(d_+).$$

The same result can be obtained by solving the terminal value problem (6.143). Indeed, the form of the pay-off function of the derivative suggests to look for solutions of (6.143) of the form $v_g(t, x, y) = u(t, x/y)$. The function $u(t, z)$ satisfies a standard Black-Scholes PDE in 1+1 dimension, whose solution is given as in (6.144). The details are left as an exercise.

Exercise 6.42. *Derive* (6.144) *by solving the terminal value problem* (6.143).

Exercise 6.43. *A two assets correlation call/put option with maturity T is the standard European derivative with pay-off*

$$Y_{\text{call}} = \begin{cases} (S_2(T) - K_2)_+, & \textit{if } S_1(T) > K_1 \\ 0 & \textit{otherwise} \end{cases}$$

$$Y_{\text{put}} = \begin{cases} (K_2 - S_2(T))_+, & \textit{if } S_1(T) < K_1 \\ 0 & \textit{otherwise} \end{cases}$$

where $K_1, K_2 > 0$ and $S_1(t), S_2(t)$ are the prices of the underlying stocks. Show that in a complete 2+1 dimensional Black-Scholes market the price of the call option is

$$\Pi_{\text{call}}(t) = S_2(t)\Phi(y_2 + \overline{\sigma}_2\sqrt{\tau}, y_1 + \rho\,\overline{\sigma}_2\sqrt{\tau}; \rho) - K_2 e^{-r\tau}\Phi(y_2, y_1; \rho),$$

where $\rho, \overline{\sigma}_i$ are given by (6.137)*, $\Phi(x, y; \rho)$ is the standard cumulative joint normal distribution, see* (2.12)*, and*

$$y_i = \frac{\log(S_i(t)/K_i) + (r - \overline{\sigma}_i^2/2)\tau)}{\overline{\sigma}_i\sqrt{\tau}}, \quad i = 1, 2.$$

Show that when the stock prices are independent, the price of the two-asset correlation call option is the product of the price of a binary option and of a standard call option. Finally derive the price of the corresponding put option and establish the put-call parity.

Exercise 6.44. *Derive the Black-Scholes price, the Black-Scholes PDE (with the terminal condition) and the put-call parity for the two-asset Asian geometric call/put option, whose pay-off, for the call, is*

$$Y_{\mathrm{call}} = \left(\exp\left(\frac{1}{T} \int_0^T \log(S_1(t)^{1/2} S_2(t)^{1/2})\, dt \right) - K \right)_+ .$$

6.10 Introduction to American derivatives

Before giving the precise definition of fair price for American derivatives, we shall present some general properties of these contracts. American derivatives can be exercised at any time prior or including maturity T. Let $Y(t)$ be the pay-off resulting from exercising the derivative at time $t \in (0, T]$. We call $Y(t)$ the **intrinsic value** of the derivative. We consider only standard American derivatives, for which we have $Y(t) = g(S(t))$, for some measurable function $g : \mathbb{R} \to \mathbb{R}$. For instance, $g(x) = (x - K)_+$ for American calls and $g(x) = (K - x)_+$ for American puts. We denote by $\widehat{\Pi}_Y(t)$ the rsik-neutral price of the American derivative with intrinsic value $Y(t)$ and by $\Pi_Y(t)$ the risk-neutral price of the European derivative with pay-off $Y = Y(T)$ at maturity time T (given by (6.10)). Even if we do not know yet how $\widehat{\pi}_Y(t)$ is defined, two obvious properties of American derivatives are the following:

(i) $\widehat{\Pi}_Y(t) \geq \Pi_Y(t)$, for all $t \in [0, T]$. In fact an American derivative gives to its owner all the rights of the corresponding European derivative plus one: the option of early exercise. Thus it is clear that the American derivative cannot be cheaper than the European one.

(ii) $\widehat{\Pi}_Y(t) \geq Y(t)$, for all $t \in [0, T]$. If not, an arbitrage opportunity would arise by purchasing the American derivative and exercising it immediately.

Any reasonable definition of fair price for American derivatives must satisfy (i)-(ii).

Definition 6.9. *A time $t \in (0, T]$ is said to be an* **optimal exercise time** *for the American derivative with intrinsic value $Y(t)$ if $\widehat{\Pi}_Y(t) = Y(t)$.*

Hence by exercising the derivative at an optimal exercise time t, the buyer takes full advantage of the derivative: the resulting pay-off equals the value of the derivative. On the other hand, if $\widehat{\Pi}_Y(t) > Y(t)$ and the buyer wants to close the (long) position on the American derivative, then the optimal strategy is to sell the derivative, thereby cashing the amount $\widehat{\Pi}_Y(t)$.

Theorem 6.27. *Assume (i) holds and let $\widehat{C}(t)$ be the price of an American call at time $t \in [0,T]$. Assume further that the underlying stock price follows a generalized geometric Brownian motion and that the interest rate $r(t)$ of the money market is strictly positive for all times. Then $\widehat{C}(t) > Y(t)$, for all $t \in [0,T)$. In particular it is never optimal to exercise the call prior to maturity.*

Proof. For $S(t) \leq K$ the claim becomes $\widehat{C}(t) > 0$ for $t \in [0,T)$, which is obvious (since $\widehat{C}(t) \geq C(t) > 0$). For $S(t) > K$ we write

$$\begin{aligned}\widehat{C}(t) \geq C(t) &= \widetilde{\mathbb{E}}[(S(T)-K)_+ D(T)/D(t)|\mathcal{F}_W(t)] \geq \widetilde{\mathbb{E}}[(S(T)-K)D(T)/D(t)|\mathcal{F}_W(t)] \\ &= \widetilde{\mathbb{E}}[S(T)D(T)/D(t)|\mathcal{F}_W(t)] - K\widetilde{\mathbb{E}}[D(T)/D(t)|\mathcal{F}_W(t)] > D(t)^{-1}\widetilde{\mathbb{E}}[S^*(T)|\mathcal{F}_W(t)] - K \\ &= S(t) - K = (S(t)-K)_+,\end{aligned}$$

where we used $D(T)/D(t) < 1$ (by the positivity of the interest rate $r(t)$) and the martingale property of the discounted price $\{S^*(t)\}_{t\in[0,T]}$ of the stock. $\square$

It follows that under the assumptions of the previous theorem the earlier exercise option of the American call is worthless, hence *American and European call options with the same strike and maturity have the same value.*

Exercise 6.45. *Generalize the previous theorem standard American derivatives with convex pay-off function.*

Remark 6.14. A notable exception to the assumed conditions in Theorem 6.27 is when the underlying stock pays a dividend. In this case it can be shown that it is optimal to exercise the American call immediately before the dividend is payed, provided the price of the stock is sufficiently high, see Theorem 6.32 below.

Definition 6.10. *Let $T \in (0,\infty)$. A random variable $\tau : \Omega \to [0,T]$ is called a* **stopping time** *for the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$ if $\{\tau \leq t\} \in \mathcal{F}_W(t)$, for all $t \in [0,T]$. We denote by Q_T the set of all stopping times for the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$.*

Think of τ as the time at which some random event takes place. Then τ is a stopping time if the occurrence of the event before or at time t can be inferred by the information available up to time t (no future information is required). For the applications that we have in mind, τ will be the optimal exercise time of an American derivative, which marks the event that the price of the derivative equals its intrinsic value.

From now on we assume that the market has constant parameters and $r > 0$. Hence the price of the stock is given by the geometric Brownian motion

$$S(t) = S(0)e^{(r-\frac{\sigma^2}{2})t+\sigma\widetilde{W}(t)}.$$

We recall that in this case the price $\Pi_Y(0,T)$ at time $t=0$ of the European derivative with pay-off $Y=g(S(T))$ at maturity time $T>0$ is given by

$$\Pi_Y(0,T)=\widetilde{\mathbb{E}}[e^{-rT}g(S(T))].$$

Now, if the writer of the American derivative were sure that the buyer would exercise at the time $u\in(0,T]$, then the fair price of the American derivative at time $t=0$ would be equal to $\Pi_Y(0,u)$. As the writer cannot anticipate when the buyer will exercise, we would be tempted to define the price of the American derivative at time zero as $\max\{\Pi_Y(0,u), 0\leq u\leq T\}$. However this definition would actually be unfair, as it does not take into account the fact that the exercise time is a stopping time, i.e., it is random and it cannot be inferred using future information. This leads us to the following definition.

Definition 6.11. *In a market with constant parameters, the risk-neutral (or fair) price at time $t=0$ of the standard American derivative with intrinsic value $Y(t)=g(S(t))$ and maturity $T>0$ is given by*

$$\widehat{\Pi}_Y(0)=\max_{\tau\in Q_T}\widetilde{\mathbb{E}}[e^{-r\tau}g(S(\tau))], \tag{6.145}$$

where $S(\tau)=S(0)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\widetilde{W}(\tau)}$.

It is not possible in general to find an closed formula for the risk-neutral price of American derivatives. A notable exception is the price of perpetual American put options and of binary American put options, which we discuss next.

Perpetual American put options

An American put option is called perpetual if it never expires, i.e., $T=\infty$. This is of course an idealization, but perpetual American puts are very useful to visualize the structure of general American put options. In this section we follow closely the discussion on [26, Section 8.3]. Definition 6.145 becomes the following.

Definition 6.12. *Let Q be the set of all stopping times for the filtration $\{\mathcal{F}_W(t)\}_{t\geq 0}$, i.e., $\tau\in Q$ iff $\tau:\Omega\to[0,\infty]$ is a random variable and $\{\tau\leq t\}\in\mathcal{F}_W(t)$, for all $t\geq 0$. The risk-neutral price at time $t=0$ of the perpetual American put with strike K is*

$$\widehat{\Pi}(0)=\max_{\tau\in Q}\widetilde{\mathbb{E}}[e^{-r\tau}(K-S(\tau))_+] \tag{6.146}$$

where $S(\tau)=S(0)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\widetilde{W}(\tau)}$.

Theorem 6.28. *There holds*

$$\widehat{\Pi}(0)=v_L(S(0)), \tag{6.147}$$

where

$$v_L(x) = \begin{cases} K - x & 0 \le x \le L \\ (K-L)\left(\frac{x}{L}\right)^{-\frac{2r}{\sigma^2}} & x > L \end{cases}$$

and

$$L = \frac{2r}{2r+\sigma^2}K.$$

Before we prove the theorem, some remarks are in order:

(i) $L < K$;

(ii) For $S(0) \le L$ we have $\widehat{\Pi}(0) = v_L(S(0)) = K - S(0) = (K - S(0))_+$. Hence when $S(0) \le L$ it is optimal to exercise the derivative.

(iii) We have $\widehat{\Pi}(0) > (K - S(0))_+$ for $S(0) > L$. In fact

$$v_L'(x) = -\frac{2r}{\sigma^2}\left(\frac{x}{L}\right)^{-\frac{2r}{\sigma^2}-1}\frac{K-L}{L},$$

hence $v_L'(L) = -1$. Moreover

$$v_L''(x) = \frac{2r}{\sigma^2}(\frac{2r}{\sigma^2}+1)\left(\frac{x}{L}\right)^{-\frac{2r}{\sigma^2}-2}\frac{K-L}{L^2},$$

which is always positive. Thus the graph of $v_L(x)$ always lies above $K - x$ for $x > L$. It follows that it is not optimal to exercise the derivative if $S(0) > L$.

(iv) In the perpetual case, any time is equivalent to $t = 0$, as the time left to maturity is always infinite. Hence

$$\widehat{\Pi}(t) = v_L(S(t)).$$

In conclusion the theorem is saying us that the buyer of the derivative should exercise as soon as the stock price falls below the threshold L. In fact we can reformulate the theorem in the following terms:

Theorem 6.29. *The maximum of* $\widetilde{\mathbb{E}}[e^{-r\tau}(K - S(\tau))_+]$ *over all possible* $\tau \in Q$ *is achieved at* $\tau = \tau_*$, *where*

$$\tau_* = \min\{t \ge 0 : S(t) = L\}.$$

Moreover $\widetilde{\mathbb{E}}[e^{-r\tau_*}(K - S(\tau_*))_+] = v_L(S(0))$.

For the proof of Theorem 6.28 we need the optional sampling theorem:

Theorem 6.30. *Let* $\{X(t)\}_{t\ge0}$ *be an adapted process and* τ *a stopping time. Let* $t \wedge \tau = \min(t,\tau)$. *If* $\{X(t)\}_{t\ge0}$ *is a martingale/supermatingale/submartingale, then* $\{X(t\wedge\tau)\}_{t\ge0}$ *is also a martingale/supermartingale/submartingale.*

We can now prove Theorem 6.28. We divide the proof in two steps, which correspond respectively to Theorem 8.3.5 and Corollary 8.3.6 in [26].

Step 1: *The stochastic process* $\{e^{-r(t\wedge\tau)}v_L(S(t\wedge\tau))\}_{t\geq 0}$ *is a super-martingale for all* $\tau\in Q$. *Moreover for* $S(0)>L$ *the stochastic process* $\{e^{-r(t\wedge\tau_*)}v_L(S(t\wedge\tau_*))\}_{t\geq 0}$ *is a martingale.* By Itô's formula,

$$d(e^{-rt}v_L(S(t))) = e^{-rt}[-rv_L(S(t)) + rS(t)v_L'(S(t)) + \frac{1}{2}\sigma^2 S(t)^2 v_L''(S(t))]dt \\ + e^{-rt}\sigma S(t)v_L'(S(t))d\widetilde{W}(t).$$

The drift term is zero for $S(t)>L$ and it is equal to $-rK\,dt$ for $S(t)\leq L$. Hence

$$e^{-rt}v_L(S(t)) = v_L(S(0)) - rK\int_0^t e^{-ru}\mathbb{I}_{S(u)\leq L}(u)\,du + \int_0^t e^{-ru}\sigma S(u)v_L'(S(u))d\widetilde{W}(u).$$

Since the drift term is non-positive, then $\{e^{-rt}v_L(t)\}_{t\geq 0}$ is a supermartingale and thus by the optional sampling theorem, the process $\{e^{-r(t\wedge\tau)}v_L(S(t\wedge\tau)\}_{t\geq 0}$ is a also a supermartingale, for all $\tau\in Q$. Now, if $S(0)>L$, then, by continuity of the paths of the geometric Brownian motion, $S(u,\omega)>L$ as long as $u<\tau_*(\omega)$. Hence by stopping the process at τ_* the stock price will never fall below L and therefore the drift term vanishes, that is

$$e^{-r(t\wedge\tau_*)}v_L(S(t\wedge\tau_*)) = v_L(S(0)) + \int_0^{t\wedge\tau_*} e^{-ru}\sigma S(u)v_L'(S(u))d\widetilde{W}(u).$$

The Itô integral is a martingale and thus the Itô integral stopped at time τ_* is also a martingale by the optional sampling theorem. The claim follows.

Step 2: *The identity* (6.147) *holds.* The supermartingale property of the process $\{e^{-r(t\wedge\tau)}v_L(S(t\wedge\tau))\}_{t\geq 0}$ implies that its expectation is non-increasing, hence

$$\widetilde{\mathbb{E}}[e^{-r(t\wedge\tau)}v_L(S(t\wedge\tau))] \leq v_{L*}(S(0)).$$

As $v_L(x)$ is bounded and continuous, the limit $t\to+\infty$ gives

$$\widetilde{\mathbb{E}}[e^{-r\tau}v_L(S(\tau))] \leq v_L(S(0)).$$

As $v_L(x)\geq (K-x)_+$ we also have

$$\widetilde{\mathbb{E}}[e^{-r\tau}(K-S(\tau))_+] \leq v_L(S(0)).$$

Taking the maximum over all $\tau\in Q$ we obtain

$$\widehat{\Pi}(0) = \max_{\tau\in Q}\widetilde{\mathbb{E}}[e^{-r\tau}(K-S(\tau))_+] \leq v_L(S(0)).$$

Now we prove the reverse inequality $\widehat{\Pi}(0) \geq v_L(S(0))$. This is obvious for $S(0) \leq L$. In fact, letting $\widetilde{\tau} = \min\{t \geq 0 : S(t) \leq L\}$, we have $\widetilde{\tau} \equiv 0$ for $S(0) \leq L$ and so $\max_{\tau \in Q} \widetilde{\mathbb{E}}[e^{-r\tau}(K - S(\tau))_+] \geq \widetilde{\mathbb{E}}[e^{-r\widetilde{\tau}}(K - S(\widetilde{\tau}))_+] = (K - S(0))_+ = v_L(S(0))$, for $S(0) \leq L$. For $S(0) > L$ we use the martingale property of the stochastic process $\{e^{-r(t\wedge\tau_*)}v_L(S(t \wedge \tau_*))\}_{t\geq 0}$, which implies

$$\widetilde{\mathbb{E}}[e^{-r(t\wedge\tau_*)}v_L(S(t \wedge \tau_*))] = v_L(S(0)).$$

Hence in the limit $t \to +\infty$ we obtain

$$v_L(S(0)) = \widetilde{\mathbb{E}}[e^{-r\tau_*}v_L(S(\tau_*))].$$

Moreover $e^{-r\tau_*}v_L(S(\tau_*)) = e^{-r\tau_*}v_L(L) = e^{-r\tau_*}(K - S(\tau_*))_+$, hence

$$v_L(S(0)) = \widetilde{\mathbb{E}}[e^{-r\tau_*}(K - S(\tau_*))_+].$$

It follows that

$$\widehat{\Pi}(0) = \max_{\tau \in Q} \widetilde{\mathbb{E}}[e^{-r\tau}(K - S(\tau))_+] \geq v_L(S(0)),$$

which completes the proof. □

Next we discuss the problem of hedging the perpetual American put with a portfolio invested in the underlying stock and the risk-free asset.

Definition 6.13. *A portfolio process $\{h_S(t), h_B(t)\}_{t\geq 0}$ is said to be replicating the perpetual American put if its value $\{V(t)\}_{t\geq 0}$ equals $\widehat{\Pi}(t)$ for all $t \geq 0$.*

Thus by setting-up a replicating portfolio, the writer of the perpetual American put is sure to always be able to afford to pay-off the buyer. Note that in the European case a self-financing hedging portfolio is trivially replicating, as the price of European derivatives has been defined as the value of such portfolios. However in the American case a replicating portfolio need not be self-financing: if the buyer does not exercise at an optimal exercise time, the writer must withdraw cash from the portfolio in order to replicate the derivative. This leads to the definition of portfolio generating a cash flow.

Definition 6.14. *A portfolio $\{h_S(t), h_B(t)\}_{t\geq 0}$ with value $\{V(t)\}_{t\geq 0}$ is said to generate a* **cash flow** *with rate $c(t)$ if $\{c(t)\}_{t\geq 0}$ is adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$ and*

$$dV(t) = h_S(t)dS(t) + h_B(t)dB(t) - c(t)dt \tag{6.148}$$

Remark 6.15. Note that the cash flow has been defined so that $c(t) > 0$ when the investor withdraws cash from the portfolio (causing a decrease of its value).

The following theorem is Corollary 8.3.7 in [26].

Theorem 6.31. *The portfolio given by*

$$h_S(t) = v_L'(S(t)), \quad h_B(t) = \frac{v_L(S(t)) - h_S(t)S(t)}{B(0)e^{rt}}$$

is replicating the perpetual American put while generating the case flow $c(t) = rK\mathbb{I}_{S(t)\leq L}$ (i.e., cash is withdrawn at the rate rK whenever $S(t) \leq L$, provided of course the buyer does not exercise the derivative).

Proof. By definition, $V(t) = h_S(t)S(t) + h_B(t)B(t) = v_L(S(t)) = \widehat{\Pi}(t)$, hence the portfolio is replicating. Moreover

$$dV(t) = d(v_L(S(t))) = h_S(t)dS(t) + \frac{1}{2}v_L''(S(t))\sigma^2 S(t)^2 dt. \tag{6.149}$$

Now, a straightforward calculation shows that $v_L(x)$ satisfies

$$-rv_L + rxv_L' + \frac{1}{2}\sigma x^2 v_L'' = -rK\mathbb{I}_{x\leq L},$$

a relation which was already used in step 1 in the proof of Theorem 6.28. It follows that

$$\begin{aligned}\frac{1}{2}v_L''(S(t))\sigma^2 S(t)^2 dt &= r(v_L(S(t)) - S(t)h_S(t))dt - rK\mathbb{I}_{S(t)\leq L}dt\\ &= h_B(t)dB(t) - rK\mathbb{I}_{S(t)\leq L}dt.\end{aligned}$$

Hence (6.149) reduces to (6.148) with $c(t) = rK\mathbb{I}_{S(t)\leq L}$, and the proof is complete. □

Remarks on American put options with finite maturity

The pricing function $v_L(x)$ of perpetual American puts satisfies

$$-rv_L + rxv_L' + \frac{1}{2}\sigma^2 x^2 v_L'' = 0 \quad \text{when } x > L, \tag{6.150}$$

$$v_L(x) = (K - x), \quad \text{for } x \leq L, \quad v_L'(L) = -1. \tag{6.151}$$

It can be shown that the pricing function of American put options with finite maturity satisfies a similar problem. Namely, letting $\widehat{P}(t)$ be the fair price at time t of the American put with strike K and maturity $T > t$, it can be shown that $\widehat{P}(t) = v(t, S(t))$, where $v(t,x)$ satisfies

$$\partial_t v + rx\partial_x v + \frac{1}{2}\sigma^2 x^2 \partial_x^2 v = rv, \quad \text{if } x > x_*(t), \tag{6.152}$$

$$v(t,x) = (K - x), \quad \text{for } x \leq x_*(t), \quad \partial_x v(t, x_*(t)) = -1, \tag{6.153}$$

$$v(T,x) = (k - x)_+, \quad x_*(T) = K, \tag{6.154}$$

which is a **free-boundary value problem**. While a numerical solution of the previous problem can be found using the finite difference method, the price of the American put option is most commonly computed using binomial tree-approximations, see for instance [5].

American calls on a dividend-paying stock

Let $\widehat{C}_a(t, S(t), K, T)$ denote the Black-Scholes price at time t of the American call with strike K and maturity T assuming that the underlying stock pays the dividend $aS(t_0^-)$ at time $t_0 \in (0, T)$. We denote by $C_a(t, S(t), K, T)$ the Black-Scholes price of the corresponding European call. We omit the subscript a to denote prices in the absence of dividends. Moreover replacing the letter C with the letter P gives the price of the corresponding put option. We say that it is optimal to exercise the American call at time t if its Black-Scholes price at this time equals the intrinsic value of the call, i.e., $\widehat{C}_a(t, S(t), K, T) = (S(t) - K)_+$.

Theorem 6.32. *Consider the American call with strike K and expiration date T and assume that the underlying stock pays the dividend $aS(t_0^-)$ at the time $t_0 \in (0, T)$. Then*

$$\widehat{C}_a(t, S(t), K, T) > (S(t) - K)_+, \quad \textit{for } t \in [t_0, T),$$

i.e., it is not optimal to exercise the American call prior to maturity after the dividend is paid. Moreover, there exists $\delta > 0$ such that, if

$$S(t_0^-) > \max(\frac{\delta}{1-a}, K),$$

then the equality

$$\widehat{C}_a(t_0^-, S(t_0^-), K, T) = (S(t_0^-) - K)_+$$

holds, and so it is optimal to exercise the American call "just before" the dividend is to be paid.

Proof. For the first claim we can assume $(S(t) - K)_+ = S(t) - K$, otherwise the American call is out of the money and so it is clearly not optimal to exercise. By Theorem 6.8 we have

$$C_a(t, S(t), K, T) = C(t, S(t), K, T), \quad P_a(t, S(t), K, T) = P(t, S(t), K, T), \quad \text{for } t \geq t_0.$$

Hence, by Theorem 6.6, the put-call parity holds after the dividend is paid:

$$C_a(t, S(t), K, T) = P_a(t, S(t), K, T) + S(t) - Ke^{-r(T-t)}, \quad t \geq t_0.$$

Thus, for $t \in [t_0, T)$,

$$\widehat{C}_a(t, S(t), K, T) \geq C_a(t, S(t), K, T) > S(t) - K = (S(t) - K)_+,$$

where we used that $P(t, S(t), K, T) > 0$ and $r \geq 0$. This proves the first part of the theorem, i.e., the fact that it is not optimal to exercise the American call prior to expiration after the dividend has been paid. In particular

$$\widehat{C}_a(t, S(t), K, T) = C_a(t, S(t), K, T), \quad \text{for } t \geq t_0. \tag{6.155}$$

Next we show that it is optimal to exercise the American call "just before the dividend is paid", i.e., $\widehat{C}_a(t_0^-, S(t_0^-), K, T) = (S(t_0^-) - K)_+$, provided the price of the stock is sufficiently high. Of course it must be $S(t_0^-) > K$. Assume first that $\widehat{C}_a(t_0^-, S(t_0^-), K, T) > S(t_0^-) - K$; then, owing to (6.155), $\widehat{C}_a(t_0^-, S(t_0^-), K, T) = C_a(t_0^-, S(t_0^-), K, T)$ (buying the American call just before the dividend is paid is not better than buying the European call, since it is never optimal to exercise the derivative prior to expiration). By Theorem 6.8 we have $C_a(t_0^-, S(t_0^-), K, T) = C(t_0^-, (1-a)S(t_0^-), K, T) = C(t_0, (1-a)S(t_0^-), K, T)$, where for the latter equality we used the continuity in time of the Black-Scholes price function in the absence of dividends. Since $(1-a)S(t_0^-) = S(t_0)$, then

$$\widehat{C}_a(t_0^-, S(t_0^-), K, T) > S(t_0^-) - K \Rightarrow \widehat{C}_a(t_0^-, S(t_0^-), K, T) = C(t_0, S(t_0), K, T).$$

Hence

$$\widehat{C}_a(t_0^-, S(t_0^-), K, T) > S(t_0^-) - K \Rightarrow C(t_0, S(t_0), K, T) > S(t_0^-) - K = S(t_0) + (1-a)S(t_0^-) - K.$$

Therefore, taking the contrapositive statement,

$$C(t_0, S(t_0), K, T) \leq S(t_0) + (1-a)S(t_0^-) - K \Rightarrow \widehat{C}_a(t_0^-, S(t_0^-), K, T) = S(t_0^-) - K. \quad (6.156)$$

Next we remark that the function $x \to c(t, x, K, T) - x$ is decreasing (since $\Delta = \partial_x c = \Phi(d_1) < 1$, see Theorem 6.7), and

$$\lim_{x \to 0^+} C(t, x, K, T) - x = 0,$$

$$\lim_{x \to +\infty} C(t, x, K, T) - x = \lim_{x \to +\infty} P(t, x, K, T) - Ke^{-r(T-t)} = -Ke^{-r(T-t)},$$

see Exercise 6.12. Thus if $(1-a)S(t_0^-) - K > -Ke^{-r(T-t)}$, i.e., $S(t_0^-) > (1-a)^{-1}K(1 - e^{-r(T-t)}$, there exists ω such that if $S(t_0) > \omega$, i.e., $S(t_0^-) > \omega/(1-a)$, then the inequality $C(t_0, S(t_0), K, T) \leq S(t_0) + (1-a)S(t_0^-) - K$ holds. It follows by (6.156) that for such values of $S(t_0^-)$ it is optimal to exercise the call "at time t_0^-". Letting $\delta = \max(\omega, K(1 - e^{-r(T-t)})$ concludes the proof of the theorem. $\square$

Appendix A

Numerical projects

A.1 A project on the Asian option

The Asian call option[1] with strike $K > 0$ and maturity $T > 0$ is the non-standard European derivative with pay-off $Y = \left(\frac{1}{T}\int_0^T S(t)\,dt - K\right)_+$ and similarly one defines the Asian put option. The Black-Scholes price of the Asian option can be computed numerically either by the Monte Carlo method or by solving the boundary value problem (6.42) using the finite difference method. The main purpose of this project is to compare the performance of the two methods. For the finite difference approach it is convenient to invert the time direction in the problem (6.42) by changing variable $t \to T - t$, thereby obtaining the system

$$-\partial_t u + \frac{\sigma^2}{2}(\gamma(t) - z)^2 \partial_z^2 u = 0, \quad t \in (0, T), z \in \mathbb{R} \tag{A.1a}$$

$$u(0, z) = z^+, \quad \lim_{z\to-\infty} u(t, z) = 0, \ \lim_{z\to\infty}(u(t, z) - z) = 0, \quad t \in (0, T], \tag{A.1b}$$

where $\gamma(t) = \frac{1-e^{-rt}}{rT}$.

Part 1

- Write a short introduction on the Asian call/put option, where you should discuss in particular its financial utility and main differences with the standard European call/put (you can find plenty of info on the web). Outline the content of the rest of the report.
- Solve Exercise 6.18.
- Write a finite difference scheme that solves the PDE (A.1) in the domain $(t, z) \in (0, T) \times (-Z, Z)$ with the appropriate initial and boundary conditions for Asian calls or puts. Use the Crank-Nicolson method (see Remark 5.11).

[1] In this project "Asian option" always means "Asian option with arithmetic average".

Part 2

- Write a Matlab function that implements the finite difference scheme derived in Part 1. The parameters S_0, r, σ, K, T, must appear as input variables of your function.
- Plot the initial price of the Asian call/put as a function of the volatility σ and of the initial price S_0. Discuss your findings, in particular the relation with the behavior of standard call/put options. Verify numerically the validity of the put-call parity.
- Compare the price obtained by the finite difference method and by the control variate Monte Carlo method for different value of σ and compare the efficiency of the two methods, e.g., by performing speed tests. Present your results using tables and discuss them.

Include your matlab codes in an appendix with a description of how they work (e.g., as comments within the codes themselves).

A.2 A project on the CEV model

In the CEV (Constant Elasticity Variance) model, the price of the stock option with maturity $T > 0$ and pay-off $Y = g(S(T))$ is given by $\Pi_Y(t) = e^{-r(T-t)}u(t, S(t))$, where $S(t)$ is the stock price at time t and u solves

$$\partial_t u + rx\partial_x u + \frac{\sigma^2}{2}x^{2\delta}\partial_x^2 u = 0,\ x > 0,\ t \in (0, T), \tag{A.2}$$

with the terminal condition $u(T, x) = g(x)$, $x > 0$. Here $r > 0$, $\sigma > 0$, $\delta > 0$ are constants; see Section 6.6 for a derivation of the model. Actually it is more convenient to work with the equivalent problem

$$-\partial_t u + rx\partial_x u + \frac{\sigma^2}{2}x^{2\delta}\partial_x^2 u = 0,\ x > 0,\ t \in (0, T), \tag{A.3a}$$

$$u(0, x) = g(x), \quad x > 0 \tag{A.3b}$$

which is obtained by the change of variable $t \to T - t$ in (A.2).

The main purpose of this project is to derive numerically some qualitative properties of the CEV model, such as the implied volatility curve.

Part 1

- Write a short introduction on the CEV model, where you should discuss in particular the applications of the CEV model (you can find plenty of info on the web) and the main differences with the Black-Scholes model. Outline the content of the rest of the report.

- Solve Exercise 6.27.
- Write a finite difference scheme for the PDE (A.3a) on the domain $(t, x) \in (0, T) \times (0, X)$ with initial data and boundary conditions corresponding to European call or put options. Use the Crank-Nicolson method (see Remark 5.11).

Part 2

- Write a Matlab function that implements the finite difference scheme derived in Part 1 in the case of call/put options. The parameters $S_0, r, K, \sigma, \delta, X, T$, must appear as input variables of your function.
- Compare the finite difference solution for $\delta = 1$ with the exact Black-Scholes solution and discuss possible sources of error and how to eliminate them.
- Plot the initial price of call and put options as a function of the initial stock price and of the volatility parameter γ. Verify numerically the validity of the put-call parity. Highlight the main differences between the Black-Scholes price and the CEV model price of call/put options.
- Compare the price obtained by the finite difference method and the Monte Carlo method for different values of γ and compare the efficiency of the two methods, e.g., by performing speed tests. Present your results using tables and discuss them. Remark: Apply the Euler-Maruyama method to generate paths of the stock price.

Include your matlab codes in an appendix with a description of how they work (e.g., as comments within the codes themselves).

Appendix B

Solutions to selected Exercises

1. Solution to Exercise 1.3. Since an event belongs to the intersection of σ-algebras if and only if it belongs to each single σ-algebra, the proof of the first statement is trivial. As an example of two σ-algebras whose union is not a σ algebra, take $\{\emptyset, \{1\}, \{2,3,4,5,6\}, \Omega\}$, $\{\emptyset, \{2\}, \{1,3,4,5,6\}, \Omega\}$ on the sample space $\Omega = \{1,2,3,4,5,6\}$.

2. Solution to Exercise 1.6. Since A and A^c are disjoint, we have

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) \Rightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

To prove 2 we notice that $A \cup B$ is the disjoint union of the sets $A \setminus B$, $B \setminus A$ and $A \cap B$. It follows that

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

Since A is the disjoint union of $A \cap B$ and $A \setminus B$, we also have

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

and similarly

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \setminus A). \tag{B.1}$$

Combining the three identities above yields the result. Moreover, from (B.1) and assuming $A \subset B$, we obtain $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) > \mathbb{P}(A)$, which is claim 3.

3. Solution to Exercise 1.8. Since for all $k = 0, \dots, N$ the number of N-tosses $\omega \in \Omega_N$ having $N_H(\omega) = k$ is given by the binomial coefficient

$$\binom{N}{k} = \frac{N!}{k!(N-k)!},$$

then

$$\sum_{\omega \in \Omega_N} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega_N} p^{N_H(\omega)}(1-p)^{N_T(\omega)} = \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k}.$$

By the binomial theorem, $(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}$, for all $a, b > 0$, hence

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega_N} \mathbb{P}(\{\omega\}) = (p + 1 - p)^N = 1.$$

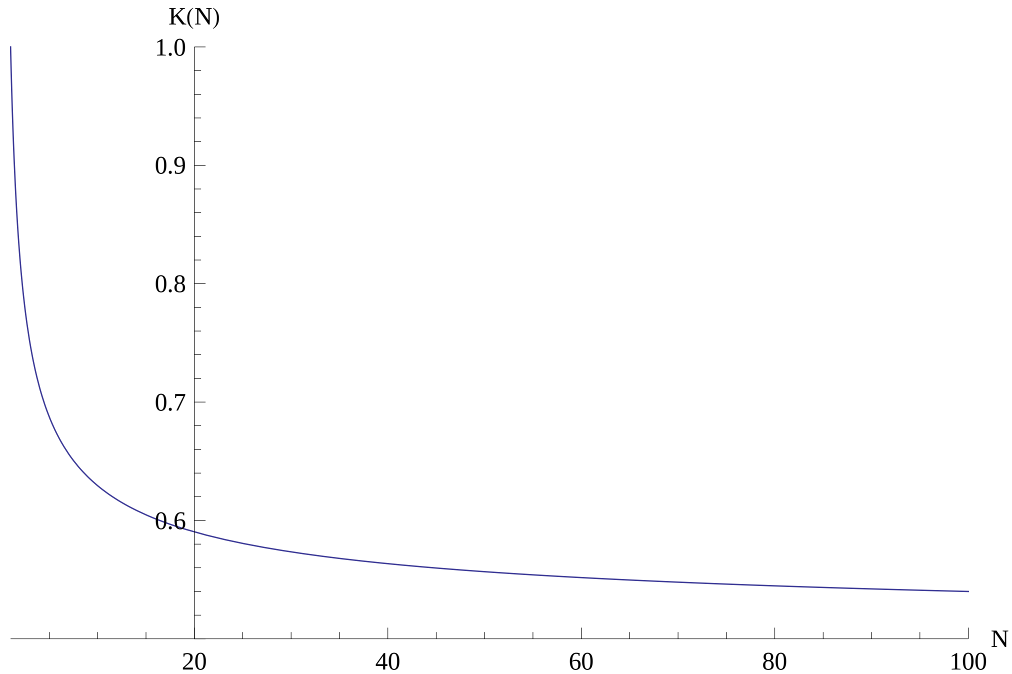


Figure B.1: A numerical solution of Exercise 1.9 for a generic odd natural number N.

4. Solution to Exercise 1.9. We expect that $\mathbb{P}(A|B) > \mathbb{P}(A)$, that is to say, the first toss being a head increases the probability that the number of heads in the complete N-toss will be larger than the number of tails. To verify this, we first observe that $\mathbb{P}(A) = 1/2$, since N is odd and thus there will be either more heads or more tails in any N-toss. Moreover, $\mathbb{P}(A|B) = \mathbb{P}(C)$, where $C \in \Omega_{N-1}$ is the event that the number of heads in a $(N-1)$-toss is larger or equal to the number of tails. Letting k be the number of heads, $\mathbb{P}(C)$ is the probability that $k \in \{(N-1)/2, \dots, N-1\}$. Since there are $\binom{N-1}{k}$ possible $(N-1)$-tosses with k-heads, then

$$\mathbb{P}(C) = \sum_{k=(N-1)/2}^{N-1} \binom{N-1}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{N-1-k} = \frac{1}{2^{N-1}} \sum_{k=(N-1)/2}^{N-1} \binom{N-1}{k}.$$

Thus proving the statement for a generic odd N is equivalent to prove the inequality

$$K(N) = \frac{1}{2^{N-1}} \sum_{k=(N-1)/2}^{N-1} \binom{N-1}{k} > \frac{1}{2}.$$

A "numerical proof" of this inequality is provided in Figure B.1. Note that the function $K(N)$ is decreasing and converges to $1/2$ as $N \to \infty$.

5. Solution to Exercise 2.1. Since $\Omega = \{X \in \mathbb{R}\}$, then $\Omega \in \sigma(X)$. $\{X \in U\}^c$ is the set of sample points $\omega \in \Omega$ such that $X(\omega) \notin U$. The latter is equivalent to $X(\omega) \in U^c$, hence $\{X \in U\}^c = \{X \in U^c\}$. Since $U^c \in \mathcal{B}(\mathbb{R})$, it follows that $\{X \in U\} \in \sigma(X)$. Finally we have to prove that $\sigma(X)$ is closed with respect to the countable union of sets. Let $\{A_k\}_{k\in\mathbb{N}} \subset \sigma(X)$. By definition of $\sigma(X)$, there exist $\{U_k\}_{k\in\mathbb{N}} \subset \mathcal{B}(\mathbb{R})$ such that $A_k = \{X \in U_k\}$. Thus we have the following chain of equivalent statements

$$\omega \in \cup_{k\in\mathbb{N}} A_k \Leftrightarrow \exists \bar{k} \in \mathbb{N} : X(\omega) \in U_{\bar{k}} \Leftrightarrow X(\omega) \in \cup_{k\in\mathbb{N}} U_k \Leftrightarrow \omega \in \{X \in \cup_{k\in\mathbb{N}} U_k\}.$$

Hence $\cup_{k\in\mathbb{N}} A_k = \{X \in \cup_{k\in\mathbb{N}} U_k\}$. Since $\cup_{k\in\mathbb{N}} U_k \in \mathcal{B}(\mathbb{R})$, then $\cup_{k\in\mathbb{N}} A_k \in \sigma(X)$.

6. Solution to Exercise 2.2. Let A be an event that is resolved by both variables X, Y. This means that there exist $U, V \in \mathcal{B}(\mathbb{R})$ such that $A = \{X \in U\} = \{Y \in V\}$. Hence, using the independence of X, Y,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(X \in U, Y \in V) = \mathbb{P}(X \in U)\mathbb{P}(Y \in V) = \mathbb{P}(A)\mathbb{P}(A) = \mathbb{P}(A)^2.$$

Therefore $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Now let a, b be two deterministic constants. Note that, for all $U \subset \mathbb{R}$,

$$\mathbb{P}(a \in U) = \begin{cases} 1 & \text{if } a \in U \\ 0 & \text{otherwise} \end{cases}$$

and similarly for b. Hence

$$\mathbb{P}(a \in U, b \in V) = \begin{cases} 1 & \text{if } a \in U \text{ and } b \in V \\ 0 & \text{otherwise} \end{cases} = \mathbb{P}(a \in U)\mathbb{P}(b \in V).$$

Finally we show that X and $Y = g(X)$ are independent if and only if Y is a deterministic constant. For the "if" part we use that

$$\mathbb{P}(a \in U, X \in V) = \begin{cases} \mathbb{P}(X \in V) & \text{if } a \in U \\ 0 & \text{otherwise} \end{cases} = \mathbb{P}(a \in U)\mathbb{P}(X \in V).$$

For the "only if" part, let $z \in \mathbb{R}$ and $U = \{g(X) \le z\} = \{X \in g^{-1}(-\infty, z]\}$. Then, using the independence of X and $Y = g(X)$,

$$\begin{aligned}\mathbb{P}(g(X) \le z) &= \mathbb{P}(g(X) \le z, g(X) \le z) = \mathbb{P}(X \in g^{-1}(-\infty, z], g(X) \le z) \\ &= \mathbb{P}(X \in g^{-1}(-\infty, z])\mathbb{P}(g(X) \le z) = \mathbb{P}(g(X) \le z)\mathbb{P}(g(X) \le z).\end{aligned}$$

Hence $\mathbb{P}(Y \le z)$ is either 0 or 1, which implies that Y is a deterministic constant.

7. Solution to Exercise 2.4. $A \in \sigma(f(X))$ if and only if $A = \{f(X) \in U\}$, for some $U \in \mathcal{B}(\mathbb{R})$. The latter is equivalent to $X(\omega) \in \{f \in U\}$, hence $A = \{X \in \{f \in U\}\}$. Similarly, $B = \{Y \in \{g \in V\}\}$, for some $V \in \mathcal{B}(\mathbb{R})$. Hence

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{X \in \{f \in U\}\} \cap \{Y \in \{g \in V\}\}).$$

As X and Y are independent, the right hand side is equal to $\mathbb{P}(\{X \in \{f \in U\}\})\mathbb{P}(\{Y \in \{g \in V\}\})$, hence $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, as claimed.

8. Solution to Exercise 2.6. Write $(-\infty, b]$ as the disjoint union of the sets $(-\infty, a]$ and $(a, b]$. Hence also $\{X \in (-\infty, a]\}$, $\{X \in (a, b]\}$ are disjoint. It follows that

$$\mathbb{P}(-\infty < X \le b) = \mathbb{P}(-\infty < X \le a) + \mathbb{P}(a < X \le b),$$

that is, $F(b) = F(a) + \mathbb{P}(a < X \le b)$, by which the claim follows. To establish that F_X is right-continuous we now show that

$$\mathbb{P}(X \le x_0 + \frac{1}{n}) \to \mathbb{P}(X \le x_0) \quad \text{as } n \to \infty, \text{ for all } x_0 \in \mathbb{R}.$$

By the first part of the exercise it suffices to show that $\mathbb{P}(x_0 < x \le x_0 + \frac{1}{n}) \to 0$ as $n \to \infty$. The intervals $A_n = (x_0, x_0 + n^{-1}]$ satisfy $A_{n+1} \subset A_n$ and $\cap_n A_n = \emptyset$. Hence by Exercise 1.7 we have $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\emptyset) = 0$. The fact that F_X is non-decreasing is obvious. The properties $\lim_{x\to\infty} F_X(x) = 1$ and $\lim_{x\to-\infty} F_X(x) = 0$ are also obvious.

9. Solution Exercise 2.7. One way to solve this exercise is to set $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(R)$, $X(\omega) = \omega$, define the probability of the open interval (a, b) as $\mathbb{P}((a, b)) = F(b) - F(a)$ and then use Catatheódory's theorem to extend this definition to a probability measure on the whole $\mathcal{B}(\mathbb{R})$. Moreover $\mathbb{P}(X \le x) = \mathbb{P}(\omega < x) = \mathbb{P}((-\infty, x)) = F(x)$.

10. Solution to Exercise 2.8. We first compute the distribution function F_Y of $Y = X^2$. Clearly, $F_Y(y) = \mathbb{P}(X^2 \le y) = 0$, if $y \le 0$. For $y > 0$ we have

$$F_Y(y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(-\sqrt{y} < X < \sqrt{y}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-x^2/2}\, dx.$$

Hence, for $y > 0$,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{\sqrt{2\pi}} \left(e^{-y/2} \frac{d}{dy}(\sqrt{y}) - e^{-y/2} \frac{d}{dy}(-\sqrt{y}) \right) = \frac{1}{\sqrt{2\pi}} \frac{e^{-y/2}}{\sqrt{y}}.$$

Since $\Gamma(1/2) = \sqrt{\pi}$, this is the claim.

11. Solution to Exercise 2.11. A 2×2 symmetric matrix

$$C = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

is positive definite if and only if $\operatorname{Tr} C = a + c > 0$ and $\det C = ac - b^2 > 0$. In particular, $a, c > 0$. Let us denote

$$a = \sigma_1^2, \quad c = \sigma_2^2, \quad \rho = \frac{b}{\sigma_1 \sigma_2}.$$

Note that $\rho^2 = \frac{b^2}{ac} < 1$. Thus

$$C = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

and so

$$C^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

Substituting into (2.9) proves (2.10).

12. Solution to Exercise 2.12. $X \in \mathcal{N}(0,1)$ means that X has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

while $Y \in \mathcal{E}(1)$ means that Y has density

$$f_Y(y) = e^{-y} \mathbb{I}_{y \geq 0}.$$

Since X, Y are independent, they have the joint density $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. Hence

$$\mathbb{P}(X \leq Y) = \iint\limits_{x \leq y} f_{X,Y}(x,y)\, dx\, dy = \int_{\mathbb{R}} dx\, e^{-\frac{x^2}{2}} \int_x^\infty dy\, \frac{1}{\sqrt{2\pi}} e^{-y} \mathbb{I}_{y \geq 0}.$$

To compute this integral, we first divide the domain of integration on the variable x in $x \leq 0$ and $x \geq 0$. So doing we have

$$\mathbb{P}(X \leq Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 dx\, e^{-\frac{x^2}{2}} \int_0^\infty dy\, e^{-y} + \frac{1}{\sqrt{2\pi}} \int_0^\infty dx\, e^{-\frac{x^2}{2}} \int_x^\infty dy\, e^{-y}.$$

Computing the integrals we find

$$\mathbb{P}(X \leq Y) = \frac{1}{2} + e^{1/2}(1 - \Phi(1)),$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-y^2/2} dy$ is the standard normal distribution.

13. Solution to Exercise 2.15. The density of $S(t)$ is given by

$$f_{S(t)}(x) = \frac{d}{dx} F_{S(t)}(x),$$

provided the distribution $F_{S(t)}$, i.e.,

$$F_{S(t)}(x) = \mathbb{P}(S(t) \leq x)$$

is differentiable. Clearly, $f_{S(t)}(x) = F_{S(t)}(x) = 0$, for $x < 0$. For $x > 0$ we use that

$$S(t) \leq x \quad \text{if and only if} \quad W(t) \leq \frac{1}{\sigma}\left(\log \frac{x}{S(0)} - \alpha t\right) := A(x).$$

Thus,

$$\mathbb{P}(S(t) \leq x) = \mathbb{P}(-\infty < W(t) \leq A(x)) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{A(x)} e^{-\frac{y^2}{2t}}\, dy,$$

where for the second equality we used that $W(t) \in N(0,t)$. Hence

$$f_{S(t)}(x) = \frac{d}{dx}\left(\frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{A(x)} e^{-\frac{y^2}{2t}}\, dy\right) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{A(x)^2}{2t}} \frac{dA(x)}{dx},$$

for $x > 0$, that is

$$f_{S(t)}(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \frac{1}{x} \exp\left\{-\frac{(\log x - \log S(0) - \alpha t)^2}{2\sigma^2 t}\right\} \mathbb{I}_{x>0}.$$

Since

$$\int_0^\infty f_{S(t)}(y) dy = 1,$$

then $p_0 = \mathbb{P}(S(t) = 0) = 0$.

14. Solution to Exercise 3.1. Let X be a binomial random variable. Then

$$\mathbb{E}[X] = \sum_{k=1}^{N} k\mathbb{P}(X = k) = \sum_{k=1}^{N} k \binom{N}{k} p^k (1-p)^{N-k}.$$

Now, by the binomial theorem

$$\sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k} = (x+y)^N,$$

for all $x, y > 0$. Differentiating with respect to x we get

$$\sum_{k=1}^{N} k \binom{N}{k} x^{k-1} y^{N-k} = N(x+y)^{N-1},$$

Letting $x = p$ and $y = 1 - p$ in the last identity we find $\mathbb{E}[X] = Np$.

15. Solution to Exercise 3.3. If $Y = 0$ almost surely, the claim is obvious. Hence we may assume that $\mathbb{E}[Y^2] > 0$. Let

$$Z = X - \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]} Y.$$

Then

$$0 \leq \mathbb{E}[Z^2] = \mathbb{E}[X^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]},$$

by which (3.3) follows.

16. Solution to Exercise 3.5. The first and second properties follow by the linearity of the expectation. In fact

$$\mathrm{Var}[\alpha X] = \mathbb{E}[\alpha^2 X^2] - \mathbb{E}[\alpha X]^2 = \alpha^2 \mathbb{E}[X^2] - \alpha^2 \mathbb{E}[X]^2 = \alpha^2 \mathrm{Var}[X],$$

and

$$\begin{aligned}\mathrm{Var}[X+Y] = \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 = \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] \\ - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y).\end{aligned}$$

For the third property, let $a \in \mathbb{R}$ and compute, using 1 and 2,

$$\mathrm{Var}[Y - aX] = a^2 \mathrm{Var}[X] + \mathrm{Var}[Y] - 2a\mathrm{Cov}(X,Y).$$

Since the variance of a random variable is always non-negative, the parabola $y(a) = a^2\mathrm{Var}[X] + \mathrm{Var}[Y] - 2a\mathrm{Cov}(X,Y)$ must always lie above the a-axis, or touch it at one single point $a = a_0$. Hence

$$\mathrm{Cov}(X,Y)^2 - \mathrm{Var}[X]\mathrm{Var}[Y] \leq 0,$$

which proves the first part of the claim 3. Moreover $\mathrm{Cov}(X,Y)^2 = \mathrm{Var}[X]\mathrm{Var}[Y]$ if and only if there exists a_0 such that $\mathrm{Var}[-a_0 X + Y] = 0$, i.e., $Y = a_0 X + b_0$ almost surely, for some constant b_0. Substituting in the definition of covariance, we see that $\mathrm{Cov}(X, a_0 X + b_0) = a_0 \mathrm{Var}[X]$, by which the second claim of property 3 follows immediately.

17. Solution Exercise 3.8. By linearity of the expectation,

$$\mathbb{E}[W_n(t)] = \frac{1}{\sqrt{n}} \mathbb{E}[M_{[nt]}] = 0,$$

where we used the fact that $\mathbb{E}[X_k] = \mathbb{E}[M_k] = 0$. Since $\mathrm{Var}[M_k] = k$, we obtain

$$\mathrm{Var}[W_n(t)] = \frac{[nt]}{n}.$$

Since $[nt] \sim nt$, as $n \to \infty$, then $\lim_{n\to\infty} \mathrm{Var}[W_n(t)] = t$. As to the covariance of $W_n(t)$ and $W_n(s)$ for $s \neq t$, we compute

$$\begin{aligned}\mathrm{Cov}[W_n(t), W_n(s)] &= \mathbb{E}[W_n(t)W_n(s)] - \mathbb{E}[W_n(t)]\mathbb{E}[W_n(s)] = \mathbb{E}[W_n(t)W_n(s)] \\ &= \mathbb{E}\left[\frac{1}{\sqrt{n}} M_{[nt]} \frac{1}{\sqrt{n}} M_{[ns]}\right] = \frac{1}{n} \mathbb{E}[M_{[nt]} M_{[ns]}]. \end{aligned} \tag{B.2}$$

Assume $t > s$ (a similar argument applies to the case $t < s$). If $[nt] = [ns]$ we have $\mathbb{E}[M_{[nt]}M_{[ns]}] = \mathrm{Var}[M_{[ns]}] = [ns]$. If $[nt] \geq 1 + [ns]$ we have

$$\mathbb{E}[M_{[nt]}M_{[ns]}] = \mathbb{E}[(M_{[nt]} - M_{[ns]})M_{[ns]}] + \mathbb{E}[M_{[ns]}^2] = \mathbb{E}[M_{[nt]} - M_{[ns]}]\mathbb{E}[M_{[ns]}] + \mathrm{Var}[M_{[ns]}] = [ns],$$

where we used that the increment $M_{[nt]} - M_{[ns]}$ is independent of $M_{[ns]}$. Replacing into (B.2) we obtain

$$\mathrm{Cov}[W_n(t), W_n(s)] = \frac{[ns]}{n}.$$

It follows that $\lim_{n\to\infty} \mathrm{Cov}[W_n(t), W_n(s)] = s$.

18. Solution to Exercise 3.9. We have

$$\mathbb{E}[X(t)] = g(t)\mathbb{E}[W(t)] - \int_0^t g'(s)\mathbb{E}[W(s)]\,ds = 0,$$

$$\begin{aligned}
\mathrm{Var}[X(t)] = \mathbb{E}[X(t)^2] = g(t)^2\mathbb{E}[W(t)^2] + \mathbb{E}\left[\left(\int_0^t g'(s)W(s)\,ds\right)^2\right] \\
- 2g(t)\mathbb{E}\left[\int_0^t g'(s)W(t)W(s)\,ds\right] \\
= g(t)^2 t + \int_0^t\int_0^t g'(s)g'(\tau)\mathrm{Cov}(W(s),W(\tau))\,d\tau\,ds \\
- 2g(t)\int_0^t g'(s)\mathrm{Cov}(W(s),W(t))\,ds.
\end{aligned}$$

Using $\mathrm{Cov}(W(s),W(t)) = \min(s,t)$, and after some technical but straightforward calculation, we obtain $\mathrm{Var}[X(t)] = \Delta(t)$. To show that $X(t)$ is normally distributed, let $\{t_0 = 0,\dots,t_n = t\}$ be a uniform partition of the interval $[0,t]$ and consider the Riemann sum approximation of $X(t)$:

$$\begin{aligned}
X_n(t) &= g(t_n)W(t_n) - \sum_{i=1}^{n}(g(t_i) - g(t_{i-1})W(t_i) \\
&= -\sum_{i=1}^{n-1} g(t_i)W(t_i) + \sum_{i=1}^{n} g(t_{i-1})W(t_i) \\
&= -\sum_{i=0}^{n-1} g(t_i)W(t_i) + \sum_{j=0}^{n-1} g(t_j)W(t_{j+1}),
\end{aligned}$$

where in the last step we used $W(t_0) = W(0) = 0$ in the first sum and made the change of index $j = i-1$ in the second sum. Hence

$$X_n(t) = \sum_{i=0}^{n-1} g(t_i)(W(t_{i+1}) - W(t_i)).$$

Thus $X_n(t)$ is normally distributed because it is a linear combination of the independent and normally distributed random variables $W(t_{i+1}) - W(t_i)$. It can be shown that this property carries over in the limit $n\to\infty$ and since $X_n(t)\to X(t)$ in this limit the proof is completed.

19. Solution to Exercise 3.11. We have

$$\mathbb{E}[X] = \int_{\mathbb{R}} \lambda x e^{-\lambda x}\mathbb{I}_{x\geq 0}\,dx = \lambda\int_0^\infty x e^{-\lambda x}\,dx = \frac{1}{\lambda},$$

and

$$\begin{aligned}\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \int_{\mathbb{R}} \lambda x^2 e^{-\lambda x} \mathbb{I}_{x\geq 0}\, dx - \frac{1}{\lambda^2}\\ &= \lambda \int_0^{\infty} x^2 e^{-\lambda x}\, dx - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.\end{aligned}$$

20. Solution to Exercise 3.12. According to Theorem 3.3, to prove that $\widetilde{\mathbb{P}}$ is equivalent to $\mathbb{P}$ we have to show that $\mathbb{E}[Z] = 1$. Using the density of exponential random variables we have

$$\mathbb{E}[Z] = \frac{\widetilde{\lambda}}{\lambda}\mathbb{E}[e^{-(\widetilde{\lambda}-\lambda)X}] = \frac{\widetilde{\lambda}}{\lambda}\int_0^{\infty} e^{-(\widetilde{\lambda}-\lambda)x}\lambda e^{-\lambda x}\, dx = 1.$$

To show that $X \in \mathcal{E}(\widetilde{\lambda})$ in the probability measure $\widetilde{\mathbb{P}}$ we compute

$$\widetilde{\mathbb{P}}(X \leq x) = \mathbb{E}[Z\mathbb{I}_{X\leq x}] = \frac{\widetilde{\lambda}}{\lambda}\mathbb{E}[e^{-(\widetilde{\lambda}-\lambda)X}\mathbb{I}_{X\leq x}] = \frac{\widetilde{\lambda}}{\lambda}\int_0^{x} e^{-(\widetilde{\lambda}-\lambda)y}\lambda e^{-\lambda y}\, dy = 1 - e^{-\widetilde{\lambda}x}.$$

21. Solution to Exercise 3.27. We have $\mathbb{E}[\mathrm{Err}] = \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = 0$. Let Y be $\mathcal{G}$-measurable and set $\mu = \mathbb{E}[Y - X]$. Then

$$\begin{aligned}\mathrm{Var}[Y-X] &= \mathbb{E}[(Y-X-\mu)^2] = \mathbb{E}[(Y-X-\mu+\mathbb{E}[X|\mathcal{G}]-\mathbb{E}[X|\mathcal{G}])^2]\\ &= \mathbb{E}\Big[(\mathbb{E}[X|\mathcal{G}]-X)^2 + (Y-\mu-\mathbb{E}[X|\mathcal{G}])^2 + 2(\mathbb{E}[X|\mathcal{G}]-X)(Y-\mu-\mathbb{E}[X|\mathcal{G}])\Big]\\ &= \mathrm{Var}[\mathrm{Err}] + \mathbb{E}[\alpha] + 2\mathbb{E}[\beta],\end{aligned}$$

where $\alpha = (Y-\mu-\mathbb{E}[X|\mathcal{G}])^2$ and $\beta = (\mathbb{E}[X|\mathcal{G}]-X)(Y-\mu-\mathbb{E}[X|\mathcal{G}])$. As $\mathbb{E}[\alpha] \geq 0$ we have $\mathrm{Var}[Y-X] \geq \mathrm{Var}[\mathrm{Err}] + 2\mathbb{E}[\beta]$. Furthermore, as $Y-\mu-\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, then

$$\mathbb{E}[\beta] = \mathbb{E}[\mathbb{E}[\beta|\mathcal{G}]] = \mathbb{E}[(Y-\mu-\mathbb{E}[X|\mathcal{G}])\mathbb{E}[(\mathbb{E}[X|\mathcal{G}]-X)|\mathcal{G}] = 0.$$

Hence $\mathrm{Var}[Y-X] \geq \mathrm{Var}[\mathrm{Err}]$, for all $\mathcal{G}$-measurable random variables Y.

22. Solution to Exercise 3.29. First we observe that claim (i) follows by claim (ii). In fact, if the compound Poisson process is a martingale, then

$$\mathbb{E}[N(t)-\lambda t|\mathcal{F}_N(s)] = N(s) - \lambda s, \quad \text{for all } 0 \leq s \leq t,$$

by which it follows that

$$\mathbb{E}[N(t)|\mathcal{F}_N(s)] = N(s) + \lambda(t-s) \geq N(s), \quad \text{for all } 0 \leq s \leq t.$$

Hence it remains to prove (ii). We have

$$\begin{aligned}\mathbb{E}[N(t)-\lambda t|\mathcal{F}_N(s)] &= \mathbb{E}[N(t)-N(s)+N(s)-\lambda t|\mathcal{F}_N(s)]\\ &= \mathbb{E}[N(t)-N(s)|\mathcal{F}_N(s)] + \mathbb{E}[N(s)|\mathcal{F}_N(s)] - \lambda t\\ &= \mathbb{E}[N(t)-N(s)] + N(s) - \lambda t = \lambda(t-s) + N(s) - \lambda t = N(s) - \lambda s.\end{aligned}$$

23. Solution to Exercise 3.31. Taking the conditional expectation of both sides of (3.25) with respect to the event $\{X(s) = x\}$ gives (3.27).

24. Solution to Exercise 4.3. Assume

$$X_i(t) = \alpha_i t + \sigma_i dW(t), \quad i = 1, 2,$$

for constants $\alpha_1, \alpha_2, \sigma_1, \sigma_2$. Then the right hand side of (4.15) is

$$\begin{aligned}
&\int_0^t [(\alpha_2 s + \sigma_2 W(s))\sigma_1 + (\alpha_1 s + \sigma_1 W(s))\sigma_2] dW(s) \\
&\quad + \int_0^t [(\alpha_2 s + \sigma_2 W(s))\alpha_1 + (\alpha_1 s + \sigma_1 W(s))\alpha_2 + \sigma_1\sigma_2] ds \\
&= \sigma_1 \int_0^t (\alpha_2 s + \sigma_2 W(s)) dW(s) + \sigma_2 \int_0^t (\alpha_1 s + \sigma_1 W(s)) dW(s) \\
&\quad + \alpha_1\alpha_2 \frac{t^2}{2} + \alpha_1\sigma_2 \int_0^t W(s) ds + \alpha_1\alpha_2 \frac{t^2}{2} + \sigma_1\alpha_2 \int_0^t W(s) ds + \sigma_1\sigma_2 t \\
&= \sigma_1\alpha_2 \int_0^t s dW(s) + 2\sigma_1\sigma_2 \int_0^t W(s) dW(s) + \sigma_2\alpha_1 \int_0^t s dW(s) \\
&\quad + \alpha_1\alpha_2 t^2 + \sigma_1\sigma_2 t + (\alpha_1\sigma_2 + \sigma_1\alpha_2) \int_0^t W(s) ds \\
&= 2\sigma_1\sigma_2 (\frac{W(t)^2}{2} - \frac{t}{2}) + \sigma_1\sigma_2 t + \alpha_1\alpha_2 t^2 + (\sigma_1\alpha_2 + \alpha_1\sigma_2)(\int_0^t s dW(s) + \int_0^t W(s) ds) \\
&= \sigma_1\sigma_2 W(t)^2 + \alpha_1\alpha_2 t^2 + (\sigma_1\alpha_2 + \alpha_1\sigma_2) t W(t) \\
&= (\alpha_1 t + \sigma_1 W(t))(\alpha_2 t + \sigma_2 W(t)) = X_1(t) X_2(t).
\end{aligned}$$

25. Solution to Exercise 4.4. We have

$$\mathrm{Cor}(W_1(t), W_2(t)) = \frac{\mathbb{E}[W_1(t)W_2(t)]}{\sqrt{\mathrm{Var}[W_1(t)]\mathrm{Var}[W_2(t)]}} = \frac{1}{t}\mathbb{E}[W_1(t)W_2(t)].$$

Hence we have to show that $\mathbb{E}[W_1(t)W_2(t)] = \rho t$. By Itô's product rule

$$d(W_1(t)W_2(t)) = W_1(t)dW_2(t) + W_2(t)dW_1(t) + dW_1(t)dW_2(t) = W_1(t)dW_2(t) + W_2(t)dW_1(t) + \rho dt.$$

Taking the expectation we find $\mathbb{E}[W_1(t)W_2(t)] = \rho t$, which concludes the first part of the exercise. As to the second part, the independent random variables $W_1(t), W_2(s)$ have the joint density

$$f_{W_1(t)W_2(s)}(x, y) = f_{W_1(t)}(x) f_{W_2(s)}(y) = \frac{1}{2\pi\sqrt{ts}} e^{-\frac{x^2}{2t} - \frac{y^2}{2s}}.$$

Hence

$$\mathbb{P}(W_1(t) > W_2(s)) = \frac{1}{2\pi\sqrt{ts}} \int_{x>y} e^{-\frac{x^2}{2t} - \frac{y^2}{2s}} \, dx \, dy = \frac{1}{2}.$$

26. Solution to Exercise 4.5. To solve the exercise we must prove that $dX(t) = \Gamma(t)dW(t)$, for some process $\{\Gamma(t)\}_{t\geq 0}$ adapted to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. In fact, by Itô's formula,

$$dX(t) = -3W(t)\,dt + (3W(t) - 3t)\,dW(t) + \frac{1}{2}6W(t)dW(t)dW(t) = 3(W(t)^2 - t)dW(t),$$

where in the last step we used that $dW(t)dW(t) = dt$.

27. Solution to Exercise 4.6. By Itô's formula, the stochastic process $\{Z(t)\}_{t\geq 0}$ satisfies

$$dZ(t) = -\theta(t)Z(t)dW(t),$$

which is the claim.

28. Solution to Exercise 4.10. We have

$$dV(t) = d(h_S(t))S(t) + h_S(t)dS(t) + dh_S(t)dS(t) + d(h_B(t))B(t) + h_B(t)dB(t)$$

where we used that $dh_B(t)dB(t) = r(t)B(t)dh_B(t)dt = 0$. Hence $dV(t) = h_S(t)dS(t) + h_B(t)dB(t)$ holds by letting $d(h_B(t)) = -B(t)^{-1}[dh_S(t)dS(t) + h_S(t)dS(t)]$.

29. Solution to Exercise 5.1. When the drift term α in (5.1) is given by (5.5), the stochastic differential equations becomes

$$dX(t) = a(b - X(t))dt + \beta(t, X(t))\,dW(t).$$

Hence $Y(t) = e^{at}X(t)$ satisfies

$$dY(t) = e^{at}(aX(t)\,dt + dX(t)) = e^{at}(ab\,dt + \beta(t, X(t))\,dW(t).$$

Integrating in the interval $[s, t]$ we obtain

$$Y(t) = xe^{as} + b(e^{at} - e^{as}) + \int_s^t e^{a\tau}\beta(\tau, X(\tau))\,dW(\tau).$$

As $\beta(\tau, X(\tau)) \leq C|X(\tau)|$ and $\{X(t)\}_{t\geq s} \in \mathbb{L}^2[\mathcal{F}(t)]$, the Itô integral in the right hand side is a martingale. Hence taking the expectation of both sides we obtain

$$\mathbb{E}[Y(t)] = e^{at}\mathbb{E}[X(t)] = xe^{as} + b(e^{at} - e^{as}).$$

30. Solution to Exercise 5.3. The solution of (5.8) could be found replacing a, b, γ=constants and $\sigma = 0$ into the general solution in Theorem 5.7 , but it is actually quicker to solve this special case independently. Letting $Y(t) = e^{-bt}X(t)$ and applying Itô's formula we find that $Y(t)$ satisfies

$$dY(t) = ae^{-bt}\,dt + \gamma e^{-bt}dW(t), \quad Y(s) = xe^{-bs}.$$

Hence

$$Y(t) = xe^{-bs} + a\int_s^t e^{-bu}\,d\tau + \gamma\int_s^t e^{-bu}\,dW(u)$$

and so

$$X(t;s,x) = xe^{b(t-s)} - \frac{a}{b}(1-e^{b(t-s)}) + \int_s^t \gamma e^{b(t-u)} dW(u).$$

Taking the expectation we obtain immediately that $\mathbb{E}[X(t;s,x)] = m(t-s,x)$. Moreover by Exercise 4.8, the Itô integral in $X(t;s,x)$ is a normal random variable with zero mean and variance $\Delta(t-s)^2$, hence the claim follows.

31. Solution to Exercise 5.4. Letting $Y(t) = e^{-\frac{t^2}{2}}X(t)$, we find that $dY(t) = e^{-\frac{t^2}{2}}dW(t)$ and $Y(0) = 1$. Thus

$$X(t) = e^{\frac{t^2}{2}} + e^{\frac{t^2}{2}} \int_0^t e^{-\frac{u^2}{2}} dW(u).$$

Note that $X(t)$ is normally distributed with mean

$$\mathbb{E}[X(t)] = e^{\frac{t^2}{2}}.$$

It follows that $\mathrm{Cov}(X(s), X(t)) = \mathbb{E}[X(s)X(t)] - \mathbb{E}[X(s)]\mathbb{E}[X(t)]$ is

$$\mathrm{Cov}(X(s),X(t)) = e^{\frac{s^2+t^2}{2}} \mathbb{E}\Big[\int_0^s e^{-\frac{u^2}{2}} dW(u) \int_0^t e^{-\frac{u^2}{2}} dW(u)\Big].$$

Assume for example that $s \leq t$. Hence

$$\mathrm{Cov}(X(s),X(t)) = e^{\frac{s^2+t^2}{2}} \mathbb{E}\Big[\int_0^t \mathbb{I}_{[0,s]} e^{-\frac{u^2}{2}} dW(u) \int_0^t e^{-\frac{u^2}{2}} dW(u)\Big].$$

Using the result of Exercise 4.2 we have

$$\begin{aligned}\mathrm{Cov}(X(s),X(t)) &= e^{\frac{s^2+t^2}{2}} \int_0^t \mathbb{I}_{[0,s]} e^{-\frac{u^2}{2}} e^{-\frac{u^2}{2}} \, du = e^{\frac{s^2+t^2}{2}} \int_0^s e^{-u^2} \, du \\ &= \sqrt{\pi} e^{\frac{s^2+t^2}{2}} \int_0^{\sqrt{2}s} e^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}} = \sqrt{\pi} e^{\frac{s^2+t^2}{2}} (\Phi(\sqrt{2}s) - \frac{1}{2}).\end{aligned}$$

For general $s, t \geq 0$ we find

$$\mathrm{Cov}(X(s),X(t)) = \sqrt{\pi} e^{\frac{s^2+t^2}{2}} (\Phi(\sqrt{2}\min(s,t)) - \frac{1}{2}).$$

32. Solution to Exercise 6.2. We have, for $t \leq t_*$,

$$\Pi_Z(t) = D(t)^{-1} \widetilde{\mathbb{E}}[D(t_*)\Pi_Y(t_*)|\mathcal{F}_W(t)] = D(t)^{-1} D(t) \Pi_Y(t) = \Pi_Y(t),$$

where we used the martingale property of the discounted price of the derivative with pay-off Y, see Theorem 6.2.

33. Solution to Exercise 6.7. The pay-off function is $g(z) = k + z\log z$. Hence the Black-Scholes price of the derivative is $\Pi_Y(t) = v(t, S(t))$, where

$$\begin{aligned}
v(t,x) &= e^{-r\tau}\int_{\mathbb{R}} g\left(xe^{\left(r-\frac{\sigma^2}{2}\right)\tau+\sigma\sqrt{\tau}y}\right)e^{-\frac{y^2}{2}}\frac{dy}{\sqrt{2\pi}}\\
&= e^{-r\tau}\int_{\mathbb{R}}\left(k + xe^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}y}(\log x + (r-\frac{\sigma^2}{2})\tau + \sigma\sqrt{\tau}y)\right)e^{-\frac{y^2}{2}}\frac{dy}{\sqrt{2\pi}}\\
&= ke^{-r\tau} + x\log x\int_{\mathbb{R}} e^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}\frac{dy}{\sqrt{2\pi}}\\
&\quad + x(r-\frac{\sigma^2}{2})\tau\int_{\mathbb{R}} e^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}\frac{dy}{\sqrt{2\pi}} + x\sigma\sqrt{\tau}\int_{\mathbb{R}} ye^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}\frac{dy}{\sqrt{2\pi}}
\end{aligned}$$

Using that

$$\int_{\mathbb{R}} e^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}\frac{dy}{\sqrt{2\pi}} = 1, \quad \int_{\mathbb{R}} ye^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}\frac{dy}{\sqrt{2\pi}} = \sigma\sqrt{\tau},$$

we obtain

$$v(t,x) = ke^{-r\tau} + x\log x + x(r+\frac{\sigma^2}{2})\tau.$$

Hence

$$\Pi_Y(t) = ke^{-r\tau} + S(t)\log S(t) + S(t)(r+\frac{\sigma^2}{2})\tau.$$

This completes the first part of the exercise. The number of shares of the stock in the hedging portfolio is given by

$$h_S(t) = \Delta(t, S(t)),$$

where $\Delta(t,x) = \frac{\partial v}{\partial x} = \log x + 1 + (r+\frac{\sigma^2}{2})\tau$. Hence

$$h_S(t) = 1 + (r+\frac{\sigma^2}{2})\tau + \log S(t).$$

The number of shares of the bond is obtained by using that

$$\Pi_Y(t) = h_S(t)S(t) + B(t)h_B(t),$$

hence

$$\begin{aligned}
h_B(t) &= \frac{1}{B(t)}(\Pi_Y(t) - h_S(t)S(t))\\
&= e^{-rt}(ke^{-r\tau} + S(t)\log S(t) + S(t)(r+\frac{\sigma^2}{2})\tau - S(t) - S(t)(r+\frac{\sigma^2}{2})\tau - S(t)\log S(t))\\
&= ke^{-rT} - S(t)e^{-rt}.
\end{aligned}$$

This completes the second part of the exercise. To compute the probability that $Y > 0$, we first observe that the pay-off function $g(z)$ has a minimum at $z = e^{-1}$ and we have

$g(e^{-1}) = k - e^{-1}$. Hence if $k \geq e^{-1}$, the derivative has probability 1 to expire in the money. If $k < e^{-1}$, there exist $a < b$ such that

$$g(z) > 0 \quad \text{if and only if} \;\; 0 < z < a \text{ or } z > b.$$

Hence for $k < e^{-1}$ we have

$$\mathbb{P}(Y > 0) = \mathbb{P}(S(T) < a) + \mathbb{P}(S(T) > b).$$

Since $S(T) = S(0)e^{\alpha T - \sigma\sqrt{T}G}$, with $G \in N(0,1)$, then

$$S(T) < a \Leftrightarrow G > \frac{\log\frac{S(0)}{a} + \alpha T}{\sigma\sqrt{T}} := A, \quad S(t) > b \Leftrightarrow G < \frac{\log\frac{S(0)}{b} + \alpha T}{\sigma\sqrt{T}} := B.$$

Thus

$$\begin{aligned}
\mathbb{P}(Y > 0) = \mathbb{P}(G > A) + \mathbb{P}(G < B) &= \int_A^{+\infty} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} + \int_{-\infty}^{B} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \\
&= 1 - \Phi(A) + \Phi(B).
\end{aligned}$$

This completes the solution of the third part of the exercise.

34. Solution to Exercise 6.8. The pay-off is

$$\begin{aligned}
&Yc = LH(S(T) - K), \quad \text{for the cash-settled binary call option,} \\
&Zc = S(T)H(S(T) - K), \quad \text{for the physically-settled binary call option,} \\
&Yp = LH(K - S(T)), \quad \text{for the cash-settled binary put option,} \\
&Zp = S(T)H(K - S(T)), \quad \text{for the physically-settled binary put option.}
\end{aligned}$$

The Black-Scholes price in all cases is given by $v(t, S(t))$, where $v(t, x)$ is the price function of the derivative given by (6.19b). Replacing $g(z) = LH(z - K)$ into (6.19b) we obtain

$$\begin{aligned}
v(t, x) &= \frac{e^{-r\tau}}{\sqrt{2\pi}} L \int_{\mathbb{R}} H(xe^{(r - \frac{\sigma^2}{2})\tau} e^{\sigma\sqrt{\tau}\, y} - K) e^{-\frac{y^2}{2}}\, dy \\
&= \frac{e^{-r\tau}}{\sqrt{2\pi}} L \int_{-d_2}^{\infty} e^{-\frac{y^2}{2}}\, dy = e^{-r\tau} L\Phi(d_2) \quad \text{(cash-settled binary call),}
\end{aligned}$$

where we recall that $d_2 = \frac{1}{\sigma\sqrt{\tau}}[\log(x/K) + (r - \frac{\sigma^2}{2})\tau]$. Similarly, using the pay-off function $g(x) = xH(x - K)$, one finds that the price function for the physically-settled binary option is

$$v(t, x) = x\Phi(d_1) \quad \text{(physically-settled binary call),}$$

where we recall that $d_1 = d_2 + \sigma\sqrt{\tau}$. For the put options one obtains

$$v(t, x) = \begin{cases} Le^{-r\tau}\Phi(-d_2) & \text{(cash-settled put),} \\ x\Phi(-d_1) & \text{(physically-settled put).} \end{cases}$$

Hence, using $\Phi(z)+\Phi(-z)=1$, the following put-call parity relations hold:

$$\Pi_{Yc}(t)+\Pi_{Yp}(t)=Le^{-r\tau}, \quad \Pi_{Zc}(t)+\Pi_{Zp}(t)=S(t).$$

The number of shares of the stock in the hedging portfolio is given by $h_S(t)=\Delta(t,S(t))$, where $\Delta(t,x)=\partial_v(t,x)$, hence

$$\Delta(t,x)=\begin{cases} \frac{Le^{-r\tau}\phi(d_2)}{x\sigma\sqrt{\tau}} & \text{(cash-settled call)},\\ \Phi(d_1)+\frac{\phi(d_1)}{\sigma\sqrt{\tau}} & \text{(physically settled call)},\end{cases}$$

and similarly for the put options. In any case the number of shares of the risk-free asset in the hedging portfolio is $h_B(t)=B(t)^{-1}[v(t,S(t))-\Delta(t,S(t))S(t)]$.

35. Solution to Exercise 6.9. The pay-off at time T_1 for this derivative is

$$Y=\max(C(T_1,S(T_1),K,T_2),P(t_1,S(T_1),K,T_2)).$$

Using the identity $\max(a,b)=a+\max(0,b-a)$ we obtain

$$Y=C(T_1,S(T_1),K,T_2)+\max(0,P(t_1,S(T_1),K,T_2)-C(t_1,S(T_1),K,T_2)).$$

By the put-call parity,

$$Y=C(T_1,S(T_1),K,T_2)+\max(0,Ke^{-r(T_2-T_1)}-S(T_1))=Z+U. \tag{B.3}$$

Hence $\Pi_Y(t)=\Pi_Z(t)+\Pi_U(t)$. Since U is the pay-off of a put option with strike $Ke^{-r(T_2-T_1)}$ expiring at time T_1 then

$$\Pi_U(t)=P(t,S(t),Ke^{-r(T_2-T_1)},T_1). \tag{B.4}$$

Applying the result of Exercise 6.2 with $Z=C(T_1,S(T_1),K,T_2)=\Pi_{(S(T_2)-K)_+}(T_1)$ we obtain

$$\Pi_Z(t)=C(t,S(t),K,T_2). \tag{B.5}$$

Replacing (B.4) and (B.5) into (B.3) we finally obtain, for the Black-Scholes price of the chooser option,

$$\Pi_Y(t)=C(t,S(t),K,T_2)+P(t,S(t),Ke^{-r(T_2-T_1)},T_1).$$

36. Solution to Exercise 6.10. First we note that for $t\geq s$, we can consider the derivative as a standard call option with maturity T and *known* strike price $K=S(s)$, hence

$$\Pi_Y(t)=C(t,S(t),S(s),T), \quad \text{for } t\geq s,$$

where $C(t,x,K,T)$ is the Black-Scholes price function of the standard call option with strike K and maturity T. For $t<s$ we write

$$\begin{aligned}\Pi_Y(t)&=e^{-r(T-t)}\widetilde{\mathbb{E}}[(S(T)-S(s))_+|\mathcal{F}_W(t)]\\ &=e^{-r(T-t)}\widetilde{\mathbb{E}}[S(t)(e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(\widetilde{W}(T)-\widetilde{W}(t))}-e^{(r-\frac{1}{2}\sigma^2)(s-t)+\sigma(\widetilde{W}(s)-\widetilde{W}(t))})_+|\mathcal{F}_W(t)]\\ &=e^{-r(T-t)}S(t)\widetilde{\mathbb{E}}[(e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(\widetilde{W}(T)-\widetilde{W}(t))}-e^{(r-\frac{1}{2}\sigma^2)(s-t)+\sigma(\widetilde{W}(s)-\widetilde{W}(t))})_+]\end{aligned}$$

where in the last step we used that $S(t)$ is measurable with respect to $\mathcal{F}_W(t)$ and that the Brownian motion increments are independent of $\mathcal{F}_W(t)$. It follows that

$$\begin{aligned}\Pi_Y(t) &= e^{-r(T-t)}S(t)\widetilde{\mathbb{E}}[e^{(r-\frac{1}{2}\sigma^2)(s-t)+\sigma(\widetilde{W}(s)-\widetilde{W}(t))}(e^{(r-\frac{1}{2}\sigma^2)(T-s)+\sigma(\widetilde{W}(T)-\widetilde{W}(s))}-1)_+]\\ &= e^{-r(T-t)}S(t)\widetilde{\mathbb{E}}[e^{(r-\frac{1}{2}\sigma^2)(s-t)+\sigma(\widetilde{W}(s)-\widetilde{W}(t))}]\widetilde{\mathbb{E}}[(e^{(r-\frac{1}{2}\sigma^2)(T-s)+\sigma(\widetilde{W}(T)-\widetilde{W}(s))}-1)_+]\end{aligned}$$

where in the last step we used that the two Brownian motion increments are independent. Computing the expectations using that $W(t_2)-W(t_1)\in\mathcal{N}(0,t_2-t_1)$, for all $t_2>t_1$, we find

$$\Pi_Y(t) = S(t)(\Phi(a+\sigma\sqrt{T-t})-e^{-r(T-s)}\Phi(a)),\quad a=\frac{(r-\frac{1}{2}\sigma^2)\sqrt{T-s}}{\sigma}$$

This completes the solution to the first part of the exercise. As to the put call parity, let $Z=(S(s)-S(T))_+$. As $(x-y)_+-(y-x)_+=x-y$, we have

$$\begin{aligned}\Pi_Y(t)-\Pi_Z(t) &= e^{-r(T-t)}\widetilde{\mathbb{E}}[(S(T)-S(s))_+|\mathcal{F}_W(t)]-e^{-r(T-t)}\widetilde{\mathbb{E}}[(S(s)-S(T))_+|\mathcal{F}_W(t)]\\ &= e^{-r(T-t)}\widetilde{\mathbb{E}}[S(T)-S(s)|\mathcal{F}_W(t)] = e^{-r(T-t)}(\widetilde{\mathbb{E}}[S(T)|\mathcal{F}_W(t)]-\widetilde{\mathbb{E}}[S(s)|\mathcal{F}_W(t)])\\ &= \begin{cases} S(t)-e^{-r(T-t)}S(s) & \text{if } s\leq t\\ S(t)-e^{-r(T-s)}S(t) & \text{if } s>t.\end{cases}\end{aligned}$$

Hence the put-call parity is

$$\Pi_Y(t)-\Pi_Z(t) = S(t)-e^{-r(T-\max(s,t))}S(\min(s,t)).$$

This completes the answer to the second question.

37. Solution to Exercise 6.13. Let $C_t(x)=C(t,x)$. We have

$$\mathbb{P}(C(t)\leq x)=\mathbb{P}(C_t(S(t))\leq x)=\mathbb{P}(S(t)\leq C_t^{-1}(x)),$$

where we used that $C_t(\cdot)$ is a monotonically increasing and thus invertible function. It follows that

$$F_{C(t)}(x)=F_{S(t)}(C_t^{-1}(x)),$$

hence

$$f_{C(t)}(x)=\frac{f_{S(t)}(C_t^{-1}(x))}{C_t'(C_t^{-1}(x))}.$$

Here $f_{S(t)}$ is the probability density of the geometric Brownian motion, see Exercise 2.15.

38. Solution to Exercise 6.15. Letting $a=1-e^{-rT}$, the price at time $t=0$ of a

European derivative with pay-off function $g(x) = (x - S(0))_+$ is

$$
\begin{aligned}
\Pi_Y(0) &= e^{-rT} \int_{\mathbb{R}^3} g((1-a)S(0)e^{(r-\sigma^2/2)T+\sigma\sqrt{T}y})e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}} \\
&= e^{-rT}S(0) \int_{\mathbb{R}^3} (e^{-\frac{\sigma^2}{2}T+\sigma\sqrt{T}y} - 1)_+ e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}} \\
&= e^{-rT}S(0) \int_{\sigma\sqrt{T}/2}^{+\infty} \left(e^{-\frac{1}{2}(y-\sigma\sqrt{T})^2} - e^{\frac{y^2}{2}}\right) \frac{dy}{\sqrt{2\pi}} \\
&= e^{-rT}S(0)[\Phi(\frac{\sigma\sqrt{T}}{2}) - \Phi(-\frac{\sigma\sqrt{T}}{2})] \\
&= S(0)e^{-rT}(2\Phi(\frac{\sigma\sqrt{T}}{2}) - 1)
\end{aligned}
$$

where Φ is the standard normal cumulative distribution.

39. Solution to Exercise 6.17. We have

$$
\begin{aligned}
\Pi_{\mathrm{AC}}(t) - \Pi_{\mathrm{AP}}(t) &= e^{-r(T-t)}(\widetilde{\mathbb{E}}[(Q(T)/T - K)_+|\mathcal{F}_W(t)] - \widetilde{\mathbb{E}}[(K - Q(T)/T)_+|\mathcal{F}_W(t)]) \\
&= e^{-r(T-t)}\widetilde{\mathbb{E}}[Q(T)/T - K|\mathcal{F}_W(t)] = \frac{e^{-r(T-t)}}{T}\widetilde{\mathbb{E}}[Q(T)|\mathcal{F}_W(t)] - Ke^{-r(T-t)}.
\end{aligned} \tag{B.6}
$$

We split Q as

$$
Q(T) = \int_0^t S(\tau)\,d\tau + \int_t^T S(\tau)\,d\tau.
$$

Accordingly

$$
\widetilde{\mathbb{E}}[Q(T)|\mathcal{F}_W(t)] = \int_0^t \widetilde{\mathbb{E}}[S(\tau)|\mathcal{F}_W(t)]\,d\tau + \int_t^T \widetilde{\mathbb{E}}[S(\tau)|\mathcal{F}_W(t)]\,d\tau.
$$

Note that $\mathcal{F}_W(t) = \mathcal{F}_{\widetilde{W}}(t)$, because the market parameters are constant. In the first integral we use that $S(\tau)$ is $\mathcal{F}_W(t)$-measurable (because $\tau \leq t$), while in the second integral we use that $e^{-r\tau}S(\tau)$ is a martingale. Hence

$$
\widetilde{\mathbb{E}}[Q(T)|\mathcal{F}_W(t)] = Q(t) + e^{-rt}S(t)\int_t^T e^{r\tau}\,d\tau = Q(t) + \frac{S(t)}{r}(e^{r(T-t)} - 1).
$$

Substituting in (B.6) concludes the exercise.

40. Solution to Exercise 6.19. By the convexity of the function $(\cdot)_+$ and the Jensen inequality in the hint, we have

$$
\begin{aligned}
\Pi_{\mathrm{AC}}(0) = e^{-rT}\mathbb{E}_q\left[\left(\frac{1}{T}\int_0^T (S(t) - K)\,dt\right)_+\right] &\leq \frac{e^{-rT}}{T}\mathbb{E}_q\left[\int_0^T (S(t) - K)_+\,dt\right] \\
= \frac{e^{-rT}}{T}\int_0^T \mathbb{E}_q[(S(t) - K)_+]\,dt &= \frac{e^{-rT}}{T}\int_0^T e^{rt}e^{-rt}\mathbb{E}_q[(S(t) - K)_+]\,dt.
\end{aligned}
$$

Since for $r \geq 0$ the (Black-Scholes) value of the call option is decreasing with maturity, then $e^{-rt}\mathbb{E}_q[(S(t)-K)_+] = C(0,S_0,K,t) < C(0,S_0,K,T)$, for $t \in [0,T)$. Thus

$$\Pi_{\mathrm{AC}}(0) < \frac{e^{-rT}}{T} C(0,S_0,K,T) \int_0^T e^{rt}\,dt = \frac{1-e^{-rT}}{rT} C(0,S_0,K,T).$$

As the function $x \to (1-e^{-x})/x$ is bounded by 1 for $x \geq 0$, we obtain in particular $\Pi_{\mathrm{AC}}(0) < C(0,S_0,K,T)$.

41. Solution to Exercise 6.29. Since we assume that the interest rate of the money market is constant, the risk-neutral value of the swap at time $t \in [0,T]$ is

$$\Pi_Y(t) = e^{-r\tau}\widetilde{\mathbb{E}}[Q_T|\mathcal{F}_W(t)], \quad Q_T = \frac{\kappa}{T}\int_0^T \sigma^2(t)\,dt - K.$$

In particular, at time $t=0$, i.e., when the contract is stipulated, we have $\Pi_Y(0) = e^{-rT}\widetilde{\mathbb{E}}[Q_T]$. The fair strike of the swap is the value of K which makes $\Pi_Y(0)=0$. We find

$$K_* = \frac{\kappa}{T}\int_0^T \widetilde{\mathbb{E}}[\sigma^2(t)]\,dt.$$

To compute K_* when $\sigma^2(t) = \sigma_0^2 S(t)$, we first compute

$$\begin{aligned} d\sigma^2(t) &= \sigma_0^2 dS(t) = \sigma_0^2(\alpha(t)S(t)dt + \sigma(t)S(t)dW(t)) \\ &= \alpha(t)\sigma^2(t)dt + \sigma^3(t)dW(t) = r\sigma^2(t)dt + \sigma^3(t)d\widetilde{W}(t), \end{aligned}$$

where $\{\widetilde{W}(t)\}_{t\geq 0}$ is a Brownian motion in the risk-neutral probability measure. It follows that

$$d(e^{-rt}\sigma^2(t)) = e^{-rt}\sigma^3(t)d\widetilde{W}(t)$$

and so $\widetilde{\mathbb{E}}[\sigma^2(t)] = \sigma^2(0)e^{rt} = \sigma_0^2 S_0 e^{rt}$. Substituting in the definition of fair strike above we find

$$K_* = \frac{\kappa}{T}\frac{\sigma_0^2 S_0}{r}(e^{rT}-1).$$

42. Solution to Exercise 6.30. Let

$$Q(t) = \int_0^t \sigma(s)^2\,ds.$$

We have $dQ(t) = \sigma(t)^2\,dt$. We compute

$$\begin{aligned} d(e^{-rt}f(t,\sigma^2(t),Q(t))) =& e^{-rt}[-rf\,dt + \partial_t f\,dt + \partial_x f\,d\sigma^2(t) + \frac{1}{2}\partial_x^2 f\,d\sigma^2(t)d\sigma^2(t) \\ &+ \partial_y f\,dQ(t) + \frac{1}{2}\partial_y^2 f\,dQ(t)dQ(t) + \partial_{xy}^2 f\,dQ(t)d\sigma^2(t)] \\ =& e^{-rt}[\partial_t f + a(b-\sigma^2(t))\partial_x f + \sigma^2(t)\partial_y f + \frac{c^2}{2}\sigma^2(t)\partial_x^2 f - rf]dt \\ &+ e^{-rt}c\sigma(t)\partial_x f d\widetilde{W}(t). \end{aligned}$$

where the function f and its derivatives are evaluated at $(t, \sigma^2(t), Q(t))$. As the discounted risk-neutral price must be a martingale in the risk-neutral probability measure, we need the drift term in the above equation to be zero. This is achieved by imposing that f satisfies the PDE

$$\partial_t f + a(b-x)\partial_x f + x\partial_y f + \frac{c^2}{2}x\partial_x^2 f = rf \tag{B.7}$$

Since $\Pi_Y(T) = Y = f(T, \sigma^2(T), Q(T))$, the terminal condition is

$$f(T,x,y) = N\left(\sqrt{\frac{\kappa}{T}y} - K\right)_+ .$$

43. Solution to Exercise 6.36. As

$$d(e^{at}r(t)) = abe^{at}dt + ce^{at}d\widetilde{W}(t),$$

we have

$$r(t) = R_0e^{-at} + b(1-e^{-at}) + \int_0^t ce^{a(s-t)}\,d\widetilde{W}(s).$$

As $r(t)$ is the sum of a deterministic function of time and the Itô integral of a deterministic function of time, then it is normally distributed. As the expectation of the Itô integral is zero, we have

$$\widetilde{\mathbb{E}}[r(t)] = R_0e^{-at} + b(1-e^{-at}).$$

Moreover,

$$\widetilde{\mathrm{Var}}[r(t)] = c^2\widetilde{\mathrm{Var}}[\int_0^t e^{a(s-t)}\,d\widetilde{W}(s)] = c^2\int_0^t e^{2a(s-t)}\,ds = \frac{c^2}{2a}(1-e^{-2at}),$$

where Itô's isometry was used for the second equality. The risk-neutral price of a zero-coupon bond with face-value 1 and expiring at time T is

$$\Pi_Y(t) = \widetilde{\mathbb{E}}[e^{-\int_t^T}r(s)\,ds|\mathcal{F}_W(t)].$$

We make the *ansatz* $\Pi_Y(t) = v(t, r(t))$; assuming that v satisfies the PDE

$$\partial_t v + a(b-x)\partial_x v + \frac{1}{2}c^2\partial_x^2 v = xv, \quad x>0, t\in(0,T) \tag{B.8}$$

we obtain that the stochastic process $\{D(t)v(t,r(t))\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale. Imposing the terminal condition $v(T,x) = 1$, for all $x>0$, we obtain

$$\begin{aligned} D(t)v(t,r(t)) &= \widetilde{\mathbb{E}}[D(T)v(T,R(T))|\mathcal{F}_W(t)] = \widetilde{\mathbb{E}}[D(T)|\mathcal{F}_W(t)] \\ &\Rightarrow v(t,R(T)) = \widetilde{\mathbb{E}}[e^{-\int_t^T r(s)\,ds}|\mathcal{F}_W(t)], \end{aligned}$$

hence $\Pi_Y(t) = v(t,r(t))$ holds. Next we assume the form (6.89) for the solution v. By straightforward calculations one obtains that $v(t,x) = e^{-xC(T-t)-A(T-t)}$ solves (B.8) if and only if the functions A, C satisfy $C' = 1 - aC$ and $A' = abC - c^2C^2/2$. Integrating with the boundary conditions $C(0) = A(0) = 0$ we obtain

$$C(\tau) = \frac{1}{a}(1 - e^{-a\tau}), \tag{B.9a}$$

$$A(\tau) = \frac{e^{-2a\tau}}{4a^3}\left(4a^2be^{a\tau}\left(e^{a\tau}(a\tau - 1) + 1\right) + c^2\left(e^{2a\tau}(3 - 2a\tau) - 4e^{a\tau} + 1\right)\right) \tag{B.9b}$$

44. Solution to Exercise 6.37. Recall that

$$F(t, T) = -\partial_T \log B(t, T) = -\partial_T \log\left(e^{-r(t)C(T-t)-A(T-t)}\right) = r(t)C'(T - t) + A'(T - t),$$

where A, C are given by (B.9). In particular, A, C satisfy

$$C' = 1 - aC, \quad A' = abC - \frac{c^2}{2}C^2. \tag{B.10}$$

Hence

$$\begin{aligned} dF(t, T) &= (abC'(T - t) - ar(t)C'(T - t) - r(t)C''(T - t) - A''(T - t))\, dt \\ &\qquad + cC'(T - t)\, d\widetilde{W}(t) \\ &= c^2C(T - t)C'(T - t)\, dt + cC'(T - t)\, d\widetilde{W}(t), \end{aligned} \tag{B.11}$$

where we used that $C'' = -aC'$ and $A'' = abC' - c^2CC'$ by (B.10). Now recall that in the HJM approach the forward rate satisfies

$$dF(t, T) = \theta(t)\sigma(t, T)\overline{\sigma}(t, T)\, dt + \sigma(t, T)\, dW(t) \tag{B.12}$$

in the physical probability and

$$dF(t, T) = \sigma(t, T)\overline{\sigma}(t, T)\, dt + \sigma(t, T)\, d\widetilde{W}(t) \tag{B.13}$$

in the risk-neutral probability. Here $\{\theta(t)\}_{t\geq 0}$ is the market price of risk used to pass from the physical to the risk-neutral probability and

$$\overline{\sigma}(t, T) = \int_t^T \sigma(t, v)\, dv$$

Comparing (B.11) and (B.13) we find $\sigma(t, T) = cC'(T - t)$ and so $\overline{\sigma}(t, T) = cC(T - t)$. Using the expression for C and replacing in (B.12), we derive the following dynamics for the forward rate:

$$dF(t, T) = \theta(t)\frac{c^2}{a}[e^{-a(T-t)} - e^{-2a(T-t)}]\, dt + ce^{-a(T-t)}\, dW(t).$$

This concludes the third part of the exercise.

45. Solution to Exercise 6.38. Letting $B(t,T) = v(t,r(t))$, and imposing that the drift of $B^*(t,T)$ is zero, we find that $v(t,x)$ satisfies the PDE

$$\partial_t v + a(t)\partial_x v + \frac{c^2}{2}\partial_x^2 v = xv, \quad x > 0, t \in (0,T)$$

with the terminal condition $v(T,x) = 1$. Looking for solutions of the form $v(t,x) = e^{-xC(T,t)-A(T,t)}$ we find that C, A satisfy $C(T,T) = 0$, $A(T,T) = 0$, and

$$\partial_t C = -1, \quad \partial_t A = \frac{c^2}{2}C^2 - Ca(t).$$

Solving the ODE's,

$$C(T,t) = (T-t), \quad A(T,t) = -c^2\frac{(T-t)^3}{6} + \int_t^T (T-s)a(s)\,ds.$$

The forward rate is given by

$$F(t,T) = -\partial_T \log B(t,T) = r(t)\partial_T C(T,t) + \partial_T A(T,t),$$

hence

$$\begin{aligned} dF(t,T) &= dr(t)\partial_T C(T,t) + r(t)\partial_t\partial_T C(T,t)\,dt + \partial_t\partial_T A(T,t)\,dt \\ &= (c^2(T-t))\,dt + c\,d\widetilde{W}(t) \end{aligned}$$

Now, the general form of the forward rate in the HJM approach in the risk-neutral probability measure is

$$dF(t,T) = \sigma(t,T)\overline{\sigma}(t,T)\,dt + \sigma(t,T)\,d\widetilde{W}(t), \quad \text{where} \quad \overline{\sigma}(t,T) = \int_t^T \sigma(t,v)\,dv.$$

Comparing with the expression above we find $\sigma(t,T) = \sigma$, so the dynamics of the forward rate in the physical probability is

$$\begin{aligned} dF(t,T) &= (\theta(t)\sigma(t,T) + \sigma(t,T)\overline{\sigma}(t,T))\,dt + \sigma(t,T)dW(t) \\ &= \sigma(\theta + \sigma(T-t))\,dt + \sigma dW(t), \end{aligned}$$

where $\theta(t)$, the market price of risk, is any adapted process (chosen by calibrating the model).

46. Solution to Exercise 6.39. The last claim follows directly by (6.124). In fact, when the interest rate of the bond is deterministic, the discount process is also deterministic and thus in particular $D(T)$ is $\mathcal{F}_W(t)$-measurable. Hence, the term in curl brackets in (6.124) satisfies

$$\Big\{\dots\Big\} = D(T)\widetilde{\mathbb{E}}[\Pi(T)|\mathcal{F}_W(t)] - D(T)\widetilde{\mathbb{E}}[\Pi(T)|\mathcal{F}_W(t)] = 0.$$

As to (6.124), we compute

$$\begin{aligned}\frac{\Pi(t)}{B(t,T)} - \widetilde{\mathbb{E}}[\Pi(T)|\mathcal{F}_W(t)] &= \frac{D(t)\Pi(t)}{\widetilde{\mathbb{E}}[D(T)|\mathcal{F}_W(t)]} - \widetilde{\mathbb{E}}[\Pi(T)|\mathcal{F}_W(t)] \\ &= \frac{\widetilde{\mathbb{E}}[D(T)\Pi(T)|\mathcal{F}_W(t)]}{\widetilde{\mathbb{E}}[D(T)|\mathcal{F}_W(t)]} - \widetilde{\mathbb{E}}[\Pi(T)|\mathcal{F}_W(t)],\end{aligned}$$

where for the last equality we used that $\{\Pi^*(t)\}_{t\in[0,T]}$ is a $\widetilde{\mathbb{P}}$-martingale relative to $\{\mathcal{F}_W(t)\}_{t\geq 0}$. The result follows.

Bibliography

[1] L. Arnold: *Stochastic differential equations. Theory and applications.* Wiley Interscience (1974)

[2] F. Black, M. Scholes: The Pricing of Options and Corporate Liabilities. The Journal of Political Economy **81**, 637–654 (1973)

[3] D. Brigo, F. Mercurio: *Interest Rate Models - Theory and Practice. Second Ed.*. Springer (2006)

[4] R. Cont, P. Tankov: *Financial Modelling With Jump Processes.* Taylor & Francis (2004)

[5] S. Calogero: *A first course in options pricing theory.* Lecture notes for the course 'Options and Mathematics", Chalmers

[6] K. C. Chan, G. A. Karolyi, F. A. Longstaff, A. B. Sanders. An empirical comparison of alternative models of the short-term interest rate. The Journal of Finance **47**, 1209–1227 (1992)

[7] R. M. Dudley: *Real Analysis and Probability.* Cambridge studies in advanced mathematics (2002)

[8] E. Ekström, P. Lötstedt, J, Tysk: Boundary Values and Finite Difference Methods for the Single Factor Term Structure Equation. Applied Math. Finance **16**, 253–259 (2009)

[9] L. C. Evans: *An introduction to stochastic differential equations.* AMS (2013)

[10] L. C. Evans: *Partial Differential Equations.* AMS (1998)

[11] W. Feller: Two singular diffusion problems. Ann. Math. **54**, 173–182 (1951)

[12] A. Friedman: *Partial Differential Equations of Parabolic type*. Dover Publications, Inc. New York (2008)

[13] J. Gatheral: *The volatility surface*. Wiley Finance (2006)

[14] S. L. Heston: A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. The Review of Financial Studies **6**, 327–343 (1993)

[15] D. J. Higham: An algorithmic introduction to numerical simulations of stochastic differential equations. SIAM review **43**, 525–546 (2001)

[16] L. Hörmander: Hypoelliptic second order differential equations. Acta Math. **119**, 147-171 (1967)

[17] I. Karatzas, S. E. Shreve. *Brownian motion and stochastic calculus.* Springer-Verlag, New York (1988)

[18] I. Karatsas, S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Second Edi- tion, Springer Verlag (1998)

[19] A. G. Z. Kemna, A. C. F. Vorst: A pricing method for options based on average asset values. Journal of Banking and Finance **14**, 113-129 (1990)

[20] D. Lamberton, B. Lapeyre: *Introduction to Stochastic Calculus applied to Finance (2nd ed.).* Chapman & Hall/CRC Financial Mathematics Series (2008)

[21] A. E. Linsday, D. R. Brecher: Simulation of the CEV model and local martingale property.

[22] B. Øksendal: *Stochastic differential equations. An introduction with applications.* 5th ed., Springer-Verlag (2000)

[23] M. M. Rao: *Probability Theory with Applications.* Academic Press (1984)

[24] D. Revuz, M. Yor: Continuous martingales and Brownian motion. Third edition, Springer (2001)

[25] M. Schroder: Computing the Constant Elasticity of Variance Option Pricing Formula. The Journal of Finance **44**, 211-219 (1989)

[26] S. E. Shreve: *Stochastic Calculus for Finance II.* Springer Finance (2008)

[27] A. Timmermann, C. W. J. Granger: Efficient market hypothesis and forecasting. International Journal of Forecasting **20**, 15-27 (2004)