MA400. September Introductory Course (Financial Mathematics and Quantitative Methods for Risk Management)

Mihail Zervos

September 2020

CONTENTS

CHAPTER

1

PROBABILITY SPACES

1.1 Preliminary considerations

1. A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ that can be described informally as follows:

   - Ω is the sample space. We can think of Ω as the set of all possible outcomes in "nature" or in a "random experiment" that we want to model. In this context, "nature" chooses exactly one point $\omega \in \Omega$, but we do not know which one, otherwise, we would have no uncertainty and we would know exactly what is going to happen.
   - $\mathcal{F}$ is a collection of event of interests. An event is a subset of Ω, so $\mathcal{F}$ is a set of subsets of Ω. We can think of $\mathcal{F}$ as all the *information* that "nature" has or all the *information* that is relevant to the modelling of a "random experiment".
   - $\mathbb{P}$ is a function that assigns a probability $\mathbb{P}(A)$ to each event $A \in \mathcal{F}$. In particular, given an event $A \in \mathcal{F}$, $\mathbb{P}(A)$ is a number in the interval $[0, 1]$ that represents our belief on how likely the event A is to occur.

   Mathematically, a probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ such that

   - Ω is a set,
   - $\mathcal{F}$ is a σ-algebra on Ω (see Definition 1.5 below), and
   - $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$ (see Definition 1.20 below).

2. **Example.** Consider tossing a coin that lands heads with probability $p \in (0,1)$ twice. In this context, we can choose the sample space, which is the set of all possible outcomes, to be the set $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, where, e.g.,

   ω_1 identifies with observing heads first and then heads,
   ω_2 identifies with observing heads first and then tails,
   ω_3 identifies with observing tails first and then heads, and
   ω_4 identifies with observing tails first and then tails.

   The family of all events of interest that can arise in this random experiment is the set

   $$\begin{aligned}\mathcal{F} = \big\{&\Omega, \emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\}, \\ &\{\omega_2, \omega_3\}, \{\omega_2, \omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}\big\}.\end{aligned}$$

   In fact, the elements of this set have a simple description in everyday language. For instance, $\{\omega_1, \omega_2\}$ is the event that we observe heads in the first toss, $\{\omega_2\}$ is the event that the coin lands heads first and then tails, Ω is the event that we observe something, and $\emptyset$ is the event that we observe nothing.

   Based on everyday intuition, we can assign a probability $\mathbb{P}(A)$ to each event $A \in \mathcal{F}$ in a consistent way, so that, e.g., $\mathbb{P}\big(\{\omega_3, \omega_4\}\big) = 1 - p$, while $\mathbb{P}\big(\{\omega_1\}\big) = p^2$.

3. **Example.** Consider drawing a number from the interval $(0,1)$ in a completely random way. In this case, we can identify the sample space Ω with $(0,1)$, and every subset A of Ω is an event: A identifies with the event that the number that we draw happens to be in the set A. Given any $a, b \in (0,1)$ such that $a < b$, intuition suggests that the probability of the event (a,b) is $b - a$, because the number that we draw is equally likely to be anywhere in $(0,1)$. In light of this simple observation, any event (i.e., subset of Ω) should have probability equal to its "length".

   The question that thus arises is: can we assign a length to every subset of $(0,1)$? The answer is no: it is not possible to assign a length to every subsets of $(0,1)$ in a consistent way. As a result, we cannot assign a probability to every subset of $\Omega \equiv (0,1)$. To develop a meaningful theory, we therefore need to restrict our attention to those subsets of Ω that do have a well-defined length.

   This example illustrates why we need to consider families $\mathcal{F}$ of events of "interest" (in the context of this example, such families should include only events that do have a well-defined length). Is this a serious restriction? Not really: it turns out that we can always choose an appropriate collection $\mathcal{F}$ of events of "interest" that contains every event of practical interest.

1.2 A subset of $(0,1)$ to which we can assign no length

4. **Example**. Suppose that we can assign a length to every subset of the real line, and denote by $L(A)$ the length of the set $A \subseteq \mathbb{R}$, so that, e.g.,

$$L\big((a,b)\big) = b - a, \quad L\big(\{a\}\big) = 0 \quad \text{and} \quad L\big((-\infty, a)\big) = L\big((a,\infty)\big) = \infty \tag{1.1}$$

for all real numbers $a < b$. Intuition suggests that the length function L should be positive, i.e., $L(A) \geq 0$ for all $A \subseteq \mathbb{R}$, increasing in the sense that, given any sets $A, B \subseteq \mathbb{R}$,

$$A \subseteq B \quad \Rightarrow \quad L(A) \leq L(B), \tag{1.2}$$

and countably additive, so that, if (A_n) is a sequence of pairwise disjoint subsets of $\mathbb{R}$, i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$L\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} L(A_n). \tag{1.3}$$

Also, the length of a set should be translation invariant, so that

$$L(A^a) = L(A) \quad \text{for all } A \subseteq \mathbb{R} \text{ and } a \in \mathbb{R}, \tag{1.4}$$

where A^a is the translation of A by a, which is defined by $A^a = \{a + x \mid \ x \in A\}$.

Now, we consider the equivalence relation $\sim$ on the real line defined by

$$x \sim y \quad \text{if} \quad x - y \in \mathbb{Q},$$

and split the interval $(0,1)$ in equivalence classes. In this context, the numbers $x, y \in (0,1)$ belong to the same equivalence class if and only if $x \sim y$, i.e., if and only if $x - y \in \mathbb{Q}$, while, if the numbers $x, y \in (0,1)$ belong to different equivalence classes, then $x \not\sim y$, i.e., $x - y \notin \mathbb{Q}$. Also, the equivalence classes are pairwise disjoint, so each number in $(0,1)$ belongs to exactly one equivalence class.

By appealing to the axiom of choice, we next consider a set C that contains exactly one representative from each equivalence class. Since C contains only one point from each equivalence class, any distinct points $x, y \in C$ belong to different equivalence classes, so $x \not\sim y$. Furthermore, given any point $z \in (0,1)$, if x is the representative in C of the equivalence class in which z belongs, then $z \sim x$, so there exists $q \in \mathbb{Q}$ such that $z = q + x$.

In view of these observations, it follows that, if we define

$$C^q = \{q + x \mid \ x \in C\}\,, \quad \text{for } q \in (-1,1) \cap \mathbb{Q},$$

then

$$C^{q_1} \cap C^{q_2} = \emptyset \quad \text{for all } q_1 \neq q_2, \tag{1.5}$$

and

$$(0,1) \subseteq \bigcup_{q\in(-1,1)\cap\mathbb{Q}} C^q \subseteq (-1,2). \tag{1.6}$$

Now, we argue by contradiction to conclude that the set C has no length. If $L(C) = 0$, then

$$1 \overset{(1.1)}{=} L\big((0,1)\big) \overset{(1.2),(1.6)}{\leq} L\left(\bigcup_{q\in(-1,1)\cap\mathbb{Q}} C^q\right) \overset{(1.3),(1.5)}{=} \sum_{q\in(-1,1)\cap\mathbb{Q}} L(C^q) \overset{(1.4)}{=} \sum_{q\in(-1,1)\cap\mathbb{Q}} L(C) = 0,$$

which is not possible. So, if $L(C)$ exists, we must have $L(C) > 0$ because L is a positive function. In this case, we can see that

$$3 \overset{(1.1)}{=} L\big((-1,2)\big) \overset{(1.2),(1.6)}{\geq} L\left(\bigcup_{q\in(-1,1)\cap\mathbb{Q}} C^q\right) \overset{(1.3),(1.5)}{=} \sum_{q\in(-1,1)\cap\mathbb{Q}} L(C^q) \overset{(1.4)}{=} \sum_{q\in(-1,1)\cap\mathbb{Q}} L(C) = \infty,$$

which cannot be true. We conclude that $L(C)$ does not exist.

1.3 σ-algebras

5. **Definition.** A *σ-algebra* on Ω is a collection $\mathcal{F}$ of subsets of Ω such that

$$\begin{aligned}
&\text{(i) } \Omega \in \mathcal{F},\\
&\text{(ii) } A \in \mathcal{F} \quad \Rightarrow \quad A^c \equiv \Omega \setminus A \in \mathcal{F},\\
&\text{(iii) } A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F} \quad \Rightarrow \quad \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.
\end{aligned}$$

6. **Example.** The *power set* $\mathcal{P}(\Omega)$ of any set Ω, namely, the collection of all subsets of Ω, is a σ-algebra on Ω.

7. **Lemma.** Given a set Ω and a σ-algebra $\mathcal{F}$ on Ω,

$$\emptyset \in \mathcal{F}, \tag{1.7}$$

$$\text{and} \quad A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F} \ \Rightarrow \ \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}. \tag{1.8}$$

In particular, *a σ-algebra is stable under countable set operations.*

Proof. Since $\emptyset = \Omega^c$, (1.7) follows immediately by properties (i) and (ii) of Definition 1.5. To prove (1.8), we consider any sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$, and we observe that

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c .$$

The event appearing on the right hand side of this expression belongs to $\mathcal{F}$ because

$$\begin{aligned}
A_n \in \mathcal{F} \text{ for all } n \geq 1 \quad &\Rightarrow \quad A_n^c \in \mathcal{F} \text{ for all } n \geq 1 && \text{(by property 5.(ii))} \\
&\Rightarrow \quad \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F} && \text{(by property 5.(iii))} \\
&\Rightarrow \quad \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{F} && \text{(by property 5.(ii))}
\end{aligned}$$

and (1.8) follows.

8. **Lemma.** Let $\{\mathcal{F}_i, \ i \in I\}$ be a family of σ-algebras on Ω indexed by a set $I \neq \emptyset$. The collection $\bigcap_{i \in I} \mathcal{F}_i$ is a σ-algebra on Ω.

Proof. We have to check the defining properties of a σ-algebra. To this end, we note that the family of events $\bigcap_{i \in I} \mathcal{F}_i$ satisfies property (iii) of Definition 1.5 because

$$\begin{aligned}
A_1, A_2, \ldots, &A_n, \ldots \in \bigcap_{i \in I} \mathcal{F}_i \\
&\Rightarrow \quad A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}_i \text{ for all } i \in I \\
&\Rightarrow \quad \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_i \text{ for all } i \in I \quad \text{(because each } \mathcal{F}_i \text{ is a } \sigma\text{-algebra)} \\
&\Rightarrow \quad \bigcup_{n=1}^{\infty} A_n \in \bigcap_{i \in I} \mathcal{F}_i .
\end{aligned}$$

Similarly, we can verify properties (i) and (ii) of Definition 1.5.

9. Given two σ-algebras $\mathcal{F}$ and $\mathcal{G}$, the collection of events $\mathcal{F} \cup \mathcal{G}$ is *not necessarily* a σ-algebra. To see this, it suffices to consider an example such as the following.

   Example. Suppose that $\Omega = \{1, 2, 3, 4\}$, and let

$$\begin{aligned}\mathcal{F} =& \big\{\Omega, \emptyset, \{1, 2\}, \{3, 4\}\big\},\\ \mathcal{G} =& \big\{\Omega, \emptyset, \{1\}, \{2, 3, 4\}\big\}.\end{aligned}$$

   Then

$$\mathcal{F} \cup \mathcal{G} = \big\{\Omega, \emptyset, \{1, 2\}, \{3, 4\}, \{1\}, \{2, 3, 4\}\big\}$$

   is *not* a σ-algebra. To see this, consider the events $\{3, 4\}$ and $\{1\}$, which both belong to $\mathcal{F} \cup \mathcal{G}$, and observe that

$$\{3, 4\} \cup \{1\} = \{1, 3, 4\} \notin \mathcal{F} \cup \mathcal{G}.$$

10. **Definition.** Given a collection $\mathcal{C}$ of subsets of Ω, the σ-algebra $\sigma(\mathcal{C})$ on Ω *generated by* $\mathcal{C}$ is the *smallest* σ-algebra on Ω containing $\mathcal{C}$. It is the intersection of all σ-algebras on Ω which have $\mathcal{C}$ as a subclass.

11. Observe that, if $\mathcal{C}$ is a family of sets and $\mathcal{H}$ is a σ-algebra, then

$$\mathcal{C} \subseteq \mathcal{H} \quad \Rightarrow \quad \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{H}) = \mathcal{H},$$

   because, by definition, $\sigma(\mathcal{C})$ is the intersection of all σ-algebras containing $\mathcal{C}$. In other words, $\sigma(\mathcal{C})$ is a subset of *any* σ-algebra containing $\mathcal{C}$.

12. **Example.** Given a set $A \subseteq \Omega$, the smallest σ-algebra containing A is $\{\Omega, \emptyset, A, A^c\}$.

13. **Example.** Suppose that $\Omega = \{1, 2, 3, 4\}$, and let

$$\mathcal{C} = \big\{\{1\}, \{1, 3, 4\}\big\}.$$

   Then

$$\sigma(\mathcal{C}) = \big\{\Omega, \emptyset, \{1\}, \{1, 3, 4\}, \{2, 3, 4\}, \{2\}, \{1, 2\}, \{3, 4\}\big\}.$$

14. **Example.** Suppose that $\Omega = \mathbb{R}$, and let

$$\mathcal{C} = \big\{(-2, 6), [0, \sqrt{3})\big\}.$$

   In this case,

$$\sigma(\mathcal{C}) = \big\{\mathbb{R}, \emptyset, A, B, C, A \cup B, A \cup C, B \cup C\big\}.$$

   where

$$A = (-\infty, -2] \cup [6, \infty), \quad B = (-2, 0) \cup [\sqrt{3}, 6) \quad \text{and} \quad C = [0, \sqrt{3}).$$

15. **Definition.** The *Borel* σ-algebra $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ is the σ-algebra on $\mathbb{R}$ generated by the family of all open intervals (a,b), i.e.,

$$\mathcal{B}(\mathbb{R}) = \sigma\big(\{(a,b) \mid \ a,b \in \mathbb{R},\ a<b\}\big).$$

More generally, consider any topological space S. The *Borel* σ-algebra $\mathcal{B}(S)$ on S is the σ-algebra on S generated by the family of all open sets, i.e.,

$$\mathcal{B}(S) = \sigma\big(\{A \subset S \mid \ A \text{ is open}\}\big).$$

The Borel σ-algebra is very important: it contains every subset of $\mathbb{R}$ that is of practical interest!

16. **Example.** $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$, where

$$\mathcal{C} = \big\{(-\infty,a] \mid \ a \in \mathbb{R}\big\}.$$

Proof. In view of the Definition 1.15 of the Borel σ-algebra on $\mathbb{R}$ and the observation in Paragraph 1.11 above, we can prove this claim as follows.

(i) $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C})$ will follow if we show that $(a,b) \in \sigma(\mathcal{C})$ for all real numbers $a<b$. This is true because

$$\begin{aligned}(a,b) &= (a,\infty) \cap (-\infty,b)\\ &= (-\infty,a]^c \cap \bigcup_{n=1}^{\infty} (-\infty, b-\tfrac{1}{n}]\,.\end{aligned}$$

(ii) $\mathcal{B}(\mathbb{R}) \supseteq \sigma(\mathcal{C})$ will follow if we show that $(-\infty,a] \in \mathcal{B}(\mathbb{R})$ for every real number a. This follows from the observation that

$$\begin{aligned}(-\infty,a] &= \bigcup_{m=1}^{\infty} (a-m,a]\\ &= \bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} (a-m, a+\tfrac{1}{n})\,.\end{aligned}$$

1.4 (Probability) measures

17. **Definition.** A pair $(\Omega, \mathcal{F})$, where Ω is a set and $\mathcal{F}$ is a σ-algebra on Ω, is called *measurable space*.

18. **Definition.** Let $(S, \mathcal{S})$ be a measurable space, so that $\mathcal{S}$ is a σ-algebra on the set S. A *measure* defined on $(S, \mathcal{S})$ is a function $\mu : \mathcal{S} \to [0, \infty]$ that is *countably additive*, i.e., it is such that

    (i) $\mu(\emptyset) = 0$, and

    (ii) if $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{S}$ is any sequence of pairwise disjoint sets (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then
    $$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

    The triplet $(S, \mathcal{S}, \mu)$ is then called a *measure space*.

19. **Definition.** Given a measure space $(S, \mathcal{S}, \mu)$, we say that

    μ is a *probability* measure if $\mu(S) = 1$,

    μ is a *finite* measure if $\mu(S) < \infty$, and

    μ is a *σ-finite* measure if there is a sequence $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{S}$ such that
    $$\mu(A_n) < \infty \text{ for all } n \geq 1 \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n = S.$$

    In this course, we will consider *only* σ-finite measures.

20. Due to its particular interest, we repeat the definition of a probability measure:

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

    (i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$, and

    (ii) if $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$ is any sequence of pairwise disjoint events (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then
    $$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

21. **Lemma.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Given any $A, B \in \mathcal{F}$,

$$\text{if } A \subseteq B, \text{ then } \mathbb{P}(A) \leq \mathbb{P}(B), \tag{1.9}$$
$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \tag{1.10}$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \tag{1.11}$$
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i). \tag{1.12}$$

Proof. Given any events $A \subseteq B$,

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A \cup (B \setminus A)) \\ &= \mathbb{P}(A) + \mathbb{P}(B \setminus A) \\ &\geq \mathbb{P}(A), \end{aligned}$$

and (1.9) follows. Also, (1.10) follows immediately from the calculations

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

Given any events A and B, if we define

$$K = A \cap B^c, \qquad L = A \cap B, \qquad M = A^c \cap B,$$

then K, L, M are pairwise disjoint,

$$A = K \cup L \quad \text{and} \quad B = L \cup M.$$

As a consequence,

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(K \cup L \cup M) \\ &= \mathbb{P}(K) + \mathbb{P}(L) + \mathbb{P}(M) \\ &= \mathbb{P}(K) + \mathbb{P}(L) + \mathbb{P}(M) + \mathbb{P}(L) - \mathbb{P}(L) \\ &= \mathbb{P}(K \cup L) + \mathbb{P}(M \cup L) - \mathbb{P}(L) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \end{aligned}$$

which proves (1.11). In view of (1.11) and the positivity of probabilities,

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

Using this inequality and a straightforward induction argument, we obtain (1.12).

22. **Lemma ("Continuity" of a measure).** Let $(S, \mathcal{S}, \mu)$ be a measure space. Given an *increasing sequence* $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ of events in $\mathcal{S}$, we can define the *limit* of the sequence by

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

In this context,

$$\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n). \tag{1.13}$$

Similarly, if $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ is a *decreasing* sequence of events in $\mathcal{S}$, the *limit* of the sequence is defined by

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

In this case, if $\mu(A_1) < \infty$, then

$$\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n). \tag{1.14}$$

Proof. Given an increasing sequence $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ of events in $\mathcal{S}$, let $B_1 = A_1$, and define recursively $B_n = A_n \setminus A_{n-1}$, for $n \geq 2$. By construction, the events $B_1, B_2, \ldots, B_n, \ldots$ are pairwise disjoint,

$$A_n = \bigcup_{k=1}^{n} B_k \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n = \bigcup_{k=1}^{\infty} B_k.$$

As a consequence,

$$\begin{aligned}
\mu\left(\lim_{n\to\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \\
&= \mu\left(\bigcup_{k=1}^{\infty} B_k\right) \\
&= \sum_{k=1}^{\infty} \mu(B_k) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} B_k\right) \\
&= \lim_{n\to\infty} \mu(A_n).
\end{aligned}$$

Consider any decreasing sequence $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ of events in $\mathcal{S}$ such that $\mu(A_1) < \infty$. Since $\emptyset \subseteq A_1 \setminus A_2 \subseteq \cdots \subseteq A_1 \setminus A_n \subseteq \cdots$,

$$\mu\left(\bigcup_{n=1}^{\infty} A_1 \setminus A_n\right) = \lim_{n\to\infty} \mu(A_1 \setminus A_n).$$

Noting that

$$\bigcup_{n=1}^{\infty} A_1 \setminus A_n = A_1 \setminus \bigcap_{n=1}^{\infty} A_n,$$

we can see that this implies that

$$\mu(A_1) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \big[\mu(A_1) - \mu(A_n)\big],$$

which establishes (1.14).

23. In the previous result, the validity of (1.14) relies heavily on the assumption $\mu(A_1) < \infty$ (in fact, on the assumption that $\mu(A_k) < \infty$, for some $k \geq 1$). To appreciate this claim, we consider the following example.

    Example. Suppose that $S = \mathbb{R}$, $\mathcal{S} = \mathcal{B}(\mathbb{R})$ and $\mu = L$, where L is the Lebesgue measure that maps each set $C \in \mathcal{B}(\mathbb{R})$ to its length $L(C)$. If we define $A_n = [n, \infty)$, for $n \geq 1$, then we can see that

$$\mu\left(\lim_{n\to\infty} A_n\right) = L\left(\bigcap_{n=1}^{\infty} [n,\infty)\right) = L(\emptyset) = 0 < \infty = \lim_{n\to\infty} L\big([n,\infty)\big) = \lim_{n\to\infty} \mu(A_n).$$

CHAPTER

2

RANDOM VARIABLES AND DISTRIBUTION FUNCTIONS

2.1 Random variables

1. Consider the random choice of a person from among N people. Assuming that all people in the group are equally likely to be chosen,

$$\Omega = \{\omega_1, \ldots, \omega_N\}, \quad \mathcal{F} = \mathcal{P}(\Omega) \quad \text{and} \quad \mathbb{P}\big(\{\omega_i\}\big) = \frac{1}{N}, \quad \text{for } i = 1, \ldots, N,$$

where ω_i is the i-th representative of the group and $\mathcal{P}(\Omega)$ is the power set of Ω (i.e., the set of all subsets of Ω), provide an appropriate probability space.

There are many quantities that can be associated with this probability space. For example, each individual $\omega \in \Omega$ is associated with their height $X(\omega)$, their weight $Y(\omega)$ or their blood type $Z(\omega)$. Each of these quantities is a *random variable*. The random variables X and Y take values in the set of positive real numbers, while the random variable Z takes values in the set of all possible blood types.

Since mathematical modelling involves mathematical objects, we concentrate our attention on random variables that take values in a space of mathematical objects such as, e.g., the real numbers $\mathbb{R}$ or the Euclidean space $\mathbb{R}^n$.

After the random choice has been made, the value of every random variable is known. On the other hand, *before* the random choice happens, every random variable is a *function* on Ω with values in the appropriate space: each individual $\omega \in \Omega$ is associated with a height $X(\omega)$, a weight $Y(\omega)$ and a blood type $Z(\omega)$.

2. Generalising the example above, a real-valued "random variable" X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a *function* mapping Ω into $\mathbb{R}$. Accordingly, each sample $\omega \in \Omega$ is associated with a unique $X(\omega) \in \mathbb{R}$.

   We view random variables as functions on the sample space Ω rather than identify them with their eventually observed value because **probability theory is concerned with the future**.

3. The distribution of a "random variable" X is of fundamental importance. In particular, we are naturally interested in knowing the probability of X taking values in a given set A. For instance, we are interested in knowing the probability of the events

$$X^{-1}(A) = \{X \in A\} = \big\{\omega \in \Omega \mid X(\omega) \in A\big\}, \quad \text{for } A \in \mathcal{B}(\mathbb{R}),$$

   or

$$X^{-1}\big((-\infty, a]\big) = \{X \leq a\} = \big\{X \in (-\infty, a]\big\} = \big\{\omega \in \Omega \mid X(\omega) \in (-\infty, a]\big\}, \quad \text{for } a \in \mathbb{R}.$$

   Since $\mathbb{P}(C)$ is well-defined only for events $C \in \mathcal{F}$, these probabilities will be well-defined only if the relevant events are in $\mathcal{F}$, which gives rise to the requirement (2.1) of the following definition.

4. **Definition.** A *real-valued random variable* X is any function $X : \Omega \to \mathbb{R}$ such that

$$X^{-1}(A) = \{X \in A\} = \big\{\omega \in \Omega \mid X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{B}(\mathbb{R}), \tag{2.1}$$

   where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

5. **Definition.** Given a measurable space $(S, \mathcal{S})$, an $(S, \mathcal{S})$-*valued random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ is a function mapping Ω into S such that

$$X^{-1}(A) = \{X \in A\} = \big\{\omega \in \Omega \mid X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{S}. \tag{2.2}$$

6. **Lemma.** Consider two measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$, and a family of sets $\mathcal{C}$ such that $\sigma(\mathcal{C}) = \mathcal{S}$. If a function $X : \Omega \to S$ satisfies

$$\{X \in A\} = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for all } A \in \mathcal{C}, \tag{2.3}$$

then X is an $(S, \mathcal{S})$-valued random variable.

Proof. We will prove that X is an $(S, \mathcal{S})$-valued random variable if we show that

$$\{X \in A\} \in \mathcal{F} \quad \text{for all } A \in \mathcal{S},$$

or equivalently, if we show that

$$\Big\{A \in \mathcal{S} \mid \ \{X \in A\} \in \mathcal{F}\Big\} = \mathcal{S}. \tag{2.4}$$

To this end, we define

$$\mathcal{H} = \Big\{A \in \mathcal{S} \mid \ \{X \in A\} \in \mathcal{F}\Big\},$$

and we note that

$$\mathcal{C} \subseteq \mathcal{H} \subseteq \mathcal{S}, \tag{2.5}$$

where the first inclusion follows thanks to (2.3).

Furthermore, we note that $\mathcal{H}$ is a σ-algebra on S, because:

(i) $S \in \mathcal{H}$ because $\{X \in S\} = \Omega \in \mathcal{F}$.

(ii) Given an event $A \in \mathcal{H}$,

$$\{X \in S \setminus A\} = \Omega \setminus \{X \in A\} \in \mathcal{F},$$

so, $S \setminus A \in \mathcal{H}$.

(iii) Given a sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{H}$,

$$\left\{X \in \bigcup_{n=1}^{\infty} A_n\right\} = \bigcup_{n=1}^{\infty} \{X \in A_n\} \in \mathcal{F},$$

so, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$.

Now, in view of the assumption that $\sigma(\mathcal{C}) = \mathcal{S}$, (2.5), and the fact that $\mathcal{H}$, $\mathcal{S}$ are σ-algebras on S, we can see that

$$\mathcal{S} = \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{H}) = \mathcal{H} \subseteq \mathcal{S},$$

which proves that $\mathcal{H} = \mathcal{S}$, and establishes (2.4).

7. **Lemma.** Suppose that X and Y are real-valued random variables defined on a measurable space $(\Omega, \mathcal{F})$, and let λ be a real number. Then, $X + Y$, XY and λX are all real-valued random variables.

Proof. In view of Lemma 2.6 and the fact that the family of sets

$$\mathcal{C}_1 = \big\{(a, \infty) \mid \ a \in \mathbb{R}\big\}$$

generates the Borel σ-algebra, i.e., $\sigma(\mathcal{C}_1) = \mathcal{B}(\mathbb{R})$, we will prove that the sum $X + Y$ of two random variables X and Y is also a random variable if we show that

$$\big\{X + Y > a\big\} = \big\{\omega \in \Omega \mid \ X(\omega) + Y(\omega) > a\big\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}. \tag{2.6}$$

To this end, we note that, given any $\omega \in \Omega$ and any $a \in \mathbb{R}$, $X(\omega) > a - Y(\omega)$ if and only if we can find a rational number q such that $X(\omega) > q > a - Y(\omega)$. Therefore,

$$\begin{aligned}\big\{\omega \in \Omega \mid \ X(\omega) + Y(\omega) > a\big\} &= \bigcup_{q \in \mathbb{Q}} \big\{\omega \in \Omega \mid \ X(\omega) > q > a - Y(\omega)\big\} \\ &= \bigcup_{q \in \mathbb{Q}} \Big(\big\{\omega \in \Omega \mid \ X(\omega) > q\big\} \cap \big\{\omega \in \Omega \mid \ Y(\omega) > a - q\big\}\Big).\end{aligned}$$

However, the expression on the right hand side of this expression is a *countable* union of events in $\mathcal{F}$ (because X and Y are random variables), and (2.6) follows.

Now, we use Lemma 2.6 and the fact that the family of sets

$$\mathcal{C}_2 = \big\{(-\infty, a] \mid \ a \in \mathbb{R}\big\}$$

generates the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ to show that, given a constant $\lambda \in \mathbb{R}$ and a random variable X, the function λX mapping Ω into $\mathbb{R}$ is a random variable by proving that

$$\{\lambda X \leq a\} = \big\{\omega \in \Omega \mid \ \lambda X(\omega) \leq a\big\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}. \tag{2.7}$$

Indeed, given any $a \in \mathbb{R}$,

$$\big\{\omega \in \Omega \mid \ \lambda X(\omega) \leq a\big\} = \begin{cases} \big\{\omega \in \Omega \mid \ X(\omega) \leq a/\lambda\big\}, & \text{if } \lambda > 0, \\ \emptyset, & \text{if } \lambda = 0 \text{ and } a < 0, \\ \Omega, & \text{if } \lambda = 0 \text{ and } a \geq 0, \\ \big\{\omega \in \Omega \mid \ X(\omega) \geq a/\lambda\big\}, & \text{if } \lambda < 0. \end{cases}$$

All of the events on the right hand side of this expression belong to $\mathcal{F}$ (because X is a random variable), and (2.7) follows.

Similarly, if X is a random variable, then, given any $a \in \mathbb{R}$,

$$\{X^2 \le a\} = \big\{\omega \in \Omega \mid \ X^2(\omega) \le a\big\} = \begin{cases} \emptyset, & \text{if } a < 0, \\ \big\{\omega \in \Omega \mid \ X(\omega) \in [-\sqrt{a}, \sqrt{a}]\big\}, & \text{if } a \ge 0. \end{cases}$$

Since either of the two events appearing on the right hand side of this expression belong to $\mathcal{F}$ (because X is a random variable), it follows that X^2 is a random variable.

Using what we have proved up to now, we can see that, given any random variables X and Y, the product XY is also a random variable because the identity

$$XY = \frac{1}{2}(X+Y)^2 - \frac{1}{2}X^2 - \frac{1}{2}Y^2$$

expresses XY as a sum of random variables.

8. **Lemma.** Suppose that $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of real-valued random variables defined on a measurable space $(\Omega, \mathcal{F})$. The functions

$$\inf_{n \ge 1} X_n, \quad \sup_{n \ge 1} X_n, \quad \liminf_{n \to \infty} X_n \quad \text{and} \quad \limsup_{n \to \infty} X_n$$

mapping Ω into $[-\infty, \infty]$, defined by

$$\left(\inf_{n \ge 1} X_n\right)(\omega) = \inf_{n \ge 1} X_n(\omega), \quad \left(\sup_{n \ge 1} X_n\right)(\omega) = \sup_{n \ge 1} X_n(\omega),$$
$$\left(\liminf_{n \to \infty} X_n\right)(\omega) = \liminf_{n \to \infty} X_n(\omega) \quad \text{and} \quad \left(\limsup_{n \to \infty} X_n\right)(\omega) = \limsup_{n \to \infty} X_n(\omega),$$

respectively, are $\big([-\infty, \infty], \mathcal{B}([-\infty, \infty])\big)$-valued random variables, where $\mathcal{B}\big([-\infty, \infty]\big)$ is the Borel σ-algebra on $[-\infty, \infty]$, so that

$$\mathcal{B}\big([-\infty, \infty]\big) = \sigma\Big(\big\{[-\infty, a] \mid \ a \in [-\infty, \infty]\big\}\Big) \supseteq \mathcal{B}(\mathbb{R}). \tag{2.8}$$

Furthermore,

$$\Big\{\omega \in \Omega \mid \ \lim_{n \to \infty} X_n(\omega) \text{ exists in } \mathbb{R}\Big\} \in \mathcal{F}. \tag{2.9}$$

Proof. In view of Lemma 2.6 and (2.8), we can see that the inclusion

$$\left\{\sup_{n \ge 1} X_n \le a\right\} = \left\{\omega \in \Omega \mid \ \sup_{n \ge 1} X_n(\omega) \le a\right\}$$
$$= \bigcap_{n=1}^{\infty} \{\omega \in \Omega \mid \ X_n(\omega) \le a\} \in \mathcal{F} \quad \text{for all } a \in [-\infty, \infty],$$

implies that $\sup_{n \ge 1} X_n$ is an $\big([-\infty, \infty], \mathcal{B}([-\infty, \infty])\big)$-valued random variable.

Recalling that, if Z is a random variable, then $-Z$ is also a random variable (see Lemma 2.7), we can see that the result we have just proved and the identity

$$\inf_{n\geq 1} X_n = -\sup_{n\geq 1}(-X_n)$$

imply that $\inf_{n\geq 1} X_n$ is an $\big([-\infty,\infty],\mathcal{B}([-\infty,\infty])\big)$-valued random variable.

If we define

$$\underline{Z}_n = \inf_{k\geq n} X_k \quad \text{and} \quad \overline{Z}_n = \sup_{k\geq n} X_k, \quad \text{for } n\geq 1,$$

then $\underline{Z}_n$ and $\overline{Z}_n$ are $\big([-\infty,\infty],\mathcal{B}([-\infty,\infty])\big)$-valued random variables for all $n\geq 1$. It follows that

$$\liminf_{n\to\infty} X_n = \lim_{n\to\infty}\inf_{k\geq n} X_k = \sup_{n\geq 1}\inf_{k\geq n} X_n = \sup_{n\geq 1}\underline{Z}_n$$

and

$$\limsup_{n\to\infty} X_n = \lim_{n\to\infty}\sup_{k\geq n} X_k = \inf_{n\geq 1}\sup_{k\geq n} X_k = \inf_{n\geq 1}\overline{Z}_n$$

are $\big([-\infty,\infty],\mathcal{B}([-\infty,\infty])\big)$-valued random variables.

Finally, we note that (2.9) follows immediately from the identity

$$\begin{aligned}&\Big\{\omega\in\Omega \mid \lim_{n\to\infty} X_n(\omega) \text{ exists in } \mathbb{R}\Big\}\\ &\quad = \Big\{\omega\in\Omega \mid \limsup_{n\to\infty} X_n(\omega) < \infty\Big\} \cap \Big\{\omega\in\Omega \mid \liminf_{n\to\infty} X_n(\omega) > -\infty\Big\}\\ &\qquad \cap \Big\{\omega\in\Omega \mid \Big(\limsup_{n\to\infty} X_n - \liminf_{n\to\infty} X_n\Big)(\omega) = 0\Big\}\end{aligned}$$

and the fact that the events on the right-hand side of this expression belong to $\mathcal{F}$.

2.2 σ-algebras generated by random variables

9. **Definition.** The σ-algebra $\sigma(X)$ generated by a real-valued random variable X, namely, the information set $\sigma(X)$ associated with the observation of X, is the σ-algebra defined by

$$\sigma(X) = \big\{\{X\in A\} \mid A\in\mathcal{B}(\mathbb{R})\big\}. \tag{2.10}$$

10. **Definition.** The σ-algebra $\sigma(X)$ generated by an $(S,\mathcal{S})$-valued random variable X, namely, the information set $\sigma(X)$ associated with the observation of X, is the collection of all sets $\{X\in A\} = \big\{\omega\in\Omega \mid X(\omega)\in A\big\}$, for $A\in\mathcal{S}$, i.e.,

$$\sigma(X) = \big\{\{X\in A\} \mid A\in\mathcal{S}\big\}. \tag{2.11}$$

11. **Lemma.** The family of events $\sigma(X)$ defined by (2.10) is indeed a σ-algebra on Ω.

Proof. We use the fact that $\mathcal{B}(\mathbb{R})$ is a σ-algebra on $\mathbb{R}$ to check that $\sigma(X)$ satisfies the three properties that characterise a σ-algebra on Ω:

(i) $\Omega \in \sigma(X)$ because $\Omega = \{X \in \mathbb{R}\}$ and $\mathbb{R} \in \mathcal{B}(\mathbb{R})$.

(ii) Let any event $C \in \sigma(X)$. We need to show that $\Omega \setminus C \in \sigma(X)$.
To this end, observe that the definition (2.10) of $\sigma(X)$ implies that there exists $A \in \mathcal{B}(\mathbb{R})$ such that

$$C = \{X \in A\} \equiv \big\{\omega \in \Omega \mid \; X(\omega) \in A\big\}.$$

Now, we calculate

$$\begin{aligned}\Omega \setminus C &= \Omega \setminus \big\{\omega \in \Omega \mid \; X(\omega) \in A\big\} \\ &= \big\{\omega \in \Omega \mid \; X(\omega) \notin A\big\} \\ &= \big\{\omega \in \Omega \mid \; X(\omega) \in \mathbb{R} \setminus A\big\} = \{X \in \mathbb{R} \setminus A\} \in \sigma(X),\end{aligned}$$

because $\mathbb{R} \setminus A \in \mathcal{B}(\mathbb{R})$.

(iii) Consider any sequence of events $C_1, C_2, \ldots, C_n, \ldots \in \sigma(X)$. We need to prove that $\bigcup_{n=1}^{\infty} C_n \in \sigma(X)$.
Since $C_n \in \sigma(X)$ for all n, the definition (2.10) of $\sigma(X)$ implies that there exists a sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{B}(\mathbb{R})$ such that

$$C_n = \{X \in A_n\} \equiv \big\{\omega \in \Omega \mid \; X(\omega) \in A_n\big\} \quad \text{for all } n = 1, 2, \ldots.$$

Now, we calculate

$$\begin{aligned}\bigcup_{n=1}^{\infty} C_n &= \bigcup_{n=1}^{\infty} \big\{\omega \in \Omega \mid \; X(\omega) \in A_n\big\} \\ &= \left\{\omega \in \Omega \mid \; X(\omega) \in \bigcup_{n=1}^{\infty} A_n\right\} = \left\{X \in \bigcup_{n=1}^{\infty} A_n\right\} \; \in \; \sigma(X),\end{aligned}$$

because $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}(\mathbb{R})$.

12. *It is worth stressing that the information set $\sigma(X)$ is associated with the random variable X and* not *with its eventually observed value*. To appreciate this comment, we consider the following example.

Suppose that $\Omega = \{1, 2, 3, 4, 5\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$. Also, let $A = \{1, 2\}$ and let X be the random variable defined by

$$X(\omega) = \mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases} \quad \text{for } \omega \in \Omega.$$

We can check that, in this case,

$$\sigma(X) = \{\Omega, \emptyset, A, A^c\} = \big\{\Omega, \emptyset, \{1,2\}, \{3,4,5\}\big\}.$$

Before observing the actual value of X, we have certainty that we *will* be able to say whether each event in this information set has occurred or not as soon as we observe X. Furthermore, there is *no* event outside this information set for which we can have such a certainty.

13. **Definition.** The σ-algebra generated by a collection of random variables $(X_i,\ i \in I)$, where $I \neq \emptyset$, namely, the information we obtain by the observation of the random variables in the family $(X_i,\ i \in I)$, is the σ-algebra

$$\sigma\big(X_i,\ i \in I\big) = \sigma\big(\sigma(X_i),\ i \in I\big) \equiv \sigma\left(\bigcup_{i\in I} \sigma(X_i)\right).$$

14. **Definition.** Given a random variable X and a σ-algebra $\mathcal{H}$ on Ω, we say that X is *$\mathcal{H}$-measurable* if $\sigma(X) \subseteq \mathcal{H}$.

15. With the terminology introduced by this definition, note that:

   Given a random variable X, $\sigma(X)$ is the *smallest* σ-algebra with respect to which X is measurable.

   Given a family of random variables $(X_i,\ i \in I)$, $\sigma\big(X_i,\ i \in I\big)$ is the *smallest* σ-algebra with respect to which every X_i is measurable.

   Informally, this definition says that a random variable X is $\mathcal{H}$-measurable if the information provided by X is a subset of the information contained in $\mathcal{H}$.

2.3 Distributions

16. **Definition.** The distribution function F_X of a real-valued random variable X is defined by

$$F_X(a) = \mathbb{P}(X \le a) \equiv \mathbb{P}\big(X \in (-\infty, a]\big), \quad \text{for } a \in \mathbb{R}.$$

   Provided there is no possibility of confusion, we often write $F(a)$ instead of $F_X(a)$.

17. **Lemma.** The following are simple properties of distribution functions:

   (i) Every distribution function F is an increasing function.

   Proof. Observing that, given any $a \le b$,

$$\{\omega \in \Omega \mid\ X(\omega) \le a\} \subseteq \{\omega \in \Omega \mid\ X(\omega) \le b\},$$

we can see that

$$F(a) = \mathbb{P}(X \le a) \le \mathbb{P}(X \le b) = F(b).$$

Here, we have used the monotonicity of a probability measure: given any $A, B \in \mathcal{F}$,

$$A \subseteq B \quad \Rightarrow \quad \mathbb{P}(A) \le \mathbb{P}(B).$$

(ii) Every distribution function F satisfies

$$\lim_{a\to-\infty} F(a) = 0 \quad \text{and} \quad \lim_{a\to\infty} F(a) = 1.$$

Proof. Since F is an increasing function, both limits exist. Therefore, we only have to show that

$$\lim_{n\to\infty} F(-n) = 0 \quad \text{and} \quad \lim_{n\to\infty} F(n) = 1.$$

To this end, we first consider the decreasing sequence of events $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ defined by

$$A_n = \{\omega \in \Omega \mid \ X(\omega) \le -n\},$$

and we observe that $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Using the "continuity" of a probability measure, we can calculate

$$\lim_{n\to\infty} F(-n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathbb{P}(\emptyset) = 0.$$

Next, we consider the increasing sequence of events $B_1 \subseteq B_2 \subseteq \cdots \subseteq B_n \subseteq \cdots$ defined by

$$B_n = \{\omega \in \Omega \mid \ X(\omega) \le n\}.$$

and we observe that $\bigcup_{n=1}^{\infty} B_n = \Omega$. In view of the "continuity" of a probability measure, it follows that

$$\lim_{n\to\infty} F(n) = \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \mathbb{P}(\Omega) = 1.$$

(iii) Every distribution function F is right-continuous.

Proof. Since F is increasing, both of the limits $\lim_{x\downarrow a} F(x)$ and $\lim_{x\uparrow a} F(x)$ exist at every point $a \in \mathbb{R}$. Therefore, to see that F is right-continuous we observe that, given any $a \in \mathbb{R}$,

$$\lim_{n\to\infty} F\left(a + \frac{1}{n}\right) = \lim_{n\to\infty} \mathbb{P}\left(X \le a + \frac{1}{n}\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \left\{X \le a + \frac{1}{n}\right\}\right) = \mathbb{P}\left(X \le a\right) = F(a).$$

18. **Example.** Suppose that we roll a fair die once. Let X be the number we observe. The distribution of X is

$$F(a) = \begin{cases} 0, & \text{if } a < 1, \\ \frac{1}{6}, & \text{if } 1 \le a < 2, \\ \frac{2}{6}, & \text{if } 2 \le a < 3, \\ \frac{3}{6}, & \text{if } 3 \le a < 4, \\ \frac{4}{6}, & \text{if } 4 \le a < 5, \\ \frac{5}{6}, & \text{if } 5 \le a < 6, \\ 1, & \text{if } 6 \le a. \end{cases}$$

19. **Example.** The distribution function of a random variable X is given by

$$F(x) = \begin{cases} 0, & \text{if } -\infty < x < 0, \\ 1 - 0.5e^{-x}, & \text{if } 0 \le x. \end{cases}$$

We can compute

$$\begin{aligned} \mathbb{P}(X = 0) &= F(0) - F(0-) \\ &= 0.5, \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}(1 < X \le 2) &= \mathbb{P}(X \le 2) - \mathbb{P}(X \le 1) \\ &= F(2) - F(1) \\ &= 0.5\left(e^{-1} - e^{-2}\right). \end{aligned}$$

20. **Definition.** The joint distribution of n random variables $X_1, \ldots, X_n$ is defined to be

$$F_{X_1 \ldots X_n}(a_1, \ldots, a_n) = \mathbb{P}(X_1 \le a_1, \ldots, X_n \le a_n) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in (-\infty, a_i]\}\right).$$

2.4 Discrete random variables

21. **Definition.** A real-valued random variable X is *discrete* if it maps Ω into a countable subset of $\mathbb{R}$. The *probability mass function* of a discrete random variable X is the collection of all pairs (x_j, p_j) such that

$$p_j = \mathbb{P}\big(X = x_j\big) > 0. \tag{2.12}$$

22. In view of (2.12), the distribution function of a discrete random variable X is given by

$$F(a) = \sum_{j \text{ such that } x_j \leq a} p_j.$$

Also,

$$p_j = F(x_j) - F(x_j-).$$

23. **Example.** Given an event $A \in \mathcal{F}$, the random variable

$$X = \mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \quad (\text{“success”}), \\ 0, & \text{if } \omega \in A^c \quad (\text{“failure”}), \end{cases}$$

is called the *indicator* of A. The probability mass function of this random variable is given by

$$p = \mathbb{P}\big(X = 1\big) = \mathbb{P}\big(\mathbf{1}_A = 1\big) = \mathbb{P}(A) \quad \text{and} \quad 1 - p = \mathbb{P}\big(X = 0\big) = \mathbb{P}\big(\mathbf{1}_A = 0\big) = \mathbb{P}(A^c).$$

We say that such a random variable X is *Bernoulli* with parameter p.

24. **Example.** A discrete random variable X has the *binomial* distribution with parameters n, p if its probability mass function is characterised by

$$p_j \equiv \mathbb{P}\big(X = j\big) = \binom{n}{j} p^j (1-p)^{n-j}, \quad \text{for } j = 0, 1, \ldots, n,$$

where

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}.$$

Here, n is a positive integer and $p \in (0, 1)$. We often write $X \sim B(n, p)$,

Suppose that a coin that lands heads with probability p is tossed n times. If we define the random variable X to be the total number of heads observed in the n tosses, then X has the binomial distribution. More generally, the total number of “successes” in a fixed number of independent trials has the binomial distribution.

This interpretation reflects the fact that a random variable $X \sim B(n, p)$ has the same distribution as $X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables, each with parameter p.

25. **Example.** A random variable X has the *Poisson* distribution with parameter $\lambda > 0$ if its probability mass function is given by

$$p_n = \mathbb{P}\big(X = n\big) = e^{-\lambda}\frac{\lambda^n}{n!}, \quad n = 0, 1, \ldots.$$

2.5 Continuous random variables

26. **Definition.** A real-valued random variable X is *continuous* if there exists a function f, called the *probability density function* of X, such that

$$\mathbb{P}\big(X \in A\big) = \int_A f(x)\,dx \quad \text{for all } A \in \mathcal{B}(\mathbb{R}). \tag{2.13}$$

27. Since probabilities are positive, every probability density function f satisfies

$$f(x) \geq 0 \quad \text{for all } x \in \mathbb{R}.$$

Since $\mathbb{P}(\Omega) \equiv \mathbb{P}(X \in \mathbb{R}) = 1$, every probability density function f satisfies

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Also, observe that (2.13) implies

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b) = \int_a^b f(x)\,dx.$$

28. **Example.** A random variable X has the *uniform* distribution if its probability density function is given by

$$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b, \\ 0, & \text{if } x < a \text{ or } b < x, \end{cases}$$

for some constants $a < b$. Given such a random variable X, we often write $X \sim \mathcal{U}(a, b)$. We say that X has the *standard uniform* distribution if $a = 0$ and $b = 1$.

29. **Example.** A random variable X has the *exponential* distribution with parameter $\mu > 0$ if its probability density function is given by

$$f(x) = \begin{cases} \mu e^{-\mu x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

30. **Example.** A random variable X has the *normal* distribution with mean m and variance σ^2 if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Here, $m \in \mathbb{R}$ and $\sigma > 0$. Given such a random variable X, we often write $X \sim \mathcal{N}(m, \sigma^2)$. Normal random variables are also called *Gaussian*. Also, we say that X has the *standard normal* distribution if $m = 0$ and $\sigma = 1$.

The probability distribution function of a normal random variable satisfies

$$F(a) = \Phi\left(\frac{a-m}{\sigma}\right),$$

where Φ is the *standard normal distribution function* defined by

$$\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}}\, dx. \tag{2.14}$$

To see this, observe first that

$$F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

If we make the change of variables $y = (x-m)/\sigma$, then

$$\begin{aligned} F(a) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-m}{\sigma}} \exp\left(-\frac{y^2}{2}\right) dy. \\ &= \Phi\left(\frac{a-m}{\sigma}\right). \end{aligned} \tag{2.15}$$

31. **Definition.** n real-valued random variables $X_1, \ldots, X_n$ are said to be *jointly continuous* if there exists a function $f : \mathbb{R}^n \to \mathbb{R}$, called the *joint probability density function* of $X_1, \ldots, X_n$, such that

$$\mathbb{P}\big(X_1 \in A_1, \ldots, X_n \in A_n\big) = \int_{A_1} \cdots \int_{A_n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

for all $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$.

CHAPTER

3

INDEPENDENCE

1. Throughout the chapter, we assume that an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is fixed. Also, we assume that every σ-algebra that we consider is a σ-algebra on Ω that is a subset of $\mathcal{F}$.

3.1 Independence of σ-algebras, random variables and events

2. **Definition.** The σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are called *independent* if

$$\mathbb{P}\big(A_1 \cap A_2 \cap \cdots \cap A_n\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big) \cdots \mathbb{P}\big(A_n\big)$$

for every choice of events $A_1 \in \mathcal{G}_1$, $A_2 \in \mathcal{G}_2$, $\ldots$, $A_n \in \mathcal{G}_n$.

3. **Definition.** The σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n, \ldots$ are called *independent* if the σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent for all $n \geq 2$.

4. **Definition.** The random variables $X_1, X_2, \ldots, X_n, \ldots$ are called *independent* if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_n), \ldots$ are independent.

5. **Definition.** The events $A_1, A_2, \ldots, A_n, \ldots$ are called *independent* if the σ-algebras $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \ldots$ are independent, where

$$\mathcal{A}_n = \big\{\Omega, \emptyset, A_n, A_n^c\big\}, \quad \text{for } n \geq 1.$$

6. Recall that the indicator of an event A is defined by

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \in A^c, \end{cases}$$

and that $\sigma(\mathbf{1}_A) = \{\Omega, \emptyset, A, A^c\}$. As a consequence, the events $A_1, A_2, \ldots, A_n, \ldots$ are independent if and only if the random variables $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \ldots, \mathbf{1}_{A_n}, \ldots$ are independent, which is true if and only if the σ-algebras $\sigma(\mathbf{1}_{A_1}), \sigma(\mathbf{1}_{A_2}), \ldots, \sigma(\mathbf{1}_{A_n})$ are independent.

7. **Example.** Two events A_1, A_2 are independent if

$$\mathbb{P}\big(A_1 \cap A_2\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big), \tag{3.1}$$

Proof. To verify that the events A_1, A_2 are independent, we have to check that

$$\mathbb{P}(C \cap D) = \mathbb{P}(C)\mathbb{P}(D) \quad \text{for all } C \in \{\Omega, \emptyset, A_1, A_1^c\} \text{ and } D \in \{\Omega, \emptyset, A_2, A_2^c\}. \tag{3.2}$$

In other words, we have to prove that (3.1) implies each of the $4 \times 4 = 16$ relations in (3.2). To this end, we calculate

$$\begin{aligned} \mathbb{P}(A_1^c \cap A_2) &= \mathbb{P}(A_2 \setminus (A_1 \cap A_2)) \\ &= \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \\ &\overset{(3.1)}{=} \big[\mathbb{P}(A_1^c) + \mathbb{P}(A_1)\big]\mathbb{P}(A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2) \\ &= \mathbb{P}(A_1^c)\mathbb{P}(A_2). \end{aligned}$$

All other identities in (3.2) are now straightforward.

8. **Example.** Similarly, we can verify that three events A_1, A_2, A_3 are independent if all of the identities

$$\begin{aligned} \mathbb{P}\big(A_1 \cap A_2\big) &= \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big), \\ \mathbb{P}\big(A_1 \cap A_3\big) &= \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_3\big), \\ \mathbb{P}\big(A_2 \cap A_3\big) &= \mathbb{P}\big(A_2\big)\mathbb{P}\big(A_3\big), \\ \mathbb{P}\big(A_1 \cap A_2 \cap A_3\big) &= \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big)\mathbb{P}\big(A_3\big). \end{aligned}$$

hold true.

9. **Lemma.** Two random variables X and Y are independent if their joint distribution function F_{XY} can be factorised in the form

$$F_{XY}(x, y) = F_X(x)F_Y(y). \tag{3.3}$$

10. **Lemma.** Real-valued random variables $X_1, \ldots, X_n$ are independent if their joint distribution function $F_{X_1,\ldots,X_n}$ can be written as the product of the associated marginal distribution functions, i.e., if

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \quad \text{for all } x_1, \ldots, x_n \in \mathbb{R}.$$

CHAPTER

4

EXPECTATION

1. Throughout the chapter, we assume that an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is fixed.

4.1 Preliminary considerations

2. Consider the toss of a coin that lands heads with probability $p \in (0, 1)$ and tails with probability $1 - p$. Also, denote by X the random variable that takes the value 1 if tails are observed and the value 0 if heads are observed. Now, consider two parties, say A and B, that bet on the coin's toss: once the coin lands, party A will pay \$$X$ to party B (i.e., A will pay B \$1 if tails occur and \$0 if heads occur). What is the value $\mathbb{E}[X]$ of this game? In other words, how much money $\mathbb{E}[X]$ should B pay to A in advance for both parties to feel that they engage in a fair game? Intuition suggest that

$$\mathbb{E}[X] = 1 \times (1 - p) + 0 \times p = 1 - p.$$

The number $\mathbb{E}[X]$ is the expectation of X.

3. Generalising the example above, the expectation $\mathbb{E}[\mathbf{1}_A]$ of the random variable

$$\mathbf{1}_A = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases}$$

where A is an event in $\mathcal{F}$, is given by

$$\mathbb{E}[\mathbf{1}_A] = 1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(A^c) = \mathbb{P}(A).$$

This idea and the requirement that expectation should be a linear operator provide the starting point of this chapter's theory.

4.2 Definitions

4. **Definition.** We say that X is a *simple random variable* if it is a discrete random variable that can take only a finite number of possible values.

   In particular, a random variable is simple if there exist distinct real numbers $x_1, x_2, \ldots, x_n$ and a measurable partition $A_1, A_2, \ldots, A_n$ of the sample space Ω (i.e., $A_1, A_2, \ldots, A_n \in \mathcal{F}$ satisfying $A_i \cap A_j = \emptyset$, for $i \neq j$, and $\bigcup_{i=1}^n A_i = \Omega$) such that

$$X(\omega) = \sum_{i=1}^n x_i 1_{A_i}(\omega) \quad \text{for all } \omega \in \Omega. \tag{4.1}$$

5. **Definition.** The *expectation* of the simple random variable X given by (4.1) is defined by

$$\mathbb{E}[X] = \sum_{i=1}^n x_i \mathbb{P}(A_i).$$

6. **Definition.** Suppose that X is a $\big([0,\infty], \mathcal{B}([0,\infty])\big)$-valued random variable. The *expectation* of X is defined by

$$\mathbb{E}[X] = \sup\big\{\mathbb{E}[Y] \mid \ Y \text{ is a simple random variable with } 0 \leq Y \leq X\big\}.$$

   Note that $\mathbb{E}[X] \geq 0$, but we may have $\mathbb{E}[X] = \infty$.

7. **Definition.** Given a real-valued random variable X, define

$$X^+ = \max(0, X) \quad \text{and} \quad X^- = -\min(0, X)$$

   and observe that X^+, X^- are positive random variables such that $X = X^+ - X^-$ and $|X| = X^+ + X^-$.

   A random variable X has *finite expectation* (is *integrable*) if both $\mathbb{E}[X^+] < \infty$ and $\mathbb{E}[X^-] < \infty$. In this case, the expectation of X is defined by

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

   We often write $\int_\Omega X(\omega)\, \mathbb{P}(d\omega)$ or $\int_\Omega X \, d\mathbb{P}$ instead of $\mathbb{E}[X]$.

8. **Definition.** We denote by $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, or just $\mathcal{L}^1$ if there is no ambiguity, the set of all integrable random variables.

   For $1 \leq p < \infty$, we denote by $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$, or just $\mathcal{L}^p$ if there is no ambiguity, the set of all random variables X such that $|X|^p \in \mathcal{L}^1$.

9. For every positive random variable X, there exists a sequence (X_n) of positive simple random variables such that X_n increases to X as n increases to infinity. An *example* of such a sequence is given by

$$X_n(\omega) = \begin{cases} k2^{-n}, & \text{if } k2^{-n} \leq X(\omega) < (k+1)2^{-n} \text{ and } 0 \leq k \leq n2^n - 1, \\ n, & \text{if } X(\omega) \geq n. \end{cases}$$

4.3 Properties of expectation

10. We say that a property holds $\mathbb{P}$-a.s. if it is true for all ω in a set of probability 1. For example, we say that $X = Y$, $\mathbb{P}$-a.s., if

$$\mathbb{P}(X = Y) \equiv \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) = Y(\omega)\}\big) = 1.$$

   Similarly, we say that a sequence of random variables (X_n) converges to a random variable X, $\mathbb{P}$-a.s., if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) \equiv \mathbb{P}\left(\left\{\omega \in \Omega \mid \ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

11. The following results hold true:

   (i) Expectation is a positive, linear operator, i.e.,

$$\begin{aligned} X \geq 0 \quad &\Rightarrow \quad \mathbb{E}[X] \geq 0, \\ X, Y \in \mathcal{L}^1 \text{ and } a, b \in \mathbb{R} \quad &\Rightarrow \quad \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]. \end{aligned}$$

   (ii) If $X = Y$, $\mathbb{P}$-a.s., then $\mathbb{E}[X] = \mathbb{E}[Y]$.

   (iii) If X and Y are independent random variables, then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

   (iv) The expectation of a discrete random variable X is given by

$$\mathbb{E}[X] = \sum_{x_i} x_i\,\mathbb{P}(X = x_i).$$

   (v) The expectation of a continuous random variable X with probability density function f is given by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx.$$

(vi) (*Jensen's inequality*) Given a random variable X and a convex function $g : \mathbb{R} \to \mathbb{R}$, such that $X, g(X) \in \mathcal{L}^1$,

$$\mathbb{E}\big[g(X)\big] \geq g\big(\mathbb{E}[X]\big).$$

(vii) (*Monotone convergence theorem*) If (X_n) is an increasing sequence of positive random variables (i.e., $0 \leq X_1 \leq X_2 \leq \cdots \leq X_n \leq \cdots$) such that $\lim_{n\to\infty} X_n = X$, $\mathbb{P}$-a.s., for some random variable X, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Note that we may have $\mathbb{E}[X] = \infty$ here.

(viii) (*Fatou's lemma*) If (X_n) is a sequence of random variables such that $X_n \geq Y$, $\mathbb{P}$-a.s., for all $n \geq 1$, for some $Y \in \mathcal{L}^1$, then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Similarly, if (X_n) is a sequence of random variables such that $X_n \leq Y$, $\mathbb{P}$-a.s., for all $n \geq 1$, for some $Y \in \mathcal{L}^1$, then

$$\mathbb{E}\left[\limsup_{n\to\infty} X_n\right] \geq \limsup_{n\to\infty} \mathbb{E}[X_n].$$

(ix) (*Dominated convergence theorem*) If (X_n) is a sequence of random variables that converges to a random variable X, $\mathbb{P}$-a.s., and is such that $|X_n| \leq Y$, $\mathbb{P}$-a.s., for all $n \geq 1$, for some $Y \in \mathcal{L}^1$, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

4.4 Moment generating functions

12. **Definition.** The moment-generating function of a random variable X is defined by

$$M_X(t) = \mathbb{E}\big[e^{tX}\big], \quad \text{for } t \in \mathbb{R}.$$

Provided there is no possibility of confusion, we often write $M(t)$ instead of $M_X(t)$.

13. Given a random variable X, suppose that there exists $\varepsilon > 0$ such that

$$M_X(t) < \infty \quad \text{for all } t \in [-\varepsilon, \varepsilon].$$

The k-th *moment* $\mathbb{E}\big[X^k\big]$ of X is equal to the k-th derivative of the moment generating function M_X evaluated at 0, namely,

$$\mathbb{E}\big[X^k\big] = M_X^{(k)}(0) \equiv \left.\frac{d^k M_X(t)}{dt^k}\right|_{t=0}. \tag{4.2}$$

Proof. Passing the derivative operator inside the expectation, we can see that

$$M_X^{(k)}(t) = \frac{d^k}{dt^k}\mathbb{E}\big[e^{tX}\big] = \mathbb{E}\left[\frac{d^k e^{tX}}{dt^k}\right] = \mathbb{E}\big[X^k e^{tX}\big].$$

Evaluating this result at $t = 0$, we obtain (4.2).

4.5 Examples

14. **Example.** Suppose that X has the binomial distribution with parameters n, p. Also recall that X has the same distribution as $X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables, each with parameter p (see Example 2.24). The moment generating function of X is given by

$$\begin{aligned} M(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\exp\left(t\sum_{i=1}^n X_i\right)\right] = \mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right] &= \prod_{i=1}^n \mathbb{E}\left[e^{tX_i}\right] \\ = \prod_{i=1}^n \Big(\mathbb{P}\left(X_i = 0\right) + e^t\mathbb{P}\left(X_i = 1\right)\Big) &= \left(1 - p + pe^t\right)^n. \end{aligned}$$

Using this expression, we calculate

$$\begin{aligned} \mathbb{E}\big[X\big] &= M'(0) = npe^t\left(1 - p + pe^t\right)^{n-1}\Big|_{t=0} \\ &= np \\ \text{and}\quad \mathbb{E}\big[X^2\big] &= M''(0) = \left\{npe^t\left(1 - p + pe^t\right)^{n-1} + n(n-1)p^2e^{2t}\left(1 - p + pe^t\right)^{n-2}\right\}\Big|_{t=0} \\ &= np + n(n-1)p^2, \end{aligned}$$

so, $\mathrm{var}(X) = \mathbb{E}\big[X^2\big] - \big(\mathbb{E}\big[X\big]\big)^2 = np(1-p)$.

15. **Example.** Suppose that X has the Poisson distribution (see also Example 2.25). We can calculate the mean of X as follows:

$$\mathbb{E}\big[X\big] = \sum_{n=0}^{\infty} ne^{-\lambda}\frac{\lambda^n}{n!} = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!} = \lambda.$$

16. **Example.** Suppose that X has the Poisson distribution. The moment generating function of X is given by

$$M(t) = \mathbb{E}\big[e^{tX}\big] = \sum_{n=0}^{\infty} e^{tn}e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{\left(\lambda e^t\right)^n}{n!} = e^{\lambda(e^t-1)}.$$

Using this result, we calculate

$$\mathbb{E}\big[X\big] = M'(0) = \lambda e^t e^{\lambda(e^t-1)}\Big|_{t=0} = \lambda$$

$$\text{and} \quad \mathbb{E}\big[X^2\big] = M''(0) = \Big\{\lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)}\Big\}\Big|_{t=0} = \lambda + \lambda^2,$$

so, $\mathrm{var}(X) = \mathbb{E}\big[X^2\big] - \big(\mathbb{E}\big[X\big]\big)^2 = \lambda$.

17. **Example.** Suppose that X is a Gaussian random variable with mean m and variance σ^2 (see also Example 2.30). We can calculate the mean of X as follows:

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m) \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\
&\quad + m \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\
&= \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{x^2}{2\sigma^2}\right) dx}_{=0} \\
&\quad + m \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx}_{=1} \\
&= m.
\end{aligned}$$

In the penultimate expression, the first integral is 0 because its integrand is an odd function, and the fact that the second integral is equal to 1 because its integrand is a probability density function.

18. **Example.** Suppose that X is a Gaussian random variable with mean m and variance σ^2. The moment generating function of X is given by

$$\begin{aligned}
M(t) &= \mathbb{E}\left[e^{tX}\right] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\
&= \exp\left(mt + \frac{1}{2}\sigma^2 t^2\right) \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\big(x-(m+\sigma^2 t)\big)^2}{2\sigma^2}\right) dx}_{=1} \\
&= \exp\left(mt + \frac{1}{2}\sigma^2 t^2\right),
\end{aligned}$$

where we have used the fact that the last integral is equal to 1 because its integrand is a probability density function.

Using this result, we calculate

$$\mathbb{E}[X] = M'(0) = (m + \sigma^2 t) \exp\left(mt + \frac{1}{2}\sigma^2 t^2\right)\bigg|_{t=0} = m$$

and

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= M''(0) \\
&= \left\{\sigma^2 \exp\left(mt + \frac{1}{2}\sigma^2 t^2\right) + (m + \sigma^2 t)^2 \exp\left(mt + \frac{1}{2}\sigma^2 t^2\right)\right\}\bigg|_{t=0} \\
&= \sigma^2 + m^2,
\end{aligned}$$

so, $\mathrm{var}(X) = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 = \sigma^2$.

19. **Example.** If X is a normal random variable with mean 0 and variance σ^2, then

$$\begin{aligned}
\mathbb{E}\left[\exp\left(-\frac{\sigma^2}{2} - X\right)\right] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\sigma^2}{2} - x\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x + \sigma^2)^2}{2\sigma^2}\right) dx \\
&= 1.
\end{aligned}$$

In these calculations, the second integral is equal to 1 because its integrand is the density of a Gaussian random variable with mean $-\sigma^2$ and variance σ^2.

20. If X and Y are independent random variables, then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ (see Property 4.11.(ii) above). The following example shows that the converse is not true.

Example. Suppose that the events $A, B, C \in \mathcal{F}$ form a partition of Ω (i.e., $A \cap B = A \cap C = B \cap C = \emptyset$ and $A \cup B \cup C = \Omega$) and have probabilities $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{3}$. Also, let X and Y be the random variables defined by

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \in B \cup C, \end{cases} \qquad Y(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 2, & \text{if } \omega \in B \\ 0, & \text{if } \omega \in C. \end{cases}$$

Combining the fact that $XY = X$ with the calculation

$$\mathbb{E}[Y] = 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 0 \times \frac{1}{3} = 1,$$

we can see that $\mathbb{E}[XY] = \mathbb{E}[X] = \mathbb{E}[X]\,\mathbb{E}[Y]$. On the other hand, X and Y are *not* independent because, e.g.,

$$\begin{aligned}
\mathbb{P}\big(\{X = 1\} \cap \{Y = 1\}\big) = \mathbb{P}(A \cap A) = \mathbb{P}(A) &= \frac{1}{3} \\
&\neq \frac{1}{9} = \mathbb{P}(A)\mathbb{P}(A) = \mathbb{P}(X = 1)\mathbb{P}(Y = 1).
\end{aligned}$$

21. The following example shows that Jensen's inequality is strict in general (see Property 4.11.(vi) above).

Example. Suppose that X has the normal distribution, namely $X \sim \mathcal{N}(0, \sigma^2)$ for $\sigma > 0$ (see also Example 2.30). Also, consider the quadratic function $g(x) = x^2$, $x \in \mathbb{R}$. Using the results of Example 2.18, we calculate

$$\mathbb{E}\left[g(X)\right] = \mathbb{E}\left[X^2\right] = \sigma^2 > 0 = \big(\mathbb{E}\left[X\right]\big)^2 = g\big(\mathbb{E}\left[X\right]\big).$$

22. The following example shows that the inequalities in Fatou's lemma can be strict (see Property 4.11.(ix) above).

Example. Suppose that $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}\big((0,1)\big)$ and $\mathbb{P}$ is the Lebesgue measure on $\big((0,1), \mathcal{B}\big((0,1)\big)\big)$. Consider the sequence $(X_n,\ n \geq 1)$ of the random variables given by

$$X_n(\omega) = (n+1)\mathbf{1}_{\left(\frac{1}{2}, \frac{1}{2} + \frac{1}{n+1}\right)}(\omega) \equiv \begin{cases} n+1, & \text{if } \omega \in \left(\frac{1}{2}, \frac{1}{2} + \frac{1}{n+1}\right), \\ 0, & \text{otherwise.} \end{cases}$$

Given any $n \geq 1$, we calculate

$$\begin{aligned} \mathbb{E}\left[X_n\right] &= (n+1)\mathbb{P}\big(\left(\tfrac{1}{2}, \tfrac{1}{2} + \tfrac{1}{n+1}\right)\big) + 0\mathbb{P}\big((0, \tfrac{1}{2}] \cup [\tfrac{1}{2} + \tfrac{1}{n+1}, 1)\big) \\ &= 1. \end{aligned}$$

Moreover, we can see that

$$\lim_{n \to \infty} X_n(\omega) = 0 \quad \text{for all } \omega \in \Omega.$$

These observations imply that

$$\lim_{n \to \infty} \mathbb{E}\left[X_n\right] = 1 > 0 = \mathbb{E}\left[\lim_{n \to \infty} X_n(\omega)\right].$$

Note that the sequence of random variables considered in this example does not satisfy the assumptions of either the monotone convergence theorem or the dominated convergence theorem.

CHAPTER

5

CONDITIONAL EXPECTATION

1. Throughout the chapter, we assume that an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting all random variables considered is fixed.

5.1 Definitions and existence

2. **Definition.** Consider a random variable X such that $\mathbb{E}\big[|X|\big] < \infty$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra on Ω. The conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ of the random variable X given the σ-algebra $\mathcal{G}$ is any random variable Y such that

   (i) Y is $\mathcal{G}$-measurable,

   (ii) $\mathbb{E}\big[|Y|\big] < \infty$, and

   (iii) for every event $C \in \mathcal{G}$,

   $$\mathbb{E}[\mathbf{1}_C Y] = \mathbb{E}[\mathbf{1}_C X].$$

   We say that a random variable Y with the properties (i)–(iii) is *a version of the conditional expectation* $\mathbb{E}[X \mid \mathcal{G}]$ of X given $\mathcal{G}$, and we write $Y = \mathbb{E}[X \mid \mathcal{G}]$, $\mathbb{P}$-a.s..

3. **Theorem.** Consider a random variable X such that $\mathbb{E}\big[|X|\big] < \infty$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. There exists a random variable Y having properties (i)–(iii) in Definition 5.2.

   Furthermore, Y is unique in the sense that, if $\tilde{Y}$ is another random variables satisfying the required properties, then $\tilde{Y} = Y$, $\mathbb{P}$-a.s..

4. **Definition.** Consider an event $B \in \mathcal{F}$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. The conditional probability of B given $\mathcal{G}$ is the *random variable* defined by

$$\mathbb{P}\left(B \mid \mathcal{G}\right) = \mathbb{E}\left[\mathbf{1}_B \mid \mathcal{G}\right].$$

5. Note that conditional expectation and probability are random variables: **probability theory is concerned with the future**.

5.2 Conditional probability given an event

6. Given an event $B \in \mathcal{F}$, $\mathbb{P}(B)$ quantifies our views on how likely it is for the event B to occur. Now, suppose that we have been informed that chance outcomes are restricted within an event $A \in \mathcal{F}$. In other words, suppose that somebody informs us that all likely to happen events are subsets of A, and all events that are subsets of A^c are impossible to occur.

   How should we modify our views, namely our probability measure, to account for this scenario? To this end, we denote by $\mathbb{P}\left(B \mid A\right)$ our modified belief on the likelihood of the event $B \in \mathcal{F}$ given the knowledge that A has occurred. Since the only new information that we possess is that chance outcomes are restricted within the event A, it is natural to postulate that $\mathbb{P}\left(B \mid A\right)$ should be proportional to $\mathbb{P}\left(B \cap A\right)$, namely

$$\mathbb{P}\left(B \mid A\right) \sim \mathbb{P}\left(B \cap A\right). \tag{5.1}$$

   However, our beliefs should "add up" to 1, so that we have a proper probability measure. This means that we should impose the requirement that $\mathbb{P}(\Omega \mid A) = 1$. Since $\mathbb{P}\left(\Omega \cap A\right) = \mathbb{P}\left(A\right)$, we conclude that we should scale the right hand side of (5.1) by $1/\mathbb{P}\left(A\right)$ to obtain

$$\mathbb{P}\left(B \mid A\right) = \frac{\mathbb{P}\left(B \cap A\right)}{\mathbb{P}\left(A\right)}. \tag{5.2}$$

   (Of course, this formula makes sense only if $\mathbb{P}\left(A\right) > 0$.)

   We can check that the function $\mathbb{P}\left(\cdot \mid A\right) : \mathcal{F} \to [0, 1]$ defined by (5.2) is indeed a probability measure on $(\Omega, \mathcal{F})$.

7. **Bayes' theorem.** Consider events $A_1, A_2, \ldots, A_n \in \mathcal{F}$ that form a partition of Ω (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^n A_i = \Omega$). Given an event $B \in \mathcal{F}$, the events $B \cap A_1, B \cap A_2, \ldots, B \cap A_n$ are pairwise disjoint and $\bigcup_{i=1}^n B \cap A_i = B$. Therefore, the additivity property of a probability measure and (5.2) imply the *total probability formula*

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \cdots + \mathbb{P}(B \cap A_n) \\ &= \mathbb{P}(B \mid A_1)\,\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\,\mathbb{P}(A_2) + \cdots + \mathbb{P}(B \mid A_n)\,\mathbb{P}(A_n)\,.\end{aligned}$$

Using this result and (5.2), we derive *Bayes' formula*

$$\begin{aligned}\mathbb{P}(A_k \mid B) &= \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(B \mid A_k)\,\mathbb{P}(A_k)}{\mathbb{P}(B \mid A_1)\,\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\,\mathbb{P}(A_2) + \cdots + \mathbb{P}(B \mid A_n)\,\mathbb{P}(A_n)}.\end{aligned}$$

8. Conditional probabilities defined as in (5.2) have an *a posteriori* character: we have been informed and we know that event A has occurred. How should we develop our theory to account for a *prior to observation* perspective? In other words, suppose that we anticipate an observation that will inform us on whether A or A^c occurs. How should we modify our views to account for this situation?

   Given the arguments in Paragraph 5.6, the natural answer is to set

$$\begin{aligned}\mathbb{P}(B \mid \text{“observation of } A \text{ or } A^c\text{”}) &= \begin{cases}\mathbb{P}(B \mid A) & \text{if } A \text{ occurs,}\\ \mathbb{P}(B \mid A^c) & \text{if } A^c \text{ occurs,}\end{cases} \\ &= \mathbb{P}(B \mid A)\,\mathbf{1}_A + \mathbb{P}(B \mid A^c)\,\mathbf{1}_{A^c}, \end{aligned} \tag{5.3}$$

   provided, of course, that $0 < \mathbb{P}(A) < 1$. Observe that our views on how likely it is for the event B to occur have now become a simple *random variable*. Given any sample $\omega \in \Omega$, the conditional probability of the event B takes the value $\mathbb{P}(B \mid A)$ if $\omega \in A$ and takes the value $\mathbb{P}(B \mid A^c)$ if $\omega \in A^c$.

9. Given events $A, B \in \mathcal{F}$ such that $0 < \mathbb{P}(A) < 1$, the random variable Y defined by

$$Y = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}\mathbf{1}_A + \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)}\mathbf{1}_{A^c} \tag{5.4}$$

   is the conditional probability of B given the σ-algebra $\{\emptyset, \Omega, A, A^c\}$. We denote this conditional probability by

$$\mathbb{P}\big(B \mid \{\emptyset, \Omega, A, A^c\}\big) = \mathbb{E}\big[\mathbf{1}_B \mid \{\emptyset, \Omega, A, A^c\}\big].$$

10. We can verify that the random variable Y defined by (5.4) is indeed the conditional probability of B given the σ-algebra $\{\emptyset, \Omega, A, A^c\}$ by checking the defining properties of conditional probability (see Definition 5.4):

(i) The simple random variable Y is clearly $\{\emptyset, \Omega, A, A^c\}$-measurable.

(ii) We calculate

$$\mathbb{E}\big[|Y|\big] = \mathbb{E}\big[Y\big] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \mathbb{P}(A) + \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)} \mathbb{P}(A^c) = \mathbb{P}(B) < \infty.$$

(iii) Let C be any event in $\{\emptyset, \Omega, A, A^c\}$. We calculate

$$\begin{aligned}
\mathbb{E}\big[\mathbf{1}_C Y\big] &= \mathbb{E}\left[\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \mathbf{1}_C \mathbf{1}_A + \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)} \mathbf{1}_C \mathbf{1}_{A^c}\right] \\
&= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \mathbb{P}(A \cap C) + \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)} \mathbb{P}(A^c \cap C) \\
&= \begin{cases} \mathbb{P}(\emptyset \cap B), & \text{if } C = \emptyset, \\ \mathbb{P}(\Omega \cap B), & \text{if } C = \Omega, \\ \mathbb{P}(A \cap B), & \text{if } C = A, \\ \mathbb{P}(A^c \cap B), & \text{if } C = A^c, \end{cases} \\
&= \mathbb{E}\big[\mathbf{1}_C \mathbf{1}_B\big].
\end{aligned}$$

5.3 Conditional expectation of a simple random variable given another simple random variable

11. Consider two simple random variables X and Z and suppose that

$$X = \sum_{i=1}^{n} x_i \mathbf{1}_{\{X = x_i\}} \quad \text{and} \quad Z = \sum_{j=1}^{m} z_j \mathbf{1}_{\{Z = z_j\}},$$

for some distinct $x_1, \ldots, x_n$ and $z_1, \ldots, z_m$. Also, assume that $\mathbb{P}(Z = z_j) > 0$ for all $j = 1, \ldots, m$.

Suppose that we have made an "experiment" that has informed us about the actual value of Z. In particular, suppose that we have been given the information that the actual value of the random variable Z is z_j, for some $j = 1, \ldots, m$. In this context where we know that the event $\{Z = z_j\}$ has occurred, we should revise our probabilities from $\mathbb{P}(\cdot)$ to $\mathbb{P}(\cdot \mid Z = z_j)$. Furthermore, we should revise the expectation of X from $\mathbb{E}[X]$ to

$$\mathbb{E}[X \mid Z = z_j] = \sum_{i=1}^{n} x_i \, \mathbb{P}(X = x_i \mid Z = z_j).$$

This conditional expectation, *the conditional expectation of X given that the random variable Z is equal to z_j*, which is a real number, has an *a posteriori* character: we have been informed that the actual value of Z is z_j.

Prior to observation, namely, before we observe the actual value of Z, it is natural to consider the random variable

$$Y = \sum_{j=1}^{m} \mathbb{E}\left[X \mid Z = z_j\right] \mathbf{1}_{\{Z=z_j\}}, \tag{5.5}$$

as the conditional expectation $\mathbb{E}\big[X \mid \sigma(Z)\big] \equiv \mathbb{E}\left[X \mid Z\right]$ of X given the σ-algebra $\sigma(Z)$, namely, given the information set $\sigma(Z)$ that is associated with the observation of the random variable Z.

12. We can verify that the random variable Y defined by (5.5) is indeed the conditional expectation of X given $\sigma(Z)$ by checking the defining properties of conditional expectation (see Definition 5.2). To this end, we first observe that the σ-algebra $\sigma(Z)$ consists of all possible unions of sets in the family $\big\{\{Z = z_1\}, \ldots, \{Z = z_m\}\big\}$, namely,

$$\sigma(Z) = \left\{ \bigcup_{k\in J} \{Z = z_k\} \;\middle|\; J \subseteq \{1, \ldots, m\} \right\}, \tag{5.6}$$

with the convention that

$$\bigcup_{k\in\emptyset} \{Z = z_k\} = \emptyset.$$

(i) In view of (5.6), we can see that Y is $\sigma(Z)$-measurable because

$$Y = \sum_{j=i}^{m} c_j \mathbf{1}_{\{Z=z_j\}},$$

where the constants c_j are given by

$$c_j = \sum_{i=1}^{n} x_i \, p_{X|Z}(x_i|z_j), \quad \text{for } j = 1, \ldots, m.$$

(ii) We calculate

$$
\begin{aligned}
\mathbb{E}\big[|Y|\big] &= \sum_{j=1}^{m} \left| \sum_{i=1}^{n} x_i \, p_{X|Z}(x_i|z_j) \right| \mathbb{P}\left(Z = z_j\right) \\
&\le \sum_{j=1}^{m} \sum_{i=1}^{n} |x_i| \frac{\mathbb{P}\left(X = x_i, Z = z_j\right)}{\mathbb{P}\left(Z = z_j\right)} \mathbb{P}\left(Z = z_j\right) \\
&= \sum_{i=1}^{n} |x_i| \sum_{j=1}^{m} \mathbb{P}\left(X = x_i, Z = z_j\right) \\
&= \sum_{i=1}^{n} |x_i| \, \mathbb{P}\left(X = x_i\right) \\
&= \mathbb{E}\big[|X|\big] < \infty.
\end{aligned}
$$

(iii) Let C be any event in $\sigma(Z)$. In view of (5.6), there exists a set $J \subseteq \{1, \ldots, m\}$ such that

$$C = \bigcup_{k \in J} \{Z = z_k\}.$$

Since $\{Z = z_1\}, \ldots, \{Z = z_m\}$ are pairwise disjoint,

$$\mathbf{1}_C = \sum_{k \in J} \mathbf{1}_{\{Z = z_k\}} \quad \text{and} \quad \mathbf{1}_{\{Z = z_k\}} \mathbf{1}_{\{Z = z_j\}} = 0, \text{ for } k \ne j.$$

In light of these observations, we calculate

$$\begin{aligned}
\mathbb{E}\big[Y\mathbf{1}_C\big] &= \mathbb{E}\left[\left(\sum_{j=1}^{m}\sum_{i=1}^{n} x_i \frac{\mathbb{P}(X=x_i, Z=z_j)}{\mathbb{P}(Z=z_j)}\mathbf{1}_{\{Z=z_j\}}\right)\left(\sum_{k\in J}\mathbf{1}_{\{Z=z_k\}}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} x_i \sum_{k\in J}\sum_{j=1}^{m} \frac{\mathbb{P}(X=x_i, Z=z_j)}{\mathbb{P}(Z=z_j)}\mathbf{1}_{\{Z=z_j\}}\mathbf{1}_{\{Z=z_k\}}\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} x_i \sum_{k\in J} \frac{\mathbb{P}(X=x_i, Z=z_k)}{\mathbb{P}(Z=z_k)}\mathbf{1}_{\{Z=z_k\}}\right] \\
&= \sum_{i=1}^{n} x_i \sum_{k\in J} \mathbb{P}(X=x_i, Z=z_k) \\
&= \sum_{i=1}^{n} x_i \sum_{k\in J} \mathbb{E}\left[\mathbf{1}_{\{X=x_i\}\cap\{Z=z_k\}}\right] \\
&= \sum_{i=1}^{n} x_i \sum_{k\in J} \mathbb{E}\left[\mathbf{1}_{\{X=x_i\}}\mathbf{1}_{\{Z=z_k\}}\right] \\
&= \sum_{i=1}^{n} x_i \mathbb{E}\left[\mathbf{1}_{\{X=x_i\}}\sum_{k\in J}\mathbf{1}_{\{Z=z_k\}}\right] \\
&= \sum_{i=1}^{n} x_i \mathbb{E}\left[\mathbf{1}_{\{X=x_i\}}\mathbf{1}_C\right] \\
&= \mathbb{E}\left[\mathbf{1}_C \sum_{i=1}^{n} x_i \mathbf{1}_{\{X=x_i\}}\right] \\
&= \mathbb{E}\left[\mathbf{1}_C X\right].
\end{aligned}$$

5.4 Conditional expectation of a continuous random variable given another continuous random variable

13. Suppose that X and Z are continuous random variables with joint probability density function f_{XZ}, so that

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XZ}(x,z)\,dx$$

is the probability density function of Z, and assume that

$$\mathbb{E}\big[|X|\big] = \int_{-\infty}^{\infty} |x| f_X(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x| f_{XZ}(x,z)\,dx\,dz < \infty.$$

We define the *conditional probability density function* of X given Z by

$$f_{X|Z}(x|z) = \begin{cases} f_{XZ}(x,z)/f_Z(z), & \text{if } f_Z(z) \neq 0, \\ 0, & \text{if } f_Z(z) = 0. \end{cases}$$

The random variable

$$Y = \int_{-\infty}^{\infty} x\, f_{X|Z}(x|Z)\, dx \tag{5.7}$$

is the conditional expectation $\mathbb{E}\big[X \mid \sigma(Z)\big] \equiv \mathbb{E}[X \mid Z]$ of X given the σ-algebra $\sigma(Z)$, namely, given the information set $\sigma(Z)$ that is associated with the observation of the random variable Z.

14. We can verify that the random variable Y defined by (5.7) is indeed the conditional expectation of X given $\sigma(Z)$ by checking the defining properties of conditional expectation (see Definition 5.2):

(i) Y is $\sigma(Z)$-measurable.

(ii) $\mathbb{E}\big[|Y|\big] < \infty$. Indeed, we note that $x \mapsto |x|$ is a convex function and we use Jensen's inequality to calculate

$$\begin{aligned} \mathbb{E}\big[|Y|\big] &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} x\, f_{X|Z}(x|z)\, dx \right| f_Z(z)\, dz \\ &\leq \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |x|\, f_{X|Z}(x|z)\, dx \right) f_Z(z)\, dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x|\, f_{X|Z}(x|z) f_Z(z)\, dx\, dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x|\, f_{XZ}(x,z)\, dx\, dz \\ &= \mathbb{E}\big[|X|\big] < \infty. \end{aligned}$$

(iii) $\mathbb{E}\big[\mathbf{1}_C Y\big] = \mathbb{E}\big[\mathbf{1}_C X\big]$ for all $C \in \sigma(Z)$. To see this claim, we first note, given any event $C \in \sigma(Z)$, there exists $A \in \mathcal{B}(\mathbb{R})$ such that

$$C = \big\{\omega \in \Omega \mid Z(\omega) \in A\big\}.$$

In view of this observation, we calculate

$$\begin{aligned}
\mathbb{E}\big[\mathbf{1}_C Y\big] &= \mathbb{E}\left[\mathbf{1}_{\{Z\in A\}}\left(\int_{-\infty}^{\infty} x\, f_{X|Z}(x|Z)\,dx\right)\right]\\
&= \int_{-\infty}^{\infty} \mathbf{1}_{\{z\in A\}}\left(\int_{-\infty}^{\infty} x\, f_{X|Z}(x|z)\,dx\right) f_Z(z)\,dz\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{\{z\in A\}}\, x\, f_{X|Z}(x|z) f_Z(z)\,dx\,dz\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{\{z\in A\}}\, x\, f_{XZ}(x,z)\,dx\,dz\\
&= \mathbb{E}\big[\mathbf{1}_{\{Z\in A\}}X\big]\\
&= \mathbb{E}\big[\mathbf{1}_C X\big].
\end{aligned}$$

5.5 Properties of conditional expectation

15. In the following list of properties of conditional expectation, we assume that *all random variables* are in $\mathcal{L}^1$, and that $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are σ-algebras on Ω.

(i) $\mathbb{E}[X \mid \{\Omega, \emptyset\}] = \mathbb{E}[X]$.
(The trivial σ-algebra $\{\Omega, \emptyset\}$ can be viewed as a model for "absence of information": we can interpret Ω as the event that "something occurs" and $\emptyset$ as the event that "nothing happens". This property reflects the idea that expectation is the same as conditional expectation given no information.)

(ii) (*Linearity*) Given constants $a_1, a_2 \in \mathbb{R}$, and random variables X_1, X_2,

$$\mathbb{E}[a_1X_1 + a_2X_2 \mid \mathcal{G}] = a_1\mathbb{E}[X_1 \mid \mathcal{G}] + a_2\mathbb{E}[X_2 \mid \mathcal{G}], \quad \mathbb{P}\text{-a.s..}$$

(iii) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$, $\mathbb{P}$-a.s..
(This property reflects the idea that "knowledge" of $\mathcal{G}$ implies "knowledge" of the actual value of X.)

(iv) (*"Taking out what is known"*) If Z is $\mathcal{G}$-measurable, then

$$\mathbb{E}[ZX \mid \mathcal{G}] = Z\,\mathbb{E}[X \mid \mathcal{G}], \quad \mathbb{P}\text{-a.s..}$$

(This property is again based on the idea that "knowledge" of $\mathcal{G}$ implies "knowledge" of the actual value of Z.)

(v) (*Independence*) If $\sigma(X)$ and $\mathcal{H}$ are independent σ-algebras,

$$\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[X], \quad \mathbb{P}\text{-a.s..}$$

(Indeed, if the random variable X is independent of the information $\mathcal{G}$, then "knowledge" of $\mathcal{G}$ provides no information about X.)

(vi) (*Tower property*) If $\mathcal{H} \subseteq \mathcal{G}$, then

$$\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\big] = \mathbb{E}[X \mid \mathcal{H}], \quad \mathbb{P}\text{-a.s.}.$$

(viii) (*Conditional Jensen's inequality*) Given a random variable X and a convex function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\big[g(X) \mid \mathcal{G}\big] \geq g\big(\mathbb{E}[X \mid \mathcal{G}]\big), \quad \mathbb{P}\text{-a.s.}.$$

(viii) (*Conditional monotone convergence theorem*) If (X_n) is an increasing sequence of positive random variables (i.e., $0 \leq X_1 \leq X_2 \leq \cdots \leq X_n \leq \cdots$) converging to the random variable X, $\mathbb{P}$-a.s., then

$$\lim_{n \to \infty} \mathbb{E}[X_n \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}], \quad \mathbb{P}\text{-a.s.}.$$

(ix) (*Conditional dominated convergence theorem*) If (X_n) is a sequence of random variables that converges to a random variable X, $\mathbb{P}$-a.s., and is such that $|X_n| \leq Z$, $\mathbb{P}$-a.s., for all $n \geq 1$, for some random variable Z, then

$$\lim_{n \to \infty} \mathbb{E}[X_n \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}], \quad \mathbb{P}\text{-a.s.}.$$

(x) (*Conditional Fatou's lemma*) If (X_n) is a sequence of random variables such that $X_n \geq Z$, $\mathbb{P}$-a.s., for all $n \geq 1$, for some random variable Z, then

$$\mathbb{E}\left[\liminf_{n \to \infty} X_n \mid \mathcal{G}\right] \leq \liminf_{n \to \infty} \mathbb{E}[X_n \mid \mathcal{G}], \quad \mathbb{P}\text{-a.s.}.$$

Similarly, if (X_n) is a sequence of random variables such that $X_n \leq Z$, $\mathbb{P}$-a.s., for all $n \geq 1$, for some random variable Z, then

$$\mathbb{E}\left[\limsup_{n \to \infty} X_n \mid \mathcal{G}\right] \geq \limsup_{n \to \infty} \mathbb{E}[X_n \mid \mathcal{G}], \quad \mathbb{P}\text{-a.s.}.$$

5.6 Examples

16. **Example.** A laboratory blood test is 95% effective in detecting a certain disease when it is present. However, the test also yields a 'false positive' result for 2% of healthy people tested. If 0.1% of the population actually have the disease, what is the probability that a person has the disease, given that his test result is positive?

To derive the answer to this question, we consider the events

$$H = \{\text{the person does not have the disease}\}$$
$$\text{and} \quad P = \{\text{the test result is positive}\},$$

so that $H^c = \{\text{the person has the disease}\}$. We are given the probabilities

$$\mathbb{P}(P \mid H^c) = 0.95, \quad \mathbb{P}(P \mid H) = 0.02 \quad \text{and} \quad \mathbb{P}(H^c) = 0.001.$$

Since the events H, H^c form a partition of the sample space, we can use Bayes' theorem (see Paragraph 5.7) to calculate

$$\begin{aligned}\mathbb{P}(H^c \mid P) &= \frac{\mathbb{P}(P \mid H^c)\,\mathbb{P}(H^c)}{\mathbb{P}(P \mid H^c)\,\mathbb{P}(H^c) + \mathbb{P}(P \mid H)\,\mathbb{P}(H)} \\ &= \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.02 \times 0.999} \simeq 0.045.\end{aligned}$$

We conclude that if the result of a person's test is positive, then there is 4.5% chance that he/she has the disease.

17. **Example.** Consider two independent Poisson random variables X and U with parameters λ and μ, respectively, and define $Z = X + U$. Also, recall that the moment generating functions of X and U are given by

$$M_X(t) = e^{-\lambda} e^{\lambda e^t} \quad \text{and} \quad M_U(t) = e^{-\mu} e^{\mu e^t}.$$

(see Example 4.16). Since X and U are independent,

$$M_Z(t) = \mathbb{E}\left[e^{t(X+U)}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tU}\right] = M_X(t) M_U(t) = e^{-(\lambda+\mu)} e^{(\lambda+\mu)e^t}.$$

which proves that the random variable $Z = X + U$ has the Poisson distribution with parameter $\lambda + \mu$.

Given any $n \geq 0$ and $m = 0, 1, 2, \ldots, n$, we calculate

$$\begin{aligned}\mathbb{P}(X = m \mid Z = n) &= \frac{\mathbb{P}(X = m,\, Z = n)}{\mathbb{P}(Z = n)} \\ &= \frac{\mathbb{P}(X = m,\, U = n - m)}{\mathbb{P}(X + U = n)} \\ &= \frac{\mathbb{P}(X = m)\,\mathbb{P}(U = n - m)}{\mathbb{P}(X + U = n)} \\ &= \frac{e^{-\lambda}\frac{\lambda^m}{m!}\, e^{-\mu}\frac{\mu^{n-m}}{(n-m)!}}{e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}} \\ &= \frac{n!}{m!\,(n-m)!}\,\frac{\lambda^m \mu^{n-m}}{(\lambda+\mu)^n} \\ &= \binom{n}{m}\left(\frac{\lambda}{\lambda+\mu}\right)^m \left(1 - \frac{\lambda}{\lambda+\mu}\right)^{n-m},\end{aligned}$$

which proves that the conditional distribution of X given that the event $\{Z = n\}$ has occurred is Binomial with parameters n and $\frac{\lambda}{\lambda+\mu}$.

In view of results in Example 4.14, we can also see that

$$\mathbb{E}[X \mid Z = n] = \frac{\lambda n}{\lambda + \mu}, \quad \text{for } n \geq 0.$$

We conclude this example with the expression

$$\mathbb{P}\big(X = m \mid \sigma(Z)\big) = \binom{Z}{m} \left(\frac{\lambda}{\lambda + \mu}\right)^m \left(1 - \frac{\lambda}{\lambda + \mu}\right)^{Z-m} \mathbf{1}_{\{Z \geq m\}}$$

for the conditional probability of $\{X = m\}$ given the information set $\sigma(Z)$ that is associated with the observation of the random variable $Z = X + U$, and the expression

$$\mathbb{E}\big[X \mid \sigma(Z)\big] = \frac{\lambda Z}{\lambda + \mu}$$

for the conditional expectation of X given the information set $\sigma(Z)$ that is associated with the observation of the random variable $Z = X + U$.

CHAPTER

6

STOCHASTIC PROCESSES

1. Throughout the chapter, we assume that all random variables considered are defined on a fixed a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

6.1 Stochastic processes

2. **Definition.** A *stochastic process* is a family of random variables $(X_t,\ t \in \mathcal{T})$ indexed by a non-empty set $\mathcal{T}$.

   When the index set $\mathcal{T}$ is understood by the context, we usually write X or (X_t) instead of $(X_t,\ t \in \mathcal{T})$.

3. In this course, we consider only stochastic processes whose index set $\mathcal{T}$ is the set of natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$ or the set of positive real numbers $\mathbb{R}_+ = [0, \infty)$. In the first instance, we are talking about *discrete time processes*, in the second one, we are talking about *continuous time processes*.

4. Stochastic processes are mathematical models for quantities that evolve randomly over time. For example, we can use a stochastic process $(X_t,\ t \geq 0)$ to model the time evolution of the stock price of a given company. In this context, assuming that present time is 0, the random variable X_t is the stock price of the company at the future time t.

6.2 Filtrations and stopping times

5. **Definition.** A *filtration* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a family $(\mathcal{F}_t,\ t \in \mathcal{T})$ of σ-algebras such that

$$\mathcal{F}_t \subseteq \mathcal{F} \text{ for all } t \in \mathcal{T}, \quad \text{and} \quad \mathcal{F}_s \subseteq \mathcal{F}_t \text{ for all } s, t \in \mathcal{T} \text{ such that } s \leq t. \tag{6.1}$$

We usually write $(\mathcal{F}_t)$ or $\{\mathcal{F}_t\}$ instead of $(\mathcal{F}_t,\ t \in \mathcal{T})$.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a filtration $(\mathcal{F}_t)$, often denoted by $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, is said to be a *filtered probability space*.

6. We have seen that σ-algebras are models for information. Accordingly, filtrations are models for *flows of information*. The inclusions in (6.1) reflect the idea that, as time progresses, more information becomes available, as well as the idea that "memory is perfect" in the sense that there is no information lost in the course of time.

7. **Definition.** The *natural filtration* $(\mathcal{F}_t^X)$ of a stochastic process (X_t) is defined by

$$\mathcal{F}_t^X = \sigma\big(X_s,\ s \in \mathcal{T},\ s \leq t\big), \quad t \in \mathcal{T}.$$

8. The natural filtration of a process (X_t) is the flow of information that the observation of the evolution in time of the process (X_t) yields, and only that.

9. **Definition.** We say that a process (X_t) is *adapted* to a filtration $(\mathcal{F}_t)$ if X_t is $\mathcal{F}_t$-measurable for all $t \in \mathcal{T}$, or equivalently, if $\mathcal{F}_t^X \subseteq \mathcal{F}_t$ for all $t \in \mathcal{T}$.

10. In the context of this definition, the information becoming available by the observation of the time evolution of an $(\mathcal{F}_t)$-adapted process (X_t) is (possibly strictly) included in the information flow modelled by $(\mathcal{F}_t)$.

11. Recalling that $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{R}_+$, a *random time* is any random variable with values in $\mathcal{T} \cup \{\infty\}$.

We often use a "random time" τ to denote the time at which a given random event occurs. In this context, the set $\{\tau = \infty\}$ represents the event that the random event never occurs.

12. **Definition.** Given a filtration $(\mathcal{F}_t)$, we say that a random time τ is an $(\mathcal{F}_t)$-*stopping time* if

$$\{\tau \leq t\} = \{\omega \in \Omega \mid \ \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \text{for all } t \in \mathcal{T}. \tag{6.2}$$

13. We can think of an $(\mathcal{F}_t)$-stopping time as a random time with the property that, given any fixed time t, we know whether the random event that it represents has occurred or not in light of the available information $\mathcal{F}_t$.

Note that the filtration $(\mathcal{F}_t)$ is essential for the definition of stopping times. Indeed, a random time can be a stopping time with respect to some filtration $(\mathcal{F}_t)$, but not with respect to some other filtration $(\mathcal{G}_t)$.

14. **Example.** Suppose that τ_1 and τ_2 are two $(\mathcal{F}_t)$-stopping times. Then the random time τ defined by $\tau = \min\{\tau_1, \tau_2\}$ is an $(\mathcal{F}_t)$-stopping time.

    Proof. The assumption that τ_1 and τ_2 are $(\mathcal{F}_t)$-stopping times implies that

    $$\{\tau_1 \le t\},\ \{\tau_2 \le t\} \in \mathcal{F}_t \quad \text{for all } t \in \mathcal{T}.$$

    Therefore

    $$\{\tau \le t\} = \{\tau_1 \le t\} \cup \{\tau_2 \le t\} \in \mathcal{F}_t \quad \text{for all } t \in \mathcal{T},$$

    which proves the claim.

6.3 Martingales

15. **Definition.** An $(\mathcal{F}_t)$-adapted stochastic process (X_t) is an $(\mathcal{F}_t)$-*supermartingale* if

    $$\begin{aligned} &\text{(i)} \quad \mathbb{E}\big[|X_t|\big] < \infty \quad \text{for all } t \in \mathcal{T}, \text{ and} \\ &\text{(ii)} \quad \mathbb{E}\big[X_t \mid \mathcal{F}_s\big] \le X_s, \quad \mathbb{P}\text{-a.s., for all } s, t \in \mathcal{T} \text{ such that } s < t. \end{aligned}$$

    An $(\mathcal{F}_t)$-adapted stochastic process (X_t) is an $(\mathcal{F}_t)$-*submartingale* if

    $$\begin{aligned} &\text{(i)} \quad \mathbb{E}\big[|X_t|\big] < \infty \quad \text{for all } t \in \mathcal{T}, \text{ and} \\ &\text{(ii)} \quad \mathbb{E}\big[X_t \mid \mathcal{F}_s\big] \ge X_s, \quad \mathbb{P}\text{-a.s., for all } s, t \in \mathcal{T} \text{ such that } s < t. \end{aligned}$$

    An $(\mathcal{F}_t)$-adapted stochastic process (X_t) is an $(\mathcal{F}_t)$-*martingale* if

    $$\begin{aligned} &\text{(i)} \quad \mathbb{E}\big[|X_t|\big] < \infty \quad \text{for all } t \in \mathcal{T}, \text{ and} \\ &\text{(ii)} \quad \mathbb{E}\big[X_t \mid \mathcal{F}_s\big] = X_s, \quad \mathbb{P}\text{-a.s., for all } s, t \in \mathcal{T} \text{ such that } s < t. \end{aligned}$$

16. A process (X_t) is a submartingale if $(-X_t)$ is a supermartingale, and vice versa, while a process (X_t) is a martingale if it is both a submartingale and a supermartingale.

    A supermartingale "decreases on average". A submartingale "increases on average".

17. **Example.** A gambler bets repeatedly on a game of chance. If we denote by X_0 the gambler's initial capital and by X_n the gambler's total wealth after their n-th bet, then $X_n - X_{n-1}$ are the gambler's *net winnings* from their n-th bet $(n \ge 1)$.

    If (X_n) is a martingale, then the game series is *fair*.

    If (X_n) is a submartingale, then the game series is *favourable* to the gambler.

    If (X_n) is a supermartingale, then the game series is *unfavourable* to the gambler.

18. **Example.** Let $X_1, X_2, \dots$ be a sequence of independent random variables in $\mathcal{L}^1$ such that $\mathbb{E}[X_n] = 0$ for all n. If we set

$$S_0 = 0, \quad S_n = X_1 + X_2 + \cdots + X_n, \quad \text{for } n \geq 1,$$
$$\mathcal{F}_0 = \{\emptyset, \Omega\} \quad \text{and} \quad \mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n), \quad \text{for } n \geq 1,$$

then the process (S_n) is an $(\mathcal{F}_n)$-martingale.

Proof. Since

$$|X_1 + X_2 + \cdots + X_n| \leq |X_1| + |X_2| + \cdots + |X_n|,$$

the assumption that $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ for all $n \geq 1$, implies that $\mathbb{E}\big[|S_n|\big] < \infty$ for all $n \geq 1$.

The assumption that $X_1, X_2, \dots$ are independent implies that

$$\mathbb{E}\big[X_i \mid \mathcal{F}_m\big] = \begin{cases} X_i, & \text{if } i \leq m, \\ \mathbb{E}[X_i], & \text{if } i > m. \end{cases}$$

It follows that, given any $m < n$,

$$\begin{aligned} \mathbb{E}\big[S_n \mid \mathcal{F}_m\big] &= \sum_{i=1}^{n} \mathbb{E}\big[X_i \mid \mathcal{F}_m\big] \\ &= \sum_{i=1}^{m} X_i + \sum_{i=m+1}^{n} 0 \\ &= S_m. \end{aligned}$$

19. **Example.** Let $(\mathcal{F}_t)$ be a filtration, and let any random variable $Y \in \mathcal{L}^1$. If we define

$$M_t = \mathbb{E}\big[Y \mid \mathcal{F}_t\big], \quad t \in \mathcal{T},$$

then M is a martingale.

Proof. By the definition of conditional expectation, $\mathbb{E}\big[|M_t|\big] < \infty$ for all $t \in \mathcal{T}$.

Given any times $s < t$, the tower property of conditional expectation implies

$$\begin{aligned} \mathbb{E}\big[M_t \mid \mathcal{F}_s\big] &= \mathbb{E}\big[\mathbb{E}\big[Y \mid \mathcal{F}_t\big] \mid \mathcal{F}_s\big] \\ &= \mathbb{E}\big[Y \mid \mathcal{F}_s\big] \\ &= M_s. \end{aligned}$$

20. **Definition.** A stochastic process (X_t) with continuous sample paths such that X_0 is a constant, $\mathbb{P}$-a.s., is an $(\mathcal{F}_t)$-*local martingale* if there exists a sequence (τ_n) of $(\mathcal{F}_t)$-stopping times such that

(i) $\lim_{n \to \infty} \tau_n = \infty, \quad \mathbb{P}\text{-a.s.},$

(ii) the process $(X_t^{\tau_n})$ defined by $X_t^{\tau_n} = X_{t \wedge \tau_n}$ is an $(\mathcal{F}_t)$-martingale.

Here, $a \wedge b = \min(a, b)$.

21. **Remark.** It is important to remember that

- every $(\mathcal{F}_t)$-martingale is an $(\mathcal{F}_t)$-local martingale;
- there are $(\mathcal{F}_t)$-local martingales that are *NOT* $(\mathcal{F}_t)$-martingales;
- $(\mathcal{F}_t)$-local martingales are occasionally "awkward" to work with, but they have important applications, e.g., in the modelling of financial bubbles.

6.4 Brownian motion

22. **Definition.** The *standard one-dimensional Brownian motion or Wiener process* (W_t) is the continuous time stochastic process described by the following properties:

(i) $W_0 = 0$.

(ii) *Continuity*: All of the sample paths $s \mapsto W_s(\omega)$ are continuous functions.

(iii) *Independent increments*: The increments of (W_t) in non-overlapping time intervals are independent random variables. Specifically, given any times $t_1 < t_2 < \cdots < t_k$, the random variables $W_{t_2} - W_{t_1}, \ldots, W_{t_k} - W_{t_{k-1}}$ are independent.

(iv) *Normality*: Given any times $s < t$, the random variable $W_t - W_s$ is normal with mean 0 and variance $t - s$, i.e., $W_t - W_s \sim \mathcal{N}(0, t-s)$.

23. Given any times $s < t$,

$$\begin{aligned}
\mathbb{E}\left[W_s W_t\right] &= \mathbb{E}\left[W_s(W_s + W_t - W_s)\right] \\
&= \mathbb{E}\left[W_s^2\right] + \mathbb{E}\left[W_s(W_t - W_s)\right] \\
&= s + \mathbb{E}\left[W_s\right]\mathbb{E}\left[W_t - W_s\right] \\
&= s.
\end{aligned}$$

Therefore, given any times s, t,

$$\mathbb{E}\left[W_s W_t\right] = \min\left(s, t\right).$$

24. **Time reversal.** The continuous time stochastic process $(B_t,\ t \in [0,T])$ defined by

$$B_t = W_T - W_{T-t}, \quad t \in [0,T],$$

is a standard Brownian motion.

Proof. We verify the requirements of the definition:

- $B_0 = W_T - W_{T-0} = 0$.
- The process (B_t) has continuous sample paths because this is true for (W_t).
- Given $0 \leq t_1 < t_2 < \cdots < t_k \leq T$, observe that $T - t_k < \cdots < T - t_2 < T - T_1$, and $B_{t_i} - B_{t_{i-1}} = W_{T-t_{i-1}} - W_{T-t_i}$. Therefore, the increments

$$B_{t_2} - B_{t_1}, \quad \ldots \quad B_{t_k} - B_{t_{k-1}}$$

are independent random variables because this is true for the random variables

$$W_{T-t_1} - W_{T-t_2}, \quad \ldots \quad W_{T-t_{k-1}} - W_{T-t_k}$$

which are increments of the Brownian motion (W_t) in non-overlapping time intervals.
- Fix any times $s < t$ and note that $W_{T-t} - W_{t-s} \sim \mathcal{N}(0, t-s)$ because (W_t) is a Brownian motion. Combining this observation with the fact that $B_t - B_s = -(W_{T-t} - W_{t-s})$, we can see that $B_t - B_s \sim \mathcal{N}(0, t-s)$.

25. **Definition.** An n-dimensional standard Brownian motion (W_t) is a (column) vector $(W_t^1, \ldots, W_t^n)'$ composed by independent standard one-dimensional Brownian motions $(W_t^1), \ldots, (W_t^n)$.

26. We often want a stochastic process to be a Brownian motion with respect to the flow of information modelled by a filtration $(\mathcal{F}_t)$, which gives rise to the following definition.

    Definition. If $(\mathcal{F}_t)$ is a filtration, then an $(\mathcal{F}_t)$-adapted stochastic process (W_t) is called an $(\mathcal{F}_t)$-*Brownian motion* if

    (i) (W_t) is a Brownian motion, and

    (ii) for every time $t \geq 0$, the process $(W_{t+s} - W_t,\ s \geq 0)$ is independent of $\mathcal{F}_t$, i.e., the σ-algebras $\sigma\left(W_{t+s} - W_t,\ s \geq 0\right)$ and $\mathcal{F}_t$ are independent.

27. **Lemma.** Every $(\mathcal{F}_t)$-Brownian motion (W_t) is an $(\mathcal{F}_t)$-martingale.

    Proof. The inequalities

$$\mathbb{E}\big[|W_t|\big] \leq 1 + \mathbb{E}\big[W_t^2\big] = 1 + t < \infty$$

imply that $W_t \in \mathcal{L}^1$ for all $t \geq 0$.

Given any times $s < t$,

$$\begin{aligned}\mathbb{E}[W_t \mid \mathcal{F}_s] &= \mathbb{E}[W_t - W_s \mid \mathcal{F}_s] + W_s \\ &= 0 + W_s \\ &= W_s,\end{aligned}$$

the second equality following because the random variable $W_t - W_s$ is independent of $\mathcal{F}_s$.

28. **Lemma.** Consider a standard $(\mathcal{F}_t)$-Brownian motion (W_t) and define

$$L_t = \exp\left(-\frac{1}{2}\vartheta^2 t - \vartheta W_t\right),$$

for some constant ϑ. The process (L_t) is an $(\mathcal{F}_t)$-martingale.

Proof. Given any times $s < t$, the random variable $\vartheta(W_t - W_s)$ is normal with mean 0 and variance $\vartheta^2(t-s)$ that is independent of $\mathcal{F}_s$. In view of these observations, we can see that

$$\begin{aligned}\mathbb{E}\left[L_t L_s^{-1} \mid \mathcal{F}_s\right] &= \mathbb{E}\left[\exp\left(-\frac{\vartheta^2(t-s)}{2} - \vartheta(W_t - W_s)\right) \mid \mathcal{F}_s\right] \\ &= \mathbb{E}\left[\exp\left(-\frac{\vartheta^2(t-s)}{2} - \vartheta(W_t - W_s)\right)\right] \\ &= 1,\end{aligned}$$

The last equality here follows from Example 4.19 with the identifications $X \to \vartheta(W_t - W_s)$ and $\sigma^2 \to \vartheta^2(t-s)$. Therefore,

$$\mathbb{E}\big[|L_t|\big] = \mathbb{E}[L_t L_0^{-1} \mid \mathcal{F}_0] = 1 < \infty \quad \text{and} \quad \mathbb{E}[L_t \mid \mathcal{F}_s] = L_s.$$

CHAPTER

7

STOCHASTIC CALCULUS

1. Throughout the chapter, we fix a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ carrying a standard $(\mathcal{F}_t)$-Brownian motion (W_t). Unless explicitly stated otherwise, we assume that the Brownian motion (W_t) is one-dimensional.

   A proper development of the material in this chapter is mathematically rather technical and involved. In what follows, we focus on some of the main ideas and useful results.

7.1 Itô integrals

2. The theory of Itô calculus presents one successful answer to how we can make sense to the integral

$$\int_0^t K_s \, dW_s.$$

   We assume that the integrand (K_t) is $(\mathcal{F}_t)$-adapted and has "reasonable" sample paths in the sense that

$$\int_0^t K_s^2 \, ds < \infty \quad \text{for all } t \geq 0, \ \mathbb{P}\text{-a.s..} \tag{7.1}$$

3. **Definition.** (K_t) is a *simple* process if there exist times $0 = t_0 < t_1 < \cdots < t_n = T$ and $\mathcal{F}_{t_j}$-measurable random variables $\bar{K}_j$, $j = 0, 1, \ldots, n-1$, such that

$$K_t = \sum_{j=0}^{n-1} \bar{K}_j \mathbf{1}_{[t_j, t_{j+1})}(t).$$

4. **Definition.** The stochastic integral of a simple process (K_t) as in Definition 7.3 is defined by

$$\int_0^T K_s \, dW_s = \sum_{j=0}^{n-1} \bar{K}_j (W_{t_{j+1}} - W_{t_j}).$$

5. One construction of the Itô integral starts from stochastic integrals of simple processes as above, and then appeals to a density argument based on the Itô isometry. In particular, if (K_t) is an integrand satisfying the assumptions discussed informally in Paragraph 7.2 above, then its stochastic integral satisfies the *Itô isometry*

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T K_s \, dW_s\right)^2\right] &= \mathbb{E}\left[\int_0^T K_s^2 \, ds\right] \\ &= \int_0^T \mathbb{E}\left[K_s^2\right] \, ds \quad \text{for all } T \geq 0. \end{aligned}$$

(Note that the terms in these identities may be equal to ∞.)

6. Consider an $(\mathcal{F}_t)$-adapted process (K_t) satisfying (7.1) and let (I_t) be the stochastic process aggregating the stochastic integrals

$$I_t = \int_0^t K_s \, dW_s, \quad t \geq 0.$$

The process (I_t) is an $(\mathcal{F}_t)$-local martingale (see Definition 6.20 and Remark 6.21).

If additionally (K_t) is such that

$$\mathbb{E}\left[\int_0^T K_s^2 \, ds\right] = \int_0^T \mathbb{E}\left[K_s^2\right] \, ds < \infty \quad \text{for all } T > 0,$$

then (I_t) is an $(\mathcal{F}_t)$-martingale. In fact, it belongs to the class of *square integrable* $(\mathcal{F}_t)$-martingales, which is a sub-class of all $(\mathcal{F}_t)$-martingales.

7.2 Martingale representation theorem

7. Suppose that (W_t) is a standard n-dimensional Brownian motion (see Definition 6.25). Also, let $(\mathcal{F}_t^W)$ be the natural filtration of (W_t). The *martingale representation theorem* states that, given any $(\mathcal{F}_t^W)$-local martingale (M_t),

$$M_t = M_0 + \int_0^t K_s \, dW_s, \tag{7.2}$$

for some $(\mathcal{F}_t^W)$-adapted row-vector process (K_t) satisfying

$$\int_0^t |K_s|^2 \, ds = \sum_{j=1}^n (K_s^j)^2 \, ds < \infty \quad \text{for all } t \geq 0, \ \mathbb{P}\text{-a.s.},$$

where

$$K_t \, dW_t = \sum_{j=1}^n K_t^j \, dW_t^j.$$

7.3 Itô's formula

8. *Itô processes* follow from the definition of stochastic integrals. The expression

$$dX_t = a_t \, dt + b_t \, dW_t \tag{7.3}$$

is short for

$$X_t = X_0 + \int_0^t a_s \, ds + \int_0^t b_s \, dW_s, \quad t \geq 0.$$

Here, we assume that (a_t) and (b_t) are processes that satisfy assumptions ensuring that the two integrals in this expression are well-defined.

Note that an Itô process is a local martingale if and only if

$$a_t = 0 \quad \text{for all } t \geq 0.$$

9. Itô's formula can be memorised by recalling Taylor's series expansion of a smooth function and using the expressions

$$(dW_t)^2 = dt, \quad dW_t \, dt = 0, \quad (dt)^2 = 0, \tag{7.4}$$

which imply that, if X is the Itô process given by (7.3), then

$$\begin{aligned}(dX_t)^2 &= a_t^2 \, (dt)^2 + 2a_t b_t \, dW_t \, dt + b_t^2 \, (dW_t)^2 \\ &= b_t^2 \, dt. \end{aligned} \tag{7.5}$$

10. Given a $C^{1,2}$ function $(t,x) \mapsto f(t,x)$ and the Itô process (X_t) given by (7.3), *Itô's lemma* states that the stochastic process (F_t) defined by $F_t = f(t, X_t)$ is also an Itô process. In particular, *Itô's formula* provides the expression

$$\begin{aligned} df(t,X_t) &= f_t(t,X_t)\,dt + f_x(t,X_t)\,dX_t + \frac{1}{2} f_{xx}(t,X_t)\,(dX_t)^2 \\ &= \left[f_t(t,X_t) + a_t f_x(t,X_t) + \frac{1}{2} b_t^2 f_{xx}(t,X_t) \right] dt + b_t f_x(t,X_t)\,dW_t, \end{aligned} \tag{7.6}$$

where

$$f_t(t,x) = \frac{\partial f(t,x)}{\partial t}, \quad f_x(t,x) = \frac{\partial f(t,x)}{\partial x} \quad \text{and} \quad f_{xx}(t,x) = \frac{\partial^2 f(t,x)}{\partial x^2}.$$

The following is a useful special case:

$$\begin{aligned} df(t,W_t) &= f_t(t,W_t)\,dt + f_x(t,W_t)\,dW_t + \frac{1}{2} f_{xx}(t,W_t)\,(dW_t)^2 \\ &= \left[f_t(t,W_t) + \frac{1}{2} f_{xx}(t,W_t) \right] dt + f_x(t,W_t)\,dW_t. \end{aligned} \tag{7.7}$$

11. If f does not depend explicitly on time, i.e., if $x \mapsto f(x)$ is a C^2 function, then Itô's formula takes the form

$$\begin{aligned} df(X_t) &= f'(X_t)\,dX_t + \frac{1}{2} f''(X_t)\,(dX_t)^2 \\ &= \left[a_t f'(X_t) + \frac{1}{2} b_t^2 f''(X_t) \right] dt + b_t f'(X_t)\,dW_t, \end{aligned} \tag{7.8}$$

where f' and f'' are the first and the second derivative of f, respectively. Also,

$$\begin{aligned} df(W_t) &= f'(W_t)\,dW_t + \frac{1}{2} f''(W_t)\,(dW_t)^2 \\ &= \frac{1}{2} f''(W_t)\,dt + f'(W_t)\,dW_t. \end{aligned} \tag{7.9}$$

12. **Example.** The solution to the stochastic differential equation (SDE)

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t, \tag{7.10}$$

where μ, σ are constants, is given by

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right). \tag{7.11}$$

We can verify this claim in two ways:

Way 1. Noting that

$$\frac{d \ln s}{ds} = \frac{1}{s} \quad \text{and} \quad \frac{d^2 \ln s}{ds^2} = -\frac{1}{s^2},$$

we can use (7.8) to calculate

$$\begin{aligned} d \ln S_t &= \frac{1}{S_t} \, dS_t + \frac{1}{2} \left(-\frac{1}{S_t^2} \right) (dS_t)^2 \\ &= \frac{1}{S_t} \big[\mu S_t \, dt + \sigma S_t \, dW_t \big] - \frac{1}{2S_t^2} (\sigma S_t)^2 \, dt \\ &= \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma \, dW_t, \end{aligned}$$

which implies that

$$\begin{aligned} \ln S_t - \ln S_0 &= \int_0^t d \ln S_u \\ &= \int_0^t \left(\mu - \frac{1}{2}\sigma^2 \right) du + \int_0^t \sigma \, dW_u. \end{aligned}$$

It follows that

$$\begin{aligned} S_t &= e^{\ln S_t} \\ &= \exp \left(\ln S_0 + \left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right), \end{aligned}$$

which establishes that the solution of (7.10) is given by (7.11).

Way 2. We consider the Itô process

$$dX_t = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma \, dW_t,$$

and we define $f(x) = S_0 e^x$, so that

$$f'(x) = f''(x) = f(x).$$

Using Itô's formula (7.8), we can see that the process (S_t) defined by (7.11) satisfies

$$\begin{aligned} dS_t &= df(X_t) \\ &= f'(X_t) \, dX_t + \frac{1}{2} f''(X_t) \, (dX_t)^2 \\ &= \left(\mu - \frac{1}{2}\sigma^2 \right) f(X_t) \, dt + \sigma f(X_t) \, dW_t + \frac{1}{2}\sigma^2 f(X_t) \, dt \\ &= \mu f(X_t) \, dt + \sigma f(X_t) \, dW_t \\ &= \mu S_t \, dt + \sigma S_t \, dW_t, \end{aligned} \tag{7.12}$$

which proves that (S_t) satisfies (7.10).

13. Another useful result of stochastic analysis is the *integration by parts formula*. Given the pair of Itô processes

$$dX_t = a_t\,dt + b_t\,dW_t,$$
$$dY_t = c_t\,dt + e_t\,dW_t,$$

the product process (X_tY_t) is again an Itô process, and

$$\begin{aligned} d\,(X_tY_t) &= X_t\,dY_t + Y_t\,dX_t + (dX_t)(dY_t) \\ &= [Y_ta_t + X_tc_t + b_te_t]\,dt + [Y_tb_t + X_te_t]\,dW_t, \end{aligned} \tag{7.13}$$

where we have used the formal expressions (7.4).

14. Itô's formula can be generalised in a straightforward way to account for multi-dimensional Itô processes.

Suppose that the Brownian motion (W_t) is n-dimensional (see Definition 6.25). Also, consider the Itô processes (X_t^1), ..., (X_t^m) given by

$$dX_t^i = a_t^i\,dt + b_t^i\,dW_t, \quad \text{for } i = 1, \dots, m,$$

where

$$b_t^i\,dW_t = \sum_{j=1}^{n} b_t^{ij}\,dW_t^j,$$

with (a_t^i) and $(b_t^i) = (b_t^{i1}, \dots, b_t^{in})$ being suitable row-vector stochastic processes.

If f is a $C^{1,2,\dots,2}$ function, then Itô's formula provides the expression

$$\begin{aligned} df(t, X_t^1, \dots, X_t^m) &= f_t(t, X_t^1, \dots, X_t^m)\,dt + \sum_{i=1}^{m} f_{x_i}(t, X_t^1, \dots, X_t^m)\,dX_t^i \\ &\quad + \frac{1}{2}\sum_{i,k=1}^{m} f_{x_ix_k}(t, X_t^1, \dots, X_t^m)\left(dX_t^i\right)\left(dX_t^k\right) \\ &= \Bigg(f_t(t, X_t^1, \dots, X_t^m) + \sum_{i=1}^{m} a_t^i f_{x_i}(t, X_t^1, \dots, X_t^m) \\ &\quad + \frac{1}{2}\sum_{i,k=1}^{m}\left(\sum_{\ell=1}^{n} b_t^{i\ell}b_t^{k\ell}\right) f_{x_ix_k}(t, X_t^1, \dots, X_t^m)\Bigg)dt \\ &\quad + \sum_{i=1}^{m} f_{x_i}(t, X_t^1, \dots, X_t^m) b_t^i\,dW_t. \end{aligned} \tag{7.14}$$

The second expression here follows immediately from the first one if we consider the formal expressions

$$(dt)^2 = 0, \quad dW_t^i\,dt = 0 \quad \text{and} \quad dW_t^i\,dW_t^j = \begin{cases} dt, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{7.15}$$

15. Similarly, the integration by parts formula can be generalised in a straightforward way to account for Itô processes driven by a multi-dimensional Brownian motion.

    Suppose that the Brownian motion (W_t) is n-dimensional. Given the pair of Itô processes

$$dX_t = a_t\,dt + b_t\,dW_t, \quad t \geq 0,$$
$$dY_t = c_t\,dt + e_t\,dW_t, \quad t \geq 0,$$

    where (a_t), $(b_t) = (b_t^1, \ldots, b_t^n)$, (c_t) and $(e_t) = (e_t^1, \ldots, e_t^n)$ are suitable row-vector stochastic processes, the product (X_tY_t) is an Itô process such that

$$\begin{aligned} d\,(X_tY_t) &= X_t\,dY_t + Y_t\,dX_t + (dX_t)(dY_t) \\ &= (Y_ta_t + X_tc_t + b_te_t')\,dt + (Y_tb_t + X_te_t)\,dW_t, \end{aligned}$$

    where we have used the formal expressions in (7.15).

7.4 Changes of probability measure

16. We can have many probability measures other than $\mathbb{P}$ defined on the measurable space $(\Omega, \mathcal{F})$. Indeed, let Y be any random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$Y > 0, \quad \mathbb{P}\text{-a.s.}, \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}[Y] = 1.$$

    Here, we write $\mathbb{E}^{\mathbb{P}}$ instead of just $\mathbb{E}$ to indicate that we compute expectations with respect to the probability measure $\mathbb{P}$. We can then define the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by

$$\mathbb{Q}(A) \equiv \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_A] = \mathbb{E}^{\mathbb{P}}[Y\mathbf{1}_A], \quad \text{for } A \in \mathcal{F}. \tag{7.16}$$

    This probability measure has the property that, given any event $A \in \mathcal{F}$,

$$\mathbb{P}(A) = 0 \;\Leftrightarrow\; \mathbb{Q}(A) = 0 \quad \text{and} \quad \mathbb{P}(A) = 1 \;\Leftrightarrow\; \mathbb{Q}(A) = 1.$$

    Any probability measures $\mathbb{P}$ and $\mathbb{Q}$ having this property are called *equivalent*.

17. **Lemma.** The function $\mathbb{Q} : \mathcal{F} \to [0, 1]$ defined by (7.16) is indeed a probability measure.

    Proof. First, we note that the identity $\mathbb{E}^{\mathbb{P}}[Y] = 1$ and the inequalities $0 \leq \mathbf{1}_A \leq 1$ imply that

$$0 \leq \mathbb{Q}(A) \leq 1 \quad \text{for all } A \in \mathcal{F},$$

    so (7.16) defines a function $\mathbb{Q} : \mathcal{F} \to [0, 1]$. In particular,

$$\mathbb{Q}(\emptyset) = \mathbb{E}^{\mathbb{P}}[Y\mathbf{1}_\emptyset] = \mathbb{E}^{\mathbb{P}}[\mathbf{1}_\emptyset] = \mathbb{P}(\emptyset) = 0 \quad \text{and} \quad \mathbb{Q}(\Omega) = \mathbb{E}^{\mathbb{P}}\big[Y\mathbf{1}_\Omega\big] = \mathbb{E}^{\mathbb{P}}\big[Y\big] = 1.$$

Furthermore, given any sequence (A_n) of pairwise disjoint events in $\mathcal{F}$, we can use the monotone convergence theorem and the linearity of expectation to calculate

$$\begin{aligned}\mathbb{Q}\left(\bigcup_{j=1}^{\infty} A_j\right) &= \mathbb{E}^{\mathbb{P}}\left[Y\mathbf{1}_{\bigcup_{j=1}^{\infty} A_j}\right] = \mathbb{E}^{\mathbb{P}}\left[Y\sum_{j=1}^{\infty}\mathbf{1}_{A_j}\right] = \mathbb{E}^{\mathbb{P}}\left[\lim_{n\to\infty}\sum_{j=1}^{n} Y\mathbf{1}_{A_j}\right] \\ &= \lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\left[\sum_{j=1}^{n} Y\mathbf{1}_{A_j}\right] = \lim_{n\to\infty}\sum_{j=1}^{n}\mathbb{E}^{\mathbb{P}}\left[Y\mathbf{1}_{A_j}\right] = \sum_{j=1}^{\infty}\mathbb{E}^{\mathbb{P}}\left[Y\mathbf{1}_{A_j}\right] \\ &= \sum_{j=1}^{\infty}\mathbb{Q}(A_j),\end{aligned}$$

which proves that $\mathbb{Q}$ is countably additive. It follows that $\mathbb{Q}$ has all of the properties that define a probability measure.

18. **Lemma.** Given a random variable Z such that the corresponding expectations are well-defined,

$$\mathbb{E}^{\mathbb{Q}}[Z] = \mathbb{E}^{\mathbb{P}}[YZ] \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}[Z] = \mathbb{E}^{\mathbb{Q}}\left[Y^{-1}Z\right]. \tag{7.17}$$

Proof. We prove this claim using the so-called "measure-theoretic induction", which is a proof technique that is taylor made for this kind of results. First, we assume that Z is a simple random variable, namely,

$$Z = \sum_{j=1}^{n} z_j \mathbf{1}_{\{Z=z_j\}},$$

for some distinct constants $z_1, \ldots, z_n$. In this case, we use the linearity of the expectation and the definition of $\mathbb{Q}$ to calculate

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[Z] &= \mathbb{E}^{\mathbb{Q}}\left[\sum_{j=1}^{n} z_j\mathbf{1}_{\{Z=z_j\}}\right] = \sum_{j=1}^{n} z_j\mathbb{E}^{\mathbb{Q}}\left[\mathbf{1}_{\{Z=z_j\}}\right] = \sum_{j=1}^{n} z_j\mathbb{Q}\left(\{Z=z_j\}\right) \\ &= \sum_{j=1}^{n} z_j\mathbb{E}^{\mathbb{P}}\left[Y\mathbf{1}_{\{Z=z_j\}}\right] = \mathbb{E}^{\mathbb{P}}\left[Y\sum_{j=1}^{n} z_j\mathbf{1}_{\{Z=z_j\}}\right] = \mathbb{E}^{\mathbb{P}}[YZ].\end{aligned}$$

Next, we assume that Z is a positive random variable and we consider any increasing sequence of positive simple random variables (Z_n) such that $\lim_{n\to\infty} Z_n = Z$, $\mathbb{P}$-a.s. (e.g., see Paragraph 4.9). The assumption that each of the random variables Z_n is simple and what we have proved above imply that

$$\mathbb{E}^{\mathbb{Q}}[Z_n] = \mathbb{E}^{\mathbb{P}}[YZ_n] \quad \text{for all } n \geq 1.$$

Combining this observation with the monotone convergence theorem, we can see that

$$\mathbb{E}^{\mathbb{Q}}[Z] = \mathbb{E}^{\mathbb{Q}}\left[\lim_{n\to\infty} Z_n\right] = \lim_{n\to\infty}\mathbb{E}^{\mathbb{Q}}[Z_n] = \lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}[YZ_n] = \mathbb{E}^{\mathbb{P}}\left[\lim_{n\to\infty} YZ_n\right] = \mathbb{E}^{\mathbb{P}}[YZ].$$

If Z is a random variable such that $\mathbb{E}^{\mathbb{Q}}\big[|Z|\big] < \infty$, then we consider the positive random variables Z^+, Z^- satisfying $Z = Z^+ - Z^-$ and $|Z| = Z^+ + Z^-$. Using the fact that the required result holds true if Z is a positive random variable, which we have proved above, we can see that

$$\mathbb{E}^{\mathbb{Q}}[Z] = \mathbb{E}^{\mathbb{Q}}[Z^+] - \mathbb{E}^{\mathbb{Q}}[Z^-] = \mathbb{E}^{\mathbb{P}}[YZ^+] - \mathbb{E}^{\mathbb{P}}[YZ^-] = \mathbb{E}^{\mathbb{P}}[YZ],$$

and the first identity in (7.17) has been established.

Finally, the second identity in (7.17) follows from the observation that

$$\mathbb{E}^{\mathbb{P}}[Z] = \mathbb{E}^{\mathbb{P}}\left[YY^{-1}Z\right] = \mathbb{E}^{\mathbb{Q}}\left[Y^{-1}Z\right].$$

19. Suppose that (L_t) is an $(\mathcal{F}_t)$-martingale with respect to the probability measure $\mathbb{P}$ such that $L_t > 0$, $\mathbb{P}$-a.s., and $\mathbb{E}^{\mathbb{P}}[L_t] = 1$ for all $t \geq 0$. Given a time $T > 0$, we define an equivalent probability measure $\mathbb{Q}$ on the measurable space $(\Omega, \mathcal{F}_T)$ by

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[L_T \mathbf{1}_A] \quad \text{for all } A \in \mathcal{F}_T.$$

Lemma. The process $\left(L_t^{-1},\, t \in [0,T]\right)$ is an $(\mathcal{F}_t)$-martingale with respect to the probability measure $\mathbb{Q}$.

Proof. Fix any times $0 \leq s \leq t \leq T$ and consider any $\mathcal{F}_s$-measurable random variable Z such that the corresponding expectations are well-defined. In view of Lemma 7.18, the tower property and the fact that (L_t) is an $(\mathcal{F}_t)$-martingale with respect to the probability measure $\mathbb{P}$, we can see that

$$\mathbb{E}^{\mathbb{Q}}\left[L_t^{-1}Z\right] = \mathbb{E}^{\mathbb{P}}\left[L_T L_t^{-1}Z\right] = \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}}[L_T \mid \mathcal{F}_t]\, L_t^{-1}Z\Big] = \mathbb{E}^{\mathbb{P}}\left[L_t L_t^{-1}Z\right] = \mathbb{E}^{\mathbb{P}}[Z].$$

Similarly, we can see that $\mathbb{E}^{\mathbb{Q}}[L_s^{-1}Z] = \mathbb{E}^{\mathbb{P}}[Z]$ and conclude that

$$\mathbb{E}^{\mathbb{Q}}\left[L_s^{-1}Z\right] = \mathbb{E}^{\mathbb{Q}}\left[L_t^{-1}Z\right] \quad \text{for all } 0 \leq s \leq t \leq T \text{ and } \mathcal{F}_s\text{-measurable } Z. \tag{7.18}$$

For $s = 0$ and $Z = 1$, this identity implies that $\mathbb{E}^{\mathbb{Q}}\left[L_t^{-1}\right] = 1$ for all $t \in [0,T]$. Furthermore, given any times $0 \leq s < t \leq T$ and any event $A \in \mathcal{F}_s$, we can use (7.18) with $Z = \mathbf{1}_A$ to obtain

$$\mathbb{E}^{\mathbb{Q}}\Big[\mathbb{E}^{\mathbb{Q}}\left[L_t^{-1} \mid \mathcal{F}_s\right] \mathbf{1}_A\Big] = \mathbb{E}^{\mathbb{Q}}\Big[\mathbb{E}^{\mathbb{Q}}\left[L_t^{-1}\mathbf{1}_A \mid \mathcal{F}_s\right]\Big] = \mathbb{E}^{\mathbb{Q}}\left[L_t^{-1}\mathbf{1}_A\right] = \mathbb{E}^{\mathbb{Q}}\left[L_s^{-1}\mathbf{1}_A\right]$$

In light of the definition of conditional expectation, we can see that

$$\mathbb{E}^{\mathbb{Q}}\left[L_t^{-1} \mid \mathcal{F}_s\right] = L_s^{-1},$$

and the result follows.

20. Given a random variable Z such that the corresponding expectations are well-defined, Lemma 7.18 implies that

$$\mathbb{E}^{\mathbb{Q}}[Z] = \mathbb{E}^{\mathbb{P}}[L_T Z] \equiv \frac{\mathbb{E}^{\mathbb{P}}[L_T Z]}{L_0} \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}[Z] = \mathbb{E}^{\mathbb{Q}}\left[L_T^{-1} Z\right] \equiv \frac{\mathbb{E}^{\mathbb{Q}}\left[L_T^{-1} Z\right]}{L_0^{-1}}.$$

The following result, which is also known as *Bayes' theorem*, generalises these identities and is very useful in relating conditional expectations under $\mathbb{P}$ with conditional expectations under $\mathbb{Q}$.

Lemma. If Z is an $\mathcal{F}_T$-measurable random variable such that the corresponding conditional expectations are well-defined, then

$$\mathbb{E}^{\mathbb{Q}}[Z \mid \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]}{L_s} \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}[Z \mid \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{Q}}\left[L_T^{-1} Z \mid \mathcal{F}_s\right]}{L_s^{-1}} \tag{7.19}$$

for all $s \in [0, T]$.

Proof. In view of Lemma 7.19 and the symmetric roles of the probability measures $\mathbb{P}$ and $\mathbb{Q}$, we only need to prove the first identity in (7.19). To this end, we consider the definition of conditional expectation, we observe that both sides of this identity are $\mathcal{F}_s$-measurable random variables in $\mathcal{L}^1$, and we note that, given any event $A \in \mathcal{F}_s$,

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[\frac{\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]}{L_s}\mathbf{1}_A\right] &= \mathbb{E}^{\mathbb{P}}\left[L_T \frac{\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]}{L_s}\mathbf{1}_A\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}\left[L_T \frac{\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]}{L_s}\mathbf{1}_A \mid \mathcal{F}_s\right]\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}[L_T \mid \mathcal{F}_s] \frac{\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]}{L_s}\mathbf{1}_A\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]\mathbf{1}_A\right] \\
&= \mathbb{E}^{\mathbb{P}}[L_T Z \mathbf{1}_A] \\
&= \mathbb{E}^{\mathbb{Q}}[Z \mathbf{1}_A] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[Z \mid \mathcal{F}_s]\mathbf{1}_A\right].
\end{aligned}$$

7.5 Girsanov's theorem

21. Given a constant ϑ, the process

$$L_t = \exp\left(-\frac{1}{2}\vartheta^2 t - \vartheta W_t\right),$$

is an $(\mathcal{F}_t)$-martingale (see Lemma 6.28). If we define the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}\left[L_T \mathbf{1}_A\right] \quad \text{for } A \in \mathcal{F}_T, \tag{7.20}$$

then *Girsanov's theorem* states that the process (W_t^ϑ) defined by

$$W_t^\vartheta = \vartheta t + W_t, \quad \text{for } t \in [0, T],$$

is a standard $(\mathcal{F}_t)$-Brownian motion with respect to the probability measure $\mathbb{Q}$.

In this context,

$$\mathbb{E}^{\mathbb{P}}\left[W_t\right] = 0, \quad \mathbb{E}^{\mathbb{Q}}\left[W_t\right] = -\vartheta t,$$
$$\mathbb{E}^{\mathbb{P}}\left[W_t^\vartheta\right] = \vartheta t \quad \text{and} \quad \mathbb{E}^{\mathbb{Q}}\left[W_t^\vartheta\right] = 0.$$

Also, if (K_t) is a process such that all associated stochastic integrals are well-defined, and all integrals with respect to the associated Brownian motions are martingales, then

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\int_0^t K_u\, dW_u\right] &= 0, \\
\mathbb{E}^{\mathbb{Q}}\left[\int_0^t K_u\, dW_u\right] &= \mathbb{E}^{\mathbb{Q}}\left[\int_0^t K_u\, dW_u^\vartheta - \int_0^t K_u \vartheta\, du\right] \\
&= -\vartheta\, \mathbb{E}^{\mathbb{Q}}\left[\int_0^t K_u\, du\right], \\
\mathbb{E}^{\mathbb{P}}\left[\int_0^t K_u\, dW_u^\vartheta\right] &= \mathbb{E}^{\mathbb{P}}\left[\int_0^t K_u\, dW_u + \int_0^t K_u \vartheta\, du\right] \\
&= \vartheta\, \mathbb{E}^{\mathbb{P}}\left[\int_0^t K_u\, du\right] \\
\text{and} \quad \mathbb{E}^{\mathbb{Q}}\left[\int_0^t K_u\, dW_u^\vartheta\right] &= 0.
\end{aligned}$$

22. In a more general context, suppose that the Brownian motion (W_t) is n-dimensional and let (X_t) be an n-dimensional $(\mathcal{F}_t)$-adapted process satisfying

$$\int_0^t |X_s|^2\, ds < \infty \quad \text{for all } t \geq 0, \quad \mathbb{P}\text{-a.s.}.$$

Under this assumption, the process (L_t) given by

$$L_t = \exp\left(-\frac{1}{2}\int_0^t |X_s|^2\, ds - \int_0^t X_s\, dW_s\right)$$

is well-defined for all t. Using Itô's formula, we can verify that

$$L_t = 1 - \int_0^t X_s L_s\, dW_s, \tag{7.21}$$

so (L_t) is an $(\mathcal{F}_t)$-local martingale.

Under appropriate conditions, the process (L_t) given by (7.21) is a martingale, in which case, $\mathbb{E}\big[L_T\big] = 1$ for all $T \geq 0$. One sufficient condition for (L_t) to be a martingale is *Novikov's condition*:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t |X_s|^2\, ds\right)\right] < \infty \quad \text{for all } t \geq 0.$$

If (L_t) is a martingale, then, given any fixed time $T > 0$, we can define a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by

$$\mathbb{Q}(A) = \mathbb{E}\left[L_T \mathbf{1}_A\right], \quad \text{for } A \in \mathcal{F}_T.$$

Girsanov's theorem states that, given any fixed time $T > 0$, the process $(\widetilde{W}_t)$ defined by

$$\widetilde{W}_t = W_t + \int_0^t X_s\, ds, \quad t \in [0, T]$$

is an n-dimensional $(\mathcal{F}_t)$-Brownian motion with respect to $\mathbb{Q}$.