Introduction to Stochastic Calculus

Matthew Lorig [1]

This version: November 28, 2021

[1]Department of Applied Mathematics, University of Washington, Seattle, WA, USA. e-mail: mlorig@uw.edu

Contents

Preface

These notes are intended to give undergraduate and masters students in computational finance an understanding of stochastic calculus. Because the focus of this course is on the *applied* aspects of this topic, we will sometimes forgo mathemical rigor, favoring instead a heuristic development. The mathematical statements in these notes should be taken as "true in spirit," but perhaps not always rigorously true in the mathematical sense. The hope is that, what the notes lack in rigor, they make up in clarity. Each chapter begins with a list of references, which the interested student can go to for rigorously true statements.

It should be noted that these notes are a work in progress. Students are encouraged to e-mail the professor if they find errors.

Acknowledgments

The author of these notes wishes to express his sincere thanks to Weston Barger and Yu-Chen Cheng for checking and writing homework solutions as well making corrections and improvements to the text.

Donations

If you find these notes useful and would like to make a donation to help me develop them further you can donate Bitcoin, Ethereum or Ethereum-based ERC20 tokens to the addresses below.

BTC address: 3QVEVC6QGJ8EmKsnpfXadkshwkDKzCrR4w

ETH address: 0x089d4d033F3E55B80Ac40D5da46812EFeFC8824b

Chapter 1

Review of probability

The notes from this chapter are taken primarily from (Shreve, 2004, Chapter 1) and (Grimmett and Stirzaker, 2001, Chapters 1–5).

1.1 Events as sets

Definition 1.1.1. The set of all possible outcomes of an experiment is called the *sample space*. We denote the sample space as Ω.

We will typically denote by ω a generic element of Ω.

Definition 1.1.2. An *event* is a subset of the sample space. We usually denote events by capital roman letters $A, B, C, \ldots$.

Example 1.1.3 (Toss a coin once). When one tosses a coin, there are two possible outcomes: heads (H) and tails (T). Thus, we have $\Omega = \{H, T\}$. One element of Ω is, e.g., $\omega = H$. Possible event: "toss a head." $A = H$.

Example 1.1.4 (Toss two distinguishable coins). $\Omega = \{(HH), (HT), (HT), (TT)\}$. One element of Ω is, e.g., $\omega = (HT)$. Possible event: ''second toss a tail." $A = \{(HT), (TT)\}$.

Example 1.1.5 (Toss two indistinguishable coins). $\Omega = \{\{HH\}, \{HT\}, \{TT\}\}$. One element of Ω is, e.g., $\{HH\}$. Possible event: ''coins match." $A = \{\{HH\}, \{TT\}\}$.

Example 1.1.6 (Roll a die). $\Omega = \{1, 2, 3, 4, 5, 6\}$. One element of Ω is, e.g., $\omega = 2$. Possible event: "roll an odd number." $A = \{1, 3, 5\}$.

Here, we use $(\cdot)$ to denote an ordered sequence and we use $\{\cdot\}$ to denote an unordered set. Thus, $(\mathrm{HT}) \neq (\mathrm{TH})$ but $\{\mathrm{HT}\} = \{\mathrm{TH}\}$.

If A and B are subsets of Ω, we can reasonably concern ourselves with events such as "not A" (A^c), "A or B" ($\mathrm{A} \cup \mathrm{B}$), "A and B" ($\mathrm{A} \cap \mathrm{B}$), etc. A σ-algebra is a mathematical way to descibe all possible sets of interest for a given sample space Ω.

DEFINITION 1.1.7. A collection $\mathcal{F}$ of subsets of Ω is called a *σ-algebra* if it satisfies

1. contains the empty set: $\emptyset \in \mathcal{F}$;
2. is closed under countable unions: $\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \ldots \in \mathcal{F} \Rightarrow \cup_i \mathrm{A}_i \in \mathcal{F}$;
3. is closed under complements: $\mathrm{A} \in \mathcal{F} \Rightarrow \mathrm{A}^c \in \mathcal{F}$;

Note that if $\mathcal{F}$ is a σ-algebra then $\Omega \in \mathcal{F}$ by items 1 and 3. Note also that $\mathcal{F}$ is closed under countable intersections since

$$\begin{aligned}
\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \ldots \in \mathcal{F} \quad &\Rightarrow \quad \mathrm{A}_1^c, \mathrm{A}_2^c, \mathrm{A}_3^c, \ldots \in \mathcal{F} \\
&\Rightarrow \quad \cup_i \mathrm{A}_i^c \in \mathcal{F} \\
&\Rightarrow \quad \Big(\bigcup_i \mathrm{A}_i^c\Big)^c \in \mathcal{F} \\
&\Rightarrow \quad \bigcap_i \mathrm{A}_i \in \mathcal{F}.
\end{aligned}$$

Alternatively, one can define of a σ-algebra $\mathcal{F}$ as a set of subsets of Ω that contains at least the empty set $\emptyset$ and is closed under countable set operations (though, *not* necessarily closed under *un*countable set operators).

EXAMPLE 1.1.8 (TRIVIAL σ-ALGEBRA). The set of subsets $\mathcal{F}_0 := \{\emptyset, \Omega\}$ of Ω is commonly referred to as the *trivial* σ-algebra.

EXAMPLE 1.1.9. If A is a subset of Ω then $\mathcal{F}_\mathrm{A} := \{\emptyset, \Omega, \mathrm{A}, \mathrm{A}^c\}$ is a σ-algebra.

EXAMPLE 1.1.10. The *power set* of Ω, written 2^Ω is the collection of all subsets of Ω. The power set $\mathcal{F} = 2^\Omega$ is a σ-algebra.

DEFINITION 1.1.11. Let $\mathcal{G}$ be a collection of subsets of Ω. The σ-algebra *generated* by $\mathcal{G}$, written $\sigma(\mathcal{G})$, is the smallest σ-algebra that contains $\mathcal{G}$.

By "smallest" σ-algebra we mean the σ-algebra with the fewest sets. One can show (although we will not do so in these notes) that $\sigma(\mathcal{G})$ is equal to the intersection of all σ-algebras that contain $\mathcal{G}$.

EXAMPLE 1.1.12. The collection of sets $\mathcal{G} = \{\emptyset, A, \Omega\}$ is not a σ-algebra because it does not contain A^c. However, we could create a σ-algebra from $\mathcal{G}$ by simply adding the set A^c. Thus, we have $\sigma(\mathcal{G}) = \{\emptyset, \Omega, A, A^c\}$.

DEFINITION 1.1.13. The pair $(\Omega, \mathcal{F})$ where Ω is a sample space and $\mathcal{F}$ is a σ-algebra of subsets of Ω is called a *measurable space*.

1.2 PROBABILITY

So far, we have not yet talked about probabilities at all – only outcomes of a random experiment (elements $\omega \in \Omega$) and events (subsets $A \subseteq \Omega$). A probability measure assigns probabilities to events.

DEFINITION 1.2.1. A *probability measure* defined on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ that satisfies

1. $\mathbb{P}(\Omega) = 1$;
2. if $A_i \cap A_j = \emptyset$ for $i \neq j$ then $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$. (countable additivity)

As we shall see in Section 1.3.2, it is very important to recognize that Item 2 holds only for *countable* unions. For an uncountable union it is *not true* that $\mathbb{P}(\cup_\alpha A_\alpha) = \sum_\alpha \mathbb{P}(A_\alpha)$.

How can we see the well-known result $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ from the above definition? Simply note that

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

A probability measure $\mathbb{P}$ does *not* need to correspond to empirically observed probabilities! For example, from experience, we know that if we toss a fair coin we have $\mathbb{P}(H) = \mathbb{P}(T) = 1/2$. However, we can always define a measure $\widetilde{\mathbb{P}}$ that assigns different probabilities $\widetilde{\mathbb{P}}(H) = p$ and $\widetilde{\mathbb{P}}(T) = 1 - p$. As long as $p \in [0, 1]$ the measure $\widetilde{\mathbb{P}}$ is a probability measure on $(\Omega, \mathcal{F})$ where $\Omega = \{H, T\}$ and $\mathcal{F} = \{\emptyset, \Omega, H, T\}$.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is often referred to as a *probability space* or *probability triple*. To review, the sample space Ω is the collection of all possible outcomes of an experiment. The σ-algebra $\mathcal{F}$ is all sets of interest of an experiment. And the probability measure $\mathbb{P}$ assigns probabilities to these sets.

When a sample space is countable $\Omega = \{\omega_1, \omega_2, \ldots\}$, we can always take the σ-algebra as the power set $\mathcal{F} = 2^\Omega$ and construct a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ by specifying the probabilities of each individual outcome $\mathbb{P}(\omega_i) = p_i$. However, when the sample space Ω is uncountable, choosing an appropriate σ-algebra $\mathcal{F}$, and constructing a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a more delicate procedure.

1.3 INFINITE PROBABILITY SPACES

In this section, we construct a probability measure on two uncountable sample spaces.

1.3.1 Uniform Lebesgue measure on $(0,1)$

In our first example, we will construct a mathematical model for choosing a number at random on the open interval $(0,1)$. Thus, we take $\Omega = (0,1)$. Note that this sample space is uncountable. A particular outcome is, e.g., $\omega = 0.5$ or $\omega = 1/\sqrt{2}$. Clearly, the number of possible outcomes is infinite. We would like to construct a probability $\mathbb{P}$ so that all numbers in the interval $(0,1)$ are equally likely to be chosen. It makes sense, then, to choose

$$\mathbb{P}(\{\omega : \omega \in (a,b)\}) = \mu((a,b)) := b - a, \qquad 0 < a \leq b < 1, \tag{1.1}$$

where we have defined the *Lebesgue measure* μ. Equation (1.1) tells us how to determine the probability that ω falls within an open interval. But, in fact, (1.1) tells us more than that. If $\mathbb{P}$ is to be a probability measure, then it must satisfy the countable additivity property given in Definition 1.2.1. Thus, we also know, for example, that

$$\begin{aligned}\mathbb{P}(\{\omega : \omega \in (a,b) \cup (c,d)\}) &= \mathbb{P}(\{\omega : \omega \in (a,b)\}) + \mathbb{P}(\{\omega : \omega \in (c,d)\}) \\ &= (b-a) + (d-c), \qquad 0 < a < b < c < d < 1.\end{aligned}$$

It is natural to ask, *what are all of the sub-sets of* $(0,1)$ *whose probabilities are determined by* (1.1) *and the properties of probability measures given in Definition 1.2.1*? Surprisingly, the answer is *not* the power set $2^{(0,1)}$. It turns out that the power set $2^{(0,1)}$ has sets whose probabilities are *not* determined by (1.1). The sets whose probabilities *are* uniquely determined by (1.1) and Definition 1.2.1 are the sets in σ-algebra generated by the open intervals

$$\mathcal{B}((0,1)) := \sigma(\mathcal{O}), \qquad \text{where} \qquad \mathcal{O} := \{A \subseteq (0,1) : A = (a,b),\ 0 \leq a < b \leq 1\}.$$

We call $\mathcal{B}((0,1))$ the *Borel σ-algebra on* $(0,1)$. Thus, the appropriate sample space for our experiment is $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}((0,1)), \mu)$.

We can generalize the notion of Borel sets to all topological spaces.

Definition 1.3.1. Let Ω be some topological space and let $\mathcal{O}(\Omega)$ be the set of open sets in Ω. We define we define the *Borel σ-algebra on* Ω, denoted $\mathcal{B}(\Omega)$ by $\mathcal{B}(\Omega) := \sigma(\mathcal{O}(\Omega))$.

Remark 1.3.2. Do not worry too much about what exactly Borel σ-algebras are. Just think of them as "reasonable" sets. In fact, you would have to think very hard to come up with a set that is not a Borel set.

1.3.2 Infinite sequence of coin tosses

In this section we consider and infinite sequence of coin tosses. We define

$$\Omega := \text{ the set of infinite sequences of Hs and Ts .}$$

Note that this set is uncountable because there is a one-to-one correspondence between Ω and the set of reals in $[0,1]$. We will denote a generic element of Ω as follows:

$$\omega = \omega_1\omega_2\omega_3\ldots$$

where ω_i is the result of the ith coin toss. We want to construct a σ-algebra for this experiment.

Let us define some σ-algebras. First, consider the trivial σ-algebra

$$\mathcal{F}_0 = \{\emptyset, \Omega\}.$$

Given no information, I can tell if ω is in the sets in $\mathcal{F}_0$ because we know $\omega \in \Omega$ and $\omega \notin \emptyset$. Next, define two sets

$$\mathrm{A_H} = \{\omega \in \Omega : \omega_1 = \mathrm{H}\}, \qquad \mathrm{A_T} = \{\omega \in \Omega : \omega_1 = \mathrm{T}\}.$$

Noting that $\mathrm{A_H} = \mathrm{A_T^c}$ we see that

$$\mathcal{F}_1 := \{\emptyset, \Omega, \mathrm{A_H}, \mathrm{A_T}\},$$

satisfies the conditions of σ-algebra. Given ω_1 it is possible to say whether or not ω is in each of the sets in $\mathcal{F}_1$. For example, if $\omega_1 = \mathrm{H}$ then $\omega \in \mathrm{A_H}$ and $\omega \in \Omega$, but $\omega \notin \mathrm{A_T}$ and $\omega \notin \emptyset$. Next define four sets

$$\mathrm{A_{HH}} := \{\omega \in \Omega : \omega_1 = \mathrm{H}, \omega_2 = \mathrm{H}\}, \qquad \mathrm{A_{HT}} := \{\omega \in \Omega : \omega_1 = \mathrm{H}, \omega_2 = \mathrm{T}\},$$
$$\mathrm{A_{TT}} := \{\omega \in \Omega : \omega_1 = \mathrm{T}, \omega_2 = \mathrm{T}\}, \qquad \mathrm{A_{TH}} := \{\omega \in \Omega : \omega_1 = \mathrm{T}, \omega_2 = \mathrm{H}\}.$$

We wish to construct a σ-algebra that contains these sets and the sets in $\mathcal{F}_1$. The smallest such σ-algebra is

$$\mathcal{F}_2 = \left\{ \begin{matrix} \emptyset, \Omega, \mathrm{A_H}, \mathrm{A_T}, \mathrm{A_{HH}}, \mathrm{A_{HT}}, \mathrm{A_{TT}}, \mathrm{A_{TH}}, \mathrm{A_{HH}^c}, \mathrm{A_{HT}^c}, \mathrm{A_{TT}^c}, \mathrm{A_{TH}^c} \\ \mathrm{A_{HH}} \cup \mathrm{A_{TH}}, \mathrm{A_{HH}} \cup \mathrm{A_{TT}}, \mathrm{A_{HT}} \cup \mathrm{A_{TH}}, \mathrm{A_{HT}} \cup \mathrm{A_{TT}} \end{matrix} \right\}.$$

Given ω_1 and ω_2, we can say if ω belongs to each of the sets in $\mathcal{F}_2$. Continuing in this way, we can define a σ-algebra $\mathcal{F}_n$ for every $n \in \mathbb{N}$. Finally, we take

$$\mathcal{F} := \sigma(\mathcal{F}_\infty), \qquad \mathcal{F}_\infty = \cup_n \mathcal{F}_n.$$

One might ask if we could have simply taken $\mathcal{F} = \mathcal{F}_\infty$? Well, $\mathcal{F}_\infty$ contains every set that can be described in terms of *finitely many* coin tosses. However, we may be interested in sets such as "sequences for which x percent of coin tosses are heads," and these sets are not in $\mathcal{F}_\infty$. It turns out such sets are in $\mathcal{F}$.

Now, we want to construct a probability measure on $\mathcal{F}$. Let us assume the coin tosses are independent (a term we will describe rigorously later on) and that the probability of a head is p. Setting $q = 1 - p$, it should be obvious that

$$\begin{aligned} \mathbb{P}(\emptyset) &= 0, & \mathbb{P}(\Omega) &= 1, & \mathbb{P}(A_H) &= p, & \mathbb{P}(A_T) &= q, \\ \mathbb{P}(A_{HH}) &= p^2, & \mathbb{P}(A_{HT}) &= pq, & \mathbb{P}(A_{TH}) &= pq, & \mathbb{P}(A_{TT}) &= q^2, \ldots \end{aligned}$$

Continuing in this way, we can define $\mathbb{P}(A)$ for every $A \in \mathcal{F}_\infty$. What about the sets that are in $\mathcal{F}$ but not in $\mathcal{F}_\infty$? It turns out that once we have defined $\mathbb{P}$ for sets in $\mathcal{F}_\infty$ there is only one way to assign probabilities to those sets that are in $\mathcal{F}$ but not in $\mathcal{F}_\infty$. We refer the interested reader to *Carathéodory's Extension Theorem* for details.

Now, let us define

$$A = \left\{ \omega : \lim_{n\to\infty} \frac{\#\text{H in first } n \text{ coin tosses}}{n} = \frac{1}{2} \right\}.$$

The strong law of large numbers (SLLN) tells us that $\mathbb{P}(A) = 1$ if $p = 1/2$ and $\mathbb{P}(A) = 0$ if $p \neq 1/2$ (if you have not yet seen the SLLN, you should be able to see this from intuition). Now it should be clear why *uncountable* additivity does *not* hold for probability measures. The probability of any given sequence of infinite coin tosses is zero: $\mathbb{P}(\omega) = 0$. If we were to attempt to compute $\mathbb{P}(A)$ by adding up the probabilities $\mathbb{P}(\omega)$ of all elements $\omega \in A$ we would find

$$\sum_{\omega\in A} \mathbb{P}(\omega) = \sum_{\omega\in A} 0 = 0 \neq 1 = \mathbb{P}(A), \qquad \text{(when } p = 1/2\text{)}.$$

Thus, uncountable additivity clearly does *not* hold.

We finish this example (we will come back to it!) with the following definition

DEFINITION 1.3.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If a set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say that the event A occurs $\mathbb{P}$ *almost surely* (written, $\mathbb{P}$-a.s.).

Note in the example above that, when $p = 1/2$ we have $\mathbb{P}(A) = 1$ and thus A occurs almost surely. But it is important to recognize that $A \neq \Omega$ and $A^c \neq \emptyset$. The elements of A^c are part of the sample space Ω, but they have zero probability of occurring.

1.4 Random variables and distributions

A random variable maps the outcome of an experiment to $\mathbb{R}$. We capture this idea with the following definition.

Definition 1.4.1. A *random variable* defined on $(\Omega, \mathcal{F})$ is a function $X : \Omega \to \mathbb{R}$ with the property that

$$\{X \in A\} := \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F},$$

for all $A \in \mathcal{B}(\mathbb{R})$.

Observe all any random variables must be defined on a measurable space $(\Omega, \mathcal{F})$, as these appear in the definition. Note, however, that the probability meausre $\mathbb{P}$ does *not* appear in the definition. Random variables are defined *independent* of a probability measure $\mathbb{P}$.

What does Definition 1.4.1 mean? Recall that a probability measure $\mathbb{P}$ defined on $(\Omega, \mathcal{F})$ maps $\mathcal{F} \to [0, 1]$. In order for us to answer the question: "what is the probability that $X \in A$?" we need for the set $\{X \in A\} \in \mathcal{F}$. And this is precisely what Definition 1.4.1 requires. Why do we only consider sets $A \in \mathcal{B}(\mathbb{R})$ rather than any set $A \subset \mathbb{R}$? The answer is rather technical and, frankly, not worth exploring at the moment.

A word on notation: the standard convention is to use capital Roman letters (typically, X, Y, Z) for random variables and lower case Roman letters (x, y, z) for real numbers.

Let us us look at some random variables.

Example 1.4.2 (Discrete time model for stock prices). Consider the infinite sequence of coin tosses in Section 1.3.2. We Define a sequence of random variables $(S_n)_{n\geq 0}$ via

$$S_0(\omega) = 1, \qquad S_{n+1}(\omega) = \begin{cases} uS_n & \text{if } \omega_n = H \\ dS_n & \text{if } \omega_n = T \end{cases} \tag{1.2}$$

Here, S_n represents the value of a stock at time n. Note that $\mathbb{P}(S_1 = u) = \mathbb{P}(A_H) = p$. Likewise $\mathbb{P}(S_2 = ud) = \mathbb{P}(A_{HT} \cup A_{TH}) = 2pq$. More generally, one can show that

$$\mathbb{P}(S_n = u^k d^{n-k}) = \binom{n}{k} p^k q^{n-k}. \tag{1.3}$$

Note if we had simply defined the random variables $(S_n)_{n\geq 1}$ as having probabilities given by (1.3) we would have no information about how, e.g., S_n relates to S_{n-1}. From the above construction (1.2), however, we know that if $S_n = u^n$ then $S_{n-1} = u^{n-1}$. Thus, the structure of a given probability space, not just the probabilities of events, is very important.

EXAMPLE 1.4.3. Let $(\Omega, \mathcal{F}) = ((0,1), \mathcal{B}((0,1))$ Define random variables $X(\omega) = \omega$ and $Y(\omega) = 1 - \omega$. Clearly, we have $X = 1 - Y$. Now, suppose we defined $\mathbb{P}(d\omega) := d\omega$. Then X and Y have the same distribution. For $x \in [0,1]$ we have

$$\mathbb{P}(X \leq x) = \mathbb{P}(\omega \leq x) = \int_0^x \mathbb{P}(d\omega) = \int_0^x d\omega = x,$$
$$\mathbb{P}(Y \leq x) = \mathbb{P}(1 - \omega \leq x) = \int_{1-x}^1 \mathbb{P}(d\omega) = \int_{1-x}^1 d\omega = x.$$

However, if we defined a new probability measure via $\widetilde{\mathbb{P}}(d\omega) = 2\omega d\omega$ then X and Y have different distributions. For $x \in [0,1]$ we have

$$\mathbb{P}(X \leq x) = \mathbb{P}(\omega \leq x) = \int_0^x \widetilde{\mathbb{P}}(d\omega) = \int_0^x 2\omega d\omega = x^2,$$
$$\mathbb{P}(Y \leq x) = \mathbb{P}(1 - \omega \leq x) = \int_{1-x}^1 \widetilde{\mathbb{P}}(d\omega) = \int_{1-x}^1 2\omega d\omega = 1 - (1-x)^2.$$

The distribution of a random variable X is most easily described through its cumulative distribution function.

DEFINITION 1.4.4. The *distribution function* $F_X : \mathbb{R} \to [0,1]$ of a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given by

$$F_X(x) := \mathbb{P}(X \leq x).$$

Observe that, while a random variable X is defined with respect to $(\Omega, \mathcal{F})$ (with no reference to $\mathbb{P}$), the distribution F_X is specific to a probability measure $\mathbb{P}$.

Note that we put the random variable X in the subscript of F_X to remind us that F_X is the distribution function corresponding to the random variable X (and not, e.g., Y). It is a good idea to do this.

We give here some obvious properties of F_X.

1. We have the following limits: $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
2. F_X is non-decreasing: $x < y$ implies $F_X(x) \leq F_X(y)$.
3. F_X is right-continuous and has left-limits: $F_X(x+) := \lim_{h \searrow 0} F_X(x+h) = F_X(x)$.

Many (but not all) random variables fall in to one of two categories: discrete and continuous. We describe these two categories below.

DEFINITION 1.4.5. A random variable X is called *discrete* if it takes values in some countable set $A := \{x_1, x_2, \ldots\} \subset \mathbb{R}$. We associate is a discrete random variable a *probability mass function* $f_X : A \to \mathbb{R}$, defined by $f_X(x_i) := \mathbb{P}(X = x_i)$.

DEFINITION 1.4.6. A random variable X is called *continuous* if its distribution function F_X can be written as

$$F_X(x) = \int_{-\infty}^{x} \mathrm{d}u\, f_X(u), \qquad x \in \mathbb{R},$$

for some $f_X : \mathbb{R} \to [0, \infty)$ called the *probability density function*.

It may help to think of the density function f_X as $f_X(x)\mathrm{d}x = \mathbb{P}(X \in \mathrm{d}x)$.

Note that for a continuous random variable X we have $f_X = F_X'$.

If X is either discrete or continuous, it is easy to compute $\mathbb{P}(X \in A)$ for any $A \in \mathcal{B}(\mathbb{R})$. We have

$$\text{discrete :} \qquad \mathbb{P}(X \in A) = \sum_{\{i : x_i \in A\}} f_X(x_i),$$

$$\text{continuous :} \qquad \mathbb{P}(X \in A) = \int_A \mathrm{d}x\, f_X(x).$$

REMARK 1.4.7. Although we have defined $F_X : \mathbb{R} \to [0, 1]$ by $F_X(x) := \mathbb{P}(X \leq x)$, it is common to also to utilize F_X as a set function $F_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$, which means $F_X(B) := \mathbb{P}(X \in B)$. It should always be clear from the argument of F_X, which of the two meanings we intend.

EXAMPLES OF DISCRETE RANDOM VARIABLES

The following discrete random variables frequently arise in applications in nature and social sciences.

EXAMPLE 1.4.8. If X is distributed as a *Bernoulli random variable* with parameter $p \in [0, 1]$, written $X \sim \mathrm{Ber}(p)$, then

$$X \in \{0, 1\}, \qquad f_X(k) = \begin{cases} 1 - p & k = 0, \\ p & k = 1. \end{cases}$$

EXAMPLE 1.4.9. If X is distributed as a *Binomial random variable* with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$, written $X \sim \mathrm{Bin}(n, p)$, then

$$X \in \{0, 1, 2, \ldots, n\}, \qquad f_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Note that if $X_i \sim \mathrm{Ber}(p)$ and independent of each other then $Y := \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n, p)$.

EXAMPLE 1.4.10. If X is distributed as a *Geometric random variable* with parameter $p \in [0, 1]$, written $X \sim \mathrm{Geo}(p)$, then

$$X \in \mathbb{N}, \qquad f_X(k) = p(1 - p)^{k-1}.$$

Note that $X_i \sim \mathrm{Ber}(p)$ $i = 1, 2, \ldots$ are independent of each other then $Y := \inf\{i : X_i = 1\} \sim \mathrm{Geo}(p)$.

EXAMPLE 1.4.11. If X is distributed as a *Poisson random variable* with parameter $\lambda \in \mathbb{R}_+$, written $X \sim \text{Poi}(\lambda)$, then

$$X \in \{0\} \cup \mathbb{N}, \qquad f_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

EXAMPLES OF CONTINUOUS RANDOM VARIABLES

Before introducing some common continuous random variables, let us introduce a useful function.

DEFINITION 1.4.12. Let A be a set in some topological space Ω (e.g., $\Omega = \mathbb{R}^d$). The *indicator function* $\mathbb{1}_A : \Omega \to \{0, 1\}$ is defined as follows

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Notation: We will sometimes write $\mathbb{1}_A(x) = \mathbb{1}_{\{x \in A\}}$.

We now introduce some continuous random variables that frequently arise in applications.

EXAMPLE 1.4.13. If X is distributed as a *Uniform random variable* on the interval $[a, b] \subset \mathbb{R}$, written $X \sim \mathcal{U}([a, b])$, then

$$X \in [a, b], \qquad f_X(x) = \mathbb{1}_{[a,b]}(x) \frac{1}{b - a}.$$

EXAMPLE 1.4.14. If X is distributed as a *Exponential random variable* with mean $\lambda > 0$, written $X \sim \mathcal{E}(\lambda)$, then

$$X \in [0, \infty), \qquad f_X(x) = \mathbb{1}_{[0,\infty)}(x) \lambda e^{-\lambda x}.$$

Note that $f_X(x) = 0$ if $x < 0$ due to the presence of the indicator function.

EXAMPLE 1.4.15. If X is distributed as a *Gaussian* or *Normal random variable* with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ (we will give a meaning for "mean" and "variance" below), written $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$X \in \mathbb{R}, \qquad f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2} \right).$$

A random variable $Z \sim \mathcal{N}(0, 1)$ is referred to as *standard normal*.

1.5 Stochastic Processes

Intuitively, we think of a stochastic process as a process that evolves randomly in time. Now that we understand what a random variable is, we can define rigourously what we mean when we say *stochastic process*.

Definition 1.5.1. A *Stochastic process* is a collection of random variables $X = (X_t)_{t\in\mathbb{T}}$ where $\mathbb{T}$ is some index set. If the index set $\mathbb{T}$ is countable (e.g., $\mathbb{T} = \mathbb{N}_0$) we say that X is a *discrete time* process. If the index set $\mathbb{T}$ is uncountable (e.g., $\mathbb{T} = \mathbb{R}_+$) we say that X is a *continuous time* process. The *State Space* S of a stochastic process X is union of the state spaces of $(X_t)_{t\in\mathbb{T}}$.

We can think of a stochastic process $X : \mathbb{T} \times \Omega \to \mathbb{R}$ in (at least) two ways. First, for any $t \in \mathbb{T}$ we have that $X_t : \Omega \to \mathbb{R}$ is random variable. Second, for any $\omega \in \Omega$, we have that $X_\cdot(\omega) : \mathbb{T} \to \mathbb{R}$ is a function of time. Both interpretations can be useful.

1.6 Expectation

When we think of averaging we think of weighting outcomes by their probabilities. The mathematical way to encode this is via the expectation.

Definition 1.6.1. Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The *expectation* of X, writtien $\mathbb{E}X$, is defined as

$$\mathbb{E}X := \int_\Omega X(\omega)\mathbb{P}(d\omega),$$

where the integral is understood in the Lebesgue sense.

1.6.1 Integration in the Lebesgue sense

For those who have not previously encountered Lesbesgue integration, we now give a brief (*very brief*!) overview of this concept.

Definition 1.6.2. Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $A \in \mathcal{F}$. The *indicator random variable*, denoted $\mathbb{1}_A$, is defined by

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

Observe that $\mathbb{1}_A \sim \text{Ber}(p)$ with $p = \mathbb{P}(A)$. For disjoint sets A and B we have

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B, \qquad A \cap B = \emptyset.$$

And, for any two sets A and B we have

$$\mathbb{1}_{A\cap B} = \mathbb{1}_A \mathbb{1}_B.$$

DEFINITION 1.6.3. A collection of non-empty sets (A_i) is said to be a *partition* of Ω if $A_i \cap A_j \neq \emptyset$ for all i and j and $\cup_i A_i = \Omega$.

DEFINITION 1.6.4. Let (A_i) be a finite partition of Ω. A non-negative random variable X, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which is of the form

$$X(\omega) = \sum_{i=1}^{n} x_i \mathbb{1}_{A_i}(\omega), \qquad x_i \geq 0, \qquad A_i \in \mathcal{F},$$

is called *simple*.

Let X be a simple random variable. We define the expectation of X as follows

$$\mathbb{E}X := \sum_{i=1}^{n} x_i \mathbb{P}(A_i). \qquad \text{(if X is simple)}$$

Note that, from this definition, we have

$$\mathbb{E}\mathbb{1}_A = \mathbb{P}(A).$$

Thus, *we can always represent probabilities of sets as expectations of indicator random variables.*

Now, consider a non-negative random variable X, which is not necessarily simple. Let $(X_n)_{n\geq 0}$ be an increasing sequence of simple random variables that converges almost surely to X. That is

$$X_i \leq X_{i+1}, \qquad \lim_{i\to\infty} X_i \to X, \qquad \mathbb{P}\text{-a.s.}.$$

We define the expectation of a non-negative random variable X as the following limit

$$\mathbb{E}X := \lim_{i\to\infty} \mathbb{E}X_i, \qquad \text{(if X is non-negative)} \tag{1.4}$$

where each of the expectations on the right-hand side are well-defined because all of the X_i are simple by construction. Finally, consider a general random variable X that could take either positive or negative values. Define

$$X^+ = \max\{X, 0\}, \qquad X^- = \max\{-X, 0\}.$$

Note that X^+ and X^- are non-negative and $X = X^+ - X^-$. With this in mind, we define

$$\mathbb{E}X := \mathbb{E}X^+ - \mathbb{E}X^-,$$

where the expectations of X^+ and X^- are defined via (1.4).

Definition 1.6.1 of $\mathbb{E}X$ makes sense if $\mathbb{E}|X| < \infty$ or if $\mathbb{E}X^{\pm} = \infty$ and $\mathbb{E}X^{\mp} < \infty$. In the latter case, we have $\mathbb{E}X = \pm\infty$. If both $\mathbb{E}X^+ = \infty$ and $\mathbb{E}X^- = \infty$, then we find ourselves in an $\infty - \infty$ situation and, in this case, $\mathbb{E}X$ is undefined.

1.6.2 Computing expectations

If X is either discrete or continuous Definition 1.6.1 reduces to the formulas one learns as an undergraduate.

$$\text{discrete :} \qquad \mathbb{E}X = \sum_i x_i \, f_X(x_i),$$

$$\text{continuous :} \qquad \mathbb{E}X = \int_{\mathbb{R}} \mathrm{d}x \, x \, f_X(x).$$

In the discrete case, the sum runs over all possible values of x. More generally, we can express the expected value of X as

$$\mathbb{E}X = \int_{\mathbb{R}} x \, F_X(\mathrm{d}x) := \lim_{\|\Pi\| \to 0} \sum_i \left(\frac{x_i + x_{i+1}}{2} \right) \Big(F_X(x_{i+1}) - F_X(x_i) \Big), \tag{1.5}$$

where Π is a partition of $\mathbb{R}$, meaning

$$\Pi = \{x_i, x_2, \ldots, x_n\}, \qquad x_i < x_{i+1} \, \forall i, \qquad \|\Pi\| := \sup_i (x_{i+1} - x_i).$$

The expression on the righ-hand side of (1.5) is known as a *Stieltjes integral*. The advantage of using the Stieltjes integral $\int x F_X(\mathrm{d}x)$ to compute an expectation is that every random variable X – whether it be discrete or continuous – has a distribution F_X. Thus, by usinig the Stieltjes integral, we avoid having to treat discrete and continuous cases separately.

Note that $\mathbb{E}$ is a linear operator. If X and Y are random variables and a and b are constants, then

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y.$$

How does one compute $\mathbb{E}g(X)$ where $g : \mathbb{R} \to \mathbb{R}$? Although we have not stated it explicitly, it should be obvious that if X is a random variable, then $Y := g(X)$ is also a random variable. [1] Thus, we have

$$\mathbb{E}Y = \mathbb{E}g(X) = \int_\Omega g(X(\omega)) \mathbb{P}(\mathrm{d}\omega),$$

[1] Rigorously, g should be a *measurable function*, meaning $g^{-1}(A) \in \mathcal{B}(\mathbb{R})$ for all $A \in \mathcal{B}(\mathbb{R})$. Do not concern yourself too much with this.

which in the discrete and continuous cases become

$$\text{discrete :} \qquad \mathbb{E}g(X) = \sum_i g(x_i) f_X(x_i),$$

$$\text{continuous :} \qquad \mathbb{E}g(X) = \int_{\mathbb{R}} dx\, g(x) f_X(x).$$

1.7 Change of measure

Consider two probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ defined on a measurable space $(\Omega, \mathcal{F})$. What is the relation between $\mathbb{P}$ and $\widetilde{\mathbb{P}}$? The following theorem answers this question.

Theorem 1.7.1. *Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $Z \geq 0$ be a random variable satisfying $\mathbb{E}Z = 1$. Define a $\widetilde{\mathbb{P}} : \mathcal{F} \to [0,1]$ by*

$$\widetilde{\mathbb{P}}(A) := \mathbb{E}Z\mathbb{1}_A. \tag{1.6}$$

Then $\widetilde{\mathbb{P}}$ is a probability measure on $(\Omega, \mathcal{F})$. Denote by $\widetilde{\mathbb{E}}$ the expectation taken with respect to $\widetilde{\mathbb{P}}$. Then

$$\widetilde{\mathbb{E}}X = \mathbb{E}ZX, \qquad \text{and if} \qquad Z > 0, \qquad \text{then} \qquad \mathbb{E}X = \widetilde{\mathbb{E}}\frac{1}{Z}X. \tag{1.7}$$

where X is a random variable defined on $(\Omega, \mathcal{F})$.

Definition 1.7.2. We call the random variable Z in Theorem 1.7.1 the *Radon-Nikodým derivative* of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$.

Proof of Theorem 1.7.1. First, we show that $\widetilde{\mathbb{P}}$ satisfies the properties of a probability measure given in Definition 1.2.1. Let us check that $\widetilde{\mathbb{P}}(\Omega) = 1$. We have

$$\widetilde{\mathbb{P}}(\Omega) = \mathbb{E}Z\mathbb{1}_\Omega = \mathbb{E}Z = 1.$$

Next, we check that $\widetilde{\mathbb{P}}$ satisfies countable additivity. Let $(A_i)_{i\geq 0}$ be a sequence of disjoint sets. We have

$$\widetilde{\mathbb{P}}(\cup_i A_i) = \mathbb{E}\mathbb{1}_{\cup_i A_i} Z = \sum_i \mathbb{E}\mathbb{1}_{A_i} Z = \sum_i \widetilde{\mathbb{P}}(A_i).$$

Note that interchanging the sum with the expectation $\mathbb{E}\sum_i \to \sum_i \mathbb{E}$ is allowed by Tonelli's Theorem. Finally, to show equation (1.7) holds, it is enough to check that it holds for simple random variables $X = \sum_i x_i \mathbb{1}_{A_i}$. We have

$$\widetilde{\mathbb{E}}X = \widetilde{\mathbb{E}}\sum_i x_i \mathbb{1}_{A_i} = \sum_i x_i \mathbb{E}Z\mathbb{1}_{A_i} = \sum_i x_i \widetilde{\mathbb{P}}(A_i),$$

which agrees with the definition of Expectation for simple random variables. Finally, if $Z > 0$, we have

$$\widetilde{\mathbb{E}}\frac{1}{Z}X = \mathbb{E}Z\frac{1}{Z}X = \mathbb{E}X.$$

□

DEFINITION 1.7.3. A probability measure $\mathbb{P}$ defined on $(\Omega, \mathcal{F})$ is *absolutely continuous* with respect to another probability measure $\widetilde{\mathbb{P}}$, written $\mathbb{P} \gg \widetilde{\mathbb{P}}$, if

$$\mathbb{P}(A) = 0 \qquad \Rightarrow \qquad \widetilde{\mathbb{P}}(A) = 0.$$

In order for there to exist a Radon-Nikodým derivative $Z = d\widetilde{\mathbb{P}}/d\mathbb{P}$ we must have $\mathbb{P} \gg \widetilde{\mathbb{P}}$. The reason for this is that the relationship between $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ is multiplicative $\widetilde{\mathbb{P}}(d\omega) = Z(\omega)\mathbb{P}(d\omega)$. If $\mathbb{P}(A) = 0$, then there is no random variable Z that would result in $\widetilde{\mathbb{P}}(A) > 0$.

DEFINITION 1.7.4. Two probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are *equivalent*, written $\mathbb{P} \sim \widetilde{\mathbb{P}}$, if

$$\mathbb{P}(A) = 0 \qquad \Leftrightarrow \qquad \widetilde{\mathbb{P}}(A) = 0.$$

Two probability measures are equivalent $\mathbb{P} \sim \widetilde{\mathbb{P}}$ if and only if the Radon-Nikodým Derivative that relates them is strictly positive $Z > 0$. Equivalent measures agree on which events will happen with probability zero (and thus, they agree on which events will happen with probability one).

EXAMPLE 1.7.5. Let us return to Example 1.4.3. We set $(\Omega, \mathcal{F}) = ((0,1), \mathcal{B}((0,1)))$. On this measure space, we define two probability measures $\mathbb{P}(d\omega) = d\omega$ and $\widetilde{\mathbb{P}}(d\omega) = 2\omega d\omega$. Note that we have

$$\widetilde{\mathbb{P}}(A) = \widetilde{\mathbb{E}}\mathbb{1}_A = \int_\Omega \mathbb{1}_A(\omega)\widetilde{\mathbb{P}}(d\omega) = \int_\Omega \mathbb{1}_A(\omega)2\omega d\omega = \int_\Omega \mathbb{1}_A 2\omega \mathbb{P}(d\omega) = \mathbb{E}\mathbb{1}_A Z, \qquad Z(\omega) := 2\omega.$$

One can easily check that $\mathbb{E}Z = 1$ and $Z > 0$. Defining $\widetilde{\mathbb{P}}$ by (1.6), one can easily check that (1.7) holds true.

It is quite common to use the notation

$$Z(\omega) = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}(\omega), \qquad \widetilde{\mathbb{P}}(d\omega) = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}(\omega)\mathbb{P}(d\omega),$$

as a reminder of how the Radon-Nikodým Derivative Z relates $\widetilde{\mathbb{P}}$ to $\mathbb{P}$. For a finite probability space, it is true that $Z(\omega) = \widetilde{\mathbb{P}}(\omega)/\mathbb{P}(\omega)$. However, for an infinite probability space, it makes no sense in general to define $Z(\omega) = \widetilde{\mathbb{P}}(\omega)/\mathbb{P}(\omega)$ since it may be that $\mathbb{P}(\omega) = 0$. Nevertheless, the heuristic $Z(\omega) = \widetilde{\mathbb{P}}(\omega)/\mathbb{P}(\omega)$ gives the correct intuition. In particular, for the special case of an infinite probability space in which $\mathbb{P}(d\omega) = p(\omega)d\omega$ and $\widetilde{\mathbb{P}}(d\omega) = \widetilde{p}(\omega)d\omega$ and $\mathbb{P} \gg \widetilde{\mathbb{P}}$, we have $Z(\omega) = \widetilde{p}(\omega)/p(\omega)$.

Example 1.7.6 (Change of measure Normal random variable). On $(\Omega, \mathcal{F}, \mathbb{P})$ let $X \sim \mathcal{N}(0,1)$ and define $Y = X + \theta$. Clearly, we have $Y \sim \mathcal{N}(\theta, 1)$. Now, define a random variable Z by

$$Z = e^{-\theta X - \frac{1}{2}\theta^2}.$$

Clearly $Z > 0$. We also have $\mathbb{E}Z = 1$. To see this, simply compute

$$\begin{aligned}\mathbb{E}Z &= \int_{\mathbb{R}} dx\, e^{-\theta x - \frac{1}{2}\theta^2} f_X(x) \\ &= \int_{\mathbb{R}} dx\, e^{-\theta x - \frac{1}{2}\theta^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\ &= \int_{\mathbb{R}} dx\, \frac{1}{\sqrt{2\pi}} e^{-(x+\theta)^2/2} = 1.\end{aligned}$$

Since $Z > 0$ and $\mathbb{E}Z = 1$, we can define a new probability measure $\widetilde{\mathbb{P}}$ with $Z = d\widetilde{\mathbb{P}}/d\mathbb{P}$ as the Radon-Nikodým derivative. Let us compute the distribution of Y under $\widetilde{\mathbb{P}}$. We have

$$\begin{aligned}\widetilde{\mathbb{P}}(Y \le b) &= \mathbb{E}Z\mathbb{1}_{\{Y \le b\}} = \mathbb{E}e^{-\theta X - \frac{1}{2}\theta^2}\mathbb{1}_{\{X \le b-\theta\}} \\ &= \int_{-\infty}^{b-\theta} dx\, e^{-\theta x - \frac{1}{2}\theta^2} f_X(x) \\ &= \int_{-\infty}^{b-\theta} dx\, \frac{1}{\sqrt{2\pi}} e^{-(x+\theta)^2/2} \\ &= \int_{-\infty}^{b} dz\, \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\end{aligned}$$

Thus, under $\widetilde{\mathbb{P}}$ we see that $Y \sim \mathcal{N}(0,1)$. The Radon-Nikodým derivative Z changes the mean of Y from θ to 0, but it does not affect the variance of Y.

Example 1.7.7 (Change of measure Exponential random variable). On $(\Omega, \mathcal{F}, \mathbb{P})$, let $X \sim \mathcal{E}(\lambda)$ and define

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

Clearly, we have $Z \ge 0$. We also have

$$\begin{aligned}\mathbb{E}Z = \mathbb{E}\frac{\mu}{\lambda}e^{-(\mu-\lambda)X} &= \int_0^\infty dx\, \frac{\mu}{\lambda} e^{-(\mu-\lambda)x} f_X(x) \\ &= \int_0^\infty dx\, \frac{\mu}{\lambda} e^{-(\mu-\lambda)x} \lambda e^{-\lambda x} = \int_0^\infty dx\, \mu e^{-\mu x} = 1.\end{aligned}$$

Thus, we can define a new probability measure $\widetilde{\mathbb{P}}$ with $Z = d\widetilde{\mathbb{P}}/d\mathbb{P}$ as the Radon-Nikodým derivative. Let us compute the distribution of X under $\widetilde{\mathbb{P}}$. We have

$$\begin{aligned}\widetilde{\mathbb{P}}(X \le b) &= \mathbb{E}Z\mathbb{1}_{\{X \le b\}} = \mathbb{E}\frac{\mu}{\lambda}e^{-(\mu-\lambda)X}\mathbb{1}_{\{X \le b\}} \\ &= \int_0^b \frac{\mu}{\lambda} e^{-(\mu-\lambda)x} f_X(x) = \int_0^b dx\, \frac{\mu}{\lambda} e^{-(\mu-\lambda)x} \lambda e^{-\lambda x} = \int_0^b dx\, \mu e^{-\mu x}\end{aligned}$$

Thus, under $\widetilde{\mathbb{P}}$, we have $X \sim \mathcal{E}(\mu)$.

1.8 Exercises

Exercise 1.1. Let $\mathcal{F}$ be a σ-algebra of Ω. Suppose $B \in \mathcal{F}$. Show that $\mathcal{G} := \{A \cap B : A \in \mathcal{F}\}$ is a σ-algebra of B.

Exercise 1.2. Let $\mathcal{F}$ and $\mathcal{G}$ be σ-algebras of Ω. (a) Show that $\mathcal{F} \cap \mathcal{G}$ is a σ-algebra of Ω. (b) Show that $\mathcal{F} \cup \mathcal{G}$ is not necessarily a σ-algebra of Ω.

Exercise 1.3. Describe the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for the following three experiments: (a) a biased coin is tossed three times; (b) two balls are drawn without replacement from an urn which originally contained two blue and two red balls; (c) a biased coin is tossed repeatedly until a head turns up.

Exercise 1.4. Suppose X is a continuous random variable with distribution F_X. Let g be a strictly increasing continuous function. Define $Y = g(X)$. (a) What is F_Y, the distribution of Y? (b) What is f_Y, the density of Y?

Exercise 1.5. Suppose X is a continuous random variable with distribution F_X. Find F_Y where Y is given by (a) X^2 (b) $\sqrt{|X|}$ (c) $\sin X$ (d) $F_X(X)$.

Exercise 1.6. Suppose X is a continuous random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let f be the density of X under $\mathbb{P}$ and assume $f > 0$. Let g be the density function of a random variable. Define $Z := g(X)/f(X)$. (a) Show that $Z \equiv d\widetilde{\mathbb{P}}/d\mathbb{P}$ defines a Radon-Nikodým derivative. (b) What is the density of X under $\widetilde{\mathbb{P}}$?

Exercise 1.7. Let X be uniformly distributed on $[0, 1]$. For what function g is the random variable $g(X)$ exponentially distributed with parameter 1 (i.e. $g(X) \sim \mathcal{E}(1)$)?

Chapter 2

Information and conditioning

The notes from this chapter are taken primarily from (Shreve, 2004, Chapter 2).

2.1 Information and σ-algebras

Let us return to the coin-toss example of Section 1.3.2. If we are given no information about ω what can we say about ω? In other words, what are the subsets of Ω for which we can say: "ω is in this set" or "ω is not in this set"? The answer is $\emptyset$ and Ω, which, together, form the trivial σ-algebra $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Now suppose we are given the value of ω_1. What are the subsets of Ω for which we can say: "ω is in this set" or "ω is not in this set"? The answer is the sets in $\mathcal{F}_0$ as well as A_H and A_T. Together, these sets form the σ-algebra $\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}$. We say the sets in $\mathcal{F}_1$ are *resolved* by the first coin toss.

Now suppose we are given the value of ω_1 and ω_2. What are the subsets of Ω for which we can say: "ω is in this set" or "ω is not in this set"? The answer is the sets in $\mathcal{F}_2$, given by

$$\mathcal{F}_2 = \left\{ \begin{matrix} \emptyset, \Omega, A_H, A_T, A_{HH}, A_{HT}, A_{TT}, A_{TH}, A_{HH}^c, A_{HT}^c, A_{TT}^c, A_{TH}^c \\ A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT} \end{matrix} \right\}.$$

The sets in $\mathcal{F}_2$ are resolved by the first two coin tosses.

Continuing in this way, for each $n \in \mathbb{N}$ we can define $\mathcal{F}_n$ as the σ-algebra containing the sets that are resolved by the first n coin tosses. Note that if a set $A \in \mathcal{F}_n$ then $A \in \mathcal{F}_{n+1}$. Thus, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. In other words, $\mathcal{F}_{n+1}$ contains more "information" than $\mathcal{F}_n$. This kind of structure is encapsulated in the following definition.

Definition 2.1.1. Let Ω be a nonempty set. Let T be a fixed positive number, and assume that for each $t \in [0, T]$ there is a σ-algebra $\mathcal{F}_t$. Assume further that if $0 \leq s \leq t \leq T$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$. Then we

call the sequence of σ-algebras $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,\mathrm{T}]}$ a *continuous time filtration.*

A *discrete time filtration* is a sequence of σ-algebras $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ that satisfies $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all n.

EXAMPLE 2.1.2. Let $\Omega = \mathrm{C}_0[0,\mathrm{T}]$, the set of continuous functions defined on $[0,\mathrm{T}]$, starting from zero. We denote by $\omega = (\omega_t)_{t\in[0,\mathrm{T}]}$ and element of Ω. Let $\mathcal{F}_t$ be the σ-algebra generated by observing ω over the interval $[0,t]$. Mathematically, we write this as

$$\mathcal{F}_t := \sigma(\omega_s, 0 \le s \le t).$$

It should be obvious that the sequence of σ-algebras $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,\mathrm{T}]}$ forms a filtration. Below, we define two sets, one of which is in $\mathcal{F}_t$, one of which is not.

$$\mathrm{A} := \{\omega : \sup_{0\le s\le t} \omega_s \le 1\} \in \mathcal{F}_t, \qquad\qquad \mathrm{B} := \{\omega : \omega_{\mathrm{T}} \le 1\} \notin \mathcal{F}_t.$$

The set A is an element of $\mathcal{F}_t$ because, given the path of ω over the interval $[0,t]$ one can answer the question: is the maximum of ω over the interval $[0,t]$ less than 1? The set B is not an element of $\mathcal{F}_t$ because one needs to know ω_{T} in order to answer the question: is $\omega_{\mathrm{T}} \le 1$?

In the above example we generated a sequence of σ-algebras by observing directly an element $\omega \in \Omega$. Suppose that, instead of observing ω we can observe only a random variable $\mathrm{X}(\omega)$. We can use this information to generate a σ-algebra as well.

DEFINITION 2.1.3. Let X be a random variable defined on a nonempty sample space Ω. The *σ-algebra generated by* X, denoted $\sigma(\mathrm{X})$, is the collection of all subsets of Ω of the form $\{\mathrm{X} \in \mathrm{A}\}$ where $\mathrm{A} \in \mathcal{B}(\mathbb{R})$.

EXAMPLE 2.1.4. Let us return to Example 1.4.2. What is $\sigma(\mathrm{S}_2)$? From the definition, we need to ask, which sets are of the form $\{\mathrm{S}_2 \in \mathrm{A}\}$? Since S_2 can only take three values, u^2, ud and d^2 we check the following sets

$$\{\mathrm{S}_2 = u^2\} = \mathrm{A}_{\mathrm{HH}}, \qquad \{\mathrm{S}_2 = ud\} = \mathrm{A}_{\mathrm{HT}} \cup \mathrm{A}_{\mathrm{TH}}, \qquad \{\mathrm{S}_2 = d^2\} = \mathrm{A}_{\mathrm{TT}}.$$

We add to these sets the sets that are necessary to form a σ-algebra (i.e., $\emptyset$, Ω and unions and complements of the above sets) to obtain

$$\sigma(\mathrm{S}_2) = \sigma(\{\mathrm{A}_{\mathrm{HH}}, \mathrm{A}_{\mathrm{TT}}, \mathrm{A}_{\mathrm{HT}} \cup \mathrm{A}_{\mathrm{TH}}\}).$$

Note, that $\sigma(\mathrm{S}_2) \subset \mathcal{F}_2$ since $\mathrm{A}_{\mathrm{HT}}, \mathrm{A}_{\mathrm{TH}} \in \mathcal{F}_2$ but $\mathrm{A}_{\mathrm{HT}}, \mathrm{A}_{\mathrm{TH}} \notin \sigma(\mathrm{S}_2)$. The reason is that, if $\mathrm{S}_2 = ud$ we cannot say if $\omega_1 = \mathrm{T}$ or $\omega_1 = \mathrm{H}$.

DEFINITION 2.1.5. Let X be a random variable defined on a nonempty sample space Ω. Let $\mathcal{G}$ be a σ-algebra of subsets of Ω. If $\sigma(\mathrm{X}) \subset \mathcal{G}$ we say that X *is $\mathcal{G}$-measurable*, and we write $\mathrm{X} \in \mathcal{G}$.

A random variable X is $\mathcal{G}$-measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X. Obviously, if $X \in \mathcal{G}$ then $g(X) \in \mathcal{G}$ (assuming is g is a measurable map from $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Eventually, we will want to consider stochastic processes $X = (X_t)_{t\in[0,T]}$ and we will want to know at each time t if X_t is measureable with respect to σ-algebra $\mathcal{F}_t$.

DEFINITION 2.1.6. Let Ω be a nonempty sample space equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$. Let $X = (X_t)_{t\in[0,T]}$ be a collection of random variables indexed by $t \in [0, T]$. We say this collection of random variables is $\mathbb{F}$-*adapted* if $X_t \in \mathcal{F}_t$ for all $t \in [0, T]$.

2.2 INDEPENDENCE

When $X \in \mathcal{G}$ this means that the information in $\mathcal{G}$ is sufficient to determine the value of X. On the other extreme, if X is independent (a term we will define soon) of $\mathcal{G}$ this means that the informtion in $\mathcal{G}$ tells us nothing about the value of X.

DEFINITION 2.2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say that two sets two sets A and B in $\mathcal{F}$ are *independent*, written $A \perp\!\!\!\perp B$, if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

EXAMPLE 2.2.2. Let us return to the coin-toss example of Section 1.3.2. Consider two sets and their intersection

$$\{\omega_1 = H\} = A_H, \qquad \{\omega_2 = H\} = A_{HH} \cup A_{TH} \qquad \{\omega_1 = H\} \cap \{\omega_2 = H\} = A_{HH}$$

Since the coin tosses are indepdendent, we should have $\{\omega_1 = H\} \perp\!\!\!\perp \{\omega_2 = H\}$ Let us verify that these events are independent according to Definition 2.2.1. We have

$$\mathbb{P}(A_H) = p, \qquad \mathbb{P}(A_{HH} \cup A_{TH}) = p^2 + qp = p, \qquad \mathbb{P}(A_{HH}) = p^2.$$

Thus, $\mathbb{P}(A_H) \cdot \mathbb{P}(A_{HH} \cup A_{TH}) = \mathbb{P}(A_{HH})$, as expected.

EXAMPLE 2.2.3. Can a set be independent of itself? Surprisingly, the answer is "yes." Suppose $A \perp\!\!\!\perp A$. Then, by the definition of independent sets, we have $\mathbb{P}(A \cap A) = \mathbb{P}(A) \cdot \mathbb{P}(A)$. We also have $\mathbb{P}(A \cap A) = \mathbb{P}(A)$. Combining these equations, we obtain $\mathbb{P}(A) \cdot \mathbb{P}(A) = \mathbb{P}(A)$. This equation has two solutions $\mathbb{P}(A) = 1$ and $\mathbb{P}(A) = 0$. Thus, a set is independent of itself if the probability of that set is zero or one.

Having defined independent sets, we can now extend to independent σ-algebras and random variables.

DEFINITION 2.2.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$ (i.e., $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$). We say these *two σ-algebras are independent*, written $\mathcal{G} \perp\!\!\!\perp \mathcal{H}$, if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \qquad \forall\, A \in \mathcal{G}, \qquad \forall\, B \in \mathcal{H}.$$

Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say these *two random variables are independent*, written $X \perp\!\!\!\perp Y$, if $\sigma(X) \perp\!\!\!\perp \sigma(Y)$. Lastly, we say the random variable X is *independent* of the σ-algebra $\mathcal{G}$, written $X \perp\!\!\!\perp \mathcal{G}$, if $\sigma(X) \perp\!\!\!\perp \mathcal{G}$.

Recall from Definition 2.1.3 that $\sigma(X)$ contains all sets of the form $\{X \in A\}$, where $A \in \mathcal{B}(\mathbb{R})$. Combining this with Definition 2.2.4 we see that

$$X \perp\!\!\!\perp Y \qquad \Leftrightarrow \qquad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B), \qquad \forall\, A, B \in \mathcal{B}(\mathbb{R}). \tag{2.1}$$

It follows from (2.1) that

$$X \perp\!\!\!\perp Y \qquad \Rightarrow \qquad \mathbb{E}XY = \mathbb{E}X \cdot \mathbb{E}Y.$$

Note that $\mathbb{E}XY = \mathbb{E}X \cdot \mathbb{E}Y$ does *not* imply $X \perp\!\!\!\perp Y$.

The above notion of independence is called *pairwise* independence. If $X \perp\!\!\!\perp Y$ and $Y \perp\!\!\!\perp Z$, this notion of independence *not* imply $X \perp\!\!\!\perp Z$ (for example, what if $Z = X$?). Thus, at times, we may need a stronger notion of independence.

DEFINITION 2.2.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ be sub-σ-algebras of $\mathcal{F}$. We say the *sequence of σ-algebras are independent*, if

$$\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i), \qquad \forall\, A_1 \in \mathcal{G}_1, \forall\, A_2 \in \mathcal{G}_2, \ldots, \forall A_n \in \mathcal{G}_n.$$

Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say the *sequence of random variables are independent* if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_n)$ are independent.

As with with a pair of random variables, a sequence of random variables $(X_i)_{i \geq 1}$ is independent if and only if

$$\mathbb{P}\left(\cap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i), \qquad \forall\, A_1 \in \mathcal{B}(\mathbb{R}), \forall\, A_2 \in \mathcal{B}(\mathbb{R}), \ldots, \forall\, A_n \in \mathcal{B}(\mathbb{R}).$$

We will often say that a sequence of random variables $(X_i)_{i \geq 0}$ is *independent and identically distributed* (iid), by which me mean all X_i have the same distribution and $(X_i)_{1 \leq i \leq n}$ are independent for every $n \in \mathbb{N}$.

EXAMPLE 2.2.6. Let us return to the coin-toss example of Section 1.3.2. Let us define a squence of random variables $(X_i)_{i\geq 1}$ via

$$X_1(\omega) = \begin{cases} 1 & \omega_1 = H, \\ 0 & \omega_1 = T, \end{cases}, \qquad X_2(\omega) = \begin{cases} 1 & \omega_2 = H, \\ 0 & \omega_2 = T, \end{cases}, \qquad X_i(\omega) = \begin{cases} X_1 & i \text{ odd}, \\ X_2 & i \text{ even}. \end{cases},$$

Clearly, since the coin tosses are independent we have $X_1 \perp\!\!\!\perp X_2$ and $X_i \perp\!\!\!\perp X_j$ if i is even and j is odd. But, the sequence $(X_i)_{1\leq i\leq n}$ is not independent for any $n \geq 3$ since $X_i = X_{i+2n}$ for any $i, n \in \mathbb{N}$.

It is not easy to verify if two random variables X and Y are independent using Expression (2.1), since the equation must be verified for *all* Borel sets $A, B \in \mathcal{B}(\mathbb{R})$. In fact, there is an easier way to check independence.

DEFINITION 2.2.7. The *joint distribution function* $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ of two random variables X and Y defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given by

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

Again, we have two special cases for jointly discrete and jointly continuous random variables.

DEFINITION 2.2.8. Two random variables X and Y are called *jointly discrete* if the pair (X, Y) takes values in some countable set $A = \{x_1, x_2, \ldots\} \times \{y_1, y_2, \ldots\} \subset \mathbb{R}^2$. We associate is a discrete random variable a *probability mass function* $f_{X,Y} : A \to \mathbb{R}$, defined by $f_{X,Y}(x_i, y_j) := \mathbb{P}(X = x_i, Y = y_j)$.

DEFINITION 2.2.9. A pair of random variables X and Y is called *jointly continuous* if its joint distribution function $F_{X,Y}$ can be written as

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} du\,dv\, f_{X,Y}(u, v), \qquad (x, y) \in \mathbb{R}^2,$$

for some $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ called the *joint probability density function*.

As in the one-dimensional case, it may help to think of the joint density function $f_{X,Y}$ as $f_{X,Y}(x, y)dxdy = \mathbb{P}(X \in dx, Y \in dy)$.

Note that for jointly continuous random variables X and Y we have $f_{X,Y}(x, y) = \partial_x \partial_y F_X(x, y)$.

If the pair (X, Y) is either jointly discrete or jointly continuous, it is easy to compute $\mathbb{P}((X, Y) \in A)$ for any $A \in \mathcal{B}(\mathbb{R}^2)$. We have

$$\text{discrete :} \qquad \mathbb{P}((X, Y) \in A) = \sum_{\{i,j:(x_i,y_j)\in A\}} f_{X,Y}(x_i, y_j),$$

$$\text{continuous :} \qquad \mathbb{P}((\mathrm{X},\mathrm{Y}) \in \mathrm{A}) = \int_{\mathrm{A}} \mathrm{d}x\mathrm{d}y\, f_{\mathrm{X},\mathrm{Y}}(x,y).$$

To recover the *marginal* distribution F_{X} from $\mathrm{F}_{\mathrm{X},\mathrm{Y}}$, simply note that

$$\mathrm{F}_{\mathrm{X}}(x) = \mathbb{P}(\mathrm{X} \le x) = \mathbb{P}(\mathrm{X} \le x, \mathrm{Y} \le \infty) = \mathrm{F}_{\mathrm{X},\mathrm{Y}}(x,\infty).$$

It follows that for the discrete and continuous cases, we have, respectively

$$\text{discrete :} \qquad f_{\mathrm{X}}(x_i) = \sum_j f_{\mathrm{X},\mathrm{Y}}(x_i, y_j),$$

$$\text{continuous :} \qquad f_{\mathrm{X}}(x) = \int_{\mathbb{R}} \mathrm{d}y\, f_{\mathrm{X},\mathrm{Y}}(x,y).$$

The following theorem gives some easy-to-check conditinos for independence.

THEOREM 2.2.10. *Let* X *and* Y *be random variables definied on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *The following conditions are equivalent (that is, if one of them holds, all of them hold)*

1. $\mathrm{X} \perp\!\!\!\perp \mathrm{Y}$.
2. $\mathrm{F}_{\mathrm{X},\mathrm{Y}}(x,y) = \mathrm{F}_{\mathrm{X}}(x)\mathrm{F}_{\mathrm{Y}}(y)$ *for every* $(x,y) \in \mathbb{R}^2$.
3. *Discrete case:* $f_{\mathrm{X},\mathrm{Y}}(x,y) = f_{\mathrm{X}}(x) f_{\mathrm{Y}}(y)$ *for every* $(x,y) \in \mathbb{R}^2$.
   Continuous case: $f_{\mathrm{X},\mathrm{Y}}(x,y) = f_{\mathrm{X}}(x) f_{\mathrm{Y}}(y)$ *for 'almost' every* $(x,y) \in \mathbb{R}^2$.
4. $\mathbb{E}[\mathrm{e}^{\mathrm{i}u\mathrm{X}+\mathrm{i}v\mathrm{Y}}] = \mathbb{E}\mathrm{e}^{u\mathrm{i}\mathrm{X}} \cdot \mathbb{E}\mathrm{e}^{\mathrm{i}v\mathrm{Y}}$ *for all* $(u,v) \in \mathbb{R}^2$.

Together with expectation, the most important statistical properties of a random variable (or pair) are the variance and co-variance.

DEFINITION 2.2.11. The *variance* of a random variable X, written $\mathbb{V}\mathrm{X}$ is defined by

$$\mathbb{V}\mathrm{X} = \mathbb{E}(\mathrm{X} - \mathbb{E}\mathrm{X})^2 = \mathbb{E}\mathrm{X}^2 - (\mathbb{E}\mathrm{X})^2,$$

whenever the expectation exists.

DEFINITION 2.2.12. The *co-variance* of two random variables X and Y, written $\mathrm{Co}\mathbb{V}[\mathrm{X},\mathrm{Y}]$ is defined by

$$\mathrm{Co}\mathbb{V}[\mathrm{X},\mathrm{Y}] = \mathbb{E}(\mathrm{X} - \mathbb{E}\mathrm{X})(\mathrm{X} - \mathbb{E}\mathrm{Y}) = \mathbb{E}\mathrm{X}\mathrm{Y} - \mathbb{E}\mathrm{X} \cdot \mathbb{E}\mathrm{Y},$$

whenever the expectation exists.

Note that $\mathrm{Co}\mathbb{V}[\mathrm{X},\mathrm{X}] = \mathbb{V}\mathrm{X}$.
Note $\mathbb{V}\cdot$ is *not* a linear operator, since

$$\mathbb{V}[a\mathrm{X} + b\mathrm{Y}] = a^2\mathbb{V}\mathrm{X} + b^2\mathbb{V}\mathrm{Y} + 2ab\,\mathrm{Co}\mathbb{V}[\mathrm{X},\mathrm{Y}].$$

where a and b are constants.

DEFINITION 2.2.13. We say two random variables are *un-correlated* if $\mathrm{Co}\mathbb{V}[\mathrm{X},\mathrm{Y}] = 0$.

Note that $\mathrm{X} \perp\!\!\!\perp \mathrm{Y}$ implies X and Y are uncorrelated. However, the converse is *not* true.

2.3 Conditional expectation

Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ algebra of $\mathcal{F}$. When $X \in \mathcal{G}$ this means that the information in $\mathcal{G}$ is sufficient to determin the value of X. When $X \perp\!\!\!\perp \mathcal{G}$, this means that the information in $\mathcal{G}$ gives us no information at all about X. Usually, however, the information in $\mathcal{G}$ gives us some information about X, but not enough to determine X exactly. And this brings us to the notion of conditioning.

Presumably, you have run across the following formula for the conditional probability of a set A given B

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \qquad \mathbb{P}(B) > 0.$$

When (X, Y) have are jointly discrete or jointly continuous, this readily leads to *conditional probability mass function*

$$\text{discrete :} \qquad f_{X|Y}(x_i, y_j) := \mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i \cap Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)},$$

$$\text{continuous :} \qquad f_{X|Y}(x, y)\mathrm{d}x := \mathbb{P}(X \in \mathrm{d}x | Y = y) = \frac{\mathbb{P}(X \in \mathrm{d}x \cap Y \in \mathrm{d}y)}{\mathbb{P}(Y \in \mathrm{d}y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}\mathrm{d}x.$$

And from this, we can define $\mathbb{E}[X|Y = y]$, the *conditional expectation of* X *given* $Y = y$

$$\text{discrete :} \qquad \mathbb{E}[X|Y = y_j] := \sum_i x_i \, f_{X|Y}(x_i, y_j),$$

$$\text{continuous :} \qquad \mathbb{E}[X|Y = y] := \int_{\mathbb{R}} \mathrm{d}x \, x \, f_{X|Y}(x, y)$$

Note that $\mathbb{E}[X|Y = y]$ is simply a function of y – there is nothing random about it.

Unfortunately, there are cases for which the pair (X, Y) are neither jointly discrete nor jointly continuous. And, for these cases we need a more general notion of conditional expectation. Here we will make two conceptual leaps:

1. We will condition with respect to a σ-algebra rather than conditioning on an event.
2. The conditional expectation will be a random variable.

We will just hop in with our new definition of conditional expectation and then we will see, through an example, that his new definition makes sense.

Definition 2.3.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let X be a random variable that is either nonnegative or integrable. The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}[X|\mathcal{G}]$, is any random variable that satisfies

1. Measurability: $\mathbb{E}[X|\mathcal{G}] \in \mathcal{G}$.
2. Partial averaging: $\mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{1}_A X]$ for all $A \in \mathcal{G}$.
   Alternatively, $\mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[ZX]$ for all $Z \in \mathcal{G}$.

When $\mathcal{G} = \sigma(Y)$ we shall often use the short-hand notation $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$.

Conditional probabilities are defined from conditional expectations using

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}].$$

Admittedly, Definition 2.3.1 is rather abstract (and, for the purposes of computation, useless). In fact, it is not at all clear from Definition 2.3.1 that $\mathbb{E}[X|\mathcal{G}]$ even exists! It does exist, though we will not prove this here.

Conditional expectation has an interesting L^2 interpretation. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ (i.e., $\mathcal{G} \subset \mathcal{F}$). Define

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) := \{X : \Omega \to \mathbb{R} \text{ s.t. } X^{-1}(B) \in \mathcal{F}\ \forall B \in \mathcal{B}(\mathbb{R}) \text{ and } \mathbb{E}|X|^2 < \infty\},$$

and likewise for $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Clearly, since $\mathcal{G} \subset \mathcal{F}$ we have $L^2(\Omega, \mathcal{G}, \mathbb{P}) \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$. Next, define an inner product

$$\langle X, Y\rangle := \mathbb{E}XY, \qquad\qquad X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}).$$

From the definition of conditional expectation, we have

$$0 = \langle X - \mathbb{E}[X|\mathcal{G}], Z\rangle = \mathbb{E}(X - \mathbb{E}[X|\mathcal{G}])Z, \qquad\qquad \forall\, Z \in \mathcal{G}.$$

Thus, $\mathbb{E}[X|\mathcal{G}]$ is the projection of $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto the subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

When conditioning on the σ-algebra generated by a random variable, it is easiest to use the following formula

$$\mathbb{E}[X|Y] = \psi(Y), \qquad\qquad \psi(y) := \mathbb{E}[X|Y = y]. \tag{2.2}$$

The following example should help to build some intuition for conditional expectation.

EXAMPLE 2.3.2. Let $\Omega = \{a, b, c, d, e, f\}$, $\mathcal{F} = 2^\Omega$ and $\mathbb{P}(\omega) = (1/6)$ for $\omega = a, b, \ldots, f$. Define two random variables X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$ as follows

$$\begin{pmatrix} \omega \\ X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} a & b & c & d & e & f \\ 1 & 3 & 3 & 3 & 5 & 7 \\ 2 & 2 & 1 & 1 & 7 & 7 \end{pmatrix}.$$

Let $\mathcal{G} = \sigma(Y)$. Next, let us compute $\mathbb{E}[X|Y]$ using (2.2) and check if it agrees with Definition 2.3.1. We have

$$\begin{pmatrix} \omega \\ Y(\omega) \\ \mathbb{E}[X|Y](\omega) \end{pmatrix} = \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 1 & 1 & 7 & 7 \\ 2 & 2 & 3 & 3 & 6 & 6 \end{pmatrix}.$$

Le us check measurability: is $\mathbb{E}[X|Y] \in \sigma(Y)$? In other words, is $\sigma(\mathbb{E}[X|Y]) \subseteq \sigma(Y)$? We have

$$\sigma(Y) = \sigma(\{a,b\},\{c,d\},\{e,f\}),$$
$$\sigma(\mathbb{E}[X|Y]) = \sigma(\{a,b\},\{c,d\},\{e,f\}).$$

So, yes, $\mathbb{E}[X|Y] \in \sigma(Y)$. Another way to think of measurability is to simply ask: given the value of Y, can one determine the value of $\mathbb{E}[X|Y]$? Clearly, in this case the answer is "yes." Next, let us check the partial averaging property: does $\mathbb{E}[\mathbb{1}_A \mathbb{E}[X|Y]] = \mathbb{E}[\mathbb{1}_A X]$ for all $A \in \sigma(Y)$. Rather than check this for *every* $A \in \sigma(Y)$, let us just check that this holds for the sets $\{a,b\}$, $\{c,d\}$ and $\{e,f\}$. We have

$$\begin{aligned}
&A = \{a,b\}, && \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|Y]] = \mathbb{P}(a)2 + \mathbb{P}(b)2 = 4/6, && \mathbb{E}[\mathbb{1}_A X] = \mathbb{P}(a)1 + \mathbb{P}(b)3 = 4/6, \\
&A = \{c,d\}, && \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|Y]] = \mathbb{P}(c)3 + \mathbb{P}(d)3 = 6/6, && \mathbb{E}[\mathbb{1}_A X] = \mathbb{P}(c)3 + \mathbb{P}(d)3 = 6/6, \\
&A = \{e,f\}, && \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|Y]] = \mathbb{P}(e)6 + \mathbb{P}(f)6 = 12/6, && \mathbb{E}[\mathbb{1}_A X] = \mathbb{P}(e)5 + \mathbb{P}(f)7 = 12/6.
\end{aligned}$$

The following properties are arguably more important to remember than the definition of conditional expectation. Memorize them!

THEOREM 2.3.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Conditional expectations satisfy the following properties.*

1. *Linearity:* $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$.
2. *Taking out what is known: if* $X \in \mathcal{G}$ *then* $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$.
3. *Iterated conditioning: if* $\mathcal{H}$ *is a sub-σ-algebra of* $\mathcal{G}$ *then* $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.
4. *Independence: if* $X \perp\!\!\!\perp \mathcal{G}$ *then* $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$.

Theorem 2.3.3 can be proved directly from Definition 2.3.1, though we will not do so here. In addition to the above properties, the following Theorem, which we state without proof is often useful:

THEOREM 2.3.4 (JENSEN'S INEQUALITY). *Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function. Then we have*

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}], \qquad \mathbb{P} - a.s. \tag{2.3}$$

In order to keep straight which direction the inequality in (2.3) goes, it is helpful to remember that $\phi(x) = x^2$ is a convex function and the conditional variance of a function satisfies

$$\mathbb{V}[X|\mathcal{G}] = \mathbb{E}[X^2|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]^2 \geq 0.$$

Now that we have defined conditional expectation and established some of its key properties, we can define "Markov process" and "martingale" – two seemingly similar, but very distinct concepts.

DEFINITION 2.3.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider an $\mathbb{F}$-adapted stochastic process $M = (M_t)_{t\in[0,T]}$. We say that M is

$$\begin{aligned}
\text{a } \textit{martingale} \text{ if} \quad & \mathbb{E}[M_t|\mathcal{F}_s] = M_s, \quad && \forall\, 0 \leq s \leq t \leq T, \\
\text{a } \textit{sub-martingale} \text{ if} \quad & \mathbb{E}[M_t|\mathcal{F}_s] \geq M_s, \quad && \forall\, 0 \leq s \leq t \leq T, \\
\text{a } \textit{super-martingale} \text{ if} \quad & \mathbb{E}[M_t|\mathcal{F}_s] \leq M_s, \quad && \forall\, 0 \leq s \leq t \leq T.
\end{aligned}$$

We have given above the definition of a *continuous time* martingale (resp. sub-, super-). We can also define *discrete-time* martingales by making the obvious modifications.

Admittedly, the definition of sub- and super-martingales seems backwards; sub-martingales tend to rise in expectation, whereas a super-martingales tend to fall in expectation.

Note: when we say that a process M is a martingale (or sub- or super-martingale) this is with respect to a fixed probability measure and filtration. If $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are two probability measures and $\mathbb{F}$ and $\mathbb{G}$ are two filtrations, it is entirely possible that a process M may be a martingale with respect to $(\mathbb{P}, \mathbb{F})$ and may not be a martingale with respect to $(\widetilde{\mathbb{P}}, \mathbb{F})$, $(\mathbb{P}, \mathbb{G})$ or $(\widetilde{\mathbb{P}}, \mathbb{G})$.

DEFINITION 2.3.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider an $\mathbb{F}$-adapted stochastic process $X = (X_t)_{t\in[0,T]}$. Assume that for all $0 \leq s \leq t \leq T$ and for every nonnegative, Borel-measurable function f, there is another Borel-measurable function g (which depends on s, t, and f) such that

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = g(X_s).$$

Then we say X is a *Markov process* or simply "X is Markov."

Identifying $g(X_s) \equiv \mathbb{E}[f(X_t)|X_s]$ we can write the Markov property as follows

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \mathbb{E}[f(X_t)|X_s]$$

A Markov process is a process for which the following holds: *given the present (i.e., X_s), the future (i.e, X_t, $t \geq s$) is independent of the past (i.e, $\mathcal{F}_s$).* What this means in practice is that

$$\mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{P}(X_t \in A|X_s), \qquad \forall s \leq t, \qquad \forall\, A \in \mathcal{B}(\mathbb{R}). \qquad \text{(if X is Markov)}$$

If X_t is a discrete or continuous random variable for every t then we have a *transition kernel*, written as P in the discrete case and Γ in the continuous case.

$$\begin{aligned} \text{discrete :} \qquad & P(s,x;t,y) := \mathbb{P}(X_t = y|X_s = x), \\ \text{continuous :} \qquad & \Gamma(s,x;t,y)dy = \mathbb{P}(X_t \in dy|X_s = x). \end{aligned}$$

If you can write the transition kernel of a process explicitly, then you have essentially proved that the process is Markov.

Note that *any process* that has independent increments is Markov since, if $X_t - X_s \perp\!\!\!\perp X_s$ for $t \geq s$, then

$$\mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{P}(X_t - X_s + X_s \in A|X_s),$$

Where $\mathcal{F}_s$ is the filtration generated by observing X up to time s.

Markov processes and Martingales are *entirely separate* concepts. A process X can be both a martingale and a Markov process, it can be a martingale but not a Markov process, it can be a Markov process but not a martingale, and it can be neither a Markov process nor a martingale. We illustrate the difference with an example.

EXAMPLE 2.3.7. Let us return to the stock price Example 1.4.2. Let us show that $S = (S_n)_{0\leq n}$ is a Markov process. Recall that $\mathcal{F}_m$ is the σ-algebra generated by observing $\omega_1, \omega_2, \ldots, \omega_m$. Observe that $S_m \in \mathcal{F}_m$. Next, note that

$$\mathbb{P}(S_{n+m} = S_m u^k d^{n-k}|S_m) = \binom{n}{k} p^k q^{n-k}.$$

Since we have written the transition kernel explicitly, we have established that S is Markov. Let us also find the function g in Definition 2.3.6. For any $f : \mathbb{R} \to \mathbb{R}$ we have

$$\mathbb{E}[f(S_{n+m})|\mathcal{F}_m] = \sum_{k=0}^{n} f(S_m u^k d^{n-k}) \cdot \binom{n}{k} p^k q^{n-q} =: g(S_m).$$

Thus, we have found g. Now, to see if S is a martingale note that

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_{n+1}|S_n] = p \cdot uS_n + q \cdot dS_n = (p \cdot u + q \cdot d)S_n.$$

Thus, if $(p \cdot u + q \cdot d) = 1$, then $\mathbb{E}[\mathrm{S}_{n+1}|\mathcal{F}_n] = \mathrm{S}_n$. Let us assume that $(p \cdot u + q \cdot d) = 1$. Then we have

$$\begin{aligned}\mathbb{E}[\mathrm{S}_{n+m}|\mathcal{F}_n] &= \mathbb{E}[\mathbb{E}[\mathrm{S}_{n+m}|\mathcal{F}_{n+m-1}]|\mathcal{F}_n] = \mathbb{E}[\mathrm{S}_{n+m-1}|\mathcal{F}_n] \\ &= \mathbb{E}[\mathbb{E}[\mathrm{S}_{n+m-1}|\mathcal{F}_{n+m-2}]|\mathcal{F}_n] = \mathbb{E}[\mathrm{S}_{n+m-2}|\mathcal{F}_n] \\ &= \dots \\ &= \mathbb{E}[\mathbb{E}[\mathrm{S}_{n+1}|\mathcal{F}_n]|\mathcal{F}_n] = \mathbb{E}[\mathrm{S}_n|\mathcal{F}_n] = \mathrm{S}_n,\end{aligned}$$

Therefore, the process S is a martingale. Similarly, if $(p \cdot u + q \cdot d) \geq 1$, then S is a sub-martingale and if $(p \cdot u + q \cdot d) \leq 1$, then S is a super-martingale.

2.4 Characteristic functions

Rather than describe a rnadom variable through its distribution, we may alternatively describe a random variable via its characteristic function.

Definition 2.4.1. The *characteristic function* of a random variable X is the function $\phi_{\mathrm{X}} : \mathbb{R} \to \mathbb{C}$ defined by

$$\phi_{\mathrm{X}}(t) := \mathbb{E}\mathrm{e}^{\mathrm{i}t\mathrm{X}}, \qquad \mathrm{i} = \sqrt{-1}.$$

The characteristic function always exists since $\mathbb{E}|\mathrm{e}^{\mathrm{i}t\mathrm{X}}| = 1$. We have the following obvious properties.

Theorem 2.4.2. *Characteristic functions have the following properties.*

1. *Let a, b be constants. Then $\phi_{a\mathrm{X}+b}(t) = \mathrm{e}^{\mathrm{i}bt}\phi(at)$.*
2. *If X and Y are independent then $\phi_{\mathrm{X}+\mathrm{Y}}(t) = \phi_{\mathrm{X}}(t)\phi_{\mathrm{Y}}(t)$.*
3. *If $(\mathrm{X}_i)_{1\leq i\leq n}$ are iid and $\mathrm{S}_n = \sum_{i=1}^n \mathrm{X}_i$, then $\phi_{\mathrm{S}_n}(t) = (\phi_{\mathrm{X}}(t))^n$.*

Example 2.4.3. Let $\mathrm{X} \sim \mathcal{E}(\lambda)$ as in Example 1.4.14. Then

$$\phi_{\mathrm{X}}(t) = \int_0^\infty \mathrm{d}x\, \mathrm{e}^{\mathrm{i}tx}\lambda\mathrm{e}^{-\lambda x} = \frac{\lambda}{\lambda - \mathrm{i}t}.$$

Example 2.4.4. Let $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$ as in Example 1.4.15. Then

$$\phi_{\mathrm{X}}(t) = \int_{-\infty}^\infty \mathrm{d}x\, \mathrm{e}^{\mathrm{i}tx}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) = \exp\left(\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2\right).$$

Characteristic functions have several uses. First, they can be used to capture the moments of a random variable when they exist.

THEOREM 2.4.5. *Let $\phi_X(t)$ be the characteristic function of a random variable X. Then*

$$\phi^{(n)}(0) = i^n \mathbb{E}X^n, \qquad \text{if } \mathbb{E}|X|^n < \infty.$$

PROOF. We have

$$\frac{d^n}{dt^n}\phi_X(t) = \mathbb{E}\frac{d^n}{dt^n}e^{itX} = i^n \mathbb{E}X^n e^{itX},$$

Now set $t = 0$ to complete the proof. □

The characteristic function uniquely determines the distribution of a random variable. In other words, there is a one-to-one correspondence between F_X and ϕ_X. We show this for a continuous random variable.

THEOREM 2.4.6 (INVERSION). *Suppose X is a continuous random variable with density f_X. Then*

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} dt\, e^{-itx}\phi_X(t),$$

for all x where f_X is differentiable.

To obtain F_X from f_X simply use $F_X(x) = \int_{-\infty}^{x} dy\, f_X(y)$.

PROOF. The proof of Theorem 2.4.6 follows from standard Fourier results. We have

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} dt\, e^{-itx}\int_{\mathbb{R}} dy\, e^{ity} f_X(y) = \frac{1}{2\pi}\int_{\mathbb{R}} dt\, e^{-itx}\mathbb{E}e^{itX} = \frac{1}{2\pi}\int_{\mathbb{R}} dt\, e^{-itx}\phi_X(t).$$

□

An inversion theorem for random variables that are not continuous also exists. Though, it is not particularly useful for the purposes of computation.

Perhaps the most important property of characteristic functions is they can be used to prove the convergence of a sequence of random variables to a limiting distribution.

DEFINITION 2.4.7. We say that a sequence of distribution functions $(F_n)_{n\geq 1}$ *converges to a distribution* F, written $F_n \to F$, if $\lim_{n\to\infty} F_n(x) = F(x)$ at all points x where F is continuous.

THEOREM 2.4.8 (CONTINUITY THEOREM). *Let $(F_n)_{n\geq 1}$ be a sequence of distributions functions with corresponding characteristic functions $(\phi_n)_{n\geq 1}$.*

1. *If $F_n \to F$ where F is a distribution function with corresponding characteristic function ϕ, then $\phi_n \to \phi$ (pointwise).*
2. *Conversely, if $\phi(t) := \lim_{n\to\infty}\phi_n(t)$ exists and is continuous at $t = 0$, then ϕ is the characteristic function of a distribution F and $F_n \to F$.*

Item 2 in Theorem 2.4.8 if particularly powerful. If F_n and ϕ_n are, respectively, the distribution and characteristic function of a sum of n independent random variables, it is often easier to compute ϕ_n than F_n. If we can compute ϕ_n and find its limit, then we can obtain F.

EXAMPLE 2.4.9. Suppose that $X \sim \text{Geo}(p)$. Then we have

$$\phi_X(t) = \mathbb{E}e^{itX} = \sum_{n=1}^{\infty} pq^{n-1}e^{itn} = pe^{it}\sum_{n=1}^{\infty}(qe^{it})^{n-1} = \frac{pe^{it}}{1-qe^{it}},$$

where we have used $\sum_{n=0}^{\infty} x^n = 1/(1-x)$ for $|x| < 1$. From this, one can easily show that $\mathbb{E}X = 1/p$. Now consider a sequence of random variables $(Y_n)_{n\geq 0}$ defined by

$$Y_n = \frac{1}{n}X_n, \qquad X_n \sim \text{Geo}(\lambda/n),$$

where $\lambda > 0$ is a fixed constant. For n large enough, $\lambda/n < 1$. Note that $\mathbb{E}Y_n = \frac{1}{n}\mathbb{E}X_n = \frac{1}{n}\frac{n}{\lambda} = \frac{1}{\lambda}$ for all n. We would like to know what the limiting distribution of the sequence $(Y_n)_{n\geq 0}$ is; we will use the Continuity Theorem 2.4.8 to do this. We have

$$\begin{aligned}\phi_{Y_n}(t) = \phi_{X_n}(t/n) &= \frac{e^{it/n}(\lambda/n)}{1-e^{it/n}(1-\lambda/n)} = \frac{e^{it/n}\lambda}{n-e^{it/n}(n-\lambda)} = \frac{e^{it/n}\lambda}{n(1-e^{it/n})-e^{it/n}\lambda} \\ &= \frac{(1+it/n)\lambda}{n(1-1-it/n-\ldots)-(1+it/n+\ldots)\lambda} \to \frac{\lambda}{\lambda-it} = \phi_Z(t),\end{aligned}$$

where $Z \sim \mathcal{E}(\lambda)$ (see Example 2.4.3). By the Continuity Theorem 2.4.8, since $\phi_{Y_n} \to \phi_Z$ it follows that $F_{Y_n} \to F_Z$.

2.5 EXERCISES

EXERCISE 2.1. Let $\Omega = \{a, b, c, d\}$ and let $\mathcal{F} = 2^{\Omega}$ (the set of all subsets of Ω). We define a probability measure $\mathbb{P}$ as follows

$$\mathbb{P}(a) = 1/6, \qquad \mathbb{P}(b) = 1/3, \qquad \mathbb{P}(c) = 1/4, \qquad \mathbb{P}(d) = 1/4,$$

Next, define three random variables

$$\begin{aligned} X(a) &= 1, & X(b) &= 1, & X(c) &= -1, & X(d) &= -1, \\ Y(a) &= 1, & Y(b) &= -1, & Y(c) &= 1, & Y(d) &= -1, \end{aligned}$$

and $Z = X + Y$. (a) List the sets in $\sigma(X)$. (b) What are the values of $\mathbb{E}[Y|X]$ for $\{a, b, c, d\}$? Verify the partial averaging property: $\mathbb{E}[\mathbb{1}_A\mathbb{E}[Y|X]] = \mathbb{E}[\mathbb{1}_A Y]$ for all $A \in \sigma(X)$. (c) What are the values of $\mathbb{E}[Z|X]$ for $\{a, b, c, d\}$? Verify the partial averaging property.

EXERCISE 2.2. Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let Y be a square integrable random variable: $\mathbb{E}Y^2 < \infty$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Show that

$$\mathbb{V}(Y - \mathbb{E}[Y|\mathcal{G}]) \leq \mathbb{V}(Y - X), \qquad \forall X \in \mathcal{G}.$$

EXERCISE 2.3. Give an example of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable X and a function f such that $\sigma(f(X))$ is strictly smaller than $\sigma(X)$ but $\sigma(f(X)) \neq \{\emptyset, \Omega\}$. Give a function g such that $\sigma(g(X)) = \{\emptyset, \Omega\}$.

EXERCISE 2.4. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ define random variables X and $Y_0, Y_1, Y_2, \ldots$ and suppose $\mathbb{E}|X| < \infty$. Define $\mathcal{F}_n := \sigma(Y_0, Y_1, \ldots, Y_n)$ and $X_n = \mathbb{E}[X|\mathcal{F}_n]$. Show that the sequence $X_0, X_1, X_2, \ldots$ is a martingale under $\mathbb{P}$ with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$.

EXERCISE 2.5. In this exercise, we will construct a discrete-time model for a stock $Z = (Z_n)_{n\geq 0}$. Let $X_0, X_1, \ldots$ be a sequence of i.i.d Bernoulli random variables with parameter p (i.e., $\mathbb{P}(X_i = 1) = p$). Define $S_n = \sum_{i=1}^n X_i$ where $S_0 = 0$. Let Z be given by

$$Z_n := r^{2S_n - n}, \qquad r > 0, \qquad n = 0, 1, 2, \ldots.$$

(a) Let $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ where $\mathcal{F}_n := \sigma(X_0, X_1, \ldots, X_n)$. For what value(s) of r is Z and $(\mathbb{P}, \mathbb{F})$ martingale, submartingale and supermartingale?
(b) Let $\mathbb{G} = (\mathcal{G}_n)_{n\geq 0}$ where $\mathcal{G}_n := \sigma(S_0, S_1, \ldots, S_n)$. Suppose r is such that Z is a $(\mathbb{P}, \mathbb{F})$ martingale. Is Z also a $(\mathbb{P}, \mathbb{G})$ martingale?
(c) Suppose that $\widetilde{\mathbb{P}}(X_i = 1) = \tilde{p} \neq p$. Suppose r is such that Z is a $(\mathbb{P}, \mathbb{F})$ martingale. Is Z also a $(\widetilde{\mathbb{P}}, \mathbb{F})$ martingale?

Chapter 3

Brownian motion

The notes from this chapter are primarily taken from (Shreve, 2004, Chapter 3). The goals of this chapter are (i) to define what we mean by "Brownian motion" and (ii) to develop important properties of Brownian motion.

3.1 Scaled random walks

In order to construct Brownian motion, we will begin with a random walk. We return to the setting of Section 1.3.2. Our sample space Ω and a generic element ω in this sample space are given by

$$\Omega = \text{ the set of infinite sequences of Hs and Ts }, \qquad \omega = \omega_1\omega_2\omega_3\ldots,$$

where ω_i is the result of the ith coin toss. We take $\mathcal{F}_n$ to be the σ-algebra generated by observing the first n coin tosses and we set $\mathcal{F} = \sigma(\cup_n \mathcal{F}_n)$. The coin tosses are assumed to be independent and we take $\mathbb{P}(\omega_i = \mathrm{H}) = \mathbb{P}(\omega_i = \mathrm{T}) = (1/2)$.

3.1.1 Symmetric random walk

We construct a *symmetric random walk* $\mathrm{M} = (\mathrm{M}_i)_{i\in\mathbb{N}_0}$ by

$$\mathrm{M}_0 = 0, \qquad \mathrm{M}_k = \sum_{j=1}^{k} \mathrm{X}_j, \qquad \mathrm{X}_i = \begin{cases} +1 & \omega_i = \mathrm{H}, \\ -1 & \omega_i = \mathrm{T}. \end{cases}$$

Let $k_i < k_{i+1}$ for all $i \in \mathbb{N}_0$. As the $(\mathrm{X}_i)_{i\in\mathbb{N}}$ are independent, we clearly have

$$\text{independent increments :} \qquad (\mathrm{M}_{k_2} - \mathrm{M}_{k_1}) \perp\!\!\!\perp (\mathrm{M}_{k_4} - \mathrm{M}_{k_3}).$$

Moreover, as $\mathbb{E}X_j = 0$ and $\mathbb{V}X_j = 1$ for all j we also have

$$\mathbb{E}(M_{k_1} - M_{k_2}) = \sum_{j=k_1+1}^{k_2} \mathbb{E}X_j = 0, \qquad \mathbb{V}(M_{k_2} - M_{k_1}) = \sum_{j=k_1+1}^{k_2} \mathbb{V}X_j = k_2 - k_1. \tag{3.1}$$

Thus, for the discrete time process $M = (M_i)_{i\geq 0}$ we see that variance accumulates at a rate of one per unit time.

Next, we note that M is a martingale with respect to the filtration $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$. To see this, let $k \leq l$ and note that

$$\begin{aligned}\mathbb{E}[M_l|\mathcal{F}_k] &= \mathbb{E}[M_l - M_k + M_k|\mathcal{F}_k] = \mathbb{E}[M_l - M_k|\mathcal{F}_k] + \mathbb{E}[M_k|\mathcal{F}_k] \\ &= \mathbb{E}(M_l - M_k) + M_k. = M_k,\end{aligned}$$

where we have used the independent increments property $(M_l - M_k) \perp\!\!\!\perp \mathcal{F}_k$, equation (3.1) and the fact that $M_k \in \mathcal{F}_k$.

The *quadratic variation* of the symmetric random walk M up to time k, denoted $[M, M]_k$, is defined as

$$[M, M]_k := \sum_{j=1}^{k} (M_j - M_{j-1})^2 = \sum_{j=1}^{k} X_j^2 = k,$$

where we have used $X_j^2 = 1$. The astute reader will notice that $[M, M]_k = \mathbb{V}M_k$. However, it is important to note that *the computation of variance and the computation of quadratic variation are different*! To see this, note that if $\mathbb{P}(\omega_i = H) = p \neq (1/2)$ then $\mathbb{V}X_i \neq 1$ and thus $\mathbb{V}M_k \neq k$. However, since $X_j^2 = 1$ is unaffected by the value of $p = \mathbb{P}(\omega_i = H)$, the computation of $[M, M]_k$ is also unaffected by p. Another way to see that $\mathbb{V}M_k$ and $[M, M]_k$ are different is to note that $\mathbb{V}M_k$ is a *statistical* quantity (i.e., it is an average over all ω) whereas $[M, M]_k$ is computed ω-by-ω (it just turns out that, for each ω we have $[M, M]_k(\omega) = k$.

3.1.2 Scaled symmetric random walk

We now fix a positive integer n. We construct a *scaled symmetric random walk* $W^{(n)} = (W_t^{(n)})_{t\geq 0}$ by

$$W_t^{(n)} = \frac{1}{\sqrt{n}} M_{nt}, \qquad \text{if } nt \in \mathbb{N}_0.$$

If nt is not an integer, we define $W_t^{(n)}$ to be the linear interpolation of $M_{\lfloor nt \rfloor}$ and $M_{\lceil nt \rceil}$. Note that M was a discrete time process whereas $W^{(n)}$ is a continuous time process.

Let $0 = t_0 < t_1 < t_2 < \ldots$ and suppose $nt_j \in \mathbb{N}$ for every j. Then we have

$$\text{independent increments :} \qquad (W_{t_2}^{(n)} - W_{t_1}^{(n)}) \perp\!\!\!\perp (W_{t_4}^{(n)} - W_{t_3}^{(n)}),$$

because non-overlapping increments depend on different coin tosses. For example

$$\begin{aligned} &W^{(100)}_{0.2} - W^{(100)}_{0.0} && \text{depends on coin tosses 1 through 20,} \\ &W^{(100)}_{0.7} - W^{(100)}_{0.2} && \text{depends on coin tosses 21 through 70.} \end{aligned}$$

Next, suppose $0 \le s \le t$ are such that $ns \in \mathbb{N}$ and $nt \in \mathbb{N}$. Then

$$\mathbb{E}(W^{(n)}_t - W^{(n)}_s) = \frac{1}{\sqrt{n}}\mathbb{E}(M_{nt} - M_{ns}) = 0, \qquad \mathbb{V}(W^{(n)}_t - W^{(n)}_s) = \frac{1}{n}\mathbb{V}(M_{nt} - M_{ns}) = t - s. \tag{3.2}$$

Now, let us define the scaled filtration $\mathbb{F}^{(n)} = (\mathcal{F}^{(n)}_t)_{t\ge 0}$ where

$$\mathcal{F}^{(n)}_t := \mathcal{F}_{nt}, \qquad \text{if } nt \in \mathbb{N}_0.$$

The scaled random walk $W^{(n)}$, restricted to the set of t for which $nt \in \mathbb{N}_0$, is a martingale with respect to the filtration $\mathbb{F}^{(n)}$. To see this, assume $0 \le s \le t$ are such that $ns \in \mathbb{N}$ and $nt \in \mathbb{N}$. Then we have

$$\begin{aligned} \mathbb{E}[W^{(n)}_t | \mathcal{F}^{(n)}_s] &= \mathbb{E}[W^{(n)}_t - W^{(n)}_s + W^{(n)}_s | \mathcal{F}^{(n)}_s] = \mathbb{E}[W^{(n)}_t - W^{(n)}_s | \mathcal{F}^{(n)}_s] + \mathbb{E}[W^{(n)}_s | \mathcal{F}^{(n)}_s] \\ &= \mathbb{E}(W^{(n)}_t - W^{(n)}_s) + W^{(n)}_s = W^{(n)}_s, \end{aligned}$$

where we have used $(W^{(n)}_t - W^{(n)}_s) \perp\!\!\!\perp \mathcal{F}^{(n)}_s$, equation (3.2) and $W^{(n)}_s \in \mathcal{F}^{(n)}_s$.

The *quadratic variation* of the scaled symmetric random walk $W^{(n)}$ up to time t, denoted $[W^{(n)}, W^{(n)}]_t$, is defined as

$$[W^{(n)}, W^{(n)}]_t := \sum_{j=1}^{nt} (W^{(n)}_{j/n} - W^{(n)}_{(j-1)/n})^2 = \sum_{j=1}^{nt} \left(\tfrac{1}{\sqrt{n}} X_j\right)^2 = \sum_{j=1}^{nt} \tfrac{1}{n} = t,$$

where we again assume $nt \in \mathbb{N}_0$. Thus, for the scaled symmetric random walk, we see that $\mathbb{V}W^{(n)}_t = [W^{(n)}, W^{(n)}]_t$. However, we emphasize one more time that the *computation of variance is a statistical average over all* ω and the *computation of quadratic variation is done* ω*-by-*ω.

THEOREM 3.1.1. *Fix* $t \ge 0$. *Define a random variable* $W_t := \lim_{n\to\infty} W^{(n)}_t$. *Then* $W_t \sim \mathcal{N}(0, t)$.

PROOF. We will use the Continuity Theorem 2.4.8. We have

$$\begin{aligned} \phi_{W^{(n)}_t}(u) &= \mathbb{E}\exp\left(\mathtt{i}uW^{(n)}_t\right) = \mathbb{E}\exp\left(\frac{\mathtt{i}u}{\sqrt{n}}\sum_{j=1}^{nt} X_j\right) \\ &= \prod_{j=1}^{nt} \mathbb{E}\exp\left(\frac{\mathtt{i}u}{\sqrt{n}} X_j\right) \\ &= \left(\tfrac{1}{2}e^{\mathtt{i}u/\sqrt{n}} + \tfrac{1}{2}e^{-\mathtt{i}u/\sqrt{n}}\right)^{nt} \to e^{-\frac{1}{2}tu^2}, \qquad \text{as } n \to \infty. \end{aligned}$$

The limit as $n \to \infty$ is in fact not easy to show. Nevertheless, the limit above is correct. From Example 2.4.4, we know that if a random variable Z is normally distributed $Z \sim \mathcal{N}(0, t)$ then its characteristic function ϕ_Z is given by

$$\phi_Z(u) = e^{-\frac{1}{2}tu^2}.$$

As $\lim_{n\to\infty} \phi_{W_t^{(n)}} \to \phi_Z$, we have from the Continuity Theorem 2.4.8 that $W_t^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, t)$, as claimed. □

3.2 Brownian motion

Definition 3.2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *Brownian motion* is a stochastic process $W = (W_t)_{t\geq 0}$ that satisfies:

1. $W_0 = 0$.
2. If $0 \leq r < s < t < u < \infty$ then $(W_u - W_t) \perp\!\!\!\perp (W_s - W_r)$.
3. If $0 \leq r < s$ then $W_s - W_r \sim \mathcal{N}(0, s - r)$.
4. The map $t \to W_t$ is continuous for every ω.

It is clear from the previous sections that we can construct a Brownian motion as a limit of a scaled symmetric random walk. Had we simply given Definition 3.2.1 at the beginning of this chapter with no further introduction, one might have legitimately aksed if there exists a process that satisfies the properties of a Brownian motion. There are other methods to prove the existence of Brownian motion. But, the scaled random walk construction is perhaps the most intuitive.

What is Ω in Definition 3.2.1? It could be an infinite series of Hs and Ts, representing movements up and down of a scaled random walk. Or, it could be $\Omega = C_0(\mathbb{R}_+)$, the set of continuous functions on $\mathbb{R}_+$, starting from zero. In this case, an element of Ω is a continuous function $t \to \omega(t)$ and one could simply take $W_t(\omega) = \omega(t)$. Whatever the sample space, the probability of any single element ω is zero: $\mathbb{P}(\omega) = 0$, but probabilities such as $\mathbb{P}(W_t \leq 0)$ are well-defined.

Let $0 < t_1 < t_2 < \ldots < t_d < \infty$. Note that the vector $W := (W_{t_1}, W_{t_2}, \ldots, W_{t_d})$ is a d-dimensional normally distributed random variable. The distribution of a normally distributed random vector is uniquely determined by its mean vector and covariance matrix. We clearly have $\mathbb{E}(W_{t_1}, W_{t_2}, \ldots, W_{t_d}) = (0, 0, \ldots, 0)$. The entries of the covariance matrix are of the following form: for $T \geq t$ we have

$$\begin{aligned}\mathrm{Co}\mathbb{V}[W_T, W_t] &= \mathbb{E}W_T W_t = \mathbb{E}(W_T - W_t + W_t)W_t \\ &= \mathbb{E}(W_T - W_t)W_t + \mathbb{E}W_t^2 = \mathbb{E}W_t\mathbb{E}[W_T - W_t|W_t] + t = t.\end{aligned}$$

Thus, $\mathrm{Co}\mathbb{V}[\mathrm{W}_s, \mathrm{W}_t] = s \wedge t$, and the covariance matrix for $(\mathrm{W}_{t_1}, \mathrm{W}_{t_2}, \ldots, \mathrm{W}_{t_d})$ is

$$\mathbf{C} = \begin{pmatrix} t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & \ldots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \ldots & t_d \end{pmatrix}.$$

DEFINITION 3.2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be on probability space on which a Brownian motion $\mathrm{W} = (\mathrm{W}_t)_{t\geq 0}$ is defined. A *filtration for the Brownian motion* W is a collection of σ-algebras $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying:

1. Information accumulates: if $0 \leq s < t$ then $\mathcal{F}_s \subset \mathcal{F}_t$.
2. Adaptivity: for all $t \geq 0$, we have $\mathrm{W}_t \in \mathcal{F}_t$.
3. Independence of future increments: if $u > t \geq 0$ then $(\mathrm{W}_u - \mathrm{W}_t) \perp\!\!\!\perp \mathcal{F}_t$.

The most natural choice for this filtration $\mathbb{F}$ is the *natural filtration* for W. That is $\mathcal{F}_t = \sigma(\mathrm{W}_u, 0 \leq u \leq t)$. In principle the filtration $(\mathcal{F}_t)_{t\geq 0}$ could contain more than the information obtained by observing W. However, the information in the filtration is not allowed to destroy the independence of future increments of Brownian motion.

Not surprisingly, if $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is a filtration for a Brownian motion W then W is a martingale with respect to this filtration. We see this, let $0 \leq s < t$ and observe that

$$\mathbb{E}[\mathrm{W}_t|\mathcal{F}_s] = \mathbb{E}[\mathrm{W}_t - \mathrm{W}_s + \mathrm{W}_s|\mathcal{F}_s] = \mathbb{E}[\mathrm{W}_t - \mathrm{W}_s|\mathcal{F}_s] + \mathbb{E}[\mathrm{W}_s|\mathcal{F}_s] = \mathbb{E}[\mathrm{W}_t - \mathrm{W}_s] + \mathrm{W}_s = \mathrm{W}_s.$$

3.3 QUADRATIC VARIATION

In Section 3.1, we computed the quadratic variation of a symetric random walk M and a scaled symmetric random walk $\mathrm{W}^{(n)}$. However, we did not rigorously define the notion of quadratic variation. In this Section we will define what we mean by "quadratic variation" and we will compute this quantity for a Brownian motion W. Before defining quadratic variation we first introduce the "first variation."

Let Π be a partition of $[0, \mathrm{T}]$ and define $\|\Pi\|$ as

$$\Pi = \{t_0, t_1, \ldots, t_n\}, \qquad 0 = t_0 < t_1 < \ldots < t_n = \mathrm{T}, \qquad \|\Pi\| = \max_i (t_{i+1} - t_i). \tag{3.3}$$

Let $f : [0, \mathrm{T}] \to \mathbb{R}$. We define the *first variation of f up to time* T, denoted $\mathrm{FV}_\mathrm{T}(f)$, by

$$\mathrm{FV}_\mathrm{T}(f) := \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|.$$

Suppose that $f \in C([0,T])$ and $f'(t)$ exists and is finite for all $t \in (0,T)$. Then, by the *Mean Value Theorem*, there exits $t_j^* \in [t_j, t_{j+1}]$ such that

$$f(t_{j+1}) - f(t_j) = f'(t_j^*)(t_{j+1} - t_j).$$

Thus, if $f \in C([0,T])$ and $f'(t)$ exists and is finite for all $t \in (0,T)$, we have

$$\mathrm{FV}_T(f) := \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |f'(t_j^*)|(t_{j+1} - t_j) = \int_0^T \mathrm{d}t\, |f'(t)|.$$

DEFINITION 3.3.1. Let $f : [0,T] \to \mathbb{R}$. We define the *quadratic variation of f up to time* T, denoted $[f,f]_T$ as

$$[f,f]_T := \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \Big[f(t_{j+1}) - f(t_j)\Big]^2,$$

where Π and $\|\Pi\|$ are as defined in (3.3).

PROPOSITION 3.3.2. *Suppose $f : [0,T] \to \mathbb{R}$ has a continuous first derivative: $f \in C^1((0,T))$. Then $[f,f]_T = 0$.*

PROOF. For any partition Π of $[0,T]$ we have

$$\sum_{j=0}^{n-1} \Big[f(t_{j+1}) - f(t_j)\Big]^2 = \sum_{j=0}^{n-1} [f'(t_j^*)]^2 (t_{j+1} - t_j)^2 \le \|\Pi\| \sum_{j=0}^{n-1} [f'(t_j^*)]^2 (t_{j+1} - t_j).$$

Inserting the above inequality into the definition of $[f,f]_T$, we obtain

$$[f,f]_T \le \lim_{\|\Pi\|\to 0} \|\Pi\| \cdot \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} [f'(t_j^*)]^2 (t_{j+1} - t_j) = 0 \cdot \int_0^T \mathrm{d}t\, |f'(t)|^2 = 0,$$

where we have used $\int_0^T \mathrm{d}t\, |f'(t)|^2 < \infty$ as, by assumption, $f \in C^1((0,T))$. □

In ordinary calculus, we typically deal with functions $f \in C^1$, and hence $[f,f]_T = 0$. For this reason, quadratic variation never arises in usual calculus. However, it turns out that for almost every $\omega \in \Omega$, we have that $t \mapsto W_t(\omega)$ is *not* differentiable. We can see this from the scaled random walk construction of Brownian motion. The slope of the scaled random walk $W_t^{(n)}$ at any t for which $\frac{\mathrm{d}}{\mathrm{d}t} W_t^{(n)}(\omega)$ is defined is

$$\lim_{\varepsilon\to 0} \frac{W_{t+\varepsilon}^{(n)} - W_t^{(n)}}{\varepsilon} = \pm\sqrt{n} \to \pm\infty \qquad \text{as } n \to \infty.$$

Thus, Brownian motion, which we constructed as a limit of a scaled random walk $W(\omega) := \lim_{n\to\infty} W^{(n)}(\omega)$ is not differentiable at any t, $\mathbb{P}$-a.s.. When a function $f \notin C^1((0,T))$, then Proposition 3.3.2 can fail. Indeed, as we will show, paths of BM have strictly positive quadratic variation. It is for this reason that stochastic calculus is different from ordinary calculus.

THEOREM 3.3.3. *Let* W *be a Brownian motion. Then, for all* $T \geq 0$ *we have* $[W,W]_T = T$ *almost surely.*

PROOF. For a fixed partition $\Pi = \{0 = t_0, t_1, \ldots, t_n = T\}$ we define the *sampled quadratic variation of* W, denoted Q_Π, by

$$Q_\Pi := \sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j})^2.$$

Note that $Q_\Pi \to [W,W]_T$ as $\|\Pi\| \to 0$. We will show that $\mathbb{E}Q_\Pi \to T$ and $\mathbb{V}Q_\Pi \to 0$. Using the fact that $W_{t_{j+1}} - W_{t_j} \sim \mathcal{N}(0, t_{j+1} - t_j)$, we compute

$$\mathbb{E}(W_{t_{j+1}} - W_{t_j})^2 = t_{j+1} - t_j, \tag{3.4}$$

$$\begin{aligned}\mathbb{V}(W_{t_{j+1}} - W_{t_j})^2 &= \mathbb{E}(W_{t_{j+1}} - W_{t_j})^4 - \left(\mathbb{E}(W_{t_{j+1}} - W_{t_j})^2\right)^2 \\ &= 3(t_{j+1} - t_j)^2 - (t_{j+1} - t_j)^2 = 2(t_{j+1} - t_j)^2 \end{aligned} \tag{3.5}$$

Using (3.4) and (3.5), we obtain

$$\begin{aligned}\mathbb{E}Q_\Pi &= \sum_{j=0}^{n-1} \mathbb{E}(W_{t_{j+1}} - W_{t_j})^2 = \sum_{j=0}^{n-1} (t_{j+1} - t_j) = T, \\ \mathbb{V}Q_\Pi &= \sum_{j=0}^{n-1} \mathbb{V}(W_{t_{j+1}} - W_{t_j})^2 = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^2 \leq \|\Pi\| \sum_{j=0}^{n-1} 2(t_{j+1} - t_j) = 2\|\Pi\| T.\end{aligned}$$

Thus, $\mathbb{E}Q_\Pi \to T$ and $\mathbb{V}Q_\Pi \to 0$ as $\|\Pi\| \to 0$, which proves that $[W,W]_T = \lim_{\|\Pi\|\to 0} Q_\Pi = T$. □

Suppose $dt \ll 1$ and define $dW_t := W_{t+dt} - W_t$. The above computations show that $\mathbb{E}((dW_t)^2) = dt$ and $\mathbb{V}((dW_t)^2) = 2dt^2$. Since dt^2 is practically zero for $dt \ll 1$, one can imagine that $(dW_t)^2$ is *almost* equal to a constant dt. *Informally*, we write this as

$$dW_t dW_t = dt. \tag{3.6}$$

This informal statement, while not rigorously correct, captures the spirit of the quadratic variation computation for W.

DEFINITION 3.3.4. Let $f, g : [0,T] \to \mathbb{R}$. We define the *covaration of* f *and* g *up to time* T, denoted $[f,g]_T$ as

$$[f,g]_T := \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \left[f(t_{j+1}) - f(t_j)\right]\left[g(t_{j+1}) - g(t_j)\right],$$

where Π and $\|\Pi\|$ are as defined in (3.3).

THEOREM 3.3.5. *Let* W *be a Brownian motion and let* Id *be the identity function:* $\mathrm{Id}(t) = t$. *Then, for all* $T \geq 0$ *we have* $[W, \mathrm{Id}]_T = 0$ *almost surely and* $[\mathrm{Id}, \mathrm{Id}]_T = 0$.

PROOF. For a fixed partition $\Pi = \{0 = t_0, t_1, \ldots, t_n = T\}$ we have

$$\begin{aligned}\left|[W, \mathrm{Id}]_T\right| &= \left| \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j}) \cdot (t_{j+1} - t_j) \right| \\ &= \lim_{\|\Pi\| \to 0} \left| \sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j}) \cdot (t_{j+1} - t_j) \right| \\ &\leq T \cdot \lim_{\|\Pi\| \to 0} \max_{0 \leq j \leq n-1} |W_{t_{j+1}} - W_{t_j}| = T \cdot 0 = 0.\end{aligned}$$

where, in the last step we are using the fact that W is continuous. To see that $[\mathrm{Id}, \mathrm{Id}]_T = 0$, simply note that $g \in C^1([0, T])$ and use Proposition 3.3.2. □

Just as (3.6) captures the spirit of the computation of $[W, W]_T$, the following equations

$$dW_t dt = 0, \qquad dt dt = 0,$$

informally capture the spirit of the $[W, \mathrm{Id}]_T$ and $[\mathrm{Id}, \mathrm{Id}]_T$ computations.

3.4 Markov property of Brownian motion

THEOREM 3.4.1. *Let* $W = (W_t)_{t \geq 0}$ *be a Brownian motion and let* $(\mathcal{F}_t)_{t \geq 0}$ *be a filtration for this Brownian motion. Then* W *is a Markov process.*

PROOF. According to Definition 2.3.6, we must show that there exists a function g such that

$$\mathbb{E}[f(W_T)|\mathcal{F}_t] = g(W_t),$$

where $T \geq t \geq 0$. Noting that $W_t \in \mathcal{F}_t$ and $W_T - W_t \perp\!\!\!\perp \mathcal{F}_t$, we have

$$\mathbb{E}[f(W_T)|\mathcal{F}_t] = \mathbb{E}[f(W_T - W_t + W_t)|\mathcal{F}_t] = \int_{\mathbb{R}} dy\, f(y + W_t)\Gamma(t, 0; T, y) =: g(W_t),$$

where $\Gamma(t, x; T, \cdot)$ is the density of a normal random variable with mean x and variance $T - t$. □

3.5 First hitting time of Brownian motion

THEOREM 3.5.1. *Let* $W = (W_t)_{t \geq 0}$ *be a Brownian motion with a filtration* $(\mathcal{F}_t)_{t \geq 0}$. *Define a process* $Z = (Z_t)_{t \geq 0}$ *by*

$$Z_t = e^{-\frac{1}{2}\sigma^2 t + \sigma W_t}. \tag{3.7}$$

Then Z *is a martingale with respect to* $(\mathcal{F}_t)_{t \geq 0}$.

PROOF. The proof is a simple computation. Fixing $T \geq t$ we have

$$\begin{aligned}\mathbb{E}[Z_T|\mathcal{F}_t] &= \mathbb{E}[e^{-\frac{1}{2}\sigma^2 T+\sigma W_T}|\mathcal{F}_t] = e^{-\frac{1}{2}\sigma^2 T+\sigma W_t}\mathbb{E}[e^{\sigma(W_T-W_t)}|\mathcal{F}_t] \\ &= e^{-\frac{1}{2}\sigma^2 T+\sigma W_t}e^{\frac{1}{2}\sigma^2(T-t)} = e^{-\frac{1}{2}\sigma^2 t+\sigma W_t} = Z_t.\end{aligned}$$

Thus, Z is a martingale. □

The process Z is sometimes referred to as an *exponential martingale*. The exponential martingale will be used to compute the distribution of

$$\tau_m := \inf\{t \geq 0 : W_t = m\}. \tag{3.8}$$

We call τ_m the *first hitting time* or *first passage time* of a Brownian motion W to level m.

THEOREM 3.5.2. *For any $m \in \mathbb{R}$ and $\alpha > 0$ we have*

$$\mathbb{E}e^{-\alpha\tau_m} = e^{-|m|\sqrt{2\alpha}}. \tag{3.9}$$

where τ_m is defined in (3.8).

PROOF. First, assume $m \geq 0$. Let $Z^{(m)} = (Z_t^{(m)})_{t\geq 0}$ be given by

$$Z_t^{(m)} := Z_{t\wedge\tau_m},$$

where Z is given by (3.7). We call $Z^{(m)}$ a *stopped* process as it remains at Z_{τ_m} forever after W hits m. As Z is a martingale it follows that the stopped process $Z^{(m)}$ is also a martingale. Thus, we have

$$1 = Z_0^{(m)} = \mathbb{E}Z_t^{(m)} = \mathbb{E}e^{-\frac{1}{2}\sigma^2 t\wedge\tau_m+\sigma W_{t\wedge\tau_m}}.$$

Now, it turns out that $\mathbb{P}(\tau_m < \infty) = 1$ (a fact that can be proved with relative ease). As a result, we have $\lim_{t\to\infty} t \wedge \tau_m = \tau_m$ and $W_{\tau_m} = m$. Thus, we obtain

$$1 = \lim_{t\to\infty} \mathbb{E}e^{-\frac{1}{2}\sigma^2 t\wedge\tau_m+\sigma W_{t\wedge\tau_m}} = \mathbb{E}\lim_{t\to\infty} e^{-\frac{1}{2}\sigma^2 t\wedge\tau_m+\sigma W_{t\wedge\tau_m}} = \mathbb{E}e^{-\frac{1}{2}\sigma^2\tau_m+\sigma m}.$$

Setting $\sigma = \sqrt{2\alpha}$ we obtain $\mathbb{E}e^{-\alpha\tau_m} = e^{-m\sqrt{2\alpha}}$, which agrees with (3.9) for $m \geq 0$. To obtain (3.9) for $m < 0$, simply note that, since Brownian motion is symmetric about zero, the distribution of τ_m is the same as the distribution of τ_{-m}. □

We can compute $\mathbb{E}\tau_m$ as follows

$$\mathbb{E}\tau_m = \lim_{\alpha\searrow 0} -\frac{d}{d\alpha}\mathbb{E}e^{-\alpha\tau_m} = \lim_{\alpha\searrow 0} -\frac{d}{d\alpha}e^{-|m|\sqrt{2\alpha}} = \lim_{\alpha\searrow 0}\frac{|m|}{\sqrt{\alpha}}e^{-|m|\sqrt{2\alpha}} = \infty, \qquad m \neq 0.$$

Thus, while $\mathbb{P}(\tau_m < \infty) = 1$ we have $\mathbb{E}\tau_m = \infty$.

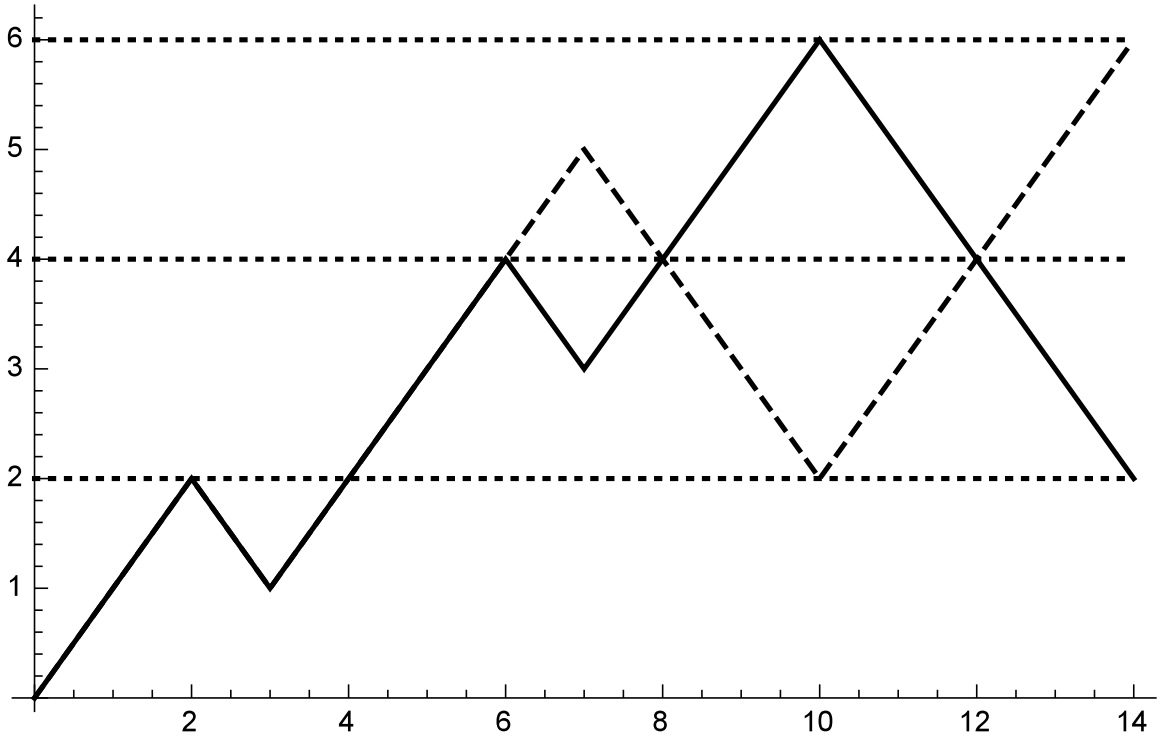

Figure 3.1: An illustration of the reflection principle for Brownian motion.

3.6 Reflection principle

For every sample path of Brownian motion that hits a level $m > 0$ prior to time t and finishes at level $W_t = w \leq m$ there is an equally likely path that finishes at a level $W_t = 2m - w$. In Figure 3.1 these two paths are represented by the solid and dashed lines, respectively. The dotted lines from top to bottom represent the levels $2m - w$, m and w. It follows that

$$\mathbb{P}(\tau_m \leq t, W_t \leq w) = \mathbb{P}(W_t \geq 2m - w), \qquad m > 0, \qquad w \leq m. \tag{3.10}$$

We will use this fact to find the density of τ_m.

Theorem 3.6.1. *For all $m \neq 0$, the first hitting time τ_m of Brownian motion to level m has a density f_{τ_m}, which given by*

$$f_{\tau_m}(t) = \mathbb{1}_{\{t \geq 0\}} \frac{|m|}{t\sqrt{2\pi t}} e^{-m^2/(2t)}. \tag{3.11}$$

Proof. Consider the case $m > 0$. Substitute $w = m$ into (3.10) to obtain

$$\mathbb{P}(\tau_m \leq t, W_t \leq m) = \mathbb{P}(W_t \geq m).$$

Now, note that $W_t \geq m$ implies that $\tau_m \leq t$. And thus

$$\mathbb{P}(\tau_m \leq t, W_t \geq m) = \mathbb{P}(W_t \geq m).$$

Adding the above two equations, we obtain

$$2\mathbb{P}(W_t \geq m) = \mathbb{P}(\tau_m \leq t, W_t \leq m) + \mathbb{P}(\tau_m \leq t, W_t \geq m) = \mathbb{P}(\tau_m \leq t).$$

The density (3.11) is obtained using

$$\begin{aligned} f_{\tau_m}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(\tau_m \leq t) = \frac{\mathrm{d}}{\mathrm{d}t} 2\mathbb{P}(W_t \geq m) = \frac{\mathrm{d}}{\mathrm{d}t} 2 \int_m^\infty \mathrm{d}x \, \frac{1}{\sqrt{2\pi t}} \exp\Big(-\frac{x^2}{2t}\Big) \\ &= \frac{\mathrm{d}}{\mathrm{d}t} 2 \int_{m/\sqrt{t}}^\infty \mathrm{d}y \, \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{y^2}{2}\Big) = \frac{m}{t\sqrt{2\pi t}} \mathrm{e}^{-m^2/(2t)} \end{aligned}$$

The case $m \leq 0$ is proved in a similar manner. □

Now, let us define the *running maximum*, denoted $\overline{W} = (\overline{W}_t)_{t\geq 0}$, of Brownian motion

$$\overline{W}_t = \max_{0\leq s\leq t} W_s.$$

Note that $\overline{W}_t \geq m$ if an only if $\tau_m \leq t$. Thus, we have from (3.10) that

$$\mathbb{P}(\overline{W}_t \geq m, W_t \leq w) = \mathbb{P}(W_t \geq 2m - w), \qquad m > 0, \qquad w \leq m. \tag{3.12}$$

We can use (3.12) to obtain the joint density of $(W_t, \overline{W}_t)$.

THEOREM 3.6.2. *For any $t > 0$, the joint density of Brownian motion W_t and its running maximum $\overline{W}_t$ is*

$$f_{W_t,\overline{W}_t}(w, m) = \frac{2(2m - w)}{t\sqrt{2\pi t}} \mathrm{e}^{-(2m-w)^2/(2t)}, \qquad m > 0, \qquad w \leq m. \tag{3.13}$$

PROOF. Note that

$$\begin{aligned} \mathbb{P}(W_t \leq w, \overline{W}_t \leq m) &= \mathbb{P}(W_t \leq w) - \mathbb{P}(W_t \leq w, \overline{W}_t \geq m) \\ &= \mathbb{P}(W_t \leq w) - \mathbb{P}(W_t \geq 2m - w) = \mathbb{P}(W_t \leq w) - \int_{2m-w}^\infty \mathrm{d}x \, f_{W_t}(x), \end{aligned}$$

where f_{W_t} is the density of W_t, which is a $\mathcal{N}(0, t)$ random variable. To obtain the density (3.13) use

$$f_{W_t,\overline{W}_t}(w, m) = \frac{\partial^2}{\partial m \partial w} \mathbb{P}(W_t \leq w, \overline{W}_t \leq m) = -\frac{\partial^2}{\partial m \partial w} \int_{2m-w}^\infty \mathrm{d}x \, f_{W_t}(x).$$

The rest of the computation is algebra. □

COROLLARY 3.6.3. *For any $t > 0$, the conditional density of $\overline{\mathrm{W}}_t$ given W_t is given by*

$$f_{\overline{\mathrm{W}}_t|\mathrm{W}_t}(m,w) = \frac{2(2m-w)}{t}\mathrm{e}^{-2m(m-w)/t}, \qquad m > 0, \qquad w \leq m. \tag{3.14}$$

PROOF. Expression (3.14) follows directly from

$$f_{\overline{\mathrm{W}}_t|\mathrm{W}_t}(m,w) = \frac{f_{\mathrm{W}_t,\overline{\mathrm{W}}_t}(w,m)}{f_{\mathrm{W}_t}(w)}.$$

where f_{W_t} is the density of W_t, which is $\mathcal{N}(0,t)$. □

3.7 EXERCISES

EXERCISE 3.1. Let W be a Brownian motion and let $\mathbb{F}$ be a filtration for W. Show that $\mathrm{W}_t^2 - t$ is a martingale with respect to the filtration $\mathbb{F}$.

EXERCISE 3.2. A *Poisson process* with intensity λ is a continuous-time process $\mathrm{N} = (\mathrm{N}_t)_{t\geq 0}$ with the property that, for any $t \geq 0$, we have

$$\mathbb{P}(\mathrm{N}_t = n) = \frac{(\lambda t)^n}{n!}\mathrm{e}^{-\lambda t}, \qquad n \in \{0,1,2,\ldots\}.$$

In other words, N_t is distributed as a *Poisson random variable* with parameter λt. Compute the characteristic function of $\mathrm{W}_{\mathrm{N}_t}$ where N is a Poisson process with intensity λ and the Brownian motion W is independent of the Poisson process N.

EXERCISE 3.3. The mth variariation of a function f, over the interval $[0,\mathrm{T}]$ is defined as

$$\mathrm{V}_{\mathrm{T}}(m,f) := \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}|f(t_{j+1}) - f(t_j)|^m, \quad \Pi = \{0 = t_0, t_1, \ldots t_n = \mathrm{T}\}, \quad \|\Pi\| = \max_j(t_{j+1} - t_j).$$

Show that $\mathrm{V}_{\mathrm{T}}(1,\mathrm{W}) = \infty$ and $\mathrm{V}_{\mathrm{T}}(3,\mathrm{W}) = 0$, where W is a Brownian motion.

EXERCISE 3.4. Define

$$\mathrm{X}_t = \mu t + \mathrm{W}_t, \qquad \tau_m := \inf\{t \geq 0 : \mathrm{X}_t = m\},$$

where $\mathrm{W} = (\mathrm{W}_t)_{t\geq 0}$ is a Brownian motion. Let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration for W. Show that Z is a martingale with respect to $\mathbb{F}$ where

$$\mathrm{Z}_t = \exp\left(\sigma\mathrm{X}_t - (\sigma\mu + \sigma^2/2)t\right).$$

Assume $\mu > 0$ and $m \geq 0$. Assume further that $\tau_m < \infty$ with probability one and the stopped process $\mathrm{Z}_{t\wedge\tau_m}$ is a martingale. Find the Laplace transform $\mathbb{E}\mathrm{e}^{-\alpha\tau_m}$.

Chapter 4

Stochastic calculus

The notes from this chapter are taken primarily from (Shreve, 2004, Chapters 4 and 5).

4.1 Itô Integrals

If a function $g : [0, T] \to \mathbb{R}$ satisfies $g \in C^1([0, T])$, then we can define

$$\int_0^T f(t)\mathrm{d}g(t) := \int_0^T f(t)g'(t)\mathrm{d}t.$$

where the right-hand side is a Riemman integral with respect to t. However, as Brownian motion is not differentiable, this method of integration will not work when we consider integrals of the form $\int_0^T \Delta_t \mathrm{d}W_t$. Our goal is to make sense of $\int_0^T \Delta_t \mathrm{d}W_t$.

Assumption 4.1.1. In what follows $W = (W_t)_{t\geq 0}$ will always represent a Brownian motion and $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ will always be a filtration for this Brownian motion. We shall assume the integrand $\Delta = (\Delta_t)_{t\geq 0}$ is adapted to $\mathbb{F}$, meaning $\Delta_t \in \mathcal{F}_t$ for all t.

Note that the process Δ *can* and, in many cases, *will* be random. However, the information available in $\mathcal{F}_t$ will always be sufficient to determine the value of Δ_t at time t. Also note, since $(W_T - W_t) \perp\!\!\!\perp \mathcal{F}_t$ for $T > t$, it follows that $(W_T - W_t) \perp\!\!\!\perp \Delta_t$. In other words, future increments of Brownian motion are independent of the Δ process.

Itô integrals for simple integrands

To begin let us assume that Δ is a *simple process*, meaning Δ is of the form

$$\Delta_t = \sum_{j=0}^{n-1} \Delta_{t_j} \mathbb{1}_{\{t_j \leq t < t_{j+1}\}}, \qquad 0 = t_0 < t_1 < \ldots t_n = T, \qquad \Delta_{t_j} \in \mathcal{F}_{t_j}.$$

Since the process Δ is constant over intervals of the form $[t_j, t_{j+1})$, it makes sense to define

$$I_T = \int_0^T \Delta_t dW_t := \sum_{j=0}^{n-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}). \qquad \text{(for } \Delta \text{ a simple process)} \tag{4.1}$$

Let us establish some properties of the process $I = (I_t)_{t\geq 0}$.

THEOREM 4.1.2. *The process* $I = (I_t)_{t\geq 0}$ *defined in* (4.1) *is a martingale with respect to the filtration* $(\mathcal{F}_t)_{t\geq 0}$.

PROOF. Without loss of generality assume $T = t_n$ and $t = t_i$ for some $0 \leq i \leq n-1$ (we can always re-define our time grid so that this is true). Then we have

$$\begin{aligned}
\mathbb{E}[I_T|\mathcal{F}_t] &= \sum_{j=0}^{n-1} \mathbb{E}[\Delta_{t_j}(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_i}] \\
&= \sum_{j=0}^{i-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \sum_{j=i}^{n-1} \mathbb{E}[\Delta_{t_j}(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_i}] \\
&= \sum_{j=0}^{i-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \sum_{j=i}^{n-1} \mathbb{E}[\Delta_{t_j}\mathbb{E}[W_{t_{j+1}} - W_{t_j}|\mathcal{F}_{t_j}]|\mathcal{F}_{t_i}] \\
&= \sum_{j=0}^{i-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) = I_t.
\end{aligned}$$

Thus, the process I is a martingale, as claimed. □

Note that $I_0 = 0$ and therefore $\mathbb{E}I_t = 0$ for all $t \geq 0$.

THEOREM 4.1.3 (ITÔ ISOMETRY). *The process* $I = (I_t)_{t\geq 0}$ *defined in* (4.1) *satisfies*

$$\mathbb{V}I_T = \mathbb{E}I_T^2 = \mathbb{E}\int_0^T \Delta_t^2 dt. \tag{4.2}$$

PROOF. We have

$$\begin{aligned}
\mathbb{E}I_T^2 &= \sum_{j=0}^{n-1}\sum_{i=0}^{n-1} \mathbb{E}\Delta_{t_i}\Delta_{t_j}(W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}) \\
&= \sum_{j-0}^{n-1} \mathbb{E}\Delta_{t_j}^2(W_{t_{j+1}} - W_{t_j})^2 + 2\sum_{j=0}^{n-1}\sum_{i=0}^{j-1} \mathbb{E}\Delta_{t_i}\Delta_{t_j}(W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}) \\
&= \sum_{j-0}^{n-1} \mathbb{E}\Delta_{t_j}^2\mathbb{E}[(W_{t_{j+1}} - W_{t_j})^2|\mathcal{F}_{t_j}] + 2\sum_{j=0}^{n-1}\sum_{i=0}^{j-1} \mathbb{E}\Delta_{t_i}\Delta_{t_j}(W_{t_{i+1}} - W_{t_i})\mathbb{E}[(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_j}] \\
&= \sum_{j-0}^{n-1} \mathbb{E}\Delta_{t_j}^2(t_{j+1} - t_j) = \mathbb{E}\int_0^T \Delta_t^2 dt.
\end{aligned}$$

□

THEOREM 4.1.4. *The process* $I = (I_t)_{t\geq 0}$ *defined in* (4.1) *satisfies*

$$[I, I]_T = \int_0^T \Delta_t^2 dt. \tag{4.3}$$

PROOF. Fix a partition Π of $[t_j, t_{j+1}]$

$$t_j = s_0 < s_1 < \ldots < s_m = t_{j+1}.$$

From the definition of Quadratic variation, we compute

$$\begin{aligned}[I, I]_{t_{j+1}} - [I, I]_{t_j} &= \lim_{\|\Pi\|\to 0} \sum_{i=0}^{m-1} \left(I_{s_{i+1}} - I_{s_i}\right)^2 = \Delta_{t_j}^2 \lim_{\|\Pi\|\to 0} \sum_{i=0}^{m-1} \left(W_{s_{i+1}} - W_{s_i}\right)^2 \\ &= \Delta_{t_j}^2 \left([W, W]_{t_{j+1}} - [W, W]_{t_j}\right) = \Delta_{t_j}^2 \left(t_{j+1} - t_j\right).\end{aligned}$$

Thus, we obtain

$$[I, I]_T = \sum_{j=0}^{n-1} \left([I, I]_{t_{j+1}} - [I, I]_{t_j}\right) = \sum_{j=0}^{n-1} \Delta_{t_j}^2 \left(t_{j+1} - t_j\right) = \int_0^T \Delta_t^2 dt,$$

as claimed. □

When we computed $\mathbb{V}W_T$ and $[W, W]_T$ we found that these two quantities were equal, even though the computations for these quantities were completely different. From (4.2) and (4.3) we now see how the variance and quadratic variation of a stochastic process can be different. Note that $\mathbb{V}I_T$ is a non-random constant, whereas $[I, I]_T$ is random.

ITÔ INTEGRALS FOR GENERAL INTEGRANDS

Clearly, it is rather restrictive to limit ourselves to integrands Δ that are simple processes. We now allow the process Δ to be any process that is adapted to $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and which satisfies the following integrability condition

$$\mathbb{E} \int_0^T \Delta_t^2 dt < \infty. \tag{4.4}$$

To construct an Itô integral with Δ as the integrand, we first approximate Δ by a simple process

$$\Delta_t \approx \Delta_t^{(n)} := \sum_{j=0}^{n-1} \Delta_{t_j} \mathbb{1}_{\{t_j \leq t < t_{j+1}\}}, \qquad 0 \leq t_0 < t_1 < \ldots < t_n = T.$$

As $n \to \infty$ the process $\Delta^{(n)}$ converges to Δ in the sense that

$$\lim_{n\to\infty} \mathbb{E} \int_0^T \left(\Delta_t - \Delta_t^{(n)}\right)^2 dt = 0. \tag{4.5}$$

We now *define* the Itô integral for a general integrand Δ by

$$I_T \equiv \int_0^T \Delta_t dW_t := \lim_{n\to\infty} \int_0^T \Delta_t^{(n)} dW_t. \tag{4.6}$$

Note that the integrals $\int_0^T \Delta_t^{(n)} dt$ are well-defined for every n, since $\Delta^{(n)}$ is a simple process. Furthermore, the condition (4.5) ensures that the limit exists in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. The Itô integral for general integrands inherits the properties we established for simple integrands.

THEOREM 4.1.5. *Let W be a Brownian motion and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration for this Brownian motion. Let $\Delta = (\Delta_t)_{0\leq t\leq T}$ be adapted to the filtration $\mathbb{F}$ and satisfy (4.5). Let $I = (I_t)_{0\leq t\leq T}$ be given by $I_t = \int_0^t \Delta_s dW_s$, where the integral is defined as in (4.6). Then the process I has the following properties.*

1. *The sample paths of I are continuous.*
2. *The process I is adapted to the filtration $\mathbb{F}$. That is, $I_t \in \mathcal{F}_t$ for all t.*
3. *If $\Gamma = (\Gamma_t)_{0\leq t\leq T}$ satisfies the same conditions as Δ, then*
$$\int_0^T (a\Delta_t + b\Gamma_t)\, dW_t = a\int_0^T \Delta_t dW_t + b\int_0^T \Gamma_t dW_t,$$
where a and b are constants.
4. *The process I is a martingale with respect to the filtration $\mathbb{F}$.*
5. *We have the Itô isometry $\mathbb{E}I_t^2 = \mathbb{E}\int_0^T \Delta_t^2 dt$.*
6. *The quadratic variation of I is given by $[I, I]_T = \int_0^T \Delta_t^2 dt$.*

4.2 Itô formula

If f and g are differentiable functions then we can compute

$$\frac{d}{dt} f(g(t)) = f'(g(t)) \cdot g'(t).$$

Thus, multiplying through by dt we obtain

$$df(g(t)) = f'(g(t)) \cdot g'(t)dt = f(g(t))dg(t), \qquad dg(t) = g'(t)dt.$$

This formula is correct in the sense that

$$f(g(T)) - f(g(0)) = \int_0^T df(g(t)) = \int_0^T f'(g(t)) \cdot g'(t)dt = \int_0^T f'(g(t))dg(t).$$

Although paths of a Brownian motion W are not differentiable, we might suspect that

$$f(W_T) - f(W_0) = \int_0^T df(W_t) \stackrel{?}{=} \int_0^T f'(W_t)dW_t. \qquad \text{(Not correct!)} \tag{4.7}$$

Unfortunately, as indicated above, equation (4.7) is not correct.

THEOREM 4.2.1. *Let* $W = (W_t)_{t\geq 0}$ *be a Brownian motion and suppose* $f : \mathbb{R} \to \mathbb{R}$ *satisfies* $f \in C^2(\mathbb{R})$. *Then, for any* $T \geq 0$ *we have*

$$f(W_T) - f(W_0) = \int_0^T f'(W_t)dW_t + \tfrac{1}{2}\int_0^T f''(W_t)dt. \tag{4.8}$$

PROOF. We shall simply sketch the proof of Theorem 4.2.1. Suppose for simplicity that f is analytic (i.e., that f is equal to its power series expansion at every point). Let $0 = t_0 < t_1 < \ldots < t_n = T$ be a partition Π of $[0, T]$. Then

$$\int_0^T df(W_t) = f(W_T) - f(W_0) = \sum_{j=0}^{n-1} \left(f(W_{t_{j+1}}) - f(W_{t_j})\right) = \sum_{j=0}^{n-1} \left(A_j + B_j + C_j\right),$$

where we have defined

$$\begin{aligned} A_j &:= f'(W_{t_j})\left(W_{t_{j+1}} - W_{t_j}\right), \\ B_j &:= \tfrac{1}{2}f''(W_{t_j})\left(W_{t_{j+1}} - W_{t_j}\right)^2, \\ C_j &:= \tfrac{1}{3!}f'''(W_{t_j})\left(W_{t_{j+1}} - W_{t_j}\right)^3 + \ldots. \end{aligned}$$

In the limit as $\|\Pi\| \to 0$ we have

$$\sum_{j=0}^{n-1} A_j \to \int_0^T f'(W_t)dW_t, \qquad \sum_{j=0}^{n-1} B_j \to \int_0^T \tfrac{1}{2}f''(W_t)dt, \qquad \sum_{j=0}^{n-1} C_j \to 0.$$

□

EXAMPLE 4.2.2. What is $\int_0^T W_t dW_t$? To answer this question, consider $f(W_t)$ with $f(x) = x^2$. According to equation (4.8) we have

$$W_T^2 - W_0^2 = \int_0^T 2W_t dW_t + \int_0^T dt$$

where we have used $f'(x) = 2x$ and $f''(x) = 2$. Noting that $W_0 = 0$ and solving for $\int_0^T W_t dW_t$ we obtain

$$\int_0^T W_t dW_t = \tfrac{1}{2}W_T^2 - \tfrac{1}{2}T.$$

Not surprisingly, there are stochastic processes that are not adequately described by Brownian motion alone. However, a large class of stochastic processes can be constructed from Brownian motion.

DEFINITION 4.2.3. Let $W = (W_t)_{t\geq 0}$ be a Brownian motion and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration for this Brownian motion. An *Itô process* is any process $X = (X_t)_{t\geq 0}$ of the form

$$X_t = X_0 + \int_0^t \Theta_s ds + \int_0^t \Delta_s dW_s, \tag{4.9}$$

where $\Theta = (\Theta_t)_{t\geq 0}$ and $\Delta = (\Delta_t)_{t\geq 0}$ are adapted to the filtration $\mathbb{F}$ and satisfy

$$\int_0^T |\Theta_t| dt < \infty, \qquad \mathbb{E}\int_0^T \Delta_t^2 dt < \infty, \qquad \forall\, T \geq 0,$$

and X_0 is not random.

We sometimes write an Itô process in *differential form*

$$dX_t = \Theta_t dt + \Delta_t dW_t. \tag{4.10}$$

Expression (4.10) literally means that X satisfies (4.9). Informally, the differential form can be understood as follows: in a small interval of time δt , the process X changes according to

$$X_{t+\delta t} - X_t \approx \Theta_t \delta t + \Delta_t \left(W_{t+\delta t} - W_t\right). \tag{4.11}$$

In fact, noting that $W_{t+\delta t} - W_t \sim \mathcal{N}(0, \delta t)$ and $W_{t+\delta t} - W_t \perp\!\!\!\perp \mathcal{F}_t$, one can use expression (4.11) to simulate the increment $X_{t+\delta t} - X_t$. This way of simulating X is called the *Euler scheme* and is the workhorse of many Monte Carlo methods.

LEMMA 4.2.4. *The quadratic variation* $[X, X]_T$ *of an Itô process* (4.9) *is given by*

$$[X, X]_T = \int_0^T \Delta_t^2 dt.$$

PROOF. We sketch the proof of Lemma 4.2.4. Let $0 = t_0 < t_1 < \ldots < t_n = T$ be a partition Π of $[0, T]$. By definition we have

$$[X, X]_T = \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \left(X_{t_{j+1}} - X_{t_j}\right)^2 = \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \left(A_j + B_j + C_j\right)$$

where we have defined

$$\begin{aligned} A_j &:= \left(I_{t_{j+1}} - I_{t_j}\right)^2, \\ B_j &:= \left(J_{t_{j+1}} - J_{t_j}\right)^2, \\ C_j &:= 2\left(I_{t_{j+1}} - I_{t_j}\right)\left(J_{t_{j+1}} - J_{t_j}\right). \end{aligned}$$

with

$$I_t = \int_0^t \Delta_s dW_s, \qquad J_t = \int_0^t \Theta_s ds.$$

In the limit as $\|\Pi\| \to 0$ we obtain

$$\sum_{j=0}^{n-1} A_j \to [I, I]_T, \qquad \sum_{j=0}^{n-1} B_j \to 0, \qquad \sum_{j=0}^{n-1} C_j \to 0.$$

The proof of these limits is similar to the proof of Theorem 4.1.4. □

DEFINITION 4.2.5. Let $X = (X_t)_{t\geq 0}$ be an Itô process, as described in Definition 4.2.3. Let $\Gamma = (\Gamma_t)_{t\geq 0}$ be adapted to the filtration of the Brownian motion $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. We define

$$\int_0^T \Gamma_t dX_t := \int_0^T \Gamma_t \Theta_t dt + \int_0^T \Gamma_t \Delta_t dW_t,$$

where we assume

$$\int_0^T |\Gamma_t \Theta_t| dt < \infty, \qquad \mathbb{E}\int_0^T (\Gamma_t \Delta_t)^2 \, dt < \infty, \qquad \forall\, T \geq 0.$$

THEOREM 4.2.6 (ITÔ FORMULA IN ONE DIMENSION). *Let $X = (X_t)_{t\geq 0}$ be an Itô process and suppose $f : \mathbb{R} \to \mathbb{R}$ satisfies $f \in C^2(\mathbb{R})$. Then, for any $T \geq 0$ we have*

$$f(X_T) - f(X_0) = \int_0^T f'(X_t) dX_t + \tfrac{1}{2}\int_0^T f''(X_t) d[X, X]_t.$$

PROOF. The proof of Theorem 4.2.6 if very similar to the proof of Theorem 4.2.1. We outline the proof here. Suppose for simplicity that f is analytic (i.e., that f is equal to its power series expansion at every point). Let $0 = t_0 < t_1 < \ldots < t_n = T$ be a partition Π of $[0, T]$. Then

$$\int_0^T df(X_t) = f(X_T) - f(X_0) = \sum_{j=0}^{n-1} \left(f(X_{t_{j+1}}) - f(X_{t_j})\right) = \sum_{j=0}^{n-1} \left(A_j + B_j + C_j\right),$$

where we have defined

$$\begin{aligned} A_j &:= f'(X_{t_j}) \left(X_{t_{j+1}} - X_{t_j}\right), \\ B_j &:= \tfrac{1}{2} f''(X_{t_j}) \left(X_{t_{j+1}} - X_{t_j}\right)^2, \\ C_j &:= \tfrac{1}{3!} f'''(X_{t_j}) \left(X_{t_{j+1}} - X_{t_j}\right)^3 + \ldots. \end{aligned}$$

In the limit as $\|\Pi\| \to 0$ we have

$$\sum_{j=0}^{n-1} A_j \to \int_0^T f'(X_t) dX_t, \qquad \sum_{j=0}^{n-1} B_j \to \int_0^T \tfrac{1}{2} f''(X_t) d[X, X]_t, \qquad \sum_{j=0}^{n-1} C_j \to 0.$$

□

In differential form, with X given by (4.10), Itô's formula becomes

$$\begin{aligned} df(X_t) &= f'(X_t) dX_t + \tfrac{1}{2} f''(X_t) d[X, X]_t \\ &= f'(X_t) \left(\Theta_t dt + \Delta_t dW_t\right) + \tfrac{1}{2} f''(X_t) \Delta_t^2 dt, \end{aligned} \tag{4.12}$$

where we have used $d[X, X]_t = \Delta_t^2 dt$. Perhaps the easiest way to remember (4.12) is to use the following two-step procedure:

1. Expand $f(X_t + dX_t) - f(X_t)$ to second order about the point X_t

$$df(X_t) = f(X_t + dX_t) - f(X_t) = f'(X_t)dX_t + \tfrac{1}{2}f''(X_t)(dX_t)^2, \tag{4.13}$$

2. Insert the differential $dX_t = \Theta_t dt + \Delta_t dW_t$ into (4.13), expand $(dX_t)^2$ and use the rules

$$dW_t dW_t = dt, \qquad dW_t dt = 0, \qquad dt dt = 0.$$

The resulting formula gives the correct expression for $df(X_t)$.

EXAMPLE 4.2.7. Let X be an Itô process with the following dynamics

$$dX_t = \mu_t X_t dt + \sigma_t X_t dW_t. \tag{4.14}$$

Assuming $\mu = (\mu_t)_{t\geq 0}$ and $\sigma = (\sigma_t)_{t\geq 0}$ are bounded above and below and $X_0 > 0$, the process X remains strictly positive. We call X a *generalized geometric Brownian motion*. The "geometric" part refers to the fact that the relative step size dX_t/X_t has dyanmics $\mu_t dt + \sigma_t dW_t$. The "generalized" part refers to the fact that the processes σ and μ are stochastic rather than constant. Define $Y_t = X_t^p$. What is dY_t? Let $f(x) = x^p$. Then $f'(x) = px^{p-1}$ and $f''(x) = p(p-1)x^{p-2}$. Thus, we have

$$\begin{aligned}
dY_t = df(X_t) &= pX_t^{p-1}dX_t + \tfrac{1}{2}p(p-1)X_t^{p-2}(dX_t)^2 \\
&= pX_t^{p-1}\left(\mu_t X_t dt + \sigma_t X_t dW_t\right) + \tfrac{1}{2}p(p-1)X_t^{p-2}\left(\mu_t X_t dt + \sigma_t X_t dW_t\right)^2 \\
&= pX_t^{p-1}\left(\mu_t X_t dt + \sigma_t X_t dW_t\right) + \tfrac{1}{2}p(p-1)X_t^{p-2}\sigma_t^2 X_t^2 dt \\
&= \left(p\mu_t + \tfrac{1}{2}p(p-1)\sigma_t^2\right)X_t^p dt + p\sigma_t X_t^p dW_t \\
&= \left(p\mu_t + \tfrac{1}{2}p(p-1)\sigma_t^2\right)Y_t dt + p\sigma_t Y_t dW_t.
\end{aligned}$$

We see from the last line that $Y = (Y_t)_{t\geq 0}$ is also a generalized geometric Brownian motion.

EXAMPLE 4.2.8. Let X have generalized geometric Brownian motion dynamics as in (4.14). We would like to find an explicit expression for X_t (i.e., an expression of the form $X_t = \ldots$ where $\ldots$ does not contain X). To this end,we let $Y_t = \log X_t$. With $f(x) = \log x$ we have $f'(x) = 1/x$ and $f''(x) = -1/x^2$. Thus, we have

$$dY_t = \frac{1}{X_t}dX_t + \frac{1}{2}\frac{-1}{X_t^2}(dX_t)^2 = \left(\mu_t - \tfrac{1}{2}\sigma_t^2\right)dt + \sigma_t dW_t.$$

Thus, we have

$$\begin{aligned}
X_T = \exp(Y_T) &= \exp\left(Y_0 + \int_0^T \left(\mu_t - \tfrac{1}{2}\sigma_t^2\right)dt + \int_0^T \sigma_t dW_t\right) \\
&= X_0 \exp\left(\int_0^T \left(\mu_t - \tfrac{1}{2}\sigma_t^2\right)dt + \int_0^T \sigma_t dW_t\right),
\end{aligned} \tag{4.15}$$

where we have used $Y_0 = \log X_0$.

PROPOSITION 4.2.9. *Let* $W = (W_t)_{t\geq 0}$ *be a Brownian motion. Suppose* $g : \mathbb{R}_+ \to \mathbb{R}_+$ *is a deterministic function. Then*

$$I_T := \int_0^T g(t)dW_t \sim \mathcal{N}(0, v(T)), \qquad v(T) = \int_0^T g^2(t)dt.$$

PROOF. Set $\mu_t = 0$ and $\sigma_t = ug(t)$ in (4.14) where u is a constant. And suppose $X_0 = 1$. Then we have

$$X_T = 1 + \int_0^T ug(t)X_t dW_t,$$

which is a martingale since Itô integrals are martingales. From (4.15), we know that X_T can be written explicitly as

$$X_T = \exp\left(-\tfrac{1}{2}\int_0^T u^2 g^2(t)dt + \int_0^T ug(t)dW_t\right) = \exp\left(-u^2\tfrac{1}{2}\int_0^T g^2(t)dt + uI_T\right).$$

Since X is a martingale, we have

$$1 = \mathbb{E}X_T = \mathbb{E}\exp\left(-u^2\tfrac{1}{2}\int_0^T g^2(t)dt + uI_T\right) \qquad \Rightarrow \qquad \mathbb{E}\exp\left(uI_T\right) = \exp\left(u^2\tfrac{1}{2}\int_0^T g^2(t)dt\right).$$

Note that $\mathbb{E}\exp(uI_T) = M_{I_T}(u)$ is the moment generating function of I_T and $\exp\left(u^2\frac{1}{2}\int_0^T g^2(t)dt\right)$ is the moment generating function of a normal random variable with mean zero and variance $v(T) = \int_0^T g^2(t)dt$. Thus, $I_T \sim \mathcal{N}(0, v(T))$, as claimed. □

4.3 MULTIVARIATE STOCHASTIC CALCULUS

DEFINITION 4.3.1. A *d-dimensional Brownian motion* is a process

$$W = (W_t^1, W_t^2, \ldots, W_t^d)_{t\geq 0}$$

with the the following properties.

1. Each $W^i = (W_t^i)_{t\geq 0}$, $i = 1, 2, \ldots d$, is a one-dimensional Brownian motion.
2. The processes $(W^i)_{1\leq i\leq d}$ are independent.

A *filtration for* W is a collection of σ-algebras $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ such that

1. Information accumulates: $\mathcal{F}_s \subset \mathcal{F}_t$ for all $0 \leq s < t$.
2. Adaptivity: $W_t \in \mathcal{F}_t$ for all $t \geq 0$.
3. Independent increments: for $0 \leq s < t$ we have $W_t - W_s \perp\!\!\!\perp \mathcal{F}_s$.

THEOREM 4.3.2. *Let* $\mathrm{W} = (\mathrm{W}^1_t, \mathrm{W}^2_t, \ldots, \mathrm{W}^d_t)_{t\geq 0}$ *be a* d*-dimensional Brownian motion. The covariation of independent components of* W *is zero:* $[\mathrm{W}^i, \mathrm{W}^j]_{\mathrm{T}} = 0$ *for all* $i \neq j$ *and* $\mathrm{T} \geq 0$.

PROOF. Let $0 = t_0 < t_1 < \ldots < t_n = \mathrm{T}$ be a partition Π of $[0, \mathrm{T}]$. The sampled covariation C_Π of W^i and W^j is given by

$$\mathrm{C}_\Pi = \sum_{k=0}^{n-1} \left(\mathrm{W}^i_{t_{k+1}} - \mathrm{W}^i_{t_k}\right)\left(\mathrm{W}^j_{t_{k+1}} - \mathrm{W}^j_{t_k}\right)$$

Since $\mathbb{E}\left(\mathrm{W}^i_{t_{k+1}} - \mathrm{W}^i_{t_k}\right)\left(\mathrm{W}^j_{t_{k+1}} - \mathrm{W}^j_{t_k}\right) = 0$, we clearly have $\mathbb{E}\mathrm{C}_\Pi = 0$. Next, we compute the variance of C_Π. We have

$$\begin{aligned}
\mathbb{V}\mathrm{C}_\Pi = \mathbb{E}\mathrm{C}^2_\Pi &= \sum_{k=0}^{n-1}\sum_{l=0}^{n-1} \mathbb{E}\left(\mathrm{W}^i_{t_{k+1}} - \mathrm{W}^i_{t_k}\right)\left(\mathrm{W}^j_{t_{k+1}} - \mathrm{W}^j_{t_k}\right)\left(\mathrm{W}^i_{t_{l+1}} - \mathrm{W}^i_{t_l}\right)\left(\mathrm{W}^j_{t_{l+1}} - \mathrm{W}^j_{t_l}\right) \\
&= \sum_{k=0}^{n-1} \mathbb{E}\left(\mathrm{W}^i_{t_{k+1}} - \mathrm{W}^i_{t_k}\right)^2\left(\mathrm{W}^j_{t_{k+1}} - \mathrm{W}^j_{t_k}\right)^2 \\
&\quad + 2\sum_{k=0}^{n-1}\sum_{l=0}^{k-1} \mathbb{E}\left(\mathrm{W}^i_{t_{k+1}} - \mathrm{W}^i_{t_k}\right)\left(\mathrm{W}^j_{t_{k+1}} - \mathrm{W}^j_{t_k}\right)\left(\mathrm{W}^i_{t_{l+1}} - \mathrm{W}^i_{t_l}\right)\left(\mathrm{W}^j_{t_{l+1}} - \mathrm{W}^j_{t_l}\right) \\
&= \sum_{k=0}^{n-1} \mathbb{E}\left(t_{k+1} - t_k\right)\left(t_{k+1} - t_k\right) \\
&\leq \|\Pi\| \sum_{k=0}^{n-1} \mathbb{E}\left(t_{k+1} - t_k\right) = \|\Pi\|\,\mathrm{T}.
\end{aligned}$$

Thus, $\mathbb{E}\mathrm{C}_\Pi \to 0$ and $\mathbb{V}\mathrm{C}_\Pi \to 0$ as $\|\Pi\| \to 0$, which proves that $[\mathrm{W}^i, \mathrm{W}^j]_{\mathrm{T}} := \lim_{\|\Pi\|\to 0} \mathrm{C}_\Pi = 0$. □

Theorem 4.3.2 can be used to derive the covariation of two Itô processes X^i and X^j.

THEOREM 4.3.3. *Let* $\mathrm{X}^i = (\mathrm{X}^i_t)_{t\geq 0}$ $i = 1, 2, \ldots, n$ *be the Itô processes given by*

$$\mathrm{d}\mathrm{X}^i_t = \Theta^i_t \mathrm{d}t + \sum_{j=1}^{d} \sigma^{ij}_t \mathrm{d}\mathrm{W}^j_t, \qquad i = 1, 2, \ldots, n, \tag{4.16}$$

where $\mathrm{W} = (\mathrm{W}^1_t, \mathrm{W}^2_t, \ldots, \mathrm{W}^d_t)_{t\geq 0}$ *is a* d*-dimensional Brownian motion. Then*

$$\mathrm{d}[\mathrm{X}^i, \mathrm{X}^j]_t = \sum_{k=1}^{d} \sigma^{ik}_t \sigma^{jk}_t \mathrm{d}t.$$

We will not prove Theorem 4.3.3. Rather, we simply remark that it can be obtained *informally* by writing

$$\mathrm{d}[\mathrm{X}^i, \mathrm{X}^j]_t = \mathrm{d}\mathrm{X}^i_t \mathrm{d}\mathrm{X}^j_t, \tag{4.17}$$

inserting expression (4.16) into (4.17) and using the multiplication rules

$$dW_t^i dW_t^j = \delta_{ij} dt, \qquad \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \qquad dW_t^j dt = 0, \qquad dt dt = 0. \tag{4.18}$$

Note that $d[X^i, X^j]_t = 0$ unless X^j and X^j are driven by at least one common one-dimensional Brownian motion.

We can now give a n-dimensional version of Itô's Lemma. We present the formula in differential form, as it is written more compactly in this way.

THEOREM 4.3.4 (ITÔ FORMULA IN TWO DIMENSIONS). *Let $X = (X_t^1, X_t^2, \ldots, X_t^n)_{t \geq 0}$ be an n-dimensional Itô process and suppose $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $f \in C^2(\mathbb{R}^n)$. Then, for any $T \geq 0$ we have*

$$df(X_t) = \sum_{i=1}^n \frac{\partial f(X_t)}{\partial x_i} dX_t^i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(X_t)}{\partial x_i \partial x_j} d[X^i, X^j]_t.$$

The proof of Theorem 4.3.4 is a straightforward extension of Theorem 4.2.6 to the n-dimensional case and will not be presented here.

To obtain an explicit expression for $df(X_t)$ in terms of $dW_t^1, dW_t^2, \ldots, dW_t^d$ and dt we can repeat the same informal procedure we used in the one-dimensional case.

1. Expand $df(X_t) = f(X_t + dX_t) - f(X_t)$ about the point X_t to second order

$$df(X_t) = \sum_{i=1}^n \frac{\partial f(X_t)}{\partial x_i} dX_t^i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(X_t)}{\partial x_i \partial x_j} dX_t^i dX_t^j. \tag{4.19}$$

2. Insert expression for dX_t^i into (4.19) and use the multiplication rules given in (4.18).

EXAMPLE 4.3.5 (PRODUCT RULE). To compute $d(X_t Y_t)$ wherer X and Y are one-dimensional Itô processes, we define $f(x, y) = xy$ and use $f_x = y$, $f_y = x$, $f_{xy} = 1$ and $f_{xx} = f_{yy} = 0$ to compute

$$d(X_t Y_t) = Y_t dX_t + X_t dY_t + d[X, Y]_t.$$

In some cases, it will be useful to see if a process is in fact a Brownian motion.

THEOREM 4.3.6 (LÉVY CHARACTERIZATION OF BROWNIAN MOTION). *Let $M = (M_t)_{t \geq 0}$ be a martingale with respect to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$. Suppose M has continuous sample paths and satisfies $M_0 = 0$ and $[M, M]_t = t$ for all $t \geq 0$. Then M is a Brownian motion.*

PROOF. Let $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be $C^{1,2}$. Note that

$$\begin{aligned} df(t, M_t) &= \partial_t f(t, M_t)dt + \partial_m f(t, M_t)dM_t + \tfrac{1}{2}\partial_m^2 f(t, M_t)d[M, M]_t \\ &= \left(\partial_t + \tfrac{1}{2}\partial_m^2\right) f(t, M_t)dt + \partial_m f(t, M_t)dM_t. \end{aligned}$$

In integral form, we have

$$f(T, M_T) = f(t, M_t) + \int_t^T \left(\partial_s + \tfrac{1}{2}\partial_m^2\right) f(s, M_s)ds + \int_t^T \partial_m f(s, M_s)dM_s.$$

Although we have not proved it, since M is a martingale (by assumption), it follows that any integral of the form $I_t := \int_0^t \Delta_s dM_s$, where $\Delta = (\Delta_t)_{t\geq 0}$ is adapted to $(\mathcal{F}_t)_{t\geq 0}$, is a martingale. Thus, using the short-hand $\mathbb{E}_t \cdot := \mathbb{E}[\cdot\,|\mathcal{F}_t]$, we have

$$\begin{aligned} \mathbb{E}_t f(T, M_T) &= f(t, M_t) + \mathbb{E}_t \int_t^T \left(\partial_s + \tfrac{1}{2}\partial_m^2\right) f(s, M_s)ds + \mathbb{E}\int_t^T \partial_m f(s, M_s)dM_s \\ &= f(t, M_t) + \mathbb{E}_t \int_t^T \left(\partial_s + \tfrac{1}{2}\partial_m^2\right) f(s, M_s)ds. \end{aligned}$$

Now, let a fix a constant u and define

$$f(t, m) = e^{um - \frac{1}{2}u^2 t}.$$

It is easy to verify that

$$\left(\partial_t + \tfrac{1}{2}\partial_m^2\right) f = 0.$$

It follows that

$$\mathbb{E}_t e^{uM_T - \frac{1}{2}u^2 T} = f(t, M_t) = e^{uM_t - \frac{1}{2}u^2 t} \qquad \Rightarrow \qquad \mathbb{E}_t e^{u(M_T - M_t)} = e^{\frac{1}{2}u^2(T-t)}. \tag{4.20}$$

By definition, $\mathbb{E}_t e^{u(M_T - M_t)}$ is the $\mathcal{F}_t$-conditional moment generating function of $(M_T - M_t)$. And $e^{\frac{1}{2}u^2(T-t)}$ is the moment generating function of a $\mathcal{N}(0, T-t)$ random variable. It follows that $M_T - M_t \sim \mathcal{N}(0, T-t)$ for all $0 \leq t \leq T < \infty$, just like a Brownian motion. Furthermore, we see that $M_T - M_t \perp\!\!\!\perp \mathcal{F}_t$, as the right-hand-side of (4.20) does not depend on $\mathcal{F}_t$. As M satisfies all properties of a Brownian motion, it must be a Brownian motion. □

EXAMPLE 4.3.7 (CORRELATED BROWNIAN MOTIONS). Let $B = (B_t)_{t\geq 0}$ be given by

$$B_t = \rho W_t^1 + \bar{\rho} W_t^2, \qquad \bar{\rho} = \sqrt{1-\rho^2}, \qquad \rho = [-1, 1],$$

where $W = (W_t^1, W_t^2)$ is a two-dimensional Brownian motion. We will use Theorem 4.3.6 to show that B is a Brownian motion. It is clear that $B_0 = 0$ and B has sample paths that are continuous. It is also

clear that B is a martingale since W^1 and W^2 are martingales. What remains is to show that $[B,B]_t = t$. We have

$$\begin{aligned}[B,B]_t &= [\rho W^1 + \bar{\rho} W^2, \rho W^1 + \bar{\rho} W^2]_t \\ &= \rho^2 [W^1, W^1]_t + 2\rho\bar{\rho}[W^1, W^2]_t + \bar{\rho}^2 [W^2, W^2]_t \\ &= \rho^2 t + 0 + \bar{\rho}^2 t = t.\end{aligned}$$

Alternatively, one can simply compute $(dB_t)^2 = (\rho dW_t^1 + \bar{\rho}^2 dW_t^2)^2$ and use $dW_t^i dW_t^j = \delta_{ij} dt$ in order to show that $dB_t = dt$. Note that

$$\begin{aligned}[B, W^1]_t &= [\rho W^1 + \bar{\rho} W^2, W^1]_t \\ &= \rho [W^1, W^1]_t + \bar{\rho}[W^2, W^1]_t = \rho t.\end{aligned}$$

where we have used $[W^i, W^j]_t = \delta_{ij} t$.

4.4 Girsanov's Theorem for a single Brownian motion

We briefly recall some results from Section 1.7. Suppose that, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we have a random variable $Z \geq 0$ that has expectation $\mathbb{E}Z = 1$. Then we can define a new probability measure $\widetilde{\mathbb{P}}$ via

$$\widetilde{\mathbb{P}}(A) = \mathbb{E} Z \mathbb{1}_A, \qquad A \in \mathcal{F},$$

and we call $Z = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}$ the Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$. If Z is strictly positive $Z > 0$, then we also have

$$\mathbb{P}(A) = \widetilde{\mathbb{E}} \frac{1}{Z} \mathbb{1}_A, \qquad A \in \mathcal{F},$$

and we call $\frac{1}{Z} = \frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}$ the Radon-Nikodým derivative of $\mathbb{P}$ with respect to $\widetilde{\mathbb{P}}$.

In Example 1.7.6, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we defined $X \sim \mathcal{N}(0,1)$ and a Radon-Nikodým derivative $Z = e^{-\theta X - \frac{1}{2}\theta^2}$. We showed that $Y := X + \theta$ was $\mathcal{N}(\theta, 1)$ under $\mathbb{P}$ and $\mathcal{N}(0,1)$ under $\widetilde{\mathbb{P}}$. Thus, Z had the effect of changing the mean of Y.

We would like to extend this idea from a static to a dynamics setting. Specifically, we would like to find a measure change that modifies the dynamics of a stochastic process $X = (X_t)_{t \geq 0}$.

Definition 4.4.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ be a filtration on this space. A *Radon-Nikodým derivative process* $(Z_t)_{0 \leq t \leq T}$ is any process of the form

$$Z_t := \mathbb{E}[Z | \mathcal{F}_t]$$

where Z is a random variable satisfying $\mathbb{E}Z = 1$ and $Z > 0$.

Note that Z in Definition 4.4.1 satisfies the conditions of a Radon-Nikodým derivative. As such, one can define a measure change $\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}$ from Z.

LEMMA 4.4.2. *A* Radon-Nikodým derivative process $Z = (Z_t)_{0\le t\le T}$ *is a martingale.*

PROOF. For $0 \le s \le t \le T$ we have

$$\mathbb{E}[Z_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}[Z|\mathcal{F}_s] = Z_s.$$

□

LEMMA 4.4.3. *Let* $(Z_t)_{0\le t\le T}$ *be a Radon-Nikodým derivative process and define* $\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = Z$. *Suppose* $Y \in \mathcal{F}_s$, *where* $s \in [0, T]$. *Then*

$$\widetilde{\mathbb{E}}Y = \mathbb{E}Z_sY.$$

PROOF. The proof is a simple exercise in iterated conditioning. We have

$$\widetilde{\mathbb{E}}Y = \mathbb{E}ZY = \mathbb{E}Y\mathbb{E}[Z|\mathcal{F}_s] = \mathbb{E}YZ_s.$$

□

LEMMA 4.4.4. *Let* $(Z_t)_{0\le t\le T}$ *be a Radon-Nikodým derivative process and define* $\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = Z$. *Suppose* $Y \in \mathcal{F}_t$ *where* $0 \le s \le t \le T$. *Then*

$$\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] = \frac{1}{Z_s}\mathbb{E}[Z_tY|\mathcal{F}_s].$$

PROOF. From Definition 2.3.1, we recall that a conditional expectation $\widetilde{\mathbb{E}}[Y|\mathcal{F}_s]$ must satisfy two properties:
(i) $\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] \in \mathcal{F}_s$.
(ii) $\widetilde{\mathbb{E}}\mathbb{1}_A\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] = \widetilde{\mathbb{E}}\mathbb{1}_AY$ for all $A \in \mathcal{F}_s$.
It is clear that property (i) is satisfied since

$$\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] = \frac{1}{Z_s}\mathbb{E}[Z_tY|\mathcal{F}_s] = \frac{1}{\mathbb{E}[Z|\mathcal{F}_s]}\mathbb{E}[Z_tY|\mathcal{F}_s] \in \mathcal{F}_s.$$

We must check property (ii). Let $A \in \mathcal{F}_s$. Then

$$\begin{aligned}\widetilde{\mathbb{E}}\mathbb{1}_A\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] &= \widetilde{\mathbb{E}}\mathbb{1}_A\frac{1}{Z_s}\mathbb{E}[Z_tY|\mathcal{F}_s] \\ &= \mathbb{E}Z_s\mathbb{1}_A\frac{1}{Z_s}\mathbb{E}[Z_tY|\mathcal{F}_s] \qquad \text{(because } \mathbb{1}_A\frac{1}{Z_s}\mathbb{E}[Z_tY|\mathcal{F}_s] \in \mathcal{F}_s \text{ and Lemma 4.4.3)}\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\mathbb{1}_A \mathbb{E}[Z_t Y|\mathcal{F}_s] \\
&= \mathbb{E}\mathbb{E}[\mathbb{1}_A Z_t Y|\mathcal{F}_s] \\
&= \mathbb{E}\mathbb{1}_A Z_t Y \\
&= \widetilde{\mathbb{E}}\mathbb{1}_A Y. \qquad \text{(because } \mathbb{1}_A Y \in \mathcal{F}_t \text{ and Lemma 4.4.3)}
\end{aligned}
$$

□

THEOREM 4.4.5 (GIRSANOV). *Let* $W = (W_t)_{0\le t\le T}$ *be a Brownian motion on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and let* $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ *be a filtration for* W. *Suppose* $\Theta = (\Theta_t)_{0\le t\le T}$ *is adapted to the filtration* $\mathbb{F}$. *Define* $(Z_t)_{0\le t\le T}$ *and* $\widetilde{W} = (\widetilde{W}_t)_{0\le t\le T}$ *by*

$$
Z_t = \exp\Big(-\int_0^t \tfrac{1}{2}\Theta_s^2 ds - \int_0^t \Theta_s dW_s\Big), \qquad d\widetilde{W}_t = \Theta_t dt + dW_t, \qquad \widetilde{W}_0 = 0.
$$

Assume that

$$
\mathbb{E}\int_0^T \Theta_t^2 Z_t^2 dt < \infty.
$$

Define a Radon-Nikodým derivative $Z \equiv \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} := Z_T$. *Then the process* $\widetilde{W}$ *is a Brownian motion under* $\widetilde{\mathbb{P}}$.

PROOF. We will use Lévy's Theorem 4.3.6 to show that $\widetilde{W}$ is a Brownian motion under $\widetilde{\mathbb{P}}$. By definition, $\widetilde{W}_0 = 0$. Also, we see that

$$
d[\widetilde{W}, \widetilde{W}]_t = (d\widetilde{W}_t)^2 = (dW_t + \Theta_t dt)^2 = (dW_t)^2 + 2\Theta_t dW_t dt + \Theta_t^2 (dt)^2 = dt.
$$

Lastly, we must verify that $\widetilde{W}$ is a martingale under $\widetilde{\mathbb{P}}$. To do this, we first show that $(Z_t)_{t\ge 0}$ is a martingale under $\mathbb{P}$. Setting

$$
Z_t = e^{X_t}, \qquad X_t = -\int_0^t \tfrac{1}{2}\Theta_s^2 ds - \int_0^t \Theta_s dW_s,
$$

we find (using the Itô formula) that

$$
\begin{aligned}
dZ_t &= e^{X_t} dX_t + \tfrac{1}{2} e^{X_t} d[X, X]_t \\
&= Z_t(-\tfrac{1}{2}\Theta_t^2 dt - \Theta_t dW_t) + \tfrac{1}{2} Z_t \Theta_t^2 dt = -Z_t \Theta_t dW_t.
\end{aligned}
$$

Since Itô integrals are martingales, it follows that $(Z_t)_{0\le t\le T}$ is a martingale under $\mathbb{P}$. In particular we have $\mathbb{E}Z = \mathbb{E}Z_T = Z_0 = 1$. We also have

$$
Z_t = \mathbb{E}[Z_T|\mathcal{F}_t] = \mathbb{E}[Z|\mathcal{F}_t],
$$

for all $0 \le t \le T$, which shows that $(Z_t)_{0\le t\le T}$ is a Radon-Nikodým derivative process. Next, we show that $(\widetilde{W}_t Z_t)_{0\le t\le T}$ is a martingale under $\mathbb{P}$. We have

$$\begin{aligned} d(\widetilde{W}_t Z_t) &= \widetilde{W}_t dZ_t + Z_t d\widetilde{W}_t + d[\widetilde{W}, Z]_t \\ &= \widetilde{W}_t(-Z_t\Theta_t dW_t) + Z_t(\Theta_t dt + dW_t) + (\Theta_t dt + dW_t)(-Z_t\Theta_t dW_t) \\ &= \widetilde{W}_t(-Z_t\Theta_t dW_t) + Z_t(\Theta_t dt + dW_t) - Z_t\Theta_t dt \\ &= (-\widetilde{W}_t\Theta_t + 1)Z_t dW_t. \end{aligned}$$

Again, since Itô integrals are martingales, it follows that $(\widetilde{W}_t Z_t)_{0\le t\le T}$ is a martingale under $\mathbb{P}$. Finally, we can show that $\widetilde{W}$ is a martingale under $\widetilde{\mathbb{P}}$. Assuming $0 \le s \le t \le T$ we have

$$\begin{aligned} \widetilde{\mathbb{E}}[\widetilde{W}_t|\mathcal{F}_s] &= \frac{1}{Z_s}\mathbb{E}[Z_t\widetilde{W}_t|\mathcal{F}_s] && \text{(by Lemma 4.4.4)} \\ &= \frac{1}{Z_s}Z_s\widetilde{W}_s = \widetilde{W}_s. \end{aligned}$$

Thus, $\widetilde{W}$ is a martingale, and therefore, a Brownian motion under $\widetilde{\mathbb{P}}$. □

We now state (without proof) Girsanov's Theorem for multi-dimensional Brownian motions.

THEOREM 4.4.6 (GIRSANOV). *Let $W = (W_t^1, W_t^2, \dots, W_t^d)_{0\le t\le T}$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ be a filtration for W. Suppose $\Theta = (\Theta_t^1, \Theta_t^2, \dots, \Theta_t^d)_{0\le t\le T}$ is adapted to the filtration $\mathbb{F}$. Define $(Z_t)_{0\le t\le T}$ and $\widetilde{W} = (\widetilde{W}_t^1, \widetilde{W}_t^2, \dots, \widetilde{W}_t^d)_{0\le t\le T}$ by*

$$Z_t = \exp\Big(-\int_0^t \tfrac{1}{2}\langle\Theta_s, \Theta_s\rangle ds - \int_0^t \langle\Theta_s, dW_s\rangle\Big), \qquad d\widetilde{W}_t = \Theta_t dt + dW_t, \qquad \widetilde{W}_0 = 0,$$

where $\langle\cdot,\cdot\rangle$ denotes a d-dimensional Euclidean inner product. Assume that

$$\mathbb{E}\int_0^T \langle\Theta_t, \Theta_t\rangle Z_t^2 dt < \infty.$$

Define a Radon-Nikodým derivative $Z \equiv \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} := Z_T$. Then the process $\widetilde{W}$ is a d-dimensional Brownian motion under $\widetilde{\mathbb{P}}$.

It is interesting to note that the components of $\widetilde{W}$ in Theorem 4.4.6 could be co-dependent under $\mathbb{P}$ (as Θ_j could depend on any of $W^1, W^2, \dots, W^d$). Nevertheless, under $\widetilde{\mathbb{P}}$, the components of $\widetilde{W}$ are independent of each other.

4.5 Fundamental Theorem of Asset Pricing

An arbitrage, roughly speaking, is a trading strategy that enables one to make money without taking any risk. We generally believe that financial markets do not admit arbitrage opportunities. As such, the models we use to describe market dynamics should not admit arbitrage. In this section we will see how Girsanov's theorem can be used in order to see if a model of a financial market admits arbitrage.

To begin, let us define what we mean by *self-financing portfolio*.

Definition 4.5.1. Consider a financial market with assets $(A_t^1, A_t^2, \dots, A_t^n)_{t\geq 0}$. A portfolio is *self-financing* if its value $X = (X_t)_{t\geq 0}$ at all times is given by

$$X_t = \sum_{i=1}^{n} \Delta_t^i A_t^i,$$

where Δ_t^i represents the number of shares of A^i held at time t, and changes to the value of the portfolio are due only to changes in the value the assets

$$dX_t = \sum_{i=1}^{n} \Delta_t^i dA_t^i.$$

Gains and/or losses of a self-financing portfolio are due only to changes in the values of the assets in the portfolio. A portfolio would not be self-financing if, for example, and investor added cash to the portfolio at different times. Now that we understand what a self-financing portfolio is, we can define (rigorously) what we mean by arbitrage.

Definition 4.5.2. An *arbitrage* is any self-financing portfolio whose value $X = (X_t)_{t\geq 0}$ satisfies

1. $X_0 = 0$,
2. $\mathbb{P}(X_T \geq 0) = 1$,
3. $\mathbb{P}(X_T > 0) > 0$,

for some $T > 0$.

From the above definition, we see that an arbitrage is a trading strategy that can be financed with zero initial investment, has no probability of losing money, and has some strictly positive probability of making money. As mentioned above, we believe markets do not admit arbitrage opportunities. As such, it would be very useful to have a theorem that would enable us to determine of a model of the market admits arbitrage. This is the subject of the next theorem.

THEOREM 4.5.3 (FUNDAMENTAL THEOREM OF ASSET PRICING). *Consider a financial market, defined under a probability measure* $\mathbb{P}$. *Let* $\mathrm{N} = (\mathrm{N}_t)_{t\geq 0}$ *be any strictly positive self-financing portfolio. A market is free of arbitrage if and only if there exists a probability measure* $\widetilde{\mathbb{P}}$, *equivalent to* $\mathbb{P}$, *under which* X/N *is martingale for all self-financing portfolios* X.

In the above Theorem, we call the portfolio N the *numéraire*. Typically, we will choose the money-market account M as numéraire. However, for some applications, it will be easier to choose a T-maturity zero coupon bond B^{T} as numéraire. The probability measure $\widetilde{\mathbb{P}}$ in the above Theorem is referred to as a *risk-neutral* or *martingale* measure. Note that the measure $\widetilde{\mathbb{P}}$ depends on the choice of numéraire. That is, If $\widetilde{\mathbb{P}}^1$ is a risk-neutral measure with N^1 as numéraire and $\widetilde{\mathbb{P}}^2$ is a risk-neutral measure with N^2 as numéraire then, in general, $\widetilde{\mathbb{P}}^1$ and $\widetilde{\mathbb{P}}^2$ will be different.

Theorem 4.5.3 is important for two reasons

1. It gives us a simple way to check if a model for the financial market contains an arbitrage. We should *never* price assets using models that admit arbitrage.
2. It provides a way for us to price derivative assets in such a way that there is no arbitrage. Specifically, the value of any asset $\mathrm{A} = (\mathrm{A}_t)_{t\geq 0}$ must satisfy

$$\frac{\mathrm{A}_t}{\mathrm{N}_t} = \widetilde{\mathbb{E}}\Big(\frac{\mathrm{A}_{\mathrm{T}}}{\mathrm{N}_{\mathrm{T}}}\Big|\mathcal{F}_t\Big), \tag{4.21}$$

   where $\widetilde{\mathbb{P}}$ is a martingale measure for numérair $\mathrm{N} = (\mathrm{N}_t)_{t\geq 0}$.

You can think of (4.21) as a pricing equation.

An important but subtle point is that, for a given market and numéraire N, there may be more than one risk-neutral/martingale measure. Suppose, for example, that both $\widetilde{\mathbb{P}}$ and $\widehat{\mathbb{P}}$ are risk-neutral/martingale measures. And let $\mathrm{V} = (\mathrm{V}_t)_{0\leq t\leq \mathrm{T}}$ be the value of a derivative asset. Then we have two possible no-arbitrage prices for V_t, namely

$$\mathrm{V}_t = \mathrm{N}_t\widetilde{\mathbb{E}}\Big(\frac{\mathrm{A}_{\mathrm{T}}}{\mathrm{N}_{\mathrm{T}}}\Big|\mathcal{F}_t\Big), \qquad \text{and} \qquad \mathrm{V}_t = \mathrm{N}_t\widehat{\mathbb{E}}\Big(\frac{\mathrm{A}_{\mathrm{T}}}{\mathrm{N}_{\mathrm{T}}}\Big|\mathcal{F}_t\Big).$$

Obviously, there can only be a single price for a given asset (otherwise there would be a clear arbitrage). As such, when more than one risk-neutral/marringale measure exists, we must assume that the market chooses a unique measure, called the *pricing measure*, under which all assets are priced.

Let us see how we can use Girsanov's theorem in order to verify if the *Black-Scholes* model admits arbitrage.

EXAMPLE 4.5.4 (BLACK-SCHOLES). Consider a market with a stock $S = (S_t)_{0\le t\le T}$ and a money market account, whose dynamics are of the form

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \qquad dM_t = rM_t dt,$$

where W is a Brownian motion under $\mathbb{P}$. Let us see if this model has an arbitrage. First we need to choose a numéraire portfolio. We will take the money market account M as numéraire, as this is typically the easiest choice. Next, we need to see if there exists a probability measure $\widetilde{\mathbb{P}}$, equivalent to $\mathbb{P}$, under which X/M is a martingale for all portfolios X. As there are only two assets in our market S and M, all porfolios must have dynamics of the form

$$\begin{aligned} dX_t &= \Delta_t dS_t + (X_t - \Delta_t S_t)\frac{1}{M_t} dM_t \\ &= \Delta_t S_t(\mu - r)dt + \Delta_t \sigma S_t dW_t + X_t r dt. \end{aligned}$$

The dynamics of X/M are then

$$\begin{aligned} d\frac{X_t}{M_t} &= X_t d\Big(\frac{1}{M_t}\Big) + \frac{1}{M_t} dX_t + d\Big[X, \frac{1}{M}\Big]_t \\ &= -r\frac{X_t}{M_t}dt + \Delta_t(\mu - r)\frac{S_t}{M_t}dt + \Delta_t\sigma\frac{S_t}{M_t}dW_t + \frac{X_t}{M_t}r dt \\ &= \Delta_t(\mu - r)\frac{S_t}{M_t}dt + \Delta_t\sigma\frac{S_t}{M_t}dW_t. \end{aligned}$$

From Girsanov's Theorem, we know that the process $\widetilde{W}$ defined by

$$d\widetilde{W}_t = \gamma_t dt + dW_t$$

is a Brownian motion under $\widetilde{\mathbb{P}}$ where

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \exp\Big(-\int_0^T \gamma_t dW_t - \tfrac{1}{2}\int_0^T \gamma_t^2 dt\Big).$$

The process $\gamma = (\gamma_t)_{0\le t\le T}$ is arbitrary. Let us write the dynamics of X/M in terms of $\widetilde{W}$. We have

$$\begin{aligned} d\frac{X_t}{M_t} &= \Delta_t(\mu - r)\frac{S_t}{M_t}dt + \Delta_t\sigma\frac{S_t}{M_t}(d\widetilde{W}_t - \gamma_t dt) \\ &= \Delta_t(\mu - r - \sigma\gamma_t)\frac{S_t}{M_t}dt + \Delta_t\sigma\frac{S_t}{M_t}d\widetilde{W}_t. \end{aligned}$$

If we choose $\gamma_t := (\mu - r)/\sigma$ then the dt-term will disappear in the dynamics of X/M and we have

$$d\frac{X_t}{M_t} = \Delta_t\sigma\frac{S_t}{M_t}d\widetilde{W}_t.$$

Thus, X/M will be a martingale under $\widetilde{\mathbb{P}}$ for all portfolios X. The market therefore does not have any arbitrage. Moreover, there is only one martingale measure. As such, if we were to consider a derivative asset that pays $\varphi(\mathrm{S_T})$ at time T, then the value $\mathrm{V} = (\mathrm{V}_t)_{0\le t\le \mathrm{T}}$ would have to satisfy

$$\frac{\mathrm{V}_t}{\mathrm{M}_t} = \widetilde{\mathbb{E}}\Big(\frac{\mathrm{V_T}}{\mathrm{M_T}}\Big|\mathcal{F}_t\Big) = \widetilde{\mathbb{E}}\Big(\frac{\varphi(\mathrm{S_T})}{\mathrm{M}_t}\Big|\mathcal{F}_t\Big),$$
$$\mathrm{V}_t = \mathrm{e}^{-r(\mathrm{T}-t)}\widetilde{\mathbb{E}}\Big(\varphi(\mathrm{S_T})\Big|\mathcal{F}_t\Big),$$

where we have used $\mathrm{M}_t = \mathrm{M}_0\mathrm{e}^{rt}$. Note that we can compute the expectation on the last line, as the dynamics of S, under $\widetilde{\mathbb{P}}$ are

$$\mathrm{dS}_t = r\mathrm{S}_t\mathrm{d}t + \sigma\mathrm{S}_t\mathrm{d}\widetilde{\mathrm{W}}_t \qquad \Rightarrow \qquad \mathrm{S_T} = \mathrm{S}_t\exp\Big((r-\sigma^2/2)(\mathrm{T}-t) + \sigma(\widetilde{\mathrm{W}}_\mathrm{T} - \widetilde{\mathrm{W}}_t)\Big),$$

and $\widetilde{\mathrm{W}}_\mathrm{T} - \widetilde{\mathrm{W}}_t \sim \mathcal{N}(0, \mathrm{T}-t)$.

4.6 Exercises

Exercise 4.1. Compute $\mathrm{d}(\mathrm{W}_t^4)$. Write W_T^4 as an integral with respect to W plus an integral with respect to t. Use this representation of W_T^4 to show that $\mathbb{E}\mathrm{W}_\mathrm{T}^4 = 3\mathrm{T}^2$. Compute $\mathbb{E}\mathrm{W}_\mathrm{T}^6$ using the same technique.

Exercise 4.2. Find an explicit expression for $\mathrm{Y_T}$ where

$$\mathrm{dY}_t = r\mathrm{d}t + \alpha\mathrm{Y}_t\mathrm{dW}_t.$$

Hint: compute $\mathrm{d}(\mathrm{Y}_t\mathrm{Z}_t)$ where $\mathrm{Z}_t := \exp(-\alpha\mathrm{W}_t + \frac{1}{2}\alpha^2 t)$.

Exercise 4.3. Suppose X, Δ and Π are given by

$$\mathrm{dX}_t = \sigma\mathrm{X}_t\mathrm{dW}_t, \qquad \Delta_t = \frac{\partial f}{\partial x}(t, \mathrm{X}_t), \qquad \Pi_t = \mathrm{X}_t\Delta_t$$

where f is some smooth function. Show that if f satisfies

$$\left(\frac{\partial}{\partial t} + \frac{1}{2}\sigma^2 x^2\frac{\partial^2}{\partial x^2}\right)f(t,x) = 0,$$

for all (t,x), then Π is a martingale with respect to a filtration $\mathcal{F}_t$ for W.

Exercise 4.4. Suppose X is given by

$$\mathrm{dX}_t = \mu(t, \mathrm{X}_t)\mathrm{d}t + \sigma(t, \mathrm{X}_t)\mathrm{dW}_t.$$

For any smooth function f define

$$M_t^f := f(t, X_t) - f(0, X_0) - \int_0^t \Big(\frac{\partial}{\partial s} + \mu(s, X_s)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(s, X_s)\frac{\partial^2}{\partial x^2} \Big) f(s, X_s) ds.$$

Show that M^f is a martingale with respect to a filtration $\mathcal{F}_t$ for W.

EXERCISE 4.5. Let $X = (X_t)_{0 \le t \le T}$ be an OU process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

$$dX_t = K(\theta - X_t)dt + \sigma dW_t.$$

Where $W = (W_t)_{0 \le t \le T}$ is a *Brownian motion* under probability measure $\mathbb{P}$. Then we can define a new probability measure $\widetilde{\mathbb{P}}$ such that the process $\widetilde{W} = (\widetilde{W}_t)_{0 \le t \le T}$ is a *Brownian motion* under $\widetilde{\mathbb{P}}$. Then the OU process $X = (X_t)_{0 \le t \le T}$ on the new probability space $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$ will be

$$dX_t = K(\theta^* - X_t)dt + \sigma d\widetilde{W}_t.$$

Find the *Radon-Nikodým* derivative $\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}$.

EXERCISE 4.6. Consider the Black-Sholes model described in Example 4.5.4. Suppose that, instead of taking the money market account M as numéraire, we take the stock price as numéraire the stock price S.

(a) Find a change of measure $d\widehat{\mathbb{P}}/d\mathbb{P}$ under which (X/S) is a $(\widehat{\mathbb{P}}, \mathbb{F})$ martingale for all self-financing portfolios X.

(b) What are the dynamics of the money market account (M/S) under $\widehat{\mathbb{P}}$? That is, find Θ_t and Δ_t such that $d(M_t/S_t) = \Theta_t dt + \Delta_t d\widehat{W}_t$ where $\widehat{W}$ is a $(\widehat{\mathbb{P}}, \mathbb{F})$ Brownian motion.

EXERCISE 4.7. Consider a market with stochastic interest rates $R = (R_t)_{0 \le t \le T}$. The *Cox-Ingersoll-Ross* (CIR) model for R is

$$dR_t = \kappa(\theta - R_t)dt + \delta\sqrt{R_t}dW_t,$$

where W is a Brownian motion under the real-world probability measure $\mathbb{P}$. Note that, in this model, $R_t \ge 0$ for all $t \in [0, T]$. Show that, under the following change of measure

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \exp\Big(-\frac{1}{2}\int_0^T \lambda^2 R_t dt - \int_0^T \lambda\sqrt{R_t}dW_t \Big),$$

the dynamics of R have the form

$$dR_t = \widetilde{\kappa}(\widetilde{\theta} - R_t)dt + \delta\sqrt{R_t}d\widetilde{W}_t,$$

and find $\widetilde{\kappa}$ and $\widetilde{\theta}$.

Chapter 5

SDEs and PDEs

The notes from this chapter are taken primarily from (Shreve, 2004, Chapter 6), (Øksendal, 2005, Chapters 5, 7, 8 and 9) and Linetsky (2007). Another good reference is (Karlin and Taylor, 1981, Chapter 15).

5.1 Stochastic differential equations

We will begin this chapter by focusing on scalar stochastic differential equations driven by a single Brownian motion. The extension to d-dimensional stochastic differential equations driven by m Brownian motions will be given in Section 5.3.

Definition 5.1.1. A *stochastic differential equation* (SDE) is an equation of the form

$$\mathrm{d}X_s = \mu(s, X_s)\mathrm{d}s + \sigma(s, X_s)\mathrm{d}W_s, \qquad X_t = x, \tag{5.1}$$

We call functions μ and σ the *drift* and *diffusion*, respectively, and we call $X_t = x$ the *initial condition*. A *(strong) solution* of an SDE is a stochastic process $X = (X_s)_{s \geq t}$ such that

$$X_T = x + \int_t^T \mu(s, X_s)\mathrm{d}s + \int_t^T \sigma(s, X_s)\mathrm{d}W_s, \tag{5.2}$$

for all $T \geq t$.

One way to envision a strong solution of an SDE is as follows: think of a sample path $W_\cdot(\omega) : [t, \infty) \to \mathbb{R}$ as input. From this input, we can construct a unique sample path $X_\cdot(\omega) : [t, \infty) \to \mathbb{R}$.

Ideally, we would like to write X_T as an *explicit* functional of the Brownian path $(W_s)_{s \geq t}$. Unfortunately, this is typically not possible. Still, it will help to build intuition if we see some explicitly solvable examples.

EXAMPLE 5.1.2 (OU PROCESS). An *Ornstein-Uhlenbeck process* (OU process, for short) is an Itô process $X = (X_t)_{t\geq 0}$ that satisfies

$$dX_t = \kappa(\theta - X_t)dt + \sigma dW_t, \tag{5.3}$$

where $W = (W_t)_{t\geq 0}$ is a one-dimensional Brownian motion and $\kappa, \theta > 0$. The OU process is *mean-reverting* in the following sense. If $X_t > \theta$ then $\kappa(\theta - X_t) < 0$ and the deterministic part of (5.3) (i.e., the dt-term) pushes the process down towards θ. If $X_t < \theta$ then $\kappa(\theta - X_t) > 0$ and the deterministic part of (5.3) pushes the process up towards θ. The OU process mean-reverts to the *long-run mean* θ. We often call κ the *rate of mean reversion*, though this nomenclature is somewhat misleading since the instantaneous rate of mean reversion is actually $\kappa(\theta - X_t)$. In quantitative finance, OU processes are very often used to model stochastic interest rates.

We will find an explicit expression for X_t and also compute $\mathbb{E}X_t$ and $\mathbb{V}X_t$. To this end, let us define $Y_t = X_t - \theta$ so that

$$dY_t = -\kappa Y_t dt + \sigma dW_t$$

Note that Y is an OU process that mean-reverts to zero. Next, we define $Z_t = f(t, Y_t) = e^{\kappa t}Y_t$. We can use the two-dimensional Itô formula to compute dZ_t. Using $f_{yy} = 0$ and the heuristic rules $dt dW_t = 0$ and $dt dt = 0$ we have

$$\begin{aligned} dZ_t &= f_t dt + f_y dY_t + \tfrac{1}{2} f_{yy} d[Y, Y]_t \\ &= \kappa e^{\kappa t} Y_t dt + e^{\kappa t} dY_t \\ &= \kappa e^{\kappa t} Y_t dt + e^{\kappa t}\left(-\kappa Y_t dt + \sigma dW_t\right) \\ &= e^{\kappa t}\sigma dW_t. \end{aligned}$$

Thus, we have obtained an expression for Z_t:

$$Z_t = Z_0 + \int_0^t e^{\kappa s}\sigma dW_s.$$

Next, we use $Y_t = e^{-\kappa t}Z_t$ and $X_t = Y_t + \theta$ to obtain

$$\begin{aligned} Y_t &= e^{-\kappa t}Y_0 + \int_0^t e^{-\kappa(t-s)}\sigma dW_s, \\ X_t &= \theta + e^{-\kappa t}(X_0 - \theta) + \int_0^t e^{-\kappa(t-s)}\sigma dW_s. \end{aligned}$$

Note that X_t has a normal distribution at every time $t > 0$ because Itô integrals with deterministic integrands are normally distributed random variables; see Proposition 4.2.9. Thus, the distribution of X_t is completely determined by its mean and variance. We have

$$\mathbb{E}X_t = \theta + e^{-\kappa t}(X_0 - \theta) + \mathbb{E}\int_0^t e^{-\kappa(t-s)}\sigma dW_s$$

$$
\begin{aligned}
&= \theta + \mathrm{e}^{-\kappa t}(\mathrm{X}_0 - \theta), \\
\mathbb{V}\mathrm{X}_t &= \mathbb{V}\int_0^t \mathrm{e}^{-\kappa(t-s)}\sigma \mathrm{d}\mathrm{W}_s \\
&= \int_0^t \left(\mathrm{e}^{-\kappa(t-s)}\sigma\right)^2 \mathrm{d}s = \frac{\sigma^2}{2\kappa}\left(1 - \mathrm{e}^{-2\kappa t}\right).
\end{aligned}
$$

where we have used Proposition 4.2.9.

Example 5.1.3 (Geometric Brownian motion). A *geometric Brownian motion* (GBM) is a process $\mathrm{Z} = (\mathrm{Z})_{t\geq 0}$ that satisfies

$$
\mathrm{d}\mathrm{Z}_t = \mu(t)\mathrm{Z}_t \mathrm{d}t + \sigma(t)\mathrm{Z}_t \mathrm{d}\mathrm{W}_t, \qquad \mathrm{Z}_0 = z,
$$

where μ and σ are *deterministic* functions of t. In quantitative finance, GBMs are very often used to model risky assets like stocks, currencies and financial indexes like the S&P500.

To solve this SDE, we consider $\mathrm{X}_t = \log \mathrm{Z}_t$. Using the Itô formula, we obtain

$$
\begin{aligned}
\mathrm{d}\mathrm{X}_t = \mathrm{d}\log \mathrm{Z}_t &= \frac{1}{\mathrm{Z}_t}\mathrm{d}\mathrm{Z}_t + \frac{1}{2}\left(\frac{-1}{\mathrm{Z}_t^2}\right)\mathrm{d}[\mathrm{Z},\mathrm{Z}]_t \\
&= \left(\mu(t) - \tfrac{1}{2}\sigma^2(t)\right)\mathrm{d}t + \sigma(t)\mathrm{d}\mathrm{W}_t.
\end{aligned}
$$

Integrating from 0 to T, we obtain

$$
\mathrm{X}_\mathrm{T} = x + \int_0^\mathrm{T} \left(\mu(t) - \tfrac{1}{2}\sigma^2(t)\right)\mathrm{d}t + \int_0^\mathrm{T} \sigma(t)\mathrm{d}\mathrm{W}_t, \qquad x = \log z.
$$

Finally, we obtain our expression for Z_T.

$$
\mathrm{Z}_\mathrm{T} = \exp\left(\mathrm{X}_\mathrm{T}\right) = z\exp\left(\int_0^\mathrm{T} \left(\mu(t) - \tfrac{1}{2}\sigma^2(t)\right)\mathrm{d}t + \int_0^\mathrm{T} \sigma(t)\mathrm{d}\mathrm{W}_t\right).
$$

Although we have mostly thrown mathematical rigor out the window in these notes, in an effort to be responsible mathematicians, we should at least state (even if we do not prove) an existence and uniqueness result.

Theorem 5.1.4 (Existence and Uniqueness of SDEs). *Consider the following SDE*

$$
\mathrm{d}\mathrm{X}_t = \mu(t, \mathrm{X}_t)\mathrm{d}t + \sigma(t, \mathrm{X}_t)\mathrm{d}\mathrm{W}_t. \tag{5.4}
$$

Suppose the functions $\mu : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ *and* $\sigma : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ *satisfy*

$$
\begin{aligned}
&\textit{At most linear growth}: && |\mu(t,x)| + |\sigma(t,x)| \leq \mathrm{C}_1(1+|x|), && \forall\, t, x, && (5.5)\\
&\textit{Lipschitz continuity}: && |\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq \mathrm{C}_2|x-y|, && \forall\, t, x, y, && (5.6)
\end{aligned}
$$

for some constants $\mathrm{C}_1, \mathrm{C}_2 < \infty$. *Then SDE* (5.4) *has a unique solution, which is adapted to to the filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ *generated by* $\mathrm{W} = (\mathrm{W}_t)_{t\geq 0}$ *and satisfies* $\mathbb{E}\int_0^\mathrm{T} \mathrm{X}_t^2 \mathrm{d}t$ *for all* $\mathrm{T} < \infty$.

REMARK 5.1.5. Theorem 5.1.4 actually refers to a *strong* solution of an SDE. There is another notion of a solution of an SDE called a *weak solution*. We will not discuss weak solutions here.

We will not prove Theorem 5.1.4. We refer the interested reader, instead, to (Øksendal, 2005, Theorem 5.21). However, we will illustrate with two examples what can go wrong if equations (5.5) and (5.6) are not satisfied.

EXAMPLE 5.1.6. Let X satisfy

$$dX_t = X_t^2 dt, \qquad X_0 = 1. \tag{5.7}$$

We identify $\mu(t,x) = x^2$, which does *not* satisfy the linear growth condition (5.5). The unique solution to (5.7) is

$$X_t = \frac{1}{1-t}, \qquad 0 \le t < 1.$$

Note that X_t blows up as $t \to 1$ and that X_t is not defined for $t \ge 1$. Thus, it is impossible to find a solution that is defined for all $t \ge 0$.

EXAMPLE 5.1.7. Let X satisfy

$$dX_t = 3X_t^{2/3} dt, \qquad X_0 = 0. \tag{5.8}$$

We identify $\mu(t,x) = 3x^{2/3}$, which does *not* satisfy the Lipschitz condition (5.6) at $x = 0$. One can check directly that any $X^{(a)}$ of the form

$$X_t^{(a)} = \begin{cases} 0 & t \le a \\ (t-a)^3 & t > a, \end{cases}$$

satisfies (5.8). Thus, solutions of (5.8) are not unique.

THEOREM 5.1.8 (MARKOV PROPERTY OF SOLUTIONS OF AN SDE). *Let $X = (X_t)_{t \ge 0}$ be the solution of an SDE of the form (5.1). The X is a Markov process. That is, for $t \le T$ and for some suitable function φ, there exists a function g (which depends on t, T and φ) such that*

$$\mathbb{E}[\varphi(X_T)|\mathcal{F}_t] = g(X_t),$$

where $\mathbb{F} = (\mathcal{F}_t)_{t \ge 0}$ is any filtration to which X is adapted.

The proof of Theorem 5.1.8 is somewhat technical and will not be given here. But, the intuitive idea for why the theorem is true is rather simple. From (5.2), we see that the value of X_T depends only on

the path of the Brownian motion over the interval $[t, \mathrm{T}]$ and the initial value $\mathrm{X}_t = x$. The path that X took to arrive at $\mathrm{X}_t = x$ plays no role. In other words, *given the present* $\mathrm{X}_t = x$*, the future* $(\mathrm{X}_\mathrm{T})_{\mathrm{T}>t}$ *is independent of the past* $\mathcal{F}_t$. With this in mind, the process X should admit a transition density

$$\mathbb{P}(\mathrm{X}_\mathrm{T} \in \mathrm{d}y | \mathrm{X}_t = x) = \Gamma(t, x; \mathrm{T}, y)\mathrm{d}y,$$

and thus, the function g should be given by

$$g(\mathrm{X}_t) = \mathbb{E}[\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_t] = \mathbb{E}[\varphi(\mathrm{X}_\mathrm{T})|\mathrm{X}_t] = \int \mathrm{d}y\, \Gamma(t, \mathrm{X}_t; \mathrm{T}, y)\varphi(y).$$

Of course, finding an explicit representation of the transition density Γ may not be possible.

5.2 Connection to partial differential equations

When X is the solution of an SDE, conditional expectations of the path of X can be related to certain PDEs. We discuss these relations below.

Theorem 5.2.1 (Kolmogorov Backward equation). *Let* X *be the solution of SDE* (5.1). *For some suitable function* φ*, define*

$$u(t, \mathrm{X}_t) := \mathbb{E}[\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_t]. \tag{5.9}$$

If the function $u \in \mathrm{C}^{1,2}$*, then it satisfies the* Kolmogorov Backward Equation *(KBE), a linear PDE of the form*

$$(\partial_t + \mathcal{A}(t))u(t, \cdot) = 0, \qquad u(\mathrm{T}, \cdot) = \varphi, \qquad \mathcal{A}(t) = \mu(t, x)\partial_x + \tfrac{1}{2}\sigma^2(t, x)\partial_x^2. \tag{5.10}$$

Proof. First, we note that the process $(u(t, \mathrm{X}_t))_{0 \le t \le \mathrm{T}}$ is a martingale since, for any $0 \le s \le t \le \mathrm{T}$, we have

$$\mathbb{E}[u(t, \mathrm{X}_t)|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}[\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_s] = u(s, \mathrm{X}_s).$$

Next we take the differential of $u(t, \mathrm{X}_t)$ and find using the Itô formula that

$$\begin{aligned}\mathrm{d}u(t, \mathrm{X}_t) &= \partial_t u(t, \mathrm{X}_t)\mathrm{d}t + \partial_x u(t, \mathrm{X}_t)\mathrm{d}\mathrm{X}_t + \tfrac{1}{2}\partial_x^2 u(t, \mathrm{X}_t)\mathrm{d}[\mathrm{X}, \mathrm{X}]_t \\ &= (\partial_t + \mathcal{A}(t))\, u(t, \mathrm{X}_t)\mathrm{d}t + \sigma(t, \mathrm{X}_t)\partial_x u(t, \mathrm{X}_t)\mathrm{d}\mathrm{W}_t.\end{aligned}$$

Integrating, we have

$$u(t, \mathrm{X}_t) = u(s, \mathrm{X}_s) + \int_s^t (\partial_r + \mathcal{A}(r))\, u(r, \mathrm{X}_r)\mathrm{d}r + \int_s^t \sigma(r, \mathrm{X}_r)\partial_x u(r, \mathrm{X}_r)\mathrm{d}\mathrm{W}_r.$$

Taking a conditional expectation and using the fact that Itô integrals are martingales, we find

$$\mathbb{E}[u(t,\mathrm{X}_t)|\mathcal{F}_s] = u(s,\mathrm{X}_s) + \int_s^t \mathbb{E}[(\partial_r + \mathcal{A}(r))\, u(r,\mathrm{X}_r)|\mathcal{F}_s]\mathrm{d}r.$$

Since $(u(t,\mathrm{X}_t))_{0\le t\le \mathrm{T}}$ is a martingale, the integral above must be zero for every $0 \le s \le t \le \mathrm{T}$ and for every possible value of X_s. The only way for this to be true is if the function u satisfies the PDE in (5.10). To see why $u(\mathrm{T},x) = \varphi(x)$ simply use the fact that $\varphi(\mathrm{X}_\mathrm{T}) \in \mathcal{F}_\mathrm{T}$ to write

$$u(\mathrm{T},\mathrm{X}_\mathrm{T}) = \mathbb{E}[\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_\mathrm{T}] = \varphi(\mathrm{X}_\mathrm{T}),$$

which must hold for all possible X_T. □

Theorem 5.2.1 tells us that the function u defined in (5.9) satisfies a PDE (5.10). Alternatively, the *Feynman-Kac formula* says that the solution u of the PDE (5.10) has the stochastic representation $u(t,x) = \mathbb{E}[\varphi(x)|\mathrm{X}_t = x]$.

The methods outlined in the proof above can be applied more generally to find PDE representations for more complicated functionals of the path of X. The basic steps are as follows

1. Find a martingale process.
2. Take the differential of it.
3. Set the $\mathrm{d}t$ term to zero.

We illustrate this method with the following Theorem

THEOREM 5.2.2. *Define*

$$u(t,\mathrm{X}_t) := \mathbb{E}\left[\mathrm{e}^{-\mathrm{A}(t,\mathrm{T})}\varphi(\mathrm{X}_\mathrm{T})\middle|\mathcal{F}_t\right], \qquad \mathrm{A}(t,s) := \int_t^s \mathrm{d}r\, \gamma(r,\mathrm{X}_r).$$

Then u solves the following PDE

$$(\partial_t + \mathcal{A}(t) - \gamma(t,x))u = 0, \qquad u(\mathrm{T},\cdot) = \varphi, \qquad \mathcal{A}(t) = \tfrac{1}{2}\sigma^2(t,x)\partial_x^2 + \mu(t,x)\partial_x. \quad (5.11)$$

PROOF. Note that the function u is *not* a martingale as, for $0 \le s \le t \le \mathrm{T}$ we have

$$\begin{aligned}\mathbb{E}[u(t,\mathrm{X}_t)|\mathcal{F}_s] &= \mathbb{E}[\mathbb{E}[\mathrm{e}^{-\mathrm{A}(t,\mathrm{T})}\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_t]|\mathcal{F}_s] \\ &= \mathbb{E}[\mathrm{e}^{-\mathrm{A}(t,\mathrm{T})}\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_s] \\ &\neq \mathbb{E}[\mathrm{e}^{-\mathrm{A}(s,\mathrm{T})}\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_s] = u(s,\mathrm{X}_s).\end{aligned}$$

However, the process $\mathrm{M} = (\mathrm{M}_t)_{0\le t\le \mathrm{T}}$, defined by

$$\mathrm{M}_t := \mathrm{e}^{-\mathrm{A}(0,t)}u(t,\mathrm{X}_t) = \mathbb{E}[\mathrm{e}^{-\mathrm{A}(0,\mathrm{T})}\varphi(\mathrm{X}_\mathrm{T})|\mathcal{F}_t]$$

is a martingale as, for $0 \le s \le t \le T$ we have

$$\mathbb{E}[M_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[e^{-A(0,T)}\varphi(X_T)|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}[e^{-A(0,T)}\varphi(X_T)|\mathcal{F}_s] = M_s.$$

Now, we take the differential of M. We compute

$$\begin{aligned} dM_t &= u(t,X_t)de^{-A(0,t)} + e^{-A(0,t)}du(t,X_t) \\ &= -e^{-A(0,t)}\gamma(t,X_t)u(t,X_t)dt + e^{-A(0,t)}\Big(\partial_t + \mathcal{A}(t)\Big)u(t,X_t)dt \\ &\quad + e^{-A(0,t)}\sigma(t,X)\partial_x u(t,X_t)dW_t \\ &= e^{-A(0,t)}\Big(\partial_t + \mathcal{A}(t) - \gamma(t,x)\Big)u(t,X_t)dt + e^{-A(0,t)}\sigma(t,X)\partial_x u(t,X_t)dW_t. \end{aligned}$$

Setting the dt term equal to zero, we obtain (5.11). The terminal condition is obtained using

$$u(T,X_T) = \mathbb{E}[e^{-A(T,T)}\varphi(X_T)|\mathcal{F}_T] = \mathbb{E}[\varphi(X_T)|\mathcal{F}_T] = \varphi(X_T),$$

completing the proof. □

5.3 Extensions to higher dimensions

So far, we have dealt only with Scalar SDEs and PDEs with one spacial dimension. To conclude this chapter, we simply restate some of the main results in higher dimensions.

Throughout this section, we consider d-dimensional diffusion process $X = (X_t)_{t\ge 0}$ that satisfies the following SDE

$$dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dW_t, \tag{5.12}$$

where W is an m-dimensional Brownian motion and

$$\mu : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d, \qquad \sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+^{d\times m}.$$

SDE (5.12) has a unique strong solution, which is square integrable for all t (i.e., $\mathbb{E}X_t^2 < \infty$) and adapted to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\ge 0}$ generated by W (i.e., $X_t \in \mathcal{F}_t$) if the following are satisfied

$$\begin{aligned} &\text{Linear Growth :} && |\mu(t,x)| + |\sigma(t,x)| \le C_1(1+|x|), \\ &\text{Lipschitz Continuous :} && |\mu(t,x)-\mu(t,y)| + |\sigma(t,x)-\sigma(t,y)| \le C_2|x-y|, \end{aligned}$$

for some constants C_1 and C_2, independent of t, x, y, where

$$|\sigma|^2 := \sum_{i=1}^{d}\sum_{j=1}^{m}\sigma_{ij}^2, \qquad |\mu|^2 := \sum_{i=1}^{d}\mu_i^2, \qquad |x|^2 := \sum_{i=1}^{d}x_i^2.$$

For any continuous-time Markov process X that lives in $\mathbb{R}^d$, we can define an *infinitesimal generator* $\mathcal{A}$ as follows

$$\mathcal{A}(t)\varphi(\mathrm{X}_t) := \lim_{\mathrm{T}\searrow t} \frac{1}{\mathrm{T}-t}\Big(\mathbb{E}(\varphi(\mathrm{X}_\mathrm{T})|\mathrm{X}_t = x) - \varphi(x)\Big),$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}$ may be any function for which the limit exists. One can show that, when X given by (5.12), the generator $\mathcal{A}$ is given by

$$\mathcal{A}(t) = \sum_{i=1}^{d} \mu_i(t,x)\partial_{x_i} + \tfrac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}(\sigma\sigma^{\mathrm{T}})_{i,j}(t,x)\partial_{x_i}\partial_{x_j}.$$

and we can write the Itô formula compactly as

$$\mathrm{d}\varphi(\mathrm{X}_t) = \mathcal{A}(t)\varphi(\mathrm{X}_t)\mathrm{d}t + (\nabla f)(\mathrm{X}_t)\sigma(t,\mathrm{X}_t)\mathrm{dW}_t.$$

Thus, we can think of $\mathcal{A}$ as the operator that appears in the $\mathrm{d}t$-term, when we take the differential $\mathrm{d}\varphi(\mathrm{X}_t)$. If we define a function $u : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ by

$$u(t,x) := \mathbb{E}\Big[\mathrm{e}^{-\mathrm{A}(t,\mathrm{T})}\varphi(\mathrm{X}_\mathrm{T})\Big|\mathrm{X}_t = x\Big],$$

where $\mathrm{A}(t,\mathrm{T})$ is given by

$$\mathrm{A}(t,\mathrm{T}) = \int_t^{\mathrm{T}} \gamma(s,\mathrm{X}_s)\mathrm{d}s, \qquad\qquad \gamma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+.$$

Then u satisfies

$$\Big(\partial_t - \gamma(t,\cdot) + \mathcal{A}(t)\Big)u(t,\cdot) = 0, \qquad\qquad u(\mathrm{T},\cdot) = \varphi.$$

5.4 Exercises

Exercise 5.1. Consider the Black-Scholes model described in Example 4.5.4. Assume for simplicity that the risk-free rate of interest is zero: $r = 0$. Consider a *power option*, which pays $\varphi(\mathrm{S}_\mathrm{T}) := p\mathrm{S}_\mathrm{T}^q$ at time T. The value V_t of this option at time $t \le \mathrm{T}$ is given by

$$\mathrm{V}_t = \widetilde{\mathbb{E}}(p\mathrm{S}_\mathrm{T}^q|\mathcal{F}_t) =: u(t,\mathrm{S}_t).$$

(a) Find the function $u(t,\mathrm{S}_t)$ by solving the SDE for S and computing the expectation above directly. *Hint*: Once you solve the SDE for S use that fact that $\widetilde{\mathrm{W}}_\mathrm{T} - \widetilde{\mathrm{W}}_t \sim \mathcal{N}(0,\mathrm{T}-t)$ under $\widetilde{\mathbb{P}}$. It may be helpful to use

$$\widetilde{\mathbb{E}}\mathrm{e}^{p\mathrm{Z}} = \mathrm{e}^{\mu p + \frac{1}{2}\sigma^2 p^2}, \qquad\qquad \mathrm{Z} \sim \mathcal{N}(\mu,\sigma^2).$$

(b) Derive a PDE and BC for the function u. Show that the expression you obtained for u in part (a) satisfies this PDE.

EXERCISE 5.2. Consider the following SDE with linear coefficients

$$\mathrm{d}X_t = (b(t) + \mu(t)X_t)\,\mathrm{d}t + (a(t) + \sigma(t)X_t)\,\mathrm{d}W_t, \qquad X_0 = x. \tag{5.13}$$

Define two processes $Z = (Z_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ by

$$\begin{aligned} \mathrm{d}Z_t &= \mu(t)Z_t\mathrm{d}t + \sigma(t)Z_t\mathrm{d}W_t, & Z_0 &= 1, \\ \mathrm{d}Y_t &= \frac{b(t)-\sigma(t)a(t)}{Z_t}\mathrm{d}t + \frac{a(t)}{Z_t}\mathrm{d}W_t, & Y_0 &= x. \end{aligned}$$

Show that $X_t := Y_tZ_t$ solves (5.13).

EXERCISE 5.3. Let W be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{F} = (\mathcal{F}_t)_{0\leq t\leq T}$ be a filtration for W. Suppose X be the solution of the following SDE

$$\mathrm{d}X_t = \kappa(\theta - X_t)\mathrm{d}t + \delta\sqrt{X_t}\mathrm{d}W_t.$$

The process X is called a *Cox-Ingersoll-Ross* (CIR) process; like the OU process, the CIR process is very often used in quantitative finance to model stochastic interest rates. Supposing that interest rates are given by X, the dynamics of a money market account are given by $\mathrm{d}M_t = X_tM_t\mathrm{d}t$. Let $B^T = (B_t^T)_{0\leq t\leq T}$ be the price of a zero-coupon bond which pays one unit of currency at time T (i.e., $B_T^T = 1$). If we interpret $\mathbb{P}$ as a the market's pricing measure with M as numéraire, then using the fundamental theorem of asset pricing, the the value of a bond must satisfy

$$\frac{B_t^T}{M_t} = \mathbb{E}\Big(\frac{B_T^T}{M_T}|\mathcal{F}_t\Big) \qquad \Rightarrow \qquad B_t^T = M_t\mathbb{E}\Big(\frac{B_T^T}{M_T}|\mathcal{F}_t\Big) = \mathbb{E}\left[\exp\left(-\int_t^T X_s\mathrm{d}s\right)\Big|X_t\right],$$

where we have used $B_T^T = 1$ and $M_t = M_0\exp(\int_0^t X_s\mathrm{d}s)$. Thus, given $X_t = x$, the price of a zero coupon bond is given by

$$u(t,x) := \mathbb{E}\left[\exp\left(-\int_t^T X_s\mathrm{d}s\right)\Big|X_t = x\right].$$

(a) Derive a PDE for the function u.
(b) To solve the PDE for u try a solution of the form

$$u(t,x) = \exp(-xA(t) - B(t)),$$

where A and B are deterministic functions of t. Show that A and B must satisfy a pair of coupled ODEs. Write these ODEs as well as their boundary conditions. You do *not* need to solve these ODEs. *Hint*: insert the expression into the PDE for u and collect terms of like powers in x. As the PDE for u must hold for all values of x, what does this tell you about the terms that are of order x^0, x^1, etc.?

EXERCISE 5.4. Let W and B be independent Brownians motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ be a filtration for (W, B). Consider a financial market in which the price of a stock $S = (S_t)_{0\le t\le T}$ is given by $S_t = e^{X_t}$ where X is given by

$$\begin{aligned} dX_t &= -\frac{1}{2}Z_t dt + \sqrt{Z_t}dW_t, \\ dZ_t &= \kappa(\theta - Z_t)dt + \sigma\sqrt{Z_t}\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right). \end{aligned}$$

In quantitative finance, the process (X, Z) is known as the *Heston model*. If we suppose that $\mathbb{P}$ is market's pricing measure with the money market account as numéraire and we suppose that interest rates are zero, then using the fundamental theorem of asset pricing, the value $V = (V_t)_{0\le t\le T}$ of a derivative that pays $\varphi(X_T)$ at time T satisfies

$$V_t = \mathbb{E}(V_T|\mathcal{F}_t) = \mathbb{E}(\varphi(X_T)|X_t, Z_t),$$

Thus, given $(X_t = x, Z_t = z)$ the value of the derivative is given by

$$u(t, x, z) := \mathbb{E}[\varphi(X_T)|X_t = x, Z_t = z].$$

(a) Derive a PDE for the function u.
(b) Let $\widehat{u}$ be the Fourier transform in u in the x variable

$$\widehat{u}(t, \xi, z) = \int_{\mathbb{R}} dx\, e^{-i\xi x}u(t, x, z).$$

Show that $\widehat{u}$ satisfies a PDE in (t, z) with a terminal condition $\widehat{u}(T, \xi, z) = \widehat{\varphi}(\xi)$ where $\widehat{\varphi}$ is the Fourier transform of φ. *Hint*: using integration by parts, we have

$$\int_{\mathbb{R}} dx\, e^{-i\xi x}\partial_x u(t, x, z) = -\int_{\mathbb{R}} dx\, u(t, x, z)\partial_x e^{-i\xi x} = i\xi\int_{\mathbb{R}} dx\, u(t, x, z)e^{-i\xi x} = i\xi\widehat{u}(t, \xi, z).$$

(c) To solve the PDE from part (b) assume that $\widehat{u}$ is of the form

$$\widehat{u}(t, \xi, z) = \widehat{\varphi}(\xi)e^{A(t,\xi)+zB(t,\xi)}.$$

where A and B are deterministic functions of t and ξ. Show that A and B satisfy a pair of coupled ODEs in t; write these ODEs as well as their boundary conditions (it may be a good idea to look at the *hint* given in Exercise 5.3). You do *not* need to solve these ODEs. Note, if one solves the ODEs for A and B, then the function u can be obtained by taking the inverse Fourier transform of $\widehat{u}$, that is

$$u(t, x, z) = \frac{1}{2\pi}\int_{\mathbb{R}} d\xi\, e^{i\xi x}\widehat{u}(t, \xi, z).$$

Typically, this integral must be computed numerically.

Chapter 6

Stochastic Control

Notes from this chapter are taken primarily from (Björk, 2004, Chapter 19) and (van Handel, 2007, Chapter 6).

6.1 Problem formulation

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and consider the following *controlled diffusion* $\mathrm{X} = (\mathrm{X}_t)_{t\geq 0}$, whose dynamics are given by

$$\mathrm{dX}_t^u = \mu(t, \mathrm{X}_t^u, u_t)\mathrm{d}t + \sigma(t, \mathrm{X}_t^u, u_t)\mathrm{dW}_t, \qquad \mathrm{X}_0^u = x \in \mathbb{R}^d, \tag{6.1}$$

where W is an n-dimensional Brownian motion with respect to $(\mathbb{P}, \mathbb{F})$ and

$$\mu : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}^d, \qquad \sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}^{d\times n}$$

Here, X^u is called the *state process* and $u = (u_t)_{t\geq 0}$ is the *control*. The control must live in some *control set* $\mathbb{U}$. Frequently, we take the control set $\mathbb{U} = \mathbb{R}^k$, but this is not required. The superscript u on X^u indicates the dependence of state process X^u on the control u (clearly, if you change u you change X^u).

Definition 6.1.1. We call a control u is *admissible* if the following holds:

1. the control is $\mathbb{F}$-adapted, that is, $u_t \in \mathcal{F}_t$ for all $t \geq 0$,
2. the control remains in the control set: $u_t(\omega) \in \mathbb{U}$ for all (t, ω),
3. the process X^u has a unique (strong) solution.

We denote by $\mathcal{U}$ the set of admissible controls.

Within the class of admissible controls is a subset of controls that we will find very useful:

DEFINITION 6.1.2. We call an admissible control u a *feedback* or *Markov* control if it is of the form: $u_t = \alpha(t, \mathrm{X}_t)$ for some function $\alpha : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{U}$.

Note Markov controls are a strict subset of admissible controls, since a Markov control at time t is only allowed to depend on (t, X_t) whereas, in general, an admissible control at time t may depend on the entire history $\mathcal{F}_t$. Nevertheless, Markov controls are important because they are easy to implement, and (as we shall see) they can actually be computed! Moreover, it often turns out that the optimal Markov control coincides with the optimal admissible control.

If we consider a Markov control $u_t = \alpha(t, \mathrm{X}_t)$ then X^u satisfies an SDE

$$\mathrm{d}\mathrm{X}^u_t = \mu(t, \mathrm{X}^u_t, \alpha(t, \mathrm{X}^u_t))\mathrm{d}t + \sigma(t, \mathrm{X}^u_t, \alpha(t, \mathrm{X}^u_t))\mathrm{d}\mathrm{W}_t, \qquad \mathrm{X}^u_0 = x \in \mathbb{R}^d.$$

Since the solution of an SDE is a Markov process, a Markov control yields a Markov state process X^u.

Below, we introduce some *cost* or *gain functionals* $\mathrm{J}(\cdot)$ that assign to each admissible control strategy u a cost or gain $\mathrm{J}(u)$. The goal of *optimal control* is to find an *optimal strategy* u^* (obviuosly!), which minimizes or maximizes this cost functional. Three common types of cost functionals are:

1. *Finite time.* Here, the cost functional is given by
$$\mathrm{J}(u) = \mathbb{E}\Big[\int_0^{\mathrm{T}} \mathrm{F}(t, \mathrm{X}^u_t, u_t)\mathrm{d}t + \varphi(\mathrm{X}^u_{\mathrm{T}})\Big], \tag{6.2}$$
where $\mathrm{F} : [0, \mathrm{T}] \times \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}$ and $\varphi : \mathbb{R}^d \to \mathbb{R}$.
2. *Indefinite time.* Here, the cost functional is given by
$$\mathrm{J}(u) = \mathbb{E}\Big[\int_0^{\tau^u} \mathrm{F}(\mathrm{X}^u_t, u_t)\mathrm{d}t + \varphi(\mathrm{X}^u_{\tau^u})\Big], \qquad \tau^u = \inf\{t \geq 0 : \mathrm{X}^u_t \notin \mathrm{A}\},$$
where $\mathrm{A} \subset \mathbb{R}^d$ is some open set, $\mathrm{F} : \mathrm{A} \times \mathbb{U} \to \mathbb{R}$ and $\varphi : \partial\mathrm{A} \to \mathbb{R}$.
3. *Infinite time.* In this case, the cost functional is given by
$$\mathrm{J}(u) = \mathbb{E}\int_0^{\infty} \mathrm{e}^{-\lambda t}\mathrm{F}(\mathrm{X}^u_t, u_t)\mathrm{d}t,$$
where $\mathrm{F} : \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}$ and $\lambda > 0$.

Note, for the Indefinite time and Infinite time functionls, the dynamics of X^u should be time-homogeneous (i.e., the coefficient functions μ and σ should not depend on t).

Of course, one might construct other cost functionals, but the above three are by far the most common. As stated above, the optimal control (if it exists) is the strategy u^* that minimizes or maximizes a given cost or gain functional $\mathrm{J}(u)$. Thus, we make the following definition:

DEFINITION 6.1.3. For a given cost functional $\mathrm{J} : \mathcal{U} \to \mathbb{R}$, we say define the *optimal control* u^*, if it exists, by

$$u^* := \underset{u\in\mathcal{U}}{\operatorname{argmax}}\, \mathrm{J}(u), \qquad \text{if our goal is to maximize } \mathrm{J}(u),$$
$$u^* := \underset{u\in\mathcal{U}}{\operatorname{argmin}}\, \mathrm{J}(u), \qquad \text{if our goal is to minimize } \mathrm{J}(u).$$

6.2 The Dynamic Programming Principle and the HJB PDE

In this Section we focus on the finite time horizon control problem and we shall attempt to find the control u^* that *maximizes* $\mathrm{J}(u)$ in (6.2). The minimization procedure is completely analogous. We shall restrict our controls to Markov controls: $u_t = \alpha(t, \mathrm{X}^u_t)$. We shall make the following assumption.

ASSUMPTION 6.2.1. We assume that there exists an admissible Markov strategy u^*, which is optimal.

Clearly, this assumption is *not* always justified. Nevertheless, we will go with it for the time being. We define the *reward-to-go* function J^u (or *cost-to-go* in the case of minimization) as

$$\mathrm{J}^u(t, \mathrm{X}^u_t) := \mathbb{E}\Big[\int_t^{\mathrm{T}} \mathrm{F}(s, \mathrm{X}^u_s, u_s)\mathrm{d}s + \varphi(\mathrm{X}^u_{\mathrm{T}})\Big|\mathcal{F}_t\Big] = \mathbb{E}\Big[\int_t^{\mathrm{T}} \mathrm{F}(s, \mathrm{X}^u_s, u_s)\mathrm{d}s + \varphi(\mathrm{X}^u_{\mathrm{T}})\Big|\mathrm{X}^u_t\Big],$$

where we have used the Markov property of X^u to replace $\mathcal{F}_t$ with the σ-algebra generated by X^u_t. We also define the *value function* V as

$$\mathrm{V}(t, x) := \mathrm{J}^{u^*}(t, x) = \max_{u\in\mathcal{U}} \mathrm{J}^u(t, x). \tag{6.3}$$

The idea of the *Dynamic Programming Principle* (DPP) is to split the optimization problem into two intervals $[t, t+\delta)$ and $[t+\delta, \mathrm{T}]$, where δ is small and positive. Note that, for any $\delta \geq 0$ we have

$$\begin{aligned}\mathrm{J}^u(t, x) &= \mathbb{E}\Big[\int_t^{t+\delta} \mathrm{F}(s, \mathrm{X}^u_s, u_s)\mathrm{d}s + \mathbb{E}\Big[\int_{t+\delta}^{\mathrm{T}} \mathrm{F}(s, \mathrm{X}^u_s, u_s)\mathrm{d}s + \varphi(\mathrm{X}^u_{\mathrm{T}})\Big|\mathcal{F}_{t+\delta}\Big]\Big|\mathrm{X}^u_t = x\Big] \\ &= \mathbb{E}\Big[\int_t^{t+\delta} \mathrm{F}(s, \mathrm{X}^u_s, u_s)\mathrm{d}s + \mathrm{J}^u(t+\delta, \mathrm{X}^u_{t+\delta})\Big|\mathrm{X}^u_t = x\Big]\end{aligned}$$

Now, fix an arbitrary control $\widehat{u}$ and set

$$u_s = \begin{cases} \widehat{u}_s & s \in [t, t+\delta), \\ u^*_s & s \in [t+\delta, \mathrm{T}]. \end{cases}$$

In words, the control u may be sub-optimal over the interval $[t, t+\delta)$ and it is optimal over the interval $[t+\delta, \mathrm{T}]$. Clearly, we have

$$\mathrm{V}(t, x) \geq \mathbb{E}\Big[\int_t^{t+\delta} \mathrm{F}(s, \mathrm{X}^u_s, u_s)\mathrm{d}s + \mathrm{V}(t+\delta, \mathrm{X}^u_{t+\delta})\Big|\mathrm{X}^u_t = x\Big]. \tag{6.4}$$

The inequality arises from the fact that the strategy u is not necessarily optimal over in interval $[t, t+\delta)$. If we had $\widehat{u} = u^*$, then we would have obtained an equality since, in this case, we have $u = u^*$. Now observe that

$$\mathrm{V}(t+\delta, \mathrm{X}^u_{t+\delta}) = \mathrm{V}(t, \mathrm{X}^u_t) + \int_t^{t+\delta} (\partial_s + \mathcal{A}^u(s))\mathrm{V}(s, \mathrm{X}^u_s)\mathrm{d}s + \text{martingale part},$$

where $\mathcal{A}^u(s)$ is the infinitesimal generator of X^u. Taking an expectation, we have

$$\mathbb{E}[\mathrm{V}(t+\delta, \mathrm{X}^u_{t+\delta})|\mathrm{X}^u_t = x] = \mathrm{V}(t,x) + \mathbb{E}\Big[\int_t^{t+\delta} (\partial_s + \mathcal{A}^u(s))\mathrm{V}(s, \mathrm{X}^u_s)\mathrm{d}s \Big| \mathrm{X}^u_t = x\Big]. \tag{6.5}$$

Inserting (6.5) into (6.4) and obtain

$$0 \geq \mathbb{E}\Big[\int_t^{t+\delta} \Big(\mathrm{F}(s, \mathrm{X}^u_s, u_s) + (\partial_s + \mathcal{A}^u(s))\mathrm{V}(s, \mathrm{X}^u_s)\Big)\mathrm{d}s \Big| \mathrm{X}^u_t = x\Big].$$

Finally, we devide by δ and take a limit as $\delta \to 0$. Assuming there is no problem with passing a limit through an expectation, we have

$$\begin{aligned}
0 &\geq \lim_{\delta\to 0} \frac{1}{\delta}\mathbb{E}\Big[\int_t^{t+\delta} \Big(\mathrm{F}(s, \mathrm{X}^u_s, u_s) + (\partial_s + \mathcal{A}^u(s))\mathrm{V}(s, \mathrm{X}_s)\Big)\mathrm{d}s \Big| \mathrm{X}^u_t = x\Big] \\
&= \mathbb{E}\Big[\lim_{\delta\to 0} \frac{1}{\delta}\int_t^{t+\delta} \Big(\mathrm{F}(s, \mathrm{X}^u_s, u_s) + (\partial_s + \mathcal{A}^u(s))\mathrm{V}(s, \mathrm{X}_s)\Big)\mathrm{d}s \Big| \mathrm{X}^u_t = x\Big] \\
&= \mathrm{F}(t,x,u) + (\partial_t + \mathcal{A}^u(t))\mathrm{V}(t,x).
\end{aligned}$$

Once again, if u is optimal, we obtain an equality. Thus, we have arrived at the *Hamilton-Jacobi-Bellman (HJB) PDE*

$$0 = \partial_t \mathrm{V} + \max_{u\in\mathbb{U}} \left(\mathrm{F}^u(t) + \mathcal{A}^u(t)\mathrm{V}\right), \qquad \mathrm{V}(\mathrm{T}, x) = \varphi(x), \tag{6.6}$$

where we have defined

$$\mathrm{F}^u(t) = \mathrm{F}(t,x,u), \qquad \mathcal{A}^u(t) = \sum_{i=1}^d \mu_i(t,x,u)\partial_{x_i} + \tfrac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \left(\sigma\sigma^{\mathrm{T}}\right)_{i,j}(t,x,u)\partial_{x_i}\partial_{x_j}.$$

We summarize this result with the following Theorem

THEOREM 6.2.2. *If (i) an optimal control u^* exists and is Markov, (ii) the value function V, defined in (6.3), satisfies $\mathrm{V} \in \mathrm{C}^{1,2}$ and (iii) the limiting procedures performed above are allowed, then V satisfies the HJB PDE (6.6) and the optimal control u^* is given by*

$$u^*_t = \alpha(t, \mathrm{X}^{u^*}_t), \qquad \textit{where} \qquad \alpha(t,x) := \operatorname*{argmax}_{u\in\mathbb{U}} \left(\mathrm{F}(t,x,u) + \mathcal{A}^u\mathrm{V}(t,x)\right). \tag{6.7}$$

Theorem 6.2.2 is so incredibly unsatisfactory! It only tells us that, *if* an optimal controls exists and is Markov, then (modulo some technical conditions) the value function V satisfies the HJB PDE and the optimal control is given by (6.7). But, this is not what we want. What we would *like*, is solve the HJB PDE (6.6) and *conclude* that the solution actually *is* the value function and u^*, given by (6.7), is the optimal control. The following Theorem gives us such a result.

THEOREM 6.2.3 (VERIFICATION THEOREM). *Suppose* $\mathrm{H} : [0,\mathrm{T}] \times \mathbb{R}^d \to \mathbb{R}$ *solves the HJB PDE* (6.6). *Define*

$$g(t,x) = \operatorname*{argmax}_{u\in\mathbb{U}} \left(\mathrm{F}(t,x,u) + \mathcal{A}^u \mathrm{H}(t,x)\right).$$

Suppose that $\mathrm{M}_t := \mathrm{H}(t,\mathrm{X}_t^u) - \int_0^t (\partial_s + \mathcal{A}^u(s))\mathrm{H}(s,\mathrm{X}_s^u)\mathrm{d}s$ *is a true martingale (not just a local martingale), where* $u_t = g(t,\mathrm{X}_t^u)$. *Then the value function* V *and optimal control* u^* *are given by*

$$\mathrm{V}(t,x) = \mathrm{H}(t,x), \qquad\qquad u_t^* = u_t.$$

The Verification Theorem tells us that, if we solve the HJB PDE, and the solution satisfies some regularity and integrability conditions, then the the solution *is* the value function and the Markov control *is* the optimal control. We will not prove the verification theorem.

6.3 SOLVING THE HJB PDE

Solving the HJB PDE essentially involves two steps

Step 1 (easy part) :	find	$u^*(t,x) = \operatorname*{argmax}_{u\in\mathbb{U}}(\mathrm{F}(t,x,u) + \mathcal{A}^u\mathrm{V}(t,x))$.	
Step 2 (hard part) :	solve	$0 = (\partial_t + \mathcal{A}^{u^*}(t))\mathrm{V} + \mathrm{F}^{u^*}$,	$\mathrm{V}(\mathrm{T},\cdot) = \varphi$.

We illustrate this process with a simple, explicitly-solvable example.

EXAMPLE 6.3.1 (LINEAR QUADARTIC REGULATOR). Suppose the dynamics of a controlled process $\mathrm{X}^u = (\mathrm{X}_t^u)_{0\le t\le \mathrm{T}}$ are given by

$$\mathrm{d}\mathrm{X}_t^u = (a\mathrm{X}_t^u + bu_t)\mathrm{d}t + c\mathrm{d}\mathrm{W}_t,$$

and consider the following cost functional, which we wish to minimize

$$\mathrm{J}(u) = \mathbb{E}\Big(\int_0^{\mathrm{T}} (q(\mathrm{X}_t^u)^2 + ru_t^2)\mathrm{d}t + h(\mathrm{X}_{\mathrm{T}}^u)^2\Big). \tag{6.8}$$

One could imagine, for example, that X^u represents the position of a particle that we are attempting to keep near the origin. We can adjust the drift of the particle through the control u, but this has a cost, which is quadratic in u. Likewise, there is a quadratic cost for allowing X^u to be away from the origin. This control problem is known as the *Linear Quadratic Regulator* (LQR) because the dynamics of X are linear in the state process X^u and control process u and the costs are quadratic in X^u and u. The HJB equation associated with cost functional (6.8) is

$$0 = V_t + \sup_{u\in\mathbb{R}}(\mathcal{A}^u V + F^u), \qquad\qquad V(t,x) = hx^2,$$

where the operator $\mathcal{A}^u$ and the function F^u are given by

$$\mathcal{A}^u = (ax+bu)\partial_x + \tfrac{1}{2}c^2\partial_x^2, \qquad\qquad F^u = qx^2 + ru^2.$$

Step one in solving the HJB PDE is finding the optimal control u^* in feedback form. To this end, we have

$$0 = \partial_u\Big[\mathcal{A}^u V + F^u\Big|_{u=u^*} = bV_x + 2ru^*, \qquad\Rightarrow\qquad u^* = \frac{-bV_x}{2r}.$$

The HJB PDE thus becomes

$$\begin{aligned} 0 &= V_t + \mathcal{A}^{u^*}V + F^{u^*} \\ &= V_t + \Big[ax + b\Big(\frac{-bV_x}{2r}\Big)\Big]V_x + \tfrac{1}{2}c^2V_{xx} + qx^2 + r\Big(\frac{-bV_x}{2r}\Big)^2 \\ &= V_t + axV_x + \tfrac{1}{2}c^2V_{xx} + qx^2 - \frac{b^2V_x^2}{4r}. \end{aligned} \tag{6.9}$$

As with all non-linear PDEs with explicit solutions, the method to solve (6.9) is to guess. In this case, the correct guess is

$$V(t,x) = P(t)x^2 + Q(t), \qquad\qquad P(T) = h, \qquad\qquad Q(T) = 0,$$

where the terminal conditions for P and Q will ensure that $V(T,x) = hx^2$. We have

$$V_t = P'x^2 + Q', \qquad\qquad V_x = 2xP, \qquad\qquad V_{xx} = 2P. \tag{6.10}$$

Inserting (6.10) into (6.9) we obtain

$$\begin{aligned} 0 &= (P'x^2 + Q') + ax(2xP) + \tfrac{1}{2}c^2(2P) + qx^2 - \frac{b^2}{4r}(2xP)^2 \\ &= x^2\Big(P' + 2aP + q - (b^2/r)P^2\Big) + \Big(Q' + c^2P\Big). \end{aligned}$$

Collecting terms of like order in x we obtain

$$\begin{aligned} \mathcal{O}(x^0): &\qquad 0 = \mathrm{Q}' + c^2\mathrm{P}, &&\qquad \mathrm{Q}(\mathrm{T}) = 0, \\ \mathcal{O}(x^2): &\qquad 0 = \mathrm{P}' + 2a\mathrm{P} + q - (b^2/r)\mathrm{P}^2, &&\qquad \mathrm{P}(\mathrm{T}) = h. \end{aligned}$$

Thus, we have obtained a coupled system of ODEs for (P, Q). The $\mathcal{O}(x^2)$ equation is a *Riccati equation*, which can be solved analytically (though, the solution is a tad messy and not worth writing down here). Once one has obtained an expression for the function P one can obtain Q as an integral

$$\mathrm{Q}(t) = c^2 \int_t^{\mathrm{T}} \mathrm{P}(s)\mathrm{d}s.$$

Finally, the optimal control is given by

$$u_t^* = u^*(t, \mathrm{X}_t^{u^*}) = \frac{-b}{2r}\mathrm{V}_x(t, \mathrm{X}_t^{u^*}) = \frac{-b}{r}\mathrm{P}(t)\mathrm{X}_t^{u^*}.$$

6.4 HJB equations associated with other cost functionals

In this section we supposed the dynamics of a controlled diffusion X^u are time-homogenous

$$\mathrm{d}\mathrm{X}_t^u = \mu(\mathrm{X}_t^u, u_t)\mathrm{d}t + \sigma(\mathrm{X}_t^u, u_t)\mathrm{d}\mathrm{W}_t. \tag{6.11}$$

In what follows, we state without proof the HJB PDEs associated with the indefinite time and infinite time cost functionals $\mathrm{J}(u)$, which were introduced introduced in Section 6.1.

1. Consider the *Indefinite time* cost functional

$$\mathrm{J}(u) = \mathbb{E}\Big[\int_0^{\tau^u} \mathrm{F}(\mathrm{X}_t^u, u_t)\mathrm{d}t + \varphi(\mathrm{X}_{\tau^u}^u)\Big], \qquad \tau^u = \inf\{t \geq 0 : \mathrm{X}_t^u \notin \mathrm{A}\},$$

   where $\mathrm{A} \subset \mathbb{R}^d$ is some open set, $\mathrm{F} : \mathrm{A} \times \mathbb{U} \to \mathbb{R}$ and $\varphi : \partial\mathrm{A} \to \mathbb{R}$. The associated HJB PDE is

$$\begin{aligned} 0 &= \sup_{u \in \mathbb{U}} (\mathcal{A}^u \mathrm{V} + \mathrm{F}^u), && x \in \mathrm{A}, \\ \mathrm{V}(x) &= \varphi(x), && x \in \partial\mathrm{A}. \end{aligned}$$

   where the operator $\mathcal{A}^u$ is the generator of the process X^u defined in (6.11) with $u_t = u$ fixed. In the one-dimensional case, we have

$$\mathcal{A}^u = \mu(x, u)\partial_x + \tfrac{1}{2}\sigma^2(x, u)\partial_x^2.$$

2. Consider the *Infinite time* cost functional

$$J(u) = \mathbb{E}\int_0^\infty e^{-\lambda t} F(X_t^u, u_t)dt,$$

where $F : \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}$ and $\lambda > 0$. The associated HJB PDE is

$$0 = \sup_{u\in\mathbb{U}} (\mathcal{A}^u V + F^u) - \lambda V. \tag{6.12}$$

We leave the formal proof of the above HJB equations as an exercise for the reader.

6.5 Exercises

Exercise 6.1. Derive the HJB equation (6.12).

Exercise 6.2 (Merton problem). The *Merton problem*, due to Nobel Prize winner Robert Merton, is a classical problem in mathematical finance whereby an investor seeks to optimize his expected utility at a fixed future date T by investing in a stock. Specifically, suppose a stock $S = (S_t)_{0\le t\le T}$ follows a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

Let $X^u = (X_t^u)_{0\le t\le T}$ be the wealth process of an investor who invests u_t dollars in S at time t and keeps the rest of his money in a bank account (which we assume pays no interest). The dynamics of X^u are given by

$$dX_t^u = \frac{u_t}{S_t} dS_t = \mu u_t dt + \sigma u_t dW_t.$$

The investor's control problem is to maximize

$$J(u) = \max_u \mathbb{E}\left[U(X_T^u)\right], \qquad U(x) = \frac{1}{1-\gamma} x^{1-\gamma}, \qquad \gamma \in (0,1)\cup(1,\infty).$$

The function U is called the investor's *utility function*; it is intended to map the investor's wealth to his happiness. Since "more money" = "more happy" the function U is strictly increasing. It is also concave since, an additional dollar means less to somebody with a wealth of 1 million dollars than it does to somebody with a wealth of 10 dollars. Find the investor's value function

$$V(t,x) := \max_u \mathbb{E}[U(X_T^u)|X_t^u = x],$$

and the optimal control u_t^*. *Hint*: to solve the HJB PDE, try a solution of the form $V(t,x) = f(t)U(x)$ with $f(T) = 1$.

Chapter 7

Jump diffusions

Notes from this chapter are taken primarily from (Øksendal and Sulem, 2005, Chapter 1). Notes for Section 7.5 on Hawkes processes follow Hawkes (1971).

7.1 Basic definitions and results on Lévy processes

Definition 7.1.1. A d-dimensional stochstic process $\eta = (\eta_t)_{t\geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is called a *Lévy process* if it satisfies

1. $\eta_0 = 0$.
2. Independent increments: for any $0 \leq t_1 < t_2 < t_3 < t_4 < \infty$, we have $\eta_{t_4} - \eta_{t_3} \perp\!\!\!\perp \eta_{t_2} - \eta_{t_1}$.
3. Stationary increments: for any $0 \leq t_1 < t_2 < \infty$, we have $\eta_{t_2} - \eta_{t_1} \sim \eta_{t_2 - t_1}$.
4. Continuity in probability: for any $\varepsilon > 0$ and $t \geq 0$, we have $\lim_{s \searrow 0} \mathbb{P}(|\eta_{t+s} - \eta_t| > \varepsilon) = 0$.

Example 7.1.2. A Brownian motion $\mathrm{W} = (\mathrm{W}_t)_{t\geq 0}$ is a continuous Lévy process with $\mathrm{W}_{t+s} - \mathrm{W}_t \sim \mathcal{N}(0, s)$.

Example 7.1.3. A *Poisson process* $\mathrm{P} = (\mathrm{P}_t)_{t\geq 0}$ with intensity λ is a pure jump Lévy process with $\mathrm{P}_{t+s} - \mathrm{P}_t \sim \mathrm{Poi}(\lambda s)$. Hence, we have

$$\mathbb{P}(\mathrm{P}_t = n) = \frac{(\lambda t)^n}{n!} \mathrm{e}^{-\lambda t}, \qquad\qquad n \in \{0, 1, 2, \ldots\},$$

We can compute the mean of P_t as follows

$$\mathbb{E}\mathrm{P}_t = \sum_{n=0} n \frac{(\lambda t)^n}{n!} \mathrm{e}^{-\lambda t} = \lambda t \sum_{n=1} \frac{(\lambda t)^{n-1}}{(n-1)!} \mathrm{e}^{-\lambda t} = \lambda t,$$

where we have used $\sum_{n=0}^{\infty} x^n / n! = \mathrm{e}^x$.

Note, Item 4 in Definition 7.1.1 does *not* mean that a Lévy process cannot jump. For example, consider a Poisson process $P = (P_t)_{t\geq 0}$ with intensity λ. A Poisson process is a jump process and yet it is easy to see that it is continuous in probability as, for any $\varepsilon \in (0,1)$, we have

$$\mathbb{P}(|P_{t+s} - P_t| > \varepsilon) = \mathbb{P}(P_{t+s} - P_t \geq 1) = 1 - \mathbb{P}(P_{t+s} - P_t = 0) = 1 - e^{-\lambda s} \to 0 \qquad \text{as } s \searrow 0.$$

Item 4 simply means that, at a fixed t, the probability that a Lévy process has a discontinuity at t is zero.

We can and do assume that any Lévy process is right-continuous with left limits (RCLL). That is

$$\lim_{s \searrow 0} \eta_{t+s} = \eta_t, \qquad \forall\, t \geq 0.$$

A process that is RCLL is sometimes called *càdlàg* (for those who speak French: *continue à droite, limite à gauche*).

A filtration for a Lévy process is defined just as a filtration for a Brownian motion.

DEFINITION 7.1.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be on probability space on which a Lévy process $\eta = (\eta_t)_{t\geq 0}$ is defined. A *filtration for the Lévy process* η is a collection of σ-algebras $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying:

1. Information accumulates: if $0 \leq s < t$ then $\mathcal{F}_s \subset \mathcal{F}_t$.
2. Adaptivity: for all $t \geq 0$, we have $\eta_t \in \mathcal{F}_t$.
3. Independence of future increments: if $u > t \geq 0$ then $(\eta_u - \eta_t) \perp\!\!\!\perp \mathcal{F}_t$.

The most natural choice for this filtration $\mathbb{F}$ is (not surprisingly) the *natural filtration* for η, that is, $\mathcal{F}_t = \sigma(\eta_s, 0 \leq s \leq t)$. In principle the filtration $\mathbb{F}$ could contain more than the information obtained by observing η. However, the information in the filtration is not allowed to destroy the independence of future increments of the Lévy process.

DEFINITION 7.1.5. The *jump* of a Lévy process η at time $t \geq 0$ is defined as

$$\Delta\eta_t := \eta_t - \eta_{t-}, \qquad \eta_{t-} := \lim_{s \nearrow t} \eta_s.$$

Obviously, if η does not experience a jump at time t, then $\Delta\eta_t = 0$.

DEFINITION 7.1.6. We define $N : \mathbb{R}_+ \times \mathcal{B}_0^d \times \Omega \to \mathbb{N}_0$ the *Poisson random measure of* η by

$$N(t, U, \omega) = \sum_{s:0<s\leq t} \mathbb{1}_{\{\Delta\eta_s(\omega) \in U\}}, \qquad U \in \mathcal{B}_0^d := \{B \in \mathcal{B}(\mathbb{R}^d) : \{0\} \notin B\}.$$

As with most random variables and stochastic processes, we will typically omit the dependence of N on ω, writing simply $\mathrm{N}(t,\mathrm{U})$ as opposed to $\mathrm{N}(t,\mathrm{U},\omega)$. The Poisson random measure $\mathrm{N}(t,\mathrm{U})$ counts the number of jumps of size $\Delta\eta_s \in \mathrm{U}$ prior to time t. It will be convenient to introduce the following differential form

$$\mathrm{N}(\mathrm{d}t,\mathrm{d}z),$$

which counts the number of jumps of size $\mathrm{d}z$ over the time interval $\mathrm{d}t$.

DEFINITION 7.1.7. Let N be the Poisson random measure of a Lévy process η. We define $\nu : \mathcal{B}_0^d \to \mathbb{R}_+$, the *Lévy measure of* η, as follows

$$\nu(\mathrm{U}) := \mathbb{E}\mathrm{N}(1,\mathrm{U}), \qquad\qquad \mathrm{U} \in \mathcal{B}_0^d.$$

THEOREM 7.1.8. *Let* $\mathrm{U} \in \mathcal{B}_0^d$. *Then the process* $(\mathrm{N}(t,\mathrm{U}))_{t\geq 0}$ *is a Poisson process with intensity* $\nu(\mathrm{U})$.

Because $\mathrm{N}(t,\mathrm{U})$ is a Poisson process with intensity $\nu(\mathrm{U})$, we have

$$\mathbb{E}\mathrm{N}(t,\mathrm{U}) = \nu(\mathrm{U})t, \qquad\qquad \Rightarrow \qquad\qquad \frac{1}{t}\mathbb{E}\mathrm{N}(t,\mathrm{U}) = \nu(\mathrm{U}).$$

Thus, $\nu(\mathrm{U})$ is the expected number of times η has a jump of size $z \in \mathrm{U}$ per unit time. Note that

$$\nu(\mathrm{U})\mathrm{d}t = \mathbb{E}\mathrm{N}(\mathrm{d}t,\mathrm{U}) = \sum_{k=0}^{\infty} k\mathbb{P}\Big(\mathrm{N}(\mathrm{d}t,\mathrm{U}) = k\Big) = \mathbb{P}(\mathrm{N}(\mathrm{d}t,\mathrm{U}) = 1) + \mathcal{O}(\mathrm{d}t^2), \qquad \text{as } \mathrm{d}t \to 0.$$

Hence, for small $\mathrm{d}t$, one can think of $\nu(\mathrm{U})\mathrm{d}t$ as the probability that η experiences a jump of size $z \in \mathrm{U}$ in the time interval $\mathrm{d}t$.

EXAMPLE 7.1.9 (COMPOUND POISSON PROCESS). Let $(\mathrm{X}_n)_{n\in\mathbb{N}}$ be a sequence of iid random vectors in $\mathbb{R}^d$ with distribution F_{X}. Let $(\mathrm{P}_t)_{t\geq 0}$ be a Poisson process with intensity λ. Assume $(\mathrm{P}_t)_{t\geq 0} \perp\!\!\!\perp (\mathrm{X}_n)_{n\in\mathbb{N}}$. We define a *compound Poisson process* $\eta = (\eta_t)_{t\geq 0}$ by

$$\eta_t = \sum_{n=1}^{\mathrm{P}_t} \mathrm{X}_n,$$

where $\sum_{n=1}^{0} = 0$. The increments of η are given by

$$\eta_t - \eta_s = \sum_{n=\mathrm{P}_s+1}^{\mathrm{P}_t} \mathrm{X}_n.$$

The distribution $F_{\eta_t-\eta_s}$ depends only on $(t-s)$ and F_X. As such we see that η is a stationary process. Also, non-overlapping increments of η are clearly independent, as they depend on different (X_i). Thus, η is a Lévy process in $\mathbb{R}^d$. Let us find the Lévy measure ν corresponding to η. For any $U \in \mathcal{B}_0^d$ we have

$$\begin{aligned}\nu(U) &= \mathbb{E}N(1,U) = \mathbb{E}\sum_{s:0<s\le 1} \mathbb{1}_{\{\Delta\eta_s\in U\}} \\ &= \mathbb{E}\sum_{n=1}^{P_1} \mathbb{1}_{\{X_n\in U\}} = \mathbb{E}\sum_{n=1}^{P_1} \mathbb{E}[\mathbb{1}_{\{X_n\in U\}}|P_1] \\ &= \mathbb{E}\sum_{n=1}^{P_1} \mathbb{E}[\mathbb{1}_{\{X_n\in U\}}] = \mathbb{E}\sum_{n=1}^{P_1} \mathbb{P}(X_n\in U) \\ &= F_X(U)\mathbb{E}P_1 = \lambda F_X(U).\end{aligned}$$

Thus, the Lévy measure of η is $\nu = \lambda F_X$.

A pure-jump Lévy process η has a finite Lévy measure $\nu(\mathbb{R}^d) < \infty$ if and only if it can be represented by a compound Poisson process. In this case, we can express η in one of two ways

$$\eta_t = \int_{\mathbb{R}^d} zN(t,dz), \qquad \text{or} \qquad \eta_t = \sum_{n=1}^{P_t} X_n. \tag{7.1}$$

However, there exist Lévy processes for which $\nu(\mathbb{R}^d) = \infty$. We call a Lévy process for which $\nu(\mathbb{R}^d) = \infty$ an *infinite activity* Lévy process. For an infinite activity Lévy process, neither of the representations in (7.1) make sense.

To write the most general form of a Lévy process, we introduce the compensated Poisson random measure.

DEFINITION 7.1.10. Let N be a Poisson random measure with associated Lévy measure ν. The *compensated Poisson random measure*, denoted $\widetilde{N}$ is defined as

$$\widetilde{N}(t,A) := N(t,A) - \nu(A)t.$$

In differential form, we have

$$\widetilde{N}(dt,dz) := N(dt,dz) - \nu(dz)dt.$$

For any $A \in \mathcal{B}_0^d$ the process $(\widetilde{N}(t,A))_{t\ge 0}$ is a martingale (with respect to its own filtration) as

$$\begin{aligned}\mathbb{E}[\widetilde{N}(t,A)|\mathcal{F}_s] &= \widetilde{N}(s,A) + \mathbb{E}[\widetilde{N}(t,A) - \widetilde{N}(s,A)|\mathcal{F}_s] \\ &= \widetilde{N}(s,A) + \mathbb{E}[N(t,A) - N(s,A)|\mathcal{F}_s] - \nu(A)(t-s)\end{aligned}$$

$$= \tilde{\mathrm{N}}(s, \mathrm{A}) + \nu(\mathrm{A})(t-s) - \nu(\mathrm{A})(t-s) = \tilde{\mathrm{N}}(s, \mathrm{A}),$$

where we have used $\mathbb{E}\mathrm{N}(t, \mathrm{A}) = t\nu(\mathrm{A})$. It follows that, for any fixed $\mathrm{R} > 0$, the the process $\mathrm{M} = (\mathrm{M}_t)_{t \geq 0}$ defined as

$$\mathrm{M}_t \equiv \int_{|z|<\mathrm{R}} z\tilde{\mathrm{N}}(t, \mathrm{d}z),$$

is also a martingale because, for any $0 \leq s \leq t < \infty$ we have

$$\mathbb{E}_s \mathrm{M}_t = \int_{|z|<\mathrm{R}} z\mathbb{E}_s\tilde{\mathrm{N}}(t, \mathrm{d}z) = \int_{|z|<\mathrm{R}} z\tilde{\mathrm{N}}(s, \mathrm{d}z) = \mathrm{M}_s,$$

where we have used the short-hand notation $\mathbb{E}_s := \mathbb{E}(\,\cdot\,|\mathcal{F}_s)$.

The following theorem gives the most general form of a Lévy process.

THEOREM 7.1.11 (ITÔ-LÉVY DECOMPOSITION). *Let $\eta = (\eta_t)_{t \geq 0}$ be a Lévy process in $\mathbb{R}^d$. Then η has the decomposition*

$$\eta_t = \mu_{\mathrm{R}} t + \sigma \mathrm{W}_t + \int_{|z|<\mathrm{R}} z\tilde{\mathrm{N}}(t, \mathrm{d}z) + \int_{|z| \geq \mathrm{R}} z\mathrm{N}(t, \mathrm{d}z), \tag{7.2}$$

for some vector $\mu_{\mathrm{R}} \in \mathbb{R}^d$, matrix $\sigma \in \mathbb{R}_+^{d \times d}$, constant $\mathrm{R} \in [0, \infty]$, Poisson random measure N on $\mathbb{R}^d$, and d-dimensional Brownian motion W that independent of N.

We can always choose $\mathrm{R} = 1$ in (7.2). It is useful to recognize when we can choose $\mathrm{R} = 0$ and $\mathrm{R} = \infty$ as, in these cases, the right-hand side of (7.2) reduces from four to three terms.

EXAMPLE 7.1.12. If the Lévy measure ν satisfies

$$\int_{|z| \geq 1} |z| \nu(\mathrm{d}z) < \infty, \tag{7.3}$$

then we can write (7.2) as

$$\begin{aligned} \eta_t &= \Big(\mu_1 + \int_{|z| \geq 1} z\nu(\mathrm{d}z)\Big)t + \sigma \mathrm{W}_t + \int_{|z|<1} z\tilde{\mathrm{N}}(t, \mathrm{d}z) + \int_{|z| \geq 1} z\left(\mathrm{N}(t, \mathrm{d}z) - \nu(\mathrm{d}z)t\right) \\ &= \mu_\infty t + \sigma \mathrm{W}_t + \int_{\mathbb{R}^d} z\tilde{\mathrm{N}}(t, \mathrm{d}z), \end{aligned}$$

where $\mu_\infty = \mu_1 + \int_{|z| \geq 1} z\nu(\mathrm{d}z)$. We call μ_∞ the *center* of η since $\mathbb{E}\eta_t = \mu_\infty t$. It is straightforward to show that if $\mu_\infty = 0$ then η is a martingale.

EXAMPLE 7.1.13. If the Lévy measure ν satisfies

$$\int_{|z|<1} |z| \nu(\mathrm{d}z) < \infty, \tag{7.4}$$

then we can write (7.2) as

$$\begin{aligned}\eta_t &= \mu_1 t + \sigma \mathrm{W}_t + \int_{|z|<1} z\mathrm{N}(t, \mathrm{d}z) - \int_{|z|<1} z\nu(\mathrm{d}z)t + \int_{|z|\geq 1} z\mathrm{N}(t, \mathrm{d}z) \\ &= \mu_0 t + \sigma \mathrm{W}_t + \int_{\mathbb{R}^d} z\mathrm{N}(t, \mathrm{d}z), \end{aligned} \tag{7.5}$$

where $\mu_0 = \mu_1 - \int_{|z|<1} z\nu(\mathrm{d}z)$. We call μ_0 the *drift* of η.

EXAMPLE 7.1.14. If the Lévy measure ν satisfies

$$\nu(\mathbb{R}^d) < \infty,$$

then (7.4) holds, and we can write (7.5) as a compound Poisson process

$$\eta_t = \mu_0 t + \sigma \mathrm{W}_t + \sum_{n=1}^{\mathrm{P}_t} \mathrm{X}_n,$$

where P is a Poisson process with intensity $\nu(\mathbb{R}^d)$ and the $(\mathrm{X}_n)_{n\in\mathbb{N}}$ are iid random vectors in $\mathbb{R}^d$ with common distribution $\mathrm{F}_\mathrm{X} = \nu/\nu(\mathbb{R}^d)$.

We will typically drop the subscript from μ_R, writing it instead simply as μ.

THEOREM 7.1.15. *Lévy processes are Markov processes.*

This should come as no surprise. We have already seen a Brownian motion and a Poisson process are Markov processes. As these processes serve as building blocks for more general Lévy processes, it follows that Lévy processes are Markov as well.

THEOREM 7.1.16 (LÉVY-KINTCHINE FORMULA). *Let $\eta = (\eta_t)_{t\geq 0}$ be the following d-dimensional Lévy process*

$$\eta_t = \mu t + \sigma \mathrm{W}_t + \int_{|z|<\mathrm{R}} z\tilde{\mathrm{N}}(t, \mathrm{d}z) + \int_{|z|\geq \mathrm{R}} z\mathrm{N}(t, \mathrm{d}z),$$

where $\mu \in \mathbb{R}^d$, $\sigma \in \mathbb{R}_+^{d\times d}$, W is a d-dimensional Brownian motion and N is a Poisson random measure on $\mathbb{R}^d$. Then the Lévy measure ν associated with N satisfies

$$\int_{\mathbb{R}^d} (1 \wedge |z|^2)\nu(\mathrm{d}z) < \infty, \tag{7.6}$$

and the characteristic function $\phi_{\eta_t} : \mathbb{R}^d \to \mathbb{C}$ is given by

$$\phi_{\eta_t}(u) := \mathbb{E}\mathrm{e}^{\mathrm{i}\langle u, \eta_t\rangle} = \mathrm{e}^{t\psi(u)}, \qquad \langle a, b\rangle := \sum_{i=1}^{d} a_i b_i. \tag{7.7}$$

where ψ, called the characteristic exponent of η, *is given by*

$$\psi(u) = \mathrm{i}\langle \mu, u\rangle - \tfrac{1}{2}\langle u, \sigma\sigma^{\mathrm{T}} u\rangle + \int_{|z|<\mathrm{R}} (\mathrm{e}^{\mathrm{i}\langle u,z\rangle} - 1 - \mathrm{i}\langle u, z\rangle)\nu(\mathrm{d}z) + \int_{|z|\geq \mathrm{R}} (\mathrm{e}^{\mathrm{i}\langle u,z\rangle} - 1)\nu(\mathrm{d}z). \quad (7.8)$$

Conversely, given triplet (μ, a, ν) with $a = \sigma\sigma^{\mathrm{T}}$ and ν satisfying (7.6), there exists a Lévy process η satisfying (7.7)-(7.8).

PROOF. We will show why (7.7)-(7.8) holds for a scalar Lévy process with jumps that are of the compound Poisson type. In this case, η has the decomposition

$$\eta_t = \mu t + \sigma \mathrm{W}_t + \sum_{n=1}^{\mathrm{P}_t} \mathrm{X}_n,$$

where W is a scalar Brownian motion, P is a Poisson process with intensity λ and the $(\mathrm{X}_n)_{n\in\mathbb{N}}$ are iid random variables with common distribution F_{X}. We compute

$$\begin{aligned}
\mathbb{E}\mathrm{e}^{\mathrm{i}u\eta_t} &= \mathbb{E}\mathrm{e}^{\mathrm{i}u\mu t + \mathrm{i}u\sigma \mathrm{W}_t + \mathrm{i}u\sum_{n=1}^{\mathrm{P}_t}\mathrm{X}_n} \\
&= \mathrm{e}^{\mathrm{i}u\mu t}\mathbb{E}\mathrm{e}^{\mathrm{i}u\sigma\mathrm{W}_t}\mathbb{E}\mathrm{e}^{\mathrm{i}u\sum_{n=1}^{\mathrm{P}_t}\mathrm{X}_n} \\
&= \mathrm{e}^{\mathrm{i}u\mu t - \frac{1}{2}\sigma^2u^2t}\mathbb{E}\mathbb{E}[\mathrm{e}^{\mathrm{i}u\sum_{n=1}^{\mathrm{P}_t}\mathrm{X}_n}|\mathrm{P}_t] \\
&= \mathrm{e}^{\mathrm{i}u\mu t - \frac{1}{2}\sigma^2u^2t}\mathbb{E}(\mathbb{E}[\mathrm{e}^{\mathrm{i}u\mathrm{X}}])^{\mathrm{P}_t} \\
&= \mathrm{e}^{\mathrm{i}u\mu t - \frac{1}{2}\sigma^2u^2t}\mathbb{E}\Big(\int \mathrm{e}^{\mathrm{i}uz}\mathrm{F}_{\mathrm{X}}(\mathrm{d}z)\Big)^{\mathrm{P}_t} \\
&= \mathrm{e}^{\mathrm{i}u\mu t - \frac{1}{2}\sigma^2u^2t}\exp\Big(\lambda t\Big(\int \mathrm{e}^{\mathrm{i}uz}\mathrm{F}_{\mathrm{X}}(\mathrm{d}z) - 1\Big)\Big) \\
&= \mathrm{e}^{\mathrm{i}u\mu t - \frac{1}{2}\sigma^2u^2t}\exp\Big(\lambda t\int (\mathrm{e}^{\mathrm{i}uz} - 1)\mathrm{F}_{\mathrm{X}}(\mathrm{d}z)\Big) \\
&= \mathrm{e}^{t\psi(u)},
\end{aligned}$$

where we have used $\mathbb{E}s^{\mathrm{P}_t} = \exp(\lambda t(s-1))$ and $\nu = \lambda \mathrm{F}_{\mathrm{X}}$. The computation for a more general Lévy process in multiple dimensions is similar. □

Let us provide a few examples of Lévy measures ν and compute the integrals that appear in expression (7.8) for the characteristic exponent ψ.

EXAMPLE 7.1.17 (DIRAC COMB). A Dirac comb Lévy measure on $\mathbb{R}$ is a measure of the form

$$\nu(\mathrm{d}z) = \sum_{i=1}^{n} \lambda_i \delta_{z_i}(z)\mathrm{d}z,$$

where λ_i is the intensity of a jump of size z_i. In this case, both (7.3) and (7.4) are satisfied, so we can choose either $R = 0$ or $R = \infty$. Suppose we choose $R = 0$. Then the third term in (7.8) disappears and the last term becomes

$$\int_{\mathbb{R}} (e^{iuz} - 1)\nu(dz) = \sum_{i=1}^{n} \lambda_i (e^{iuz_i} - 1).$$

EXAMPLE 7.1.18 (GAUSSIAN JUMPS). A Gaussian Lévy measure on $\mathbb{R}$ is a measure of the form

$$\nu(dz) = \lambda \frac{1}{\sqrt{2\pi s^2}} \exp\left(\frac{-(z-m)^2}{2s^2}\right) dz,$$

where m is the mean jump size, s^2 is the variance of the jumps, and the intensity of all jumps is $\nu(\mathbb{R}) = \lambda$. In this case, both (7.3) and (7.4) are satisfied, so we can choose either $R = 0$ or $R = \infty$. Suppose we choose $R = 0$. Then the third term in (7.8) disappears and the last term becomes

$$\int_{\mathbb{R}} (e^{iuz} - 1)\nu(dz) = \lambda\left(e^{ium - \frac{1}{2}\sigma^2 u^2} - 1\right).$$

7.2 Lévy-Itô processes and the Itô Formula

ASSUMPTION 7.2.1. From this point onward we assume $\mathbb{E}|\eta_t|^2 < \infty$, and thus $\mathbb{E}|\eta_t| < \infty$. Thus, we can take $R = \infty$ and express η as follows

$$\eta_t = \mu t + \sigma W_t + \int_{\mathbb{R}^d} z\tilde{N}(t, dz). \tag{7.9}$$

Lévy process are *semimartingales*, which (roughly speaking) form the class of stochastic processes that can be used to construct an Itô integral. In Section 4.2 we defined Itô integral of an $(\mathcal{F}_t)_{t\geq 0}$-adapted process $\Gamma = (\Gamma_t)_{t\geq 0}$ with respect to a Brownian motion $W = (W_t)_{t\geq 0}$ by introducing a sequence of simple processes $\Gamma^{(n)} = (\Gamma_t^{(n)})_{t\geq 0}$ that converged to Γ in the followingg sense

$$\Gamma_t^{(n)} := \sum_{j=0}^{n-1} \Gamma_{t_j} \mathbb{1}_{\{t_j \leq t < t_{j+1}\}}, \qquad 0 \leq t_0 < t_1 < \ldots < t_n = T, \qquad \lim_{n\to\infty} \mathbb{E}\int_0^T |\Gamma_t - \Gamma_t^{(n)}|^2 dt = 0.$$

We then by defined the Itô integral as the following limit

$$\int_0^T \Gamma_t dW_t := \lim_{n\to\infty} \int_0^T \Gamma_t^{(n)} dW_t = \lim_{n\to\infty} \sum_{j=0}^{n-1} \Gamma_{t_j}(W_{t_{j+1}} - W_{t_j}).$$

We can define the Itô integral of Γ with respect to a Lévy process η in the same manner

$$\int_0^T \Gamma_t d\eta_t := \lim_{n\to\infty} \int_0^T \Gamma_t^{(n)} d\eta_t = \lim_{n\to\infty} \sum_{j=0}^{n-1} \Gamma_{t_j}(\eta_{t_{j+1}} - \eta_{t_j}).$$

Here, we must add a technical condition that the integrand Γ be a *left-continuous* process. If a process Γ is not left-continuous, then we can still define a stochastic integral as follows

$$I_T = \int_0^T \Gamma_{t-} d\eta_t, \qquad \Gamma_{t-} = \lim_{s \nearrow t} \Gamma_s,$$

where $(\Gamma_{t-})_{t\geq 0}$ is the left-continuous version of $(\Gamma_t)_{t\geq 0}$. Note that, if Γ is left-continuous then $\Gamma_t = \Gamma_{t-}$ for all $t \geq 0$. The process $(I_t)_{t\geq 0}$ will still be a right-continuous process as η is right-continuous. In view of (7.9), for a left-continuous process Γ, we can seperate an Itô integral into three terms

$$\int_0^T \Gamma_t d\eta_t = \int_0^T \Gamma_t \mu dt + \int_0^T \Gamma_t \sigma dW_t + \int_0^T \int \Gamma_t z \tilde{N}(dt, dz). \tag{7.10}$$

Expression (7.10) suggests that we consider general processes of the form

$$dX_t = \mu_t dt + \sigma_t dW_t + \int \gamma_t(z) \tilde{N}(dt, dz), \tag{7.11}$$

where the processes $(\mu_t)_{t\geq 0}$, $(\sigma_t)_{t\geq 0}$ and $(\gamma_{t-}(z))_{t\geq 0}$, must be adapted to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ obtained by observing the processes W and $\int z N(\cdot, dz)$. Note that we have written things in *differential form*; to obtain the integral form, simply integrate and add an initial condition. We have been somewhat vague about the dimension of the various objects appearing in (7.11). In general, for $X \in \mathbb{R}^d$, we could have

$$\mu : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d, \qquad \sigma : \mathbb{R}_+ \times \Omega \to \mathbb{R}^{d\times n}, \qquad \gamma : \mathbb{R}_+ \times \mathbb{R}^m \times \Omega \to \mathbb{R}^{d\times k}.$$

where W is an n-dimensional Brownian motion (with $W^{(i)} \perp\!\!\!\perp W^{(j)}$ for $i \neq j$) and $\tilde{N}$ is a k-dimensional Poisson random measure (with $N^{(i)} \perp\!\!\!\perp N^{(j)}$ for $i \neq j$) on $\mathbb{R}^m$. Component-wise, we have

$$dX_t^{(i)} = \mu_t^{(i)} dt + \sum_{j=1}^n \sigma_t^{(i,j)} dW_t^{(j)} + \sum_{j=1}^k \int_{\mathbb{R}^m} \gamma_{t-}^{(i,j)}(z) \tilde{N}^{(j)}(dt, dz), \qquad i = 1, 2, \ldots, d.$$

We call any process of the form (7.11) a *Lévy-Itô process*.

We showed (for the scalar case) in Theorem 4.1.5 that if σ satisfies $\mathbb{E}\int_0^T |\sigma_t|^2 dt < \infty$ then the Itô integral $I_t := \int_0^t \sigma_s dW_s$ is a martingale. Likewise, if

$$\mathbb{E}\int_0^T \int_{\mathbb{R}} |\gamma_t(z)|^2 \nu(dz) dt < \infty,$$

then the process $M = (M_t)_{0\leq t\leq T}$, defined by

$$M_t := \int_0^t \int_{\mathbb{R}} \gamma_s(z) \tilde{N}(ds, dz),$$

is a martingale. It follows that if $\mu = 0$ in (7.11), and σ and $\gamma(z)$ satisfy the above square integrability conditions, then the process $X = (X_t)_{t\geq 0}$ is a martingale.

THEOREM 7.2.2 (ONE-DIMENSIONAL ITÔ FORMULA FOR A LÉVY-ITÔ PROCESSES). *Suppose* $X = (X_t)_{t\geq 0}$ *is a one-dimensional Lévy-Itô process of the form*

$$dX_t = \mu_t dt + \sigma_t dW_t + \int_{\mathbb{R}} \gamma_{t-}(z)\tilde{N}(dt, dz). \tag{7.12}$$

where the drift μ *volatility* σ *and* γ *are maps*

$$\mu : \mathbb{R}_+ \times \Omega \to \mathbb{R}, \qquad \sigma : \mathbb{R}_+ \times \Omega \to \mathbb{R}_+, \qquad \gamma : \mathbb{R}_+ \times \Omega \times \mathbb{R} \to \mathbb{R},$$

and N *is a one-dimensional Poisson random measure on* $\mathbb{R}$*, which is independent of a one-dimensional Brownian motion* W*. Let* $f : \mathbb{R} \to \mathbb{R}$ *satisfy* $f \in C^2(\mathbb{R})$*. Then*

$$\begin{aligned} df(X_t) = &\Big(\mu_t f'(X_t) + \tfrac{1}{2}\sigma_t^2 f''(X_t)\Big)dt + \sigma_t f'(X_t)dW_t \\ &+ \int_{\mathbb{R}} \Big(f\left(X_{t-} + \gamma_t(z)\right) - f\left(X_{t-}\right)\Big)\tilde{N}(dt, dz) \\ &+ \int_{\mathbb{R}} \Big(f\left(X_{t-} + \gamma_t(z)\right) - f\left(X_{t-}\right) - \gamma_t(z)f'(X_t)\Big)\nu(dz)dt. \end{aligned} \tag{7.13}$$

We are not going to prove Theorem 7.2.2. However, we will attempt to understand why the formula is correct. Suppose for simplicity that $\nu(\mathbb{R}) < \infty$. In this case we have a finite activity Lévy process and we can separate the compensated Poisson random measure $\tilde{N}$ into two parts $N(dt, dz)$ and $\nu(dz)dt$. In this case, we can write dX_t as

$$\begin{aligned} dX_t &= \mu_t dt + \sigma_t dW_t + \int_{\mathbb{R}} \gamma_t(z)N(dt, dz) - \int_{\mathbb{R}} \gamma_t(z)\nu(dz)dt \\ &= \Big(\mu_t - \int_{\mathbb{R}} \gamma_t(z)\nu(dz)\Big)dt + \sigma_t dW_t + \int_{\mathbb{R}} \gamma_t(z)N(dt, dz) \\ &= \mu'_t dt + \sigma_t dW_t + \int_{\mathbb{R}} \gamma_t(z)N(dt, dz), \\ \mu'_t &:= \mu_t - \int_{\mathbb{R}} \gamma_t(z)\nu(dz), \end{aligned}$$

where we can now identify the drift of X as μ'_t. Similarly, we can write $df(X_t)$ as

$$\begin{aligned} df(X_t) = &\Big(\mu_t f'(X_t) + \tfrac{1}{2}\sigma_t^2 f''(X_t)\Big)dt + \sigma_t f'(X_t)dW_t \\ &+ \int_{\mathbb{R}} \Big(f\left(X_{t-} + \gamma_t(z)\right) - f\left(X_{t-}\right)\Big)N(dt, dz) \\ &- \int_{\mathbb{R}} \Big(f\left(X_{t-} + \gamma_t(z)\right) - f\left(X_{t-}\right)\Big)\nu(dz)dt \\ &+ \int_{\mathbb{R}} \Big(f\left(X_{t-} + \gamma_t(z)\right) - f\left(X_{t-}\right) - \gamma_t(z)f'(X_t)\Big)\nu(dz)dt \\ = &\left(\Big(\mu_t - \int_{\mathbb{R}} \gamma_t(z)\nu(dz)\Big)f'(X_t) + \tfrac{1}{2}\sigma_t^2 f''(X_t)\right)dt + \sigma_t f'(X_t)dW_t \end{aligned}$$

$$
\begin{aligned}
&+\int_{\mathbb{R}}\Big(f\left(X_{t-}+\gamma_t(z)\right)-f\left(X_{t-}\right)\Big)N(dt,dz)\\
&=\Big(\mu_t' f'(X_t)+\tfrac{1}{2}\sigma_t^2 f''(X_t)\Big)dt+\sigma_t f'(X_t)dW_t\\
&+\int_{\mathbb{R}}\Big(f\left(X_{t-}+\gamma_t(z)\right)-f\left(X_{t-}\right)\Big)N(dt,dz). \qquad (7.14)
\end{aligned}
$$

Things are looking a bit more familiar now. The non-integral terms in (7.14) arise from the $\mu_t' dt+\sigma_t dW_t$ part of dX_t. To understand the integral term in (7.14), suppose there is a jump of size y at time t. Since $\nu(\mathbb{R})<\infty$, there can only be a single jump at time t and thus

$$N(dt,dz)=\delta_y(z)dz.$$

It follows that

$$\Delta X_t := X_t - X_{t-} = \int_{\mathbb{R}}\gamma_t(z)N(dt,dz)=\int_{\mathbb{R}}\gamma_t(z)\delta_y(z)dz=\gamma_t(y),$$

and finally, that

$$
\begin{aligned}
\Delta f(X_t)&=f(X_t)-f(X_{t-})=f(X_{t-}+\gamma_t(y))-f(X_{t-})\\
&=\int_{\mathbb{R}}\Big(f(X_{t-}+\gamma_t(z))-f(X_{t-})\Big)\delta_y(z)dz\\
&=\int_{\mathbb{R}}\Big(f(X_{t-}+\gamma_t(z))-f(X_{t-})\Big)N(dt,dz).
\end{aligned}
$$

This last expression agrees with the integral term in (7.14). Of course, if there is no jump at time t then $N(dt,dz)=0$ for all dz and in this case $\Delta X_t=\Delta f(X_t)=0$.

EXAMPLE 7.2.3. Consider the following *Geometric Lévy process*

$$dX_t=bX_t dt+aX_t dW_t+\int_{\mathbb{R}}\left(e^{c(z)}-1\right)X_{t-}\widetilde{N}(dt,dz). \qquad (7.15)$$

Note, if $X_0>0$ then this process remains strictly positive since

$$X_{t-}+\left(e^{c(z)}-1\right)X_{t-}=e^{c(z)}X_{t-}>0.$$

Comparing (7.15) with (7.12) we identify

$$\mu_t=bX_t, \qquad \sigma_t=aX_t, \qquad \gamma_t(z)=\left(e^{c(z)}-1\right)X_{t-}.$$

We see an explicit expression for X_t. To this end, we define

$$f(X_t)=\log X_t, \qquad f'(X_t)=\frac{1}{X_t}, \qquad f''(X_t)=-\frac{1}{X_t^2},$$

Plugging all of the above into (7.13) we obtain

$$\begin{aligned}
\mathrm{d}f(\mathrm{X}_t) &= \mathrm{d}\log \mathrm{X}_t \\
&= \Big(b\mathrm{X}_t\frac{1}{\mathrm{X}_t} - \tfrac{1}{2}a^2\mathrm{X}_t^2\frac{1}{\mathrm{X}_t^2}\Big)\mathrm{d}t + a\mathrm{X}_t\frac{1}{\mathrm{X}_t}\mathrm{d}\mathrm{W}_t \\
&\quad + \int_{\mathbb{R}}\Big(\log\big(\mathrm{X}_{t-} + \big(\mathrm{e}^{c(z)}-1\big)\mathrm{X}_{t-}\big) - \log\left(\mathrm{X}_{t-}\right)\Big)\widetilde{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z) \\
&\quad + \int_{\mathbb{R}}\Big(\log\big(\mathrm{X}_{t-} + \big(\mathrm{e}^{c(z)}-1\big)\mathrm{X}_{t-}\big) - \log\left(\mathrm{X}_{t-}\right) - \big(\mathrm{e}^{c(z)}-1\big)\mathrm{X}_t\frac{1}{\mathrm{X}_t}\Big)\nu(\mathrm{d}z)\mathrm{d}t \\
&= \Big(b - \tfrac{1}{2}a^2\Big)\mathrm{d}t + a\mathrm{d}\mathrm{W}_t + \int_{\mathbb{R}} c(z)\widetilde{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z) + \int_{\mathbb{R}}\Big(c(z) - \mathrm{e}^{c(z)} + 1\Big)\nu(\mathrm{d}z)\mathrm{d}t \\
&= b'\mathrm{d}t + a\mathrm{d}\mathrm{W}_t + \int_{\mathbb{R}} c(z)\widetilde{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z), \\
b' &= b - \tfrac{1}{2}a^2 - \int_{\mathbb{R}}\big(\mathrm{e}^{c(z)} - 1 - c(z)\big)\nu(\mathrm{d}z).
\end{aligned}$$

Integrating from 0 to t, we obtain

$$\log \mathrm{X}_t - \log \mathrm{X}_0 = b't + a\mathrm{W}_t + \int_{\mathbb{R}} c(z)\widetilde{\mathrm{N}}(t,\mathrm{d}z).$$

Finally, adding $\log \mathrm{X}_0$ to both sides and exponentiating, we have

$$\mathrm{X}_t = \mathrm{X}_0 \exp\left(b't + a\mathrm{W}_t + \int_{\mathbb{R}} c(z)\widetilde{\mathrm{N}}(t,\mathrm{d}z)\right).$$

THEOREM 7.2.4 (LÉVY-ITÔ ISOMETRY). *Let* $\mathrm{X} = (\mathrm{X}_t)_{t\geq 0}$ *satisfy*

$$\mathrm{d}\mathrm{X}_t = \sigma_t\mathrm{d}\mathrm{W}_t + \int_{\mathbb{R}}\gamma_t(z)\widetilde{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z), \qquad \mathrm{X}_0 = 0,$$

and assume that $\sigma = (\sigma_t)_{t\geq 0}$ *and* $\gamma(z) = (\gamma_t(z))_{t\geq 0}$ *are* $\mathbb{F}$*-adapted. Then*

$$\mathbb{E}\mathrm{X}_{\mathrm{T}}^2 = \mathbb{E}\int_0^{\mathrm{T}}\Big(\sigma_t^2 + \int_{\mathbb{R}}\gamma_t^2(z)\nu(\mathrm{d}z)\Big)\mathrm{d}t,$$

assuming the right-hand side is finite.

PROOF. One could prove Theorem 7.2.4 directly from the definition of the Itô integral. However, we will use the Itô formula instead. Setting $f(x) = x^2$ and noting that

$$f'(\mathrm{X}_t) = 2\mathrm{X}_t, \qquad f''(\mathrm{X}_t) = 2, \qquad f(\mathrm{X}_{t-} + \gamma_t(z)) - f(\mathrm{X}_{t-}) = 2\mathrm{X}_{t-}\gamma_t(z) + \gamma_t^2(z),$$

we obtain from (7.13) that

$$\begin{aligned}
\mathrm{d}f(\mathrm{X}_t) &= \tfrac{1}{2}\sigma_t^2\cdot 2\mathrm{d}t + \sigma_t 2\mathrm{X}_t\mathrm{d}\mathrm{W}_t \\
&\quad + \int_{\mathbb{R}}\Big(2\mathrm{X}_{t-}\gamma_t(z) + \gamma_t^2(z)\Big)\widetilde{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z)
\end{aligned}$$

$$
\begin{aligned}
&+ \int_{\mathbb{R}} \Big(2\mathrm{X}_{t-}\gamma_t(z) + \gamma_t^2(z) - \gamma_t(z) \cdot 2\mathrm{X}_{t-} \Big) \nu(\mathrm{d}z)\mathrm{d}t \\
&= \Big(\sigma_t^2 + \int_{\mathbb{R}} \gamma_t^2(z)\nu(\mathrm{d}z) \Big)\mathrm{d}t + (\cdots)\mathrm{dW}_t + \int_{\mathbb{R}} (\cdots)\widetilde{\mathrm{N}}(\mathrm{d}t, \mathrm{d}z),
\end{aligned}
$$

Integrating from 0 to T and taking an expectation, we obtain

$$
\begin{aligned}
\mathbb{E}\mathrm{X}_{\mathrm{T}}^2 &= \mathbb{E}\int_0^{\mathrm{T}} \Big(\sigma_t^2 + \int_{\mathbb{R}} \gamma_t^2(z)\nu(\mathrm{d}z) \Big)\mathrm{d}t + \mathbb{E}\int_0^{\mathrm{T}} (\cdots)\mathrm{dW}_t + \mathbb{E}\int_0^{\mathrm{T}} \int_{\mathbb{R}} (\cdots)\widetilde{\mathrm{N}}(\mathrm{d}t, \mathrm{d}z) \\
&= \mathbb{E}\int_0^{\mathrm{T}} \Big(\sigma_t^2 + \int_{\mathbb{R}} \gamma_t^2(z)\nu(\mathrm{d}z) \Big)\mathrm{d}t,
\end{aligned}
$$

where we have used $\mathbb{E}\int_0^{\mathrm{T}} (\cdots)\mathrm{dW}_t = 0$ and $\mathbb{E}\int_0^{\mathrm{T}} \int_{\mathbb{R}} (\cdots)\widetilde{\mathrm{N}}(\mathrm{d}t, \mathrm{d}z) = 0$. □

We now provide the multi-dimensional version of the Itô formula.

THEOREM 7.2.5. *Let* $\mathrm{X} = (\mathrm{X}_t)_{t \geq 0}$ *be a* d*-dimensional Lévy-Itô process with components satisfying*

$$
\mathrm{dX}_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{dW}_t + \int_{\mathbb{R}^m} \gamma_t(z)\widetilde{\mathrm{N}}(\mathrm{d}t, \mathrm{d}z),
$$

where W *is an* n*-dimensional Brownian motion,* $\widetilde{\mathrm{N}}(\mathrm{d}t, \mathrm{d}z) = \mathrm{N}(\mathrm{d}t, \mathrm{d}z) - \nu(\mathrm{d}z)\mathrm{d}t$ *is* k*-dimensional compensated Poisson random measure on* $\mathbb{R}^m$ *and*

$$
\mu : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d, \qquad \sigma : \mathbb{R}_+ \times \Omega \to \mathbb{R}^{d \times n}, \qquad \gamma : \mathbb{R}_+ \times \Omega \times \mathbb{R}^m \to \mathbb{R}^{d \times k}.
$$

Component-wise, we can express the dynamics of X *as*

$$
\mathrm{dX}_t^{(i)} = \mu_t^{(i)}\mathrm{d}t + \sum_{j=1}^{n} \sigma_t^{(i,j)}\mathrm{dW}_t^{(j)} + \sum_{j=1}^{k} \int_{\mathbb{R}^m} \gamma_t^{(i,j)}(z)\widetilde{\mathrm{N}}^{(j)}(\mathrm{d}t, \mathrm{d}z), \qquad i = 1, 2, \ldots, d,
$$

Let $f : \mathbb{R}^d \to \mathbb{R}$ *satisfy* $f \in \mathrm{C}^2(\mathbb{R}^d)$*. Then*

$$
\begin{aligned}
\mathrm{d}f(\mathrm{X}_t) = \Big(&\sum_{i=1}^{d} \mu_t^{(i)} \partial_{x_i} f(\mathrm{X}_t) + \tfrac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (\sigma_t \sigma_t^T)_{i,j} \partial_{x_i} \partial_{x_j} f(\mathrm{X}_t) \Big)\mathrm{d}t + \sum_{i=1}^{d} \sum_{j=1}^{n} \sigma_t^{(i,j)} \partial_{x_i} f(\mathrm{X}_t)\mathrm{dW}_t^{(j)} \\
&+ \sum_{j=1}^{k} \int_{\mathbb{R}^m} \Big(f(\mathrm{X}_{t-} + \gamma_t^{(\cdot,j)}(z)) - f(\mathrm{X}_{t-}) \Big)\widetilde{\mathrm{N}}^{(j)}(\mathrm{d}t, \mathrm{d}z) + \\
&+ \sum_{j=1}^{k} \int_{\mathbb{R}^m} \Big(f(\mathrm{X}_{t-} + \gamma_t^{(\cdot,j)}(z)) - f(\mathrm{X}_{t-}) - \gamma_t^{(\cdot,j)}(z) \cdot \nabla f(\mathrm{X}_{t-}) \Big)\nu^{(j)}(\mathrm{d}z)\mathrm{d}t.
\end{aligned}
$$

I would not recommend memorizing the above formula. Simply book-mark this page. If you work with Lévy-Itô processes for a sufficient amount of time (about 5 years for the author of these notes), you will eventually have the above formula fixed in your mind.

THEOREM 7.2.6 (QUADRATIC COVARIATION). *Let* $X = (X_t)_{t\geq 0}$ *and* $Y = (Y_t)_{t\geq 0}$ *be given by*

$$\begin{aligned} dX_t &= \mu_t dt + \sum_{j=1}^{n} \sigma_t^{(j)} dW_t^{(j)} + \sum_{j=1}^{k} \int_{\mathbb{R}^m} \gamma_t^{(j)}(z)\tilde{N}^{(j)}(dt, dz), \\ dY_t &= b_t dt + \sum_{j=1}^{n} a_t^{(j)} dW_t^{(j)} + \sum_{j=1}^{k} \int_{\mathbb{R}^m} g_t^{(j)}(z)\tilde{N}^{(j)}(dt, dz). \end{aligned} \tag{7.16}$$

Here, X *and* Y *are both one-dimensional processes,* W *is a* n*-dimensional Brownian motion,* $\tilde{N}$ *is an* k*-dimensional compensated Poisson random measure on* $\mathbb{R}^m$ *and*

$$\mu_t, b_t \in \mathbb{R}, \qquad \sigma_t, a_t \in \mathbb{R}^n, \qquad \gamma_t(\cdot), g_t(\cdot) : \mathbb{R}^m \to \mathbb{R}^k.$$

Recall the Definition 3.3.4 of quadratic covariation. The quadratic covaration of X *and* Y *up to time* T*, denoted* $[X, Y]_T$*, is given by*

$$\begin{aligned} [X, Y]_T &= \int_0^T \sum_{j=1}^{n} \sigma_t^{(j)} a_t^{(j)} dt + \int_0^T \sum_{j=1}^{k} \int_{\mathbb{R}^m} \gamma_t^{(j)}(z) \cdot g_t^{(j)}(z) N^{(j)}(dt, dz) \\ &= \int_0^T \Big(\sum_{j=1}^{n} \sigma_t^{(j)} a_t^{(j)} + \sum_{j=1}^{k} \int_{\mathbb{R}^m} \gamma_t^{(j)}(z) \cdot g_t^{(j)}(z) \nu^{(j)}(dz) \Big) dt \\ &\quad + \int_0^T \sum_{j=1}^{k} \int_{\mathbb{R}^m} \gamma_t^{(j)}(z) \cdot g_t^{(j)}(z) \tilde{N}^{(j)}(dt, dz). \end{aligned}$$

We will not prove Theorem 7.2.6. Rather, we simply note that the above result relies on the following facts

$$[W^{(i)}, N^{(j)}(\cdot, A)]_t = 0, \qquad [N^{(i)}(\cdot, A), N^{(j)}(\cdot, B)]_t = \delta_{i,j} N^{(i)}(t, A \cap B), \qquad [N^{(i)}(\cdot, B), Id]_t = 0,$$

in addition to the previously derived facts

$$[W^{(i)}, W^{(j)}]_t = \delta_{i,j} t, \qquad [W^{(i)}, Id]_t = 0, \qquad [Id, Id]]_t = 0.$$

Heurisitcally, one can derive Theorem 7.2.6 by writing $d[X, Y]_t = dX_t dY_t$ and using the rules

$$\begin{aligned} dW_t^{(i)} dW_t^{(j)} &= \delta_{i,j} dt, & \tilde{N}^{(i)}(dt, dz)\tilde{N}^{(j)}(dt, dy) &= \delta_{i,j}\delta(z - y) N^{(i)}(dt, dz) dy, \\ dW_t^{(i)} dt &= 0, & dW_t^{(i)} \tilde{N}^{(j)}(dt, dz) &= 0, \\ dt\, dt &= 0, & \tilde{N}^{(i)}(dt, dz)\, dt &= 0. \end{aligned}$$

EXAMPLE 7.2.7. Suppose X is given by (7.16). Then, from Theorem 7.2.6, the quadratic variation of X, denoted $[X, X]_T$, is given by

$$[X, X]_T = \int_0^T \sum_{j=1}^{n} (\sigma_t^{(j)})^2 dt + \int_0^T \sum_{j=1}^{k} \int_{\mathbb{R}^m} (\gamma_t^{(j)}(z))^2 N^{(j)}(dt, dz)$$

$$= \int_0^T \Big(\sum_{j=1}^n (\sigma_t^{(j)})^2 + \sum_{j=1}^k \int_{\mathbb{R}^m} (\gamma_t^{(j)}(z))^2 \nu^{(j)}(dz) \Big) dt$$
$$+ \int_0^T \sum_{j=1}^k \int_{\mathbb{R}^m} (\gamma_t^{(j)}(z))^2 \widetilde{N}^{(j)}(dt, dz).$$

EXAMPLE 7.2.8. Some texts *define* quadratic covaration of two semimartingales X and Y as the unique process $[X, Y]_T$ that satisfies

$$X_T Y_T = X_0 Y_0 + \int_0^T X_{t-} dY_t + \int_0^T Y_{t-} dX_t + [X, Y]_T.$$

Let us check that this definition yields the same result as Theorem 7.2.6. For simplicity, let us take

$$dX_t = \mu_t dt + \sigma_t dW_t + \int_{\mathbb{R}^m} \gamma_t(z) \widetilde{N}(dt, dz),$$
$$dY_t = b_t dt + a_t dW_t + \int_{\mathbb{R}^m} g_t(z) \widetilde{N}(dt, dz),$$

where W is a one-dimensional Brownian motion and $\widetilde{N}$ is a one-dimensional compensated Poisson random measure on $\mathbb{R}^m$. Let $f(X_t, Y_t) = X_t Y_t$ and observe that

$$\partial_x f(X_t, Y_t) = Y_t, \quad \partial_y f(X_t, Y_t) = X_t, \quad \partial_x \partial_y f(X_t, Y_t) = 1, \quad \partial_x^2 f(X_t, Y_t) = 0, \quad \partial_y^2 f(X_t, Y_t) = 0,$$

and also that

$$f(X_{t-} + \gamma_t(z), Y_{t-} + g_t(z)) - f(X_{t-}, Y_{t-}) = X_{t-} g_t(z) + Y_{t-} \gamma_t(z) + \gamma_t(z) g_t(z).$$

Plugging all of this into the multidimensional Itô formula yields

$$\begin{aligned}
d(X_t Y_t) &= Y_t (\mu_t dt + \sigma_t dW_t) + X_t (b_t dt + a_t dW_t) \\
&\quad + \int_{\mathbb{R}^m} \Big(X_{t-} g_t(z) + Y_{t-} \gamma_t(z) + \gamma_t(z) g_t(z) \Big) \widetilde{N}(dt, dz) \\
&\quad + \int_{\mathbb{R}^m} \Big(X_{t-} g_t(z) + Y_{t-} \gamma_t(z) + \gamma_t(z) g_t(z) - X_{t-} g_t(z) - Y_{t-} \gamma_t(z) \Big) \nu(dz) dt \\
&= Y_{t-} \Big(\mu_t dt + \sigma_t dW_t + \int_{\mathbb{R}^m} \gamma_t(z) \widetilde{N}(dt, dz) \Big) + X_{t-} \Big(b_t dt + a_t dW_t + \int_{\mathbb{R}^m} g_t(z) \widetilde{N}(dt, dz) \Big) \\
&\quad + \int_{\mathbb{R}^m} \gamma_t(z) g_t(z) \widetilde{N}(dt, dz) + \int_{\mathbb{R}^m} \gamma_t(z) g_t(z) \nu(dz) dt \\
&= Y_{t-} dX_t + X_{t-} dY_t + d[X, Y]_t.
\end{aligned}$$

Thus, we have verified the new definition of quadratic covariation agrees (at least, for the processes in this example) with Theorem 7.2.6.

7.3 Lévy-Itô SDE

We now introduce a *Lévy-Itô Stochastic Differential equation (SDE)*, which is an equation of the form

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t + \int_{\mathbb{R}^m} \gamma(t, X_{t-}, z)\tilde{N}(\mathrm{d}t, \mathrm{d}z), \qquad X_0 = x \in \mathbb{R}^d, \tag{7.17}$$

where W is a n-dimensional Brownian motion, $\tilde{N}(\mathrm{d}t, \mathrm{d}z)$ is a k-dimensional compensated Poisson random measure on $\mathbb{R}^m$, and

$$\mu : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d, \qquad \sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times n}, \qquad \gamma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^{d\times k}.$$

Dor $i = 1, 2, \ldots, d$, the dynamics of the ith component of X are then given by

$$\mathrm{d}X_t^{(i)} = \mu^{(i)}(t, X_t)\mathrm{d}t + \sum_{j=1}^{n} \sigma^{(i,j)}(t, X_t)\mathrm{d}W_t^{(j)} + \sum_{j=1}^{k} \int_{\mathbb{R}^m} \gamma^{(i,j)}(t, X_{t-}, z)\tilde{N}^{(j)}(\mathrm{d}t, \mathrm{d}z),$$

with $X_0^{(i)} = x_i \in \mathbb{R}^d$.

Definition 7.3.1. A *solution* (specifically, a *strong solution*) of (7.17) is a functional F of the Brownian motion W and compensated Poisson random measure $\tilde{N}$

$$X_t = F(W_s, \tilde{N}(s, \mathrm{d}z); 0 \le s \le t),$$

such that

$$X_T = x + \int_0^T \mu(t, X_t)\mathrm{d}t + \int_0^T \sigma(t, X_t)\mathrm{d}W_t + \int_0^T \int_{\mathbb{R}^m} \gamma(t, X_{t-}, z)\tilde{N}(\mathrm{d}t, \mathrm{d}z), \qquad \forall\, T \ge 0.$$

Theorem 7.3.2. *Consider the one-dimensional process X driven by a single Brownian motion and a single Poisson random measure on $\mathbb{R}$. If there exist constants C_1 and C_2 such that the* linear growth *condition*

$$\mu^2(t, x) + \sigma^2(t, x) + \int_{\mathbb{R}} \gamma^2(t, x, z)\nu(\mathrm{d}z) < C_1(1 + x^2), \qquad \forall\, t \ge 0,$$

and the Lipschitz continuiety *condition*

$$\begin{aligned} &|\mu(t, x) - \mu(t, x)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 \\ &\quad + \int_{\mathbb{R}} |\gamma(t, x, z) - \gamma(t, y, z)|^2 \nu(\mathrm{d}z) < C_2|x - y|^2, \qquad \forall\, t \ge 0,\ x, y \in \mathbb{R}, \end{aligned}$$

then SDE (7.17) has a unique strong solution adapted to $\mathcal{F}_t := \sigma(X_s : 0 \le s \le t)$ (i.e., the solution X of SDE (7.17) is a Markov process). The conditions for higher dimensions are analogous.

Since X, the solution (7.17) is a Markov we can define a corresponding semigroup $\mathcal{P}(t,s)$ and infinitesimal generator $\mathcal{A}(t)$. For a sufficiently nice function $\varphi : \mathbb{R}^d \to \mathbb{R}$, we have

$$\mathcal{P}(t,s)\varphi(x) := \mathbb{E}[\varphi(X_s)|X_t = x], \qquad 0 \le t \le s < \infty,$$
$$\mathcal{A}(t)\varphi(x) := \lim_{s \searrow t} \frac{1}{s-t}\Big(\mathcal{P}(t,s)\varphi(x) - \varphi(x)\Big).$$

THEOREM 7.3.3. *Suppose* X *is given by SDE* (7.17). *Then*

$$\mathcal{A}(t) = \sum_{i=1}^{d} \mu_i(t,x)\partial_{x_i} + \tfrac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}(\sigma\sigma^T)_{i,j}(t,x)\partial_{x_i}\partial_{x_j}$$
$$+ \sum_{j=1}^{k}\int_{\mathbb{R}^m} \nu^{(j)}(dz)\Big(\theta_{\gamma^{(\cdot,j)}(t,x,z)} - 1 - \gamma^{(\cdot,j)}(t,x,z)\cdot\nabla\Big),$$

where θ_γ *is the* shift operator*:* $\theta_\gamma f(x) := f(x+\gamma)$ *and* $C_0^2(\mathbb{R}^d) \in dom(\mathcal{A}(t))$.

PROOF. The proof is the same as in the no-jump case. First, write $\varphi(X_s) = \varphi(x) + \int_t^s d\varphi(X_u)$. Next, use the Itô formula to write $d\varphi(X_u)$ explicitly as terms involving du, dW_u and $\widetilde{N}(du,dz)$. Finally, take an expectation, note that integrals with respect to W and $\widetilde{N}$ are martingales, and send $s \searrow t$. We omit the details. □

We can now write the Itô formula for the solution X of Lévy-Itô SDE (7.17) in a more compact form

$$d\varphi(X_t) = \mathcal{A}(t)\varphi(X_t)dt + \nabla\varphi\cdot\sigma(t,X_t)dW_t + \int_{\mathbb{R}^d}\Big(\varphi(X_{t-}+\gamma(t,X_{t-},z)) - \varphi(X_{t-})\Big)\widetilde{N}(dt,dz).$$

THEOREM 7.3.4. *Let* $X = (X_t)_{t\ge 0}$ *satisfy*

$$dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dW_t + \int_{\mathbb{R}^d}\gamma(t,X_{t-},z)\widetilde{N}(dt,dz).$$

Then the function $u(t,x) := \mathbb{E}[\varphi(X_T)|X_t = x]$ *satisfies the* Kolmogorov Backward Equation

$$(\partial_t + \mathcal{A}(t))u(t,\cdot) = 0, \qquad u(T,\cdot) = \varphi.$$

PROOF. The proof is exactly analogous to the diffusion case. First, show that $\mathbb{E}[\varphi(X_T)|\mathcal{F}_t] = u(t,X_t)$ is a martingale. Take the differential of $u(t,X_t)$ and set the dt-term equal to zero to obtain the partial integro-differential equation that must be satisfied by u. The terminal condition is obtained from $\mathbb{E}[\varphi(X_T)|\mathcal{F}_T] = \varphi(X_T) = u(T,X_T)$. □

EXAMPLE 7.3.5. Suppose $X = (X_t)_{t\ge 0}$ is the following one-dimensional Lévy process

$$dX_t = \mu dt + \sigma dW_t + \int_{\mathbb{R}} z\widetilde{N}(dt,dz),$$

where W is a one-dimensional Brownian motion and $\widetilde{N}$ is a one-dimensional compensated Poisson random measure on $\mathbb{R}$. We wish to find an expression for $u(t,x) := \mathbb{E}[\varphi(X_T)|X_t = x]$, which satisfies the KBE; in this case

$$(\partial_t + \mathcal{A})u(t,\cdot) = 0, \qquad u(T,\cdot) = \varphi, \qquad \mathcal{A} = \mu\partial_x + \tfrac{1}{2}\sigma^2\partial_x^2 + \int_{\mathbb{R}} \nu(dz)(\theta_z - 1 - z\partial_x).$$

It will be useful to define the *Fourier* and *inverse Fourier transofrms*

$$\begin{aligned} &\text{Fourier Tranform :} && \mathbf{F}[\varphi](\xi) = \widehat{\varphi}(\xi) := \int_{\mathbb{R}} e^{-i\xi x}\varphi(x)dx, \\ &\text{Inverse Tranform :} && \mathbf{F}^{-1}[\widehat{\varphi}](x) := \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\xi x}\widehat{\varphi}(\xi)d\xi. \end{aligned}$$

Now, note that

$$\begin{aligned} \mathbf{F}[\partial_t u(t,\cdot)](\xi) &= \int_{\mathbb{R}} e^{-i\xi x}\partial_t u(t,x)dx = \partial_t \int_{\mathbb{R}} e^{-i\xi x}u(t,x)dx = \partial_t \widehat{u}(t,\xi) \\ \mathbf{F}[\mathcal{A}u(t,\cdot)](\xi) &= \int_{\mathbb{R}} e^{-i\xi x}\mathcal{A}u(t,x)dx = \int_{\mathbb{R}} u(t,x)\mathcal{A}^* e^{-i\xi x}dx \\ &= \int_{\mathbb{R}} u(t,x)\psi(\xi)e^{-i\xi x}dx = \psi(\xi)\widehat{u}(t,\xi). \end{aligned}$$

where $\mathcal{A}^*$ is the $L^2(\mathbb{R},dx)$ adjoint of $\mathcal{A}$ and $\psi(\xi)$ is the characteristic exponent of X. Specifically

$$\begin{aligned} \mathcal{A}^* &= -\mu\partial_x + \tfrac{1}{2}\sigma^2\partial_x^2 + \int_{\mathbb{R}} \nu(dz)(\theta_{-z} - 1 + z\partial_x), \\ \psi(\xi) &= \mu i\xi - \tfrac{1}{2}\sigma^2\xi^2 + \int_{\mathbb{R}} \nu(dz)(e^{i\xi z} - 1 - i\xi z), \end{aligned}$$

(check that you can derive $\mathcal{A}^*$ and $\psi(\xi)$ for yourself!). Taking the Fourier transform of the KBE and using the above results, we obtain an ODE in t for $\widehat{u}(t,\xi)$

$$(\partial_t + \psi(\xi))\widehat{u}(t,\xi) = 0, \qquad \widehat{u}(T,\xi) = \widehat{\varphi}(\xi).$$

The solution of this ODE is

$$\widehat{u}(t,\xi) = e^{(T-t)\psi(\xi)}\widehat{\varphi}(\xi).$$

Finally, we obtain u as the inverse Fourier transform of $\widehat{u}$

$$u(t,x) = \mathbf{F}^{-1}[\widehat{u}(t,\cdot)](x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\xi x + (T-t)\psi(\xi)}\widehat{\varphi}(\xi).$$

where, we have assumed, for simplicity, that $\varphi = \mathbf{F}^{-1}[\widehat{\varphi}]$. This is usually satisfied if φ is continuous and in $L^1(\mathbb{R},dx)$.

7.4 Change of measure

Assumption 7.4.1. Throughout Section 7.4, W is a one-dimensional Brownian motion and $\widetilde{N}$ is a one-dimensional compensated Poisson random measure on $\mathbb{R}$. All processes are scalar, unless specifically stated otherwise.

Review of the Girsanov change of measure for an Itô process

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration on this space. In Section 1.7 we showed that if Z is a random variable that satisfies

$$Z \geq 0, \qquad \mathbb{E}Z = 1,$$

Then we can define a new probability measure $\widetilde{\mathbb{P}}$ using

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}Z\mathbb{1}_A, \qquad \forall\, A \in \mathcal{F},$$

and we call $Z = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}$ the *Radon-Nikodým derivative of* $\widetilde{\mathbb{P}}$ *with respect to* $\mathbb{P}$. Moreover, if $Z > 0$, then probabilities under $\mathbb{P}$ can be obtained from probabilities under $\widetilde{\mathbb{P}}$ using

$$\mathbb{P}(A) = \widetilde{\mathbb{E}}\frac{1}{Z}\mathbb{1}_A, \qquad \forall\, A \in \mathcal{F},$$

and thus, we identify $\frac{1}{Z} = \frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}$ as the Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$.

In Chapter 4, we defined a *Radon-Nikodým derivative process* (Definition 4.4.1) by fixing a time horizon T and setting

$$Z_t := \mathbb{E}[Z|\mathcal{F}_t], \qquad 0 \leq t \leq T.$$

We showed (Lemma 4.4.4) that

$$\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] = \frac{1}{Z_s}\mathbb{E}[Z_t Y|\mathcal{F}_s], \qquad Y \in \mathcal{F}_t, \qquad 0 \leq s \leq t \leq T.$$

In particular if a process $Y = (Y_t)_{0\leq t\leq T}$ is adapted to $(\mathcal{F}_t)_{0\leq t\leq T}$ and $(Y_t Z_t)_{0\leq t\leq T}$ is a martingale under $\mathbb{P}$ then the process $(Y_t)_{0\leq t\leq T}$ is a martingale under $\widetilde{\mathbb{P}}$ since

$$\widetilde{\mathbb{E}}[Y_t|\mathcal{F}_s] = \frac{1}{Z_s}\mathbb{E}[Z_t Y_t|\mathcal{F}_s] = \frac{1}{Z_s}Z_s Y_s = Y_s. \tag{7.18}$$

As a specific example of this machinery, we showed (Girsanov's Theorem 4.4.5) that if we defined

$$Z = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = Z_T, \qquad Z_t = \exp\Big(-\int_0^t \tfrac{1}{2}\Theta_s^2 ds - \int_0^t \Theta_s dW_s\Big),$$

then, under $\widetilde{\mathbb{P}}$, the process

$$\widetilde{W}_t = W_t + \int_0^t \Theta_s ds, \qquad 0 \le t \le T,$$

is a martingale (in fact, a Brownian motion). We are now going to apply this machinary to a Lévy-Itô process.

Girsanov change of measure for Lévy-Itô processes

As changes of measure are more complicated for Lévy-Itô processes than they are for continuous processes, we present Girsanov's Theorem in three parts.

Theorem 7.4.2 (Girsanov Theorem for Lévy-Itô process (part I)). *On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ be a filtration and let $X = (X_t)_{0\le t\le T}$ be an Lévy-Itô process of the form*

$$dX_t = \mu_t dt + \sigma_t dW_t + \int_{\mathbb{R}} \gamma_t(z)\widetilde{N}(dt, dz), \tag{7.19}$$

where W is a Brownian motion and $\widetilde{N}$ is a compensated Poisson random meaure with associated Lévy measure ν. Suppose there exist $\mathbb{F}$-adapted processes $\beta = (\beta_t)_{0\le t\le T}$ and $\eta(z) = (\eta_t(z))_{0\le t\le T}$ such that

$$\mu_t = \sigma_t\beta_t - \int_{\mathbb{R}} \gamma_t(z)(e^{\eta_t(z)} - 1)\nu(dz). \tag{7.20}$$

Define a process $(Z_t)_{0\le t\le T}$ by

$$\begin{aligned} Z_t &:= \exp\Big(\int_0^t \alpha_s ds - \int_0^t \beta_s dW_s + \int_0^t \int_{\mathbb{R}} \eta_t(z)\widetilde{N}(dt, dz)\Big), \\ \alpha_t &= -\tfrac{1}{2}\beta_t^2 - \int_{\mathbb{R}} \left(e^{\eta_t(z)} - 1 - \eta_t(z)\right)\nu(dz), \end{aligned} \tag{7.21}$$

and assume that (β, η) are such that $\mathbb{E}Z_T < \infty$. Set $Z = \frac{d\widehat{\mathbb{P}}}{d\mathbb{P}} = Z_T$. Then X is a local martingale under $\widehat{\mathbb{P}}$.

Proof. First, we observe that $(Z_t)_{0\le t\le T}$ is a martingale. To see this, simply note that

$$dZ_t = Z_{t-}\Big(-\beta_s dW_s + \int_{\mathbb{R}} \left(e^{\eta_t(z)} - 1\right)\widetilde{N}(dt, dz)\Big).$$

Thus, we have

$$1 = Z_0 = \mathbb{E}Z_T = \mathbb{E}Z, \qquad Z_t = \mathbb{E}[Z_T|\mathcal{F}_t] = \mathbb{E}[Z|\mathcal{F}_t].$$

Since, $\mathbb{E}Z = 1$ and, by construction, $Z > 0$, the random variable Z defines a Radon-Nikodým derivative $\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}$. Also, since $Z_t = \mathbb{E}[Z|\mathcal{F}]$ we see that $(Z_t)_{0\le t\le T}$ is a Radon-Nikodým derivative process. Now, note that $(X_t)_{0\le t\le T}$ is adapted to the filtration $\mathbb{F}$. In light of (7.18), to show that $(X_t)_{0\le t\le T}$ is a martingale under $\widehat{\mathbb{P}}$ we need only to show that $(X_tZ_t)_{0\le t\le T}$ is a martingale under $\mathbb{P}$. To this end we compute

$$\begin{aligned}
d(X_tZ_t) &= X_t dZ_t + Z_t dX_t + d[X,Z]_t \\
&= X_t dZ_t + Z_{t-}\Big(\mu_t dt + \sigma_t dW_t + \int_{\mathbb{R}} \gamma_t(z)\tilde{N}(dt,dz)\Big) \\
&\quad + Z_{t-}\Big(-\sigma_t\beta_t + \int_{\mathbb{R}} \gamma_t(z)\left(e^{\eta_t(z)}-1\right)\nu(dz)\Big)dt + Z_{t-}\int_{\mathbb{R}} \gamma_t(z)\left(e^{\eta_t(z)}-1\right)\tilde{N}(dt,dz) \\
&= X_t dZ_t + Z_{t-}\Big(\sigma_t dW_t + \int_{\mathbb{R}} \gamma_t(z)\tilde{N}(dt,dz)\Big) \\
&\quad + Z_{t-}\int_{\mathbb{R}} \gamma_t(z)\left(e^{\eta_t(z)}-1\right)\tilde{N}(dt,dz).
\end{aligned}$$

Since integrals with respect to dZ_t, dW_t and $\tilde{N}(dt,dz)$ are martingales under $\mathbb{P}$, we see that $(X_tZ_t)_{t\ge 0}$ is a martingale under $\mathbb{P}$. And thus, $(X_t)_{0\le t\le T}$ is a martingale under $\widehat{\mathbb{P}}$. □

THEOREM 7.4.3 (GIRSANOV THEOREM FOR LÉVY-ITÔ PROCESS (PART II)). *On a probability space $(\Omega,\mathcal{F},\mathbb{P})$, let $Z = \frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}$ and $(Z_t)_{t\ge 0}$ be as defined in Theorem 7.4.2. Define*

$$d\widehat{W}_t := \beta_t dt + dW_t, \tag{7.22}$$

$$\hat{N}(dt,dz) = \tilde{N}(dt,dz) - \left(e^{\eta_t(z)}-1\right)\nu(dz)dt = N(dt,dz) - e^{\eta_t(z)}\nu(dz)dt. \tag{7.23}$$

Then $\widehat{W}$ is a Brownian motion under $\widehat{\mathbb{P}}$ and $\hat{N}$ is a compensated Poisson random measure under $\widehat{\mathbb{P}}$ in the sense that the process $M = (M_t)_{0\le t\le T}$, defined by

$$M_t := \int_0^t \int_{\mathbb{R}} \gamma_t(z)\hat{N}(dt,dz),$$

is a martingale for all $\mathcal{F}_t$-adapted processes $(\gamma_t(z))_{0\le t\le T}$ satisfying

$$\int_0^T \int_{\mathbb{R}} \gamma_t^2(z)e^{\eta_t(z)}\nu(dz) < \infty.$$

PROOF. For any $\varepsilon \in [0,1]$ we define

$$X_t^\varepsilon := \varepsilon\widehat{W}_t + M_t.$$

Writing all terms explicitly, we have

$$\begin{aligned}
dX_t^\varepsilon &= \varepsilon d\widehat{W}_t + \int_{\mathbb{R}} \gamma_t(z)\hat{N}(dt,dz) \\
&= \mu_t^\varepsilon dt + \varepsilon dW_t + \int_{\mathbb{R}} \gamma_t(z)\tilde{N}(dt,dz),
\end{aligned}$$

$$\mu_t^\varepsilon = \varepsilon\beta_t - \int_{\mathbb{R}} \gamma_t(z)\left(e^{\eta_t(z)} - 1\right)\nu(dz).$$

By Theorem 7.4.2, we see that, for all $\varepsilon \in [0,1]$ the process X^ε is a martingale under $\widehat{\mathbb{P}}$. In particular $X^0 = M$ is a martingale under $\widehat{\mathbb{P}}$, as claimed. Next, note that

$$X_t^1 - M_t = \widehat{W}_t,$$

is a martingale under $\widehat{\mathbb{P}}$ (since X^1 and M are martingales under $\widehat{\mathbb{P}}$). Moreover, $\widehat{W}$ is continuous and $[\widehat{W},\widehat{W}]_t = t$. Thus by Lévy's characterization of Brownian motion (Theorem 4.3.6), we conclude that $\widehat{W}$ is a Brownian motion under $\widehat{\mathbb{P}}$. □

THEOREM 7.4.4 (GIRSANOV THEOREM FOR LÉVY-ITÔ PROCESS (PART III)). *On a probability space $(\Omega,\mathcal{F},\mathbb{P})$, equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$, let X and $Z = \frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}$ be as in equations (7.19) and (7.21), respectively. Then X has the representation*

$$dX_t = \sigma_t d\widehat{W}_t + \int_{\mathbb{R}} \gamma_t(z)\widehat{N}(dt,dz), \tag{7.24}$$

where $\widehat{W}$ and $\widehat{N}$ are defined in (7.22) and (7.23) and X is a martingale under $\widehat{\mathbb{P}}$.

PROOF. To show that X has the representation (7.24) we simply note that

$$\begin{aligned} dX_t &= \mu_t dt + \sigma_t dW_t + \int_{\mathbb{R}} \gamma_t(z)\tilde{N}(dt,dz) \\ &= \mu_t dt + \sigma_t\left(d\widehat{W}_t - \beta_t dt\right) + \int_{\mathbb{R}} \gamma_t(z)\left(\widehat{N}(dt,dz) + \left(e^{\eta_t(z)} - 1\right)\nu(dz)dt\right) \\ &= \sigma_t d\widehat{W}_t + \int_{\mathbb{R}} \gamma_t(z)\widehat{N}(dt,dz) \end{aligned}$$

where, in the second line, we used (7.22) and (7.23), and in the last line we used (7.20). Since $\widehat{W}$ is a Brownian motion under $\widehat{\mathbb{P}}$ and $\widehat{N}$ is a compensated Poisson measure under $\widehat{\mathbb{P}}$ with associated Lévy measure $e^{\eta_t(z)}\nu(dz)$, it follows that X is a martingale under $\widehat{\mathbb{P}}$. □

EXAMPLE 7.4.5. Consider a pure-jump Lévy process $X = (X_t)_{0\le t\le T}$

$$dX_t = \mu dt + \int_{\mathbb{R}} z\tilde{N}(dt,dz), \qquad \tilde{N}(dt,dz) = N(dt,dz) - \nu(dz)dt.$$

We wish to find a change of measure $Z = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}$ such that, under $\widehat{\mathbb{P}}$ jumps of size dz arrive with intensity $\widehat{\nu}(dz)$. From Theorem 7.4.2, equation (7.21), and the time-homogeneity of X, we see that Z must be of the form

$$Z_t := \exp\left(\int_0^t \alpha ds + \int_0^t\int_{\mathbb{R}} \eta(z)\tilde{N}(ds,dz)\right), \qquad \alpha = -\int_{\mathbb{R}}\left(e^{\eta(z)} - 1 - \eta(z)\right)\nu(dz). \tag{7.25}$$

With Z given by (7.25) we have that

$$\widehat{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z)=\widetilde{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z)-\left(\mathrm{e}^{\eta_t(z)}-1\right)\nu(\mathrm{d}z)\mathrm{d}t=\mathrm{N}(\mathrm{d}t,\mathrm{d}z)-\mathrm{e}^{\eta_t(z)}\nu(\mathrm{d}z)\mathrm{d}t.$$

is a compensated Poisson random measure. We want

$$\mathrm{e}^{\eta_t(z)}\nu(\mathrm{d}z)=\widehat{\nu}(\mathrm{d}z).$$

Thus, we identify

$$\eta(z)=\log\frac{\widehat{\nu}(\mathrm{d}z)}{\nu(\mathrm{d}z)}.$$

Note that, in order for η to be well-defined we must have $0<\frac{\widehat{\nu}(\mathrm{d}z)}{\nu(\mathrm{d}z)}<\infty$, which holds if an only if $\widehat{\nu}\sim\nu$ (i.e., the measures are equivalent). Re-writing the dynamics of X as follows

$$\begin{aligned}
\mathrm{d}\mathrm{X}_t&=\mu\mathrm{d}t+\int_{\mathbb{R}}z\left(\widehat{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z)+\left(\mathrm{e}^{\eta_t(z)}-1\right)\nu(\mathrm{d}z)\mathrm{d}t\right)\\
&=\mu\mathrm{d}t+\int_{\mathbb{R}}z\left(\widehat{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z)+\left(\frac{\widehat{\nu}(\mathrm{d}z)}{\nu(\mathrm{d}z)}-1\right)\nu(\mathrm{d}z)\mathrm{d}t\right)\\
&=\mu\mathrm{d}t+\int_{\mathbb{R}}z\left(\widehat{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z)+\left(\widehat{\nu}(\mathrm{d}z)-\nu(\mathrm{d}z)\right)\mathrm{d}t\right)\\
&=\widehat{\mu}\mathrm{d}t+\int_{\mathbb{R}}z\widehat{\mathrm{N}}(\mathrm{d}t,\mathrm{d}z)\\
\widehat{\mu}&=\mu+\int_{\mathbb{R}}z\left(\widehat{\nu}(\mathrm{d}z)-\nu(\mathrm{d}z)\right),
\end{aligned}$$

we see that X will be a martingale under $\widehat{\mathbb{P}}$ if and only if $\widehat{\mu}=0$. Suppose, for example, that $\nu=\lambda\mathrm{F}$, where $\lambda=\nu(\mathbb{R})$ is the intensity of all jumps (assumed finite) and $\mathrm{F}(\mathrm{d}z)=\nu(\mathrm{d}z)/\nu(\mathbb{R})$ is the jump distribution. If we wish to change only the intensity of jumps (say, from λ to $\widehat{\lambda}$) and not the distribution of jumps F, then, if we want X to be a martingale under $\widehat{\mathbb{P}}$ we must have

$$\widehat{\lambda}-\lambda=\mu\cdot\left(\int_{\mathbb{R}}z\mathrm{F}(\mathrm{d}z)\right)^{-1}.$$

Note that, had we included a Brownian component in the $\mathbb{P}$ dynamics of X, we would have complete freedom to change the jump measure since any non-zero drift could be absorbed into the drift of the Brownian motion under the change of measure.

7.5 Hawkes processes

Recall that if $\mathrm{P}=(\mathrm{P}_t)_{t\geq0}$ is a Poisson process with intensity λ, then we have

$$\mathbb{E}[\mathrm{d}\mathrm{P}_t|\mathcal{F}_t]=\mathbb{E}\mathrm{d}\mathrm{P}_t=\lambda\mathrm{d}t.$$

Note that the probability of a jump $\mathbb{P}(\mathrm{dP}_t = 1) = \mathbb{E}\mathrm{dP}_t$ in the instant $\mathrm{d}t$ is entirely *independent* of this history $\mathcal{F}_t$ of P. This memorylessness property is convenient from a computational point of view and is necessary in order for P to be Markov. But, from a modeling perspective, limiting our analysis to Poisson processes (or more generally, Poisson random measures) is very restrictive. For example, if we want to model the arrival of earthquakes in Seattle, it is well-known that the occurrence of an earthquake today will increase the probability of an earthquake tomorrow. Thus, it would be useful to have some way of modeling events whose arrival intensities are history-*dependent*. This is precisely what a Hawkes process will allow us to do.

DEFINITION 7.5.1. A *Hawkes process* $\mathrm{N} = (\mathrm{N}_t)_{t\geq 0}$ is a simple counting process satisfying

$$\mathbb{E}[\mathrm{dN}_t|\mathcal{F}_t] = \lambda_t \mathrm{d}t, \qquad \lambda_t := \Lambda\Big(\int_0^{t-} h(t-s)\mathrm{dN}_s\Big), \tag{7.26}$$

where $h(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is the *kernel* (sometimes also called an *exciting function*) and $\Lambda(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is locally integrable. The intensity λ_t is an $\mathcal{F}_t$-measurable random variable where $\mathcal{F}_t$ is the natural filtration for N. When the function $\Lambda(\cdot)$ is linear, the process N_t is known as the *linear Hawkes process*, otherwise, it is called a *nonlinear Hawkes process*.

If we denote by $\tau_1, \tau_2, \ldots$ the (random) times that N jumps, then the intensity at time t is given by

$$\lambda_t = \Lambda\Big(\sum_{\tau_i < t} h(t-\tau_i)\Big).$$

Typically, h is a strictly decreasing function and Λ is increasing. Under these assumptions, the intensity λ_t is decreasing over the intervals of the form $[\tau_i, \tau_{i+1})$.

For reason of analytic tractability, in what follows, we concentrate on the linear Hawkes processes. Specifically, let us assume

$$\Lambda(z) = \nu + z,$$

where $\nu > 0$ is a strictly positive constant. Thus, from (7.34), we have

$$\lambda_t = \nu + \int_0^{t-} h(t-s)\mathrm{dN}_s \tag{7.27}$$

$$= \nu + \int_0^{t-} h(t-s)\mathrm{d}\widetilde{\mathrm{N}}_s + \int_0^{t-} h(t-s)\lambda_s \mathrm{d}s, \tag{7.28}$$

where we have introduced the *compensated Hawkes process*, whose differential is given by

$$\mathrm{d}\widetilde{\mathrm{N}}_s = \mathrm{dN}_s - \lambda_s \mathrm{d}s.$$

As, by construction, the expected infinitesimal change of $\tilde{N}$ is zero

$$\mathbb{E}[d\tilde{N}_t|\mathcal{F}_t] = \mathbb{E}[dN_t|\mathcal{F}_t] - \lambda_t dt = 0,$$

it follows that the compensated Hawkes process is a martingale

$$\mathbb{E}[\tilde{N}_t|\mathcal{F}_s] = \int_s^r \mathbb{E}[\mathbb{E}(d\tilde{N}_u|\mathcal{F}_u)|\mathcal{F}_s] + \tilde{N}_s = \tilde{N}_s, \qquad 0 \le s \le t < \infty.$$

Thus, integrals with respect to the compensated Hawkes process are martingales as well.

Suppose the intensity λ_t of a Hawkes process approaches a stationary distribution as $t \to \infty$. This distribution must have a mean, which we shall call $\bar{\lambda}$. If we start the process from time $-\infty$, which is equivalent to looking at the behavior of the process as $t \to \infty$, then we have $\mathbb{E}\lambda_t = \bar{\lambda}$ for all $t > -\infty$ and thus, taking an expectation of (7.28), we find

$$\begin{aligned}
\mathbb{E}\lambda_t &= \nu + \int_{-\infty}^{t-} h(t-s)\mathbb{E}d\tilde{N}_s + \int_{-\infty}^{t-} h(t-s)\mathbb{E}\lambda_s ds \\
&= \nu + \int_{-\infty}^{t-} h(t-s)\mathbb{E}\lambda_s ds, \\
\bar{\lambda} &= \nu + \bar{\lambda}\int_0^\infty h(s)ds.
\end{aligned}$$

Hence, we find that the mean $\bar{\lambda}$ of the stationary distribution of the intensity is given by

$$\bar{\lambda} = \nu \cdot \left(1 - \int_0^\infty h(s)ds\right)^{-1}.$$

Clearly, a stationary distribution only exists if h satisfies

$$\int_0^\infty h(s)ds < 1. \tag{7.29}$$

Thus, some texts *require* that h satisfy (7.29)

In general, a given Hawkes process $N = (N_t)_{t\ge 0}$ and the associated intensity process $\lambda = (\lambda_t)_{t\ge 0}$ are non-Markovian. However, when the kernel $h(\cdot)$ has an exponential form

$$h(t) = ae^{-bt}, \qquad a, b > 0, \tag{7.30}$$

then the intensity process λ by itself and the pair (λ, N) are Markov. To see this, observe from (7.27) and (7.30) that

$$\lambda_t = \nu + \int_0^{t-} ae^{-b(t-s)}dN_t.$$

Taking the differential of λ_t we find

$$d\lambda_t = -b\left(\int_0^{t-} ae^{-b(t-s)}dN_t\right)dt + adN_t$$

$$
\begin{aligned}
&= -b(\lambda_t - \nu)\mathrm{d}t + a\mathrm{d}\mathrm{N}_t, \\
&= -b(\lambda_t - \nu)\mathrm{d}t + a\mathrm{d}\widetilde{\mathrm{N}}_t + a\lambda_t \mathrm{d}t \\
&= \Big((a-b)\lambda_t + b\nu\Big)\mathrm{d}t + a\mathrm{d}\widetilde{\mathrm{N}}_t.
\end{aligned}
$$

Since the intensity of N_t is λ_t, we see that the future dynamics of λ depend only on the value of λ_t and not the entire history $\mathcal{F}_t$. Thus, the process λ is Markov and its generator is

$$
\begin{aligned}
\mathcal{A} &= \Big((a-b)\lambda + b\nu\Big)\partial_\lambda + \lambda\left(\theta_a - 1 - a\partial_\lambda\right) \\
&= b(\nu - \lambda)\partial_\lambda + \lambda\left(\theta_a - 1\right),
\end{aligned}
$$

where, we remind the reader that θ_a is the shift operator: $\theta_a f(\lambda) = f(\lambda + a)$. Since λ alone is Markov, it follows that the pair (λ, N) is Markov and has generator

$$
\mathcal{A} = b(\nu - \lambda)\partial_\lambda + \lambda\left(\theta_{(a,1)} - 1\right), \tag{7.31}
$$

where $\theta_{(a,1)} f(\lambda, n) = f(\lambda + a, n + 1)$.

We cannot find the probability mass function of N_T. However, we can find its Laplace transform $\mathbb{E}\mathrm{e}^{-\eta \mathrm{N}_\mathrm{T}}$. Let us define

$$
u(t, \lambda, n) = \mathbb{E}[\mathrm{e}^{-\eta \mathrm{N}_\mathrm{T}} | \lambda_t = \lambda, \mathrm{N}_t = n].
$$

The function u satisfies the Kolmogorov backward equation

$$
0 = (\partial_t + \mathcal{A})u, \qquad\qquad u(\mathrm{T}, \lambda, n) = \mathrm{e}^{-\eta n}. \tag{7.32}
$$

where $\mathcal{A}$ is given by (7.31). We seek a solution to (7.32) in exponential affine form

$$
u(t, \lambda, n) = \mathrm{e}^{-\eta n + \mathrm{A}(t) + \mathrm{B}(t)\lambda}. \tag{7.33}
$$

Inserting (7.33) into (7.32), we obtain

$$
0 = u(t, \lambda, n)\Big((\mathrm{A}'(t) + \lambda \mathrm{B}'(t)) + b(\nu - \lambda)\mathrm{B}(t) + \lambda(\mathrm{e}^{-\eta + a\mathrm{B}(t)} - 1)\Big), \qquad \mathrm{e}^{-\eta n + \mathrm{A}(\mathrm{T}) + \mathrm{B}(\mathrm{T})\lambda} = \mathrm{e}^{-\eta n}.
$$

This must hold for all λ. Thus we collect terms of like order in λ and obtain a pair of coupled ODEs for A and B

$$
\begin{aligned}
&\mathcal{O}(\lambda^0): && 0 = \mathrm{A}'(t) + b\nu\mathrm{B}(t), && \mathrm{A}(\mathrm{T}) = 0, \\
&\mathcal{O}(\lambda^1): && 0 = \mathrm{B}'(t) - b\mathrm{B}(t) + (\mathrm{e}^{-\eta + a\mathrm{B}(t)} - 1), && \mathrm{B}(\mathrm{T}) = 0.
\end{aligned}
$$

These ODEs must be solved numerically.

We finish this Section by defining a d-dimensional Hawkes process.

DEFINITION 7.5.2. A *d-dimensional Hawkes process* $\mathrm{N} = (\mathrm{N}_t^{(1)}, \mathrm{N}_t^{(2)}, \ldots, \mathrm{N}_t^{(d)})_{t\geq 0}$ is a simple counting process satisfying

$$\mathbb{E}[\mathrm{dN}_t^{(i)}|\mathcal{F}_t] = \lambda_t^{(i)}\mathrm{d}t, \qquad \lambda_t^{(i)} := \Lambda^{(i)}\Big(\sum_{j=1}^{d}\int_0^{t-} h^{(i,j)}(t-s)\mathrm{dN}_s^{(j)}\Big), \tag{7.34}$$

where $h^{(i,j)}(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ and $\Lambda^{(i)}(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is locally integrable. The intensity $\lambda_t^{(i)}$ is an $\mathcal{F}_t$-measurable random variable where $\mathcal{F}_t$ is the natural filtration for the d-dimenional process $(\mathrm{N}^{(i)})_{i=1}^d$. When the function $\Lambda^{(i)}(\cdot)$ is linear for all i, the process N_t is known as the *linear d-dimensional Hawkes process*, otherwise, it is called a *nonlinear d-dimensional Hawkes process*.

As in the one-dimensional case, linear d-dimensional Hawkes processes with exponential kernels are Markov and admit many analytically tractable results.

7.6 EXERCISES

EXERCISE 7.1. Consider a financial market defined on a probability space $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$ consisting of two assets: a stock S and a money market account M with

$$\mathrm{M}_t = \mathrm{M}_0\mathrm{e}^{rt}, \qquad \mathrm{S}_t = \mathrm{S}_0\mathrm{e}^{\mathrm{X}_t}, \qquad \mathrm{X}_t = \mu t + \sigma\widetilde{\mathrm{W}}_t + \int_{\mathbb{R}} z\widetilde{\mathrm{N}}(t, \mathrm{d}z),$$

where $\widetilde{\mathrm{W}}$ is a Brownian motion under $\widetilde{\mathbb{P}}$ and $\widetilde{\mathrm{N}}(\mathrm{d}t, \mathrm{d}z) = \mathrm{N}(\mathrm{d}t, \mathrm{d}z) - \nu(\mathrm{d}z)\mathrm{d}t$ with $\widetilde{\mathbb{E}}\mathrm{N}(\mathrm{d}t, \mathrm{d}z) = \nu(\mathrm{d}z)\mathrm{d}t$.
(a) Suppose that $\widetilde{\mathbb{P}}$ is the market's pricing measure with the money market account M as numéraire. Then S/M must be a martingale under $\widetilde{\mathbb{P}}$. Find μ.
(b) Let V_t be the price of a derivative asset that pays $\mathrm{S}_{\mathrm{T}}^p$ at time T. Show that $\mathrm{V}_t = v(t, \mathrm{S}_t)$ and find the function v.

Bibliography

Björk, T. (2004). *Arbitrage theory in continuous time*. Oxford university press.

Grimmett, G. and D. Stirzaker (2001). *Probability and random processes*. Oxford university press.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*(1), 83–90.

Karlin, S. and H. Taylor (1981). *A second course in stochastic processes*. Academic Press.

Linetsky, V. (2007). Chapter 6 spectral methods in derivatives pricing. In J. R. Birge and V. Linetsky (Eds.), *Financial Engineering*, Volume 15 of *Handbooks in Operations Research and Management Science*, pp. 223 – 299. Elsevier.

Øksendal, B. (2005). *Stochastic Differential Equations: An Introduction with Applications* (6 ed.). Springer-Verlag.

Øksendal, B. and A. Sulem (2005). *Applied stochastic control of jump diffusions*. Springer Verlag.

Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*, Volume 11. Springer Science & Business Media.

van Handel, R. (2007). *Stochastic Calculus, Filtering, and Stochastic Control*.