P1. Probability Spaces

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 10 below), and
   - $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$ (see Definition 29 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.,

   $$\begin{aligned}
   &\omega_1 \text{ identifies with observing heads first and then tails,}\\
   &\omega_2 \text{ identifies with observing heads first and then heads,}\\
   &\omega_3 \text{ identifies with observing tails first and then tails, and}\\
   &\omega_4 \text{ identifies with observing tails first and then heads.}
   \end{aligned}$$

   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, $\{\omega_1\}$ is the event that the coin lands heads first and then tails, $\{\omega_1, \omega_3, \omega_4\}$ is the event that either we observe tails in the second toss or we observe tails first and then heads, 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 $\mathbb{P}\big(\{\omega_1\}\big) = p(1 - p)$.

3. Plainly, the probabilistic aspects of tossing a coin twice can be described with more simplicity than above. In this example, the family of all events of interest $\mathcal{F}$ is the set of all subsets of Ω, so, its introduction may seem unnecessary. Indeed, just the sample space Ω and the probability measure $\mathbb{P}$ assigning the appropriate value in $[0, 1]$ to each subset of Ω provide a sufficient probabilistic model.

   However, this is not the case when one addresses the probabilistic modelling of more complex random situations arising in several areas, including finance. To appreciate this claim, we consider the following example.

4. *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: in Example 7 below, we construct a subset of $(0, 1)$ to which it is not possible to assign a length. 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 the example, such families should include only events that do have a 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.

5. It is worth keeping in mind that there are many different probability spaces: we have seen two different ones in Examples 2 and 4 above. The probabilistic modelling of different random situations may require probability spaces of different structures. Indeed, the probability space that we considered in Example 2 is not sufficiently rich to model the random drawing of a number from the interval $(0, 1)$. On the other hand, a probability space that is appropriate for the example discussed in Example 4 could be viewed as too complex to use for the modelling of tossing a coin twice. In this context, the abstract nature of the probability theory we consider is one of its most appealing features.

   We should also note that we can have many different probability measures $\mathbb{P}$ defined on the same pair $(\Omega, \mathcal{F})$. For example, each value in $(0, 1)$ of the parameter p appearing in Example 2 is associated with a different probability measure. In fact, different consistent belief systems concerning the same random situation correspond to different probability measures. For instance, two financial agents might consistently assign different probabilities to the same market events (e.g., the two agents might be a bull and a bear).

6. It is extremely important to note that the probability theory we consider is concerned with *countably* many operations. “Finite” is too restrictive relative to real life applications. Indeed, every property that holds in the limit involves infinite considerations and / or quantities. On the other hand, we should always keep in mind that the theory we consider in this course is incompatible with uncountably many operations.

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

7. *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.,

$$L\big((a,b)\big) = b-a, \quad L\big(\{a\}\big) = 0 \quad \text{and} \quad L\big((-\infty,a)\big) = L\big((a,\infty)\big) = \infty \tag{1}$$

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 $A, B \subseteq \mathbb{R}$,

$$A \subseteq B \quad \Rightarrow \quad L(A) \leq L(B), \tag{2}$$

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

$$L\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} L(A_n). \tag{3}$$

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

$$L(A^a) = L(A) \quad \text{for all } A \subseteq \mathbb{R} \text{ and } a \in \mathbb{R}, \tag{4}$$

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

$$C^{q_1} \cap C^{q_2} = \emptyset \quad \text{for all } q_1 \neq q_2, \tag{5}$$

and

$$(0,1) \subseteq \bigcup_{q \in (-1,1) \cap \mathbb{Q}} C^q \subseteq (-1,2). \tag{6}$$

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

$$1 \overset{(1)}{=} L\big((0,1)\big) \overset{(2),(6)}{\leq} L\left(\bigcup_{q \in (-1,1) \cap \mathbb{Q}} C^q\right) \overset{(3),(5)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C^q) \overset{(4)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C) = 0,$$

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

$$3 \overset{(1)}{=} L\big((-1,2)\big) \overset{(2),(6)}{\geq} L\left(\bigcup_{q \in (-1,1) \cap \mathbb{Q}} C^q\right) \overset{(3),(5)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C^q) \overset{(4)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C) = \infty,$$

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

π-systems, algebras and σ-algebras

8. *Definition.* A collection $\mathcal{I}$ of subsets of Ω is a *π-system* if it is stable under finite intersections, i.e.,

$$A_1, A_2 \in \mathcal{I} \quad \Rightarrow \quad A_1 \cap A_2 \in \mathcal{I}.$$

9. *Definition.* An *algebra* on Ω is a collection $\mathcal{A}$ of subsets of Ω such that

$$\begin{aligned} &\text{(i) } \Omega \in \mathcal{A}, \\ &\text{(ii) } A \in \mathcal{A} \quad \Rightarrow \quad A^c \equiv \Omega \setminus A \in \mathcal{A}, \\ &\text{(iii) } A_1, A_2 \in \mathcal{A} \quad \Rightarrow \quad A_1 \cup A_2 \in \mathcal{A}. \end{aligned}$$

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

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

11. A simple induction argument reveals that:

    - If $\mathcal{I}$ is a π-system, then

$$A_1, A_2, \ldots, A_n \in \mathcal{I} \quad \Rightarrow \quad \bigcap_{j=1}^{n} A_j \in \mathcal{I}.$$

    - If $\mathcal{A}$ is an algebra, then

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

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

    This should be obvious!

13. *Example.* Suppose that $\Omega = \mathbb{R}$. We can easily verify that the family of sets

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

    is a π-system. However, $\mathcal{I}$ is neither an algebra nor a σ-algebra because, e.g.,

$$(-\infty, a]^c \equiv \mathbb{R} \setminus (-\infty, a] = (a, \infty) \notin \mathcal{I}.$$

14. *Example.* Suppose that $\Omega = \mathbb{R}$. We can easily verify that the family of sets

$$\mathcal{I} = \big\{(a, b) \mid \ a, b \in \mathbb{R},\ a \leq b\big\},$$

    where we adopt the convention that $(a, a) = \emptyset$, is a π-system.

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

$$\emptyset \in \mathcal{F}, \tag{7}$$

and

$$A_1, A_2, \dots, A_n, \dots \in \mathcal{F} \;\Rightarrow\; \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}. \tag{8}$$

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

Similar properties are true for an algebra $\mathcal{A}$ on Ω. In particular, *an algebra is stable under finite set operations*.

Proof. Fix any set Ω and any σ-algebra $\mathcal{F}$ on Ω. Since $\emptyset = \Omega^c$, (7) follows immediately by properties (i) and (ii) of Definition 10. To prove (8), we consider any sequence of events $A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$, 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

$$\begin{aligned}
A_n \in \mathcal{F} \text{ for all } n \geq 1 \quad &\Rightarrow \quad A_n^c \in \mathcal{F} \text{ for all } n \geq 1 && \text{(by property 10.(ii))} \\
&\Rightarrow \quad \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F} && \text{(by property 10.(iii))} \\
&\Rightarrow \quad \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{F} && \text{(by property 10.(ii))}
\end{aligned}$$

and (8) follows.

16. Every σ-algebra is an algebra. However, an algebra is *not necessarily* a σ-algebra.

Example. Suppose that $\Omega = [0, \infty)$, and let

$$\mathcal{A} = \big\{ A \subseteq \Omega \mid \text{ either } A \text{ or } A^c \text{ is a finite set} \big\}.$$

Here, we adopt the convention that the empty set is “finite”.

We can verify that $\mathcal{A}$ is an algebra by checking the properties (i), (ii) and (iii) of Definition 9 as follows:

(i) Since $\Omega = \emptyset^c$ and we have assumed that the empty set $\emptyset$ is finite, $\Omega \in \mathcal{A}$.

(ii) By the definition of $\mathcal{A}$, given is any event $A \in \mathcal{A}$, we also have $A^c \in \mathcal{A}$.

(iii) Fix any events $A_1, A_2 \in \mathcal{A}$. To verify that property (iii) of Definition 9 is satisfied, we have to prove that $A_1 \cup A_2 \in \mathcal{A}$. To this end, we have to distinguish between two possibilities:

- Each of the sets A_1 and A_2 is finite. In this case, the set $A_1 \cup A_2$ is also finite, and so, $A_1 \cup A_2 \in \mathcal{A}$.
- At least one of the sets A_1, A_2, say the set A_1, is infinite. In this case, we observe that

$$(A_1 \cup A_2)^c = A_1^c \cap A_2^c \subseteq A_1^c.$$

Since A_1^c is finite (because $A_1 \in \mathcal{A}$ and A_1 is infinite), it follows that $(A_1 \cup A_2)^c$ has a finite number of elements, and so, $A_1 \cup A_2 \in \mathcal{A}$.

On the other hand, $\mathcal{A}$ is *not* a σ-algebra. To see this, consider, e.g., the sequence of events defined by

$$A_n = \{n\}, \quad n = 1, 2, \ldots.$$

Clearly, $A_n \in \mathcal{A}$ for all $n \geq 1$. However,

$$\bigcup_{n=1}^{\infty} A_n = \mathbb{N}^* \notin \mathcal{A},$$

because neither $\mathbb{N}^*$ nor $\mathbb{N}^{*c} = [0, \infty) \setminus \mathbb{N}^*$ has a finite number of elements. This shows that property (iii) of Definition 10 is not necessarily satisfied.

Generating σ-algebras

17. *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 10 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 10 [*Exercise*].

18. 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

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

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

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

19. *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.

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

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

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 $\mathcal{C}$.

21. In Definition 19, the σ-algebra $\sigma(\mathcal{C})$ generated by a given family of events $\mathcal{C}$ is "constructed" from "above" or from "outside". It should be stressed that attempting to construct $\sigma(\mathcal{C})$ from "inside" by repeated or countable operations may fail! Indeed, constructions from "inside" are in general possible only if the family $\mathcal{C}$ is "small", e.g., if it has countable elements (see the following two examples).

22. *Example*. Given a subset A of Ω, the smallest σ-algebra containing A is $\{\Omega, \emptyset, A, A^c\}$.

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

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

Then

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

24. *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.,

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

Borel σ-algebras are very important: they contain every event of practical interest!

25. *Example.* $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I})$, where $\mathcal{I}$ is the π-system of Example 13.

Proof. In view of the Definition 24 of the Borel σ-algebra on $\mathbb{R}$ and the observation in Paragraph 20 above, we can prove this claim as follows.

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

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

Measures

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

27. *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.*

28. *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.

29. 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

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

30. *Lemma.* Fix a measure space $(\Omega, \mathcal{F}, \mu)$. Given a finite number of pairwise disjoint events $A_1, A_2, \ldots, A_n \in \mathcal{F}$,

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i). \tag{9}$$

Proof. Given a finite number of pairwise disjoint events $A_1, \ldots, A_n \in \mathcal{F}$, we consider the sequence of events $B_1, \ldots, B_n, B_{n+1}, \ldots \in \mathcal{F}$ defined by

$$B_i = \begin{cases} A_i & \text{if } i = 1, \ldots, n, \\ \emptyset & \text{if } i = n+1, \ldots. \end{cases}$$

Noting that the events in the sequence (B_n) are pairwise disjoint, we can use the countable additivity of the measure μ (see Definition 27) to calculate

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{n} A_i\right) &= \mu\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum_{i=1}^{\infty} \mu(B_i) \\ &= \sum_{i=1}^{n} \mu(A_i) + \sum_{i=n+1}^{\infty} \mu(\emptyset) \\ &= \sum_{i=1}^{n} \mu(A_i). \end{aligned}$$

31. *Lemma*. Let $(S, \mathcal{S}, \mathbb{P})$ be a probability space. Then,

$$\text{if } A \subseteq B, \text{ then } \mathbb{P}(A) \le \mathbb{P}(B), \tag{10}$$
$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \tag{11}$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \tag{12}$$
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i). \tag{13}$$

[*Exercise*: How do these properties change if we consider a more general measure space $(S, \mathcal{S}, \mu)$?]

Proof. In view of (9) that we proved above, (10) is true because, if $A \subseteq B$, then

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

Also, (11) follows immediately from the calculations

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

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 (12).

In view of (12) and the positivity of probabilities,

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

Using this inequality and a straightforward induction argument, we can then obtain inequality (13) [*Exercise*].

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

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

In the presence of this definition,

$$\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n). \tag{14}$$

Similarly, if $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ is a *decreasing* sequence of events in $\mathcal{H}$, 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

$$\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n). \tag{15}$$

Proof. Given an increasing sequence $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ of events in $\mathcal{H}$, 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 $\mathcal{H}$ 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

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

which establishes (15).

[*Exercise*: Where is the assumption $\mu(A_1) < \infty$ needed in the proof of (15) above?]

33. In the previous result, the validity of (15) relies heavily on the assumption that $\mu(A_1) < \infty$ (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 $\Omega = \mathbb{R}$, $\mathcal{H} = \mathcal{B}(\mathbb{R})$ and $\mu = L$, where L is the Lebesgue measure that maps each set $C \in \mathcal{B}(\mathbb{R})$ to its length $L(C)$ (see also Paragraph 37 below). If we define $A_n = [n, \infty)$, for $n \geq 1$, then we can see that

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

Construction and uniqueness of measures

34. Although σ-algebras often do not include every possible subset of the underlying sample spaces, they can be sufficiently "rich" to accommodate all events of practical interest. This implies that σ-algebras can be awkward to handle. It is for this reason that we consider "coarser" classes of events such as π-systems and algebras.

35. *Theorem* (*Uniqueness of extension*). Suppose that $\mathcal{I}$ is a π-system on a set Ω, and let $\mathcal{F} = \sigma(\mathcal{I})$. Also, let μ_1 and μ_2 be two measures on $(\Omega, \mathcal{F})$ such that

$$\mu_1(A) = \mu_2(A) \quad \text{for all } A \in \mathcal{I},$$

and suppose that there exists a sequence $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{I}$ such that

$$\mu_1(A_n) = \mu_2(A_n) < \infty \text{ for all } n \geq 1, \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n = \Omega.$$

Then

$$\mu_1(A) = \mu_2(A) \quad \text{for all } A \in \mathcal{F},$$

Informally, *if two measures agree and are σ-finite on a π-system, then they agree on the σ-algebra generated by that π-system.*

36. *Theorem* (*Carathéodory's extension theorem*). Suppose that $\mathcal{A}$ is an algebra on a set Ω, and let $\mathcal{F} = \sigma(\mathcal{A})$. Also, suppose that a function $\mu_0 : \mathcal{A} \to [0, \infty]$ is countably additive in the sense that

    (i) $\mu_0(\emptyset) = 0$, and

    (ii) if $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{A}$ is any sequence of pairwise disjoint sets such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, then

    $$\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n).$$

    Then there exists a measure μ on $(\Omega, \mathcal{F})$ such that

    $$\mu(A) = \mu_0(A) \quad \text{for all } A \in \mathcal{A}.$$

37. *The Lebesgue measure.* The Lebesgue measure maps sets $C \subseteq \mathbb{R}$ to which we can assign a length to their length $L(C)$ (recall Examples 4 and 7).

    Let $\mathcal{A}$ be the collection of sets $C \subseteq \mathbb{R}$ that admit the representation

    $$C = (a_1, b_1] \cup \cdots \cup (a_k, b_k] \tag{16}$$

    or

    $$C = (-\infty, b_0] \cup (a_1, b_1] \cup \cdots \cup (a_k, b_k] \tag{17}$$

    or

    $$C = (a_1, b_1] \cup \cdots \cup (a_k, b_k] \cup (a_{k+1}, \infty) \tag{18}$$

    or

    $$C = (-\infty, b_0] \cup (a_1, b_1] \cup \cdots \cup (a_k, b_k] \cup (a_{k+1}, \infty), \tag{19}$$

    for some $k \geq 1$ and reals $-\infty < b_0 < a_1 < b_1 < \cdots < a_k < b_k < a_{k+1} < \infty$, together with $\mathbb{R}$ and $\emptyset$. We can check that this collection is an algebra on $\mathbb{R}$ and that $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ [*Exercise*].

    We define the function $L_0 : \mathcal{A} \to [0, \infty]$ by associating each set $C \in \mathcal{A}$ with its length, so that

    $$L_0(C) = \infty \quad \text{if } C \text{ is as in (17), (18) or (17)}$$

and

$$L_0(C) = \sum_{j=1}^{k} b_j - a_j \quad \text{if } C \text{ is as in (16).}$$

It can be shown that L_0 is countably additive. By appealing to the the Carathéodory's extension Theorem 36 and the uniqueness Theorem 35 (with the π-system of Example 14), we can then see that L_0 has a unique extension to a measure L on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the Lebesgue measure.

Given an interval (α, β), where $-\infty \leq \alpha < \beta \leq \infty$, we can follow the same steps to construct the Lebesgue measure on $\big((\alpha, \beta), \mathcal{B}\big((\alpha, \beta)\big)\big)$.

P2. Random Variables and Distribution Functions

Preliminary considerations

1. Let us consider choosing a person randomly from within a given population of N people. Assuming that all people in the population are equally likely to be chosen, the choices

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

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 for the modelling of this random situation.

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 $Y(\omega)$, their gender $Z(\omega)$ or their cultural background $U(\omega)$. Each of these quantities is a *random variable*. The random variables X and Y take values in the set of positive real numbers, the random variable Z takes values in the set of all possible genders, say $\{\text{male}, \text{female}\}$, while the random variable U takes values in the set of all possible cultural backgrounds. Unlike U, the random variables X, Y and Z can be quantified in practice.

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 $\mathbb{R}$, the Euclidean space $\mathbb{R}^n$, or the set $C([0, \infty))$ of all continuous real-valued functions on $[0, \infty)$. In the context of the example that we discuss here, we therefore focus on random variables such as X, Y, and Z with a definition such as the following one:

$$Z(\omega) = \begin{cases} \sqrt{2}, & \text{if } \omega \text{ is female,} \\ -\pi, & \text{if } \omega \text{ is male.} \end{cases}$$

After the random choice of a person from within the given population, every random variable is a known quantity, e.g., a given real number. 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 gender $Z(\omega)$. Of course, this story involves a degree of idealisation: if somebody is weighed before and after a big lunch, then the scale's readings are most likely to be different numbers. Having observed this, we note that all this discussion is intended as a simple motivating example rather than a serious mathematical modelling.

2. Generalising the example above, a ("useful") random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function mapping Ω into a set of (mathematical) objects S. Accordingly, each sample $\omega \in \Omega$ is associated with a unique (mathematical) object $X(\omega) \in S$.

However intuitive such a definition may be at this stage, it is too general to provide a useful working ground: see Paragraph 6 below for the appropriate definition.

3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space of Paragraph 1 and let X be the random variable that associates a person's rounded at the first decimal point height in metres $X(\omega)$ with each person $\omega \in \Omega$. For instance, $X(\omega) = 1.7\text{m}$ if the actual height of person ω is in the range $[1.65\text{m}, 1.75\text{m})$. Accordingly, the random variable X takes values in the set $\{\ldots, 1.5, 1.6, 1.7, 1.8, \ldots\}$.

   On the measurable space $(S, \mathcal{S})$ given by

$$S = \{\ldots, 1.5, 1.6, 1.7, 1.8, \ldots\} \quad \text{and} \quad \mathcal{S} = \mathcal{P}(S),$$

   which is associated with the set S in which X takes values, we can define a function $\overline{\mathbb{P}} : \mathcal{S} \to [0, 1]$ by

$$\begin{aligned} \overline{\mathbb{P}}(\emptyset) &= 0, \\ \overline{\mathbb{P}}\big(\{x\}\big) &= \frac{\text{number of people } \omega \in \Omega \text{ with height in } [x - 0.5, x + 0.5)}{N} \\ &= \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) = x\}\big), \qquad \text{for } x \in S, \end{aligned}$$

   and

$$\overline{\mathbb{P}}(A) = \sum_{x_j \in A} \overline{\mathbb{P}}\big(\{x_j\}\big), \quad \text{for } A \subseteq S \text{ with } A \neq \emptyset.$$

   We can then verify that $\overline{\mathbb{P}}$ is a probability measure on $(S, \mathcal{S})$. This probability measure is called the *distribution* of X, and, given any $A \in \mathcal{S}$, $\overline{\mathbb{P}}(A)$ is the answer to the question "what is the probability that the recorded value of X turns out to be in the range given by the set A? ".

4. To proceed further, let us consider Example 4 of Chapter P1. In particular, let us consider drawing a number from the interval $(0, 1)$ in a completely random way, so that the probability of the event that the number drawn is in the set $A \subseteq (0, 1)$ is equal to the length (i.e., the Lebesgue measure) $L(A)$ of A. Also, let C be a subset of $(0, 1)$ to which we can assign no length, so that the probability $L(C)$ of C cannot be defined (such an event exists: see Paragraph 7 of Chapter P1).

   In this context, let X be the "random variable" that takes the value 1 if the number drawn turns out to be inside C anad 0 otherwise, which is defined by

$$X(\omega) = \mathbf{1}_C(\omega) = \begin{cases} 1, & \text{if } \omega \in C, \\ 0, & \text{if } \omega \notin C. \end{cases}$$

   This "random variable" X does not have a distribution. Indeed, the probability that X turns out to be 1 is equal to the probability $L(C)$ of the event C, which is not defined.

5. Motivated by the discussion above, let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a function $X : \Omega \to S$, where S is a set. If $\mathcal{S}$ is a σ-algebra on S such that

$$X^{-1}(C) \equiv \{\omega \in \Omega \mid \ X(\omega) \in C\} \in \mathcal{F} \quad \text{for all } C \in \mathcal{S}, \tag{20}$$

then we can define a probability measure $\overline{\mathbb{P}}$ on the measurable space $(S, \mathcal{S})$ by

$$\overline{\mathbb{P}}(C) = \mathbb{P}\left(X^{-1}(C)\right) \equiv \mathbb{P}\left(\{\omega \in \Omega \mid X(\omega) \in C\}\right), \quad \text{for } C \in \mathcal{S}. \tag{21}$$

This probability measure, which is often denoted by $\mathbb{P}X^{-1}$ is called the *distribution* of X, or the *image law* of $\mathbb{P}$ under X, or just the *law* of X.

6. *The definition of $\overline{\mathbb{P}}$ in (21) is well-posed because (20) is satisfied.*

   Since the distributions of random variables are of fundamental importance in probability theory, this observation suggests that the following is the appropriate definition of a real-valued random variable:

   Definition. A *random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ and with values in a measurable space $(S, \mathcal{S})$ is any function $X : \Omega \to S$ such that

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

   In particular, a *real-valued random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ is any function $X : \Omega \to \mathbb{R}$ such that

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

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

7. Let us now consider the following issue: what information will we get if we observe a given random variable? To fix ideas, let us consider the probability space of Example 1 above, and let us denote by X the random variable considered in Paragraph 3.

   What do we learn by observing X? In other words, what information do we get if a measurement is made and we are reported the value of X only? Assuming that the population size is large (e.g., $N \geq 100$), the observed value of X does not necessarily identify ω: for instance, there may be several people with height in the interval $[1.65\text{m}, 1.75\text{m})$, and all of them are associated with the same reading 1.7 of X. However, the value of X can tell us whether or not the person involved in the measurement belongs to several classes, namely subsets of Ω. Indeed, as soon as we are reported the observed value of X, we can provide a clear yes or no as answer to each question of the form: does the person involved in the measurement belong to the set of people with an X-measurement in the set $A \subseteq \mathbb{R}$. For instance, if we are reported that the value of X is somewhere in the range $(1.367, 1.5] \cup [1.77,\ 1.823]$, then we are certain that person involved in the measurement belongs to the family of those people with height in the range $[1.35\text{m}, 1.55\text{m}) \cup [1.75\text{m}, 1.85)$. Therefore, we can see that learning the value of X will inform us on whether the person chosen for the measurement belongs or not to any of the groups of people in the collection

$$\left\{\{\omega \in \Omega \mid X(\omega) \in A\} \;\middle|\; A \in \mathcal{B}(\mathbb{R})\right\} \subseteq \mathcal{F},$$

   which is a σ-algebra on Ω.

8. Generalising the previous example, we can identify the information we obtain by the observation of an $(S, \mathcal{S})$-valued random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the σ-algebra

$$\sigma(X) = \left\{ \{\omega \in \Omega \mid \; X(\omega) \in A\} \,\middle|\, A \in \mathcal{S} \right\} \subseteq \mathcal{F}. \tag{22}$$

In particular, we can identify the information we obtain by the observation of a real-valued random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the σ-algebra

$$\sigma(X) = \left\{ \{\omega \in \Omega \mid \; X(\omega) \in A\} \,\middle|\, A \in \mathcal{B}(\mathbb{R}) \right\} \subseteq \mathcal{F}, \tag{23}$$

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

9. *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, let us consider the following very simple 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.

   Of course, *after* we observe X, we have more information. Indeed, if we have been reported that $X = 1$, then the information we possess is

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

   while, if we have been reported that $X = 0$, then we know that each event in

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

   has occurred or not.

10. Building on the ideas discussed above, we can now ask the following question: what is the information set we get if we observe two random variables, say X and Y? It might be tempting to say that this could be identified with $\sigma(X) \cup \sigma(Y)$, where $\sigma(\cdot)$ is a σ-algebra defined as in (23). However, a closer scrutiny shows that this is not a good idea.

   To fix ideas, suppose that a fair coin that lands heads (H) or tails (T) is tossed three times. The choices

$$\Omega = \big\{\mathrm{HHH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{TTH}, \mathrm{THT}, \mathrm{HTT}, \mathrm{TTT}\big\} \quad \text{and} \quad \mathcal{F} = \mathcal{P}(\Omega)$$

   provide a measurable space that is appropriate for the modelling of this random situation. Now, let X be the outcome of the first toss, and let Y be the total number of heads observed in the three coin tosses: the random variable X takes values in the set $\{\mathrm{H}, \mathrm{T}\}$, while the random variable Y takes values in the set $\{0, 1, 2, 3\}$. We can check that

$$\sigma(X) = \big\{\Omega, \emptyset, \{\mathrm{HHH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{HTT}\}, \{\mathrm{THH}, \mathrm{TTH}, \mathrm{THT}, \mathrm{TTT}\}\big\}$$

   and

$$\begin{aligned}\sigma(Y) = \big\{&\Omega, \emptyset, \{\mathrm{TTT}\}, \{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}\}, \{\mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\}, \{\mathrm{HHH}\}\\ &\{\mathrm{TTT}, \mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}\}, \{\mathrm{TTT}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\}, \{\mathrm{TTT}, \mathrm{HHH}\},\\ &\{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\}, \{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{HHH}\},\\ &\{\mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{HHH}\}, \{\mathrm{TTT}, \mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\},\\ &\{\mathrm{TTT}, \mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{HHH}\}, \{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{HHH}\}\big\}.\end{aligned}$$

   In this context, the choice $\sigma(X) \cup \sigma(Y)$ for the information set generated by the observation of the random variables X and Y is not appropriate. Indeed, we can be sure that we will be able to say whether the event $\{\mathrm{HHT}, \mathrm{HTH}\}$ has occurred or not once we observe X and Y:

$$\begin{aligned}&\text{if } X = \mathrm{H} \text{ and } Y = 0 \text{ or } 1 \text{ or } 3, \text{ then } \{\mathrm{HHT}, \mathrm{HTH}\} \text{ has not occurred,}\\ &\text{if } X = \mathrm{T} \text{ and } Y = 0 \text{ or } 1 \text{ or } 2, \text{ then } \{\mathrm{HHT}, \mathrm{HTH}\} \text{ has not occurred}\end{aligned}$$

   and

$$\text{if } X = \mathrm{H} \text{ and } Y = 2, \text{ then } \{\mathrm{HHT}, \mathrm{HTH}\} \text{ has occurred.}$$

   The event $\{\mathrm{HHT}, \mathrm{HTH}\}$ does not belong to $\sigma(X) \cup \sigma(Y)$. However, it does belong to the σ-algebra generated by $\sigma(X) \cup \sigma(Y)$, namely, $\sigma\big(\sigma(X) \cup \sigma(Y)\big)$.

11. In light of the discussion in the previous paragraph, we can identify the information we obtain by the observation of the random variables in a given family $(X_i,\ i \in I)$, where $I \neq \emptyset$ is an index set, with the σ-algebra

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

Random variables

12. *Definition.* A *real-valued random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ is a real-valued function on Ω (i.e., $X : \Omega \to \mathbb{R}$) such that

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{B}(\mathbb{R}), \tag{24}$$

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

Note that we often write $\{X \in A\}$ to denote the event $X^{-1}(A)$.

13. *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 (i.e., $X : \Omega \to S$) such that

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{S}. \tag{25}$$

14. *Definition.* The σ-algebra $\sigma(X)$ generated by a real-valued random variable X is the collection of all sets $X^{-1}(A) = \{\omega \in \Omega \mid \ X(\omega) \in A\}$, where $A \in \mathcal{B}(\mathbb{R})$, i.e.,

$$\sigma(X) = \big\{X^{-1}(A) \mid \ A \in \mathcal{B}(\mathbb{R})\big\}. \tag{26}$$

15. *Lemma.* The family of events $\sigma(X)$ defined by (26) is indeed a σ-algebra.

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 $\Omega = X^{-1}(\mathbb{R})$ 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 (26) of $\sigma(X)$ implies that there exists $A \in \mathcal{B}(\mathbb{R})$ such that

$$C = X^{-1}(A) \equiv \{\omega \in \Omega \mid \ X(\omega) \in A\}.$$

Now, we calculate

$$\begin{aligned} \Omega \setminus C &= \Omega \setminus \{\omega \in \Omega \mid \ X(\omega) \in A\} \\ &= \{\omega \in \Omega \mid \ X(\omega) \notin A\} \\ &= \{\omega \in \Omega \mid \ X(\omega) \in \mathbb{R} \setminus A\} \\ &= X^{-1}(\mathbb{R} \setminus A) \ \in \ \sigma(X), \end{aligned}$$

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 (26) 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

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

Now, we calculate

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

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

16. *Definition*. The σ-algebra $\sigma(X)$ generated by an $(S, \mathcal{S})$-valued random variable X is the collection of all sets $X^{-1}(A) = \{\omega \in \Omega \mid \ X(\omega) \in A\}$, where $A \in \mathcal{S}$, i.e.,

$$\sigma(X) = \big\{ X^{-1}(A) \mid \ A \in \mathcal{S} \big\}. \tag{27}$$

17. *Definition*. The σ-algebra generated by a collection of random variables $(X_i, \ i \in I)$, where $I \neq \emptyset$, is

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

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

19. 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.

As we have discussed above, σ-algebras can be considered as models of information. 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 $\mathcal{H}$.

20. *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

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for all } A \in \mathcal{C}, \tag{28}$$

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

[To appreciate the significance of this result, do compare (28) with (25)!]

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

$$X^{-1}(A) \in \mathcal{F} \quad \text{for all } A \in \mathcal{S},$$

or equivalently, if we show that

$$\Big\{A \in \mathcal{S} \mid \ X^{-1}(A) \in \mathcal{F}\Big\} = \mathcal{S}. \tag{29}$$

To this end, we define

$$\mathcal{H} = \Big\{A \in \mathcal{S} \mid \ X^{-1}(A) \in \mathcal{F}\Big\},$$

and we note that

$$\mathcal{C} \subseteq \mathcal{H} \subseteq \mathcal{S}, \tag{30}$$

where the first inclusion follows thanks to (28).

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

(i) $S \in \mathcal{H}$ because $X^{-1}(S) = \Omega \in \mathcal{F}$.

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

$$X^{-1}(S \setminus A) = \Omega \setminus X^{-1}(A) \in \mathcal{F},$$

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

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

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

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

Now, in view of the assumption that $\sigma(\mathcal{C}) = \mathcal{S}$, (30), 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 (29).

21. *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 20 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

$$\{\omega \in \Omega \mid \ X(\omega) + Y(\omega) > a\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}. \tag{31}$$

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,

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

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 (31) follows.

Now, we use Lemma 20 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

$$\{\omega \in \Omega \mid \ \lambda X(\omega) \le a\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}. \tag{32}$$

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

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

All of the events on the right hand side of this expression belong to $\mathcal{F}$ (because X is a random variable), and (32) follows.

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

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

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

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

expresses XY as a sum of random variables.

22. *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

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

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

$$\Big(\inf_{n\geq 1} X_n\Big)(\omega) = \inf_{n\geq 1} X_n(\omega), \qquad \Big(\sup_{n\geq 1} X_n\Big)(\omega) = \sup_{n\geq 1} X_n(\omega),$$
$$\Big(\liminf_{n\to\infty} X_n\Big)(\omega) = \liminf_{n\to\infty} X_n(\omega) \quad \text{and} \quad \Big(\limsup_{n\to\infty} X_n\Big)(\omega) = \limsup_{n\to\infty} X_n(\omega),$$

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

$$\mathcal{B}\big([-\infty,\infty]\big) = \sigma\big(\big\{[-\infty,a] \mid \ a \in [-\infty,\infty]\big\}\big) \supseteq \mathcal{B}\big(\mathbb{R}\big). \tag{33}$$

Furthermore,

$$\big\{\omega \in \Omega \mid \ \lim_{n\to\infty} X_n(\omega) \text{ exists in } \mathbb{R}\big\} \in \mathcal{F}. \tag{34}$$

Proof. In view of Lemma 20 and (33), we can see that the inclusion

$$\begin{aligned}\Big(\sup_{n\geq 1} X_n\Big)^{-1}\big([-\infty,a]\big) &= \Big\{\omega\in\Omega \mid \ \sup_{n\geq 1} X_n(\omega) \leq a\Big\} \\ &= \bigcap_{n=1}^{\infty} \{\omega\in\Omega \mid \ X_n(\omega)\leq a\} \in \mathcal{F} \quad \text{for all } a\in[-\infty,\infty],\end{aligned}$$

implies that $\sup_{n\geq 1} X_n$ 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 21), 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.

Now, 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 (34) follows immediately from the identity

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

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

23. *Lemma.* Suppose that X and Y are real-valued random variables such that Y is $\sigma(X)$-measurable. Then there exists a Borel function $g : \mathbb{R} \to \mathbb{R}$ such that $Y = g(X)$.

    More generally, given n real-valued random variables $X_1,\ldots,X_n$, suppose that a random variable Y is $\sigma(X_1,\ldots,X_n)$-measurable. Then there exist a Borel function $g : \mathbb{R}^n \to \mathbb{R}$ such that $Y = g(X_1,\ldots,X_n)$.

Distributions

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

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

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

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

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

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

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

we can see that

$$\begin{aligned} F(a) &= \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) \leq a\}\big) \\ &\leq \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) \leq b\}\big) \\ &= F(b). \end{aligned}$$

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

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

(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

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

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

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

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

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

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

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

(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 $a \in \mathbb{R}$,

$$\begin{aligned}\lim_{n\to\infty} F\left(a + \frac{1}{n}\right) &= \lim_{n\to\infty} \mathbb{P}\left(\left\{\omega \in \Omega \mid \ X(\omega) \leq a + \frac{1}{n}\right\}\right) \\ &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \left\{\omega \in \Omega \mid \ X(\omega) \leq a + \frac{1}{n}\right\}\right) \\ &= \mathbb{P}\left(\{\omega \in \Omega \mid \ X(\omega) \leq a\}\right) \\ &= F(a).\end{aligned}$$

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

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

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

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

We can compute

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

and

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

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

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

Discrete random variables

29. *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

$$p_j = \mathbb{P}\big(X = x_j\big) > 0. \tag{35}$$

30. In view of (35), the distribution function of a discrete random variable X is given by

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

Also,

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

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

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

is called the *indicator* of A. We say that the random variable $\mathbf{1}_A$ is *Bernoulli*. The probability mass function of the random variable $\mathbf{1}_A$ is given by

$$p_1 = \mathbb{P}\big(\mathbf{1}_A = 1\big) \equiv \mathbb{P}(A) \quad \text{and} \quad p_0 = \mathbb{P}\big(\mathbf{1}_A = 0\big) \equiv \mathbb{P}(A^c).$$

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

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

where

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

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.

33. *Example.* A random variable X has the *Poisson* distribution with parameter λ if its probability mass function is given by

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

Here, $\lambda > 0$. In such a case, we write $X \sim \mathbb{P}(\lambda)$.

Continuous random variables

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

$$\mathbb{P}\big(X \in A\big) = \int_A f(x)\,dx \quad \text{for all } A \in \mathcal{B}(\mathbb{R}). \tag{36}$$

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

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

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 (36) implies

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

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

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

Here, $m \in \mathbb{R}$ and $\sigma > 0$. We often write $X \sim N(m, \sigma^2)$. Normal random variables are also called *Gaussian.*

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

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

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

$$\begin{aligned} F(a) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-m}{\sigma}} \exp\left(-\frac{y^2}{2}\right) dy. \\ &= \Phi\left(\frac{a-m}{\sigma}\right). \end{aligned} \tag{37}$$

37. *Example.* A random variable X has the *exponential* distribution with parameter μ if its probability density function is given by

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

Here, $\mu > 0$.

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

The Monotone Class Theorem

39. *Theorem.* Let $\mathcal{M}$ be a family of bounded real-valued functions on a set Ω, i.e., a collection of bounded functions $f : \Omega \to \mathbb{R}$, such that

(i) $\mathcal{M}$ is a vector space, i.e.,

$$a_1, a_2 \in \mathbb{R} \text{ and } f_1, f_2 \in \mathcal{M} \quad \Rightarrow \quad a_1 f_1 + a_2 f_2 \in \mathcal{M},$$

(ii) $\mathcal{M}$ contains the constant functions, i.e., $\mathbf{1}_\Omega \in \mathcal{M}$, and

(iii) if $f_1, f_2, \ldots, f_n, \ldots$ is an increasing sequence of positive elements of $\mathcal{M}$ (i.e., $0 \leq f_1 \leq f_2 \leq \cdots \leq f_n \leq \cdots$) such that $f = \sup_{n \geq 1} f_n$ is bounded, then $f \in \mathcal{M}$.

If there exists a subset $\mathcal{C}$ of $\mathcal{M}$ that is stable under pairwise multiplication, i.e.,

$$f_1, f_2 \in \mathcal{C} \quad \Rightarrow \quad f_1 f_2 \in \mathcal{C},$$

then $\mathcal{M}$ contains all bounded $\sigma(\mathcal{C})$-measurable functions.

P3. Independence

1. In what follows, 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 $\mathcal{F}$.

2. *Definition.* The σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n, \ldots$ are called *independent* if, for every sequence of events $A_i \in \mathcal{G}_i$, $i \geq 1$, and any distinct $i_1, \ldots, i_n$,

$$\mathbb{P}\big(A_{i_1} \cap \cdots \cap A_{i_n}\big) = \prod_{k=1}^{n} \mathbb{P}\big(A_{i_k}\big).$$

3. This definition implies that the σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent if, for any choice of events $A_1 \in \mathcal{G}_1$, $A_2 \in \mathcal{G}_2$, $\ldots$, $A_n \in \mathcal{G}_n$,

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

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

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

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

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

then, clearly $\big\{\Omega, \emptyset, A, A^{\mathrm{c}}\big\} = \sigma(\mathbf{1}_A)$. As a consequence, the events $A_1, A_2, \ldots$ are independent if and only if the random variables $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \ldots$ are independent.

7. *Example.* Three events A_1, A_2, A_3 are independent if

$$\mathbb{P}\big(A_1 \cap A_2 \cap A_3\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big)\mathbb{P}\big(A_3\big), \tag{38}$$

$$\mathbb{P}\big(A_1 \cap A_2\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big), \tag{39}$$

$$\mathbb{P}\big(A_1 \cap A_3\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_3\big), \tag{40}$$

$$\mathbb{P}\big(A_2 \cap A_3\big) = \mathbb{P}\big(A_2\big)\mathbb{P}\big(A_3\big). \tag{41}$$

Proof. To verify that the events A_1, A_2, A_3 are independent, we have to check that

$$\mathbb{P}(C \cap D \cap E) = \mathbb{P}(C)\mathbb{P}(D)\mathbb{P}(E),$$
$$\text{for all } C \in \big\{\Omega, \emptyset, A_1, A_1^{\mathrm{c}}\big\},\ D \in \big\{\Omega, \emptyset, A_2, A_2^{\mathrm{c}}\big\} \text{ and } E \in \big\{\Omega, \emptyset, A_3, A_3^{\mathrm{c}}\big\}. \tag{42}$$

In other words, we have to prove that (38)–(41) imply each of the $4 \times 4 \times 4 = 64$ relations in (42). To this end, we calculate, for example,

$$\begin{aligned}
\mathbb{P}(A_1^c \cap \Omega \cap A_3) &= \mathbb{P}(A_3 \setminus (A_1 \cap A_3)) \\
&= \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_3) \\
&\overset{(40)}{=} \big[\mathbb{P}(A_1^c) + \mathbb{P}(A_1)\big]\mathbb{P}(A_3) - \mathbb{P}(A_1)\mathbb{P}(A_3) \\
&= \mathbb{P}(A_1^c)\mathbb{P}(\Omega)\mathbb{P}(A_3), \qquad (43)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A_1^c \cap A_2 \cap A_3) &= \mathbb{P}\big((A_2 \cap A_3) \setminus (A_1 \cap A_2 \cap A_3)\big) \\
&= \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_2 \cap A_3) \\
&\overset{(41),(38)}{=} \mathbb{P}(A_2)\mathbb{P}(A_3) - \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3) \\
&= \big[\mathbb{P}(A_1^c) + \mathbb{P}(A_1)\big]\mathbb{P}(A_2)\mathbb{P}(A_3) - \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3) \\
&= \mathbb{P}(A_1^c)\mathbb{P}(A_2)\mathbb{P}(A_3), \qquad (44)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A_1^c \cap A_2^c \cap A_3) &= \mathbb{P}\big((A_1^c \cap A_3) \setminus (A_1^c \cap A_2 \cap A_3)\big) \\
&= \mathbb{P}(A_1^c \cap A_3) - \mathbb{P}(A_1^c \cap A_2 \cap A_3) \\
&\overset{(43),(44)}{=} \mathbb{P}(A_1^c)\mathbb{P}(A_3) - \mathbb{P}(A_1^c)\mathbb{P}(A_2)\mathbb{P}(A_3) \\
&= \mathbb{P}(A_1^c)\big[\mathbb{P}(A_2^c) + \mathbb{P}(A_2)\big]\mathbb{P}(A_3) - \mathbb{P}(A_1^c)\mathbb{P}(A_2)\mathbb{P}(A_3) \\
&= \mathbb{P}(A_1^c)\mathbb{P}(A_2^c)\mathbb{P}(A_3).
\end{aligned}$$

All other possibilities can be proved using similar arguments (convince yourselves!).

8. *Lemma.* The 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

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

P4. Expectation

1. In what follows, we assume that an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is fixed.

Preliminary considerations

2. Suppose that we consider the toss of a coin that lands heads with probability $p \in (0,1)$ and tails with probability $1-p$. Let us 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.

Definitions

4. *Definition.* We say that X is a *simple random variable* 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., events $A_1, A_2, \ldots, A_n \in \mathcal{F}$ satisfying $A_i \cap A_j = \emptyset$, for $i \neq j$, and $\bigcup_{i=1}^n A_i = \Omega$) such that

$$X(\omega) = \sum_{i=1}^n x_i 1_{A_i}(\omega) \quad \text{for all } \omega \in \Omega. \tag{45}$$

5. *Definition.* The *expectation* of the simple random variable X given by (45) is defined by

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

6. *Definition.* Suppose that X is a *positive* random variable (i.e., $X(\omega) \in [0, \infty]$ for all $\omega \in \Omega$; here, we allow for $X(\omega)$ to be $+\infty$). The *expectation* of X is defined by

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

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

7. *Definition.* Let X be any random variable, and let

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

   Then $X = X^+ - X^-$, $|X| = X^+ + X^-$, and X^+, X^- are positive random variables.

   A random variable X has *finite expectation* (is *integrable*) if both of $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are finite. 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 $1 \le p < \infty$, 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

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

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) $\mathcal{L}^1$ is a vector space, i.e.,

$$X, Y \in \mathcal{L}^1 \text{ and } a, b \in \mathbb{R} \quad \Rightarrow \quad aX + bY \in \mathcal{L}^1.$$

Expectation is a positive, linear map on $\mathcal{L}^1$, i.e.,

$$X \geq 0 \quad \Rightarrow \quad \mathbb{E}[X] \geq 0,$$

and

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

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

(iii) (*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.

(iv) (*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

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

(v) (*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

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

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

12. *Example*. Suppose that $\Omega = (0,1)$ and $\mathcal{F} = \mathcal{B}\big((0,1)\big)$, and let $\mathbb{P}$ be 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.

Also, this example shows that the inequalities in Fatou's lemma can be strict.

13. *Lemma*. The expectation of a discrete random variable X is given by

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

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

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

14. *Example*. Suppose that X has the Poisson distribution. We can calculate the mean of X as follows:

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

15. *Example*. Suppose that X is a Gaussian random variable with mean m and variance σ^2. We can calculate the mean of X as follows:

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

Here, we have used the fact that the first integral is 0 because the integrand is an odd function, and the fact that the second integral is equal to one because the integrand is a probability density function.

P5. Conditional Expectation

1. In what follows, we assume that an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is fixed.

Preliminary considerations on conditional probability

2. 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 somebody informs us 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. A situation such as this one arises whenever we are provided with information in a random environment, e.g., when we observe the outcome of an "experiment".

   How should we modify our views, namely our probability measure, to account for this scenario? Let us denote by $\mathbb{P}(B \mid A)$ 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 $\mathbb{P}(B \mid A)$ should be proportional to $\mathbb{P}(B \cap A)$, namely

$$\mathbb{P}(B \mid A) \sim \mathbb{P}(B \cap A). \tag{46}$$

   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 $\mathbb{P}(\Omega \cap A) = \mathbb{P}(A)$, we conclude that we should scale the right hand side of (46) by $1/\mathbb{P}(A)$ to obtain

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}. \tag{47}$$

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

   We can check that the function $\mathbb{P}(\cdot \mid A) : \mathcal{F} \to [0, 1]$ defined (47) is indeed a probability measure on $(\Omega, \mathcal{F})$.

3. *Law of total probability.* Suppose that the events $A_1, A_2, \ldots, A_n \in \mathcal{F}$ 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 (47) imply

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \cdots + \mathbb{P}(B \cap A_n) \\ &= \mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B \mid A_n)\mathbb{P}(A_n). \end{aligned} \tag{48}$$

4. *Bayes' theorem.* Suppose that the events $A_1, A_2, \ldots, A_n \in \mathcal{F}$ form a partition of Ω. Given an event $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$, and any $k = 1, 2, \ldots, n$, (47) and the law of total probability (48) imply

$$\begin{aligned} \mathbb{P}(A_k \mid B) &= \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(B \mid A_k)\mathbb{P}(A_k)}{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B \mid A_n)\mathbb{P}(A_n)}. \end{aligned} \tag{49}$$

5. Conditional probabilities defined as in (47) 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 2 above, the natural answer is to set

$$\begin{aligned}\mathbb{P}(B \mid \text{“observation of } A \text{ or } A^c\text{”}) &= \begin{cases}\mathbb{P}(B \mid A) & \text{if } A \text{ occurs,}\\ \mathbb{P}(B \mid A^c) & \text{if } A^c \text{ occurs,}\end{cases}\\ &= \mathbb{P}(B \mid A)\mathbf{1}_A + \mathbb{P}(B \mid A^c)\mathbf{1}_{A^c},\end{aligned} \tag{50}$$

   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*. Assuming that “nature” decides on $\omega \in \Omega$, the conditional probability of 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 $\omega \in A^c$.

   This *prior to observation of A or A^c* point of view underlies the modern definition of conditional probability and expectation.

Preliminary considerations on conditional expectation

6. let us consider two simple random variables X and Z, so 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 real numbers $x_1, \ldots, x_n$, and $z_1, \ldots, z_m$. Also, let us 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

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

   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 value of the random variable Z, it is natural to define *the conditional expectation of X given the information we can* *obtain from the observation of the random variable Z* by

$$\mathbb{E}\left[X \mid \sigma(Z)\right] \equiv \mathbb{E}\left[X \mid Z\right] = \sum_{j=1}^{m} \mathbb{E}\left[X \mid Z = z_j\right] \mathbf{1}_{\{Z=z_j\}},$$

which is a random variable.

7. Recall from the chapter where we defined random variables that σ-algebras are mathematical models for information. A further intuitive idea behind conditional expectation is the following. Suppose that we are interested in a random variable X, say the price of a certain stock at some future date. Now, suppose that the only information that becomes available to us is provided by a σ-algebra $\mathcal{G}$. Then $\mathbb{E}[X \mid \mathcal{G}]$ is the expected value of X if we are given this information.

   To appreciate more this point, we consider the following "extreme" cases.

   - Suppose that the information $\mathcal{G}$ contains the "knowledge" of the actual value of X, i.e., X is $\mathcal{G}$-measurable ($\sigma(X) \subseteq \mathcal{G}$). In this case, $\mathbb{E}[X \mid \mathcal{G}] = X$ (see 14.(iii) below), which reflects the idea that "knowledge" of $\mathcal{G}$ implies "knowledge" of the actual value of X.
   - Suppose that X is independent of the information $\mathcal{G}$. In this case, $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$ (see 14.(xi) below), which conveys the idea that "knowledge" of $\mathcal{G}$ provides no information about 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". In this context, the result $\mathbb{E}[X \mid \{\Omega, \emptyset\}] = \mathbb{E}[X]$ (see 14.(i) below), which expresses expectation as a "special case" of conditional expectation is only the "expected" one.

Definitions and existence

8. *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 $C \in \mathcal{G}$,

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

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

9. *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 8.

   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..

10. *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

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

Special cases and examples

11. *Example.* Let $A \in \mathcal{F}$ be an event such that $0 < \mathbb{P}(A) < 1$. The conditional probability of an event $B \in \mathcal{F}$ given $\{\emptyset, \Omega, A, A^c\}$ is given by

    $$\begin{aligned} \mathbb{P}\big(B \mid \{\emptyset, \Omega, A, A^c\}\big)(\omega) &= \begin{cases} \mathbb{P}(A \cap B)/\mathbb{P}(A), & \text{if } \omega \in A, \\ \mathbb{P}(A^c \cap B)/\mathbb{P}(A^c), & \text{if } \omega \in A^c, \end{cases} \\ &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \mathbf{1}_A + \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)} \mathbf{1}_{A^c}. \end{aligned} \tag{51}$$

    [Do observe that this is the same as (50).]

    Proof. We can check that the random variable Y given by

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

    satisfies the defining properties of conditional probability (see Definition 10 above) as follows.

(i) It is straightforward to see that the simple random variable Y is $\{\emptyset, \Omega, A, A^c\}$-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

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

12. *Lemma.* Suppose that X and Z are simple random variables, so 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 real numbers $x_1, \dots, x_n$ and $z_1, \dots, z_m$, and that $\mathbb{P}(Z = z_j) > 0$ for all $j = 1, \dots, m$. We define the *conditional probability mass function* of X given Z by

$$p_{X|Z}(x_i|z_j) = \frac{\mathbb{P}(X = x_i, Z = z_j)}{\mathbb{P}(Z = z_j)}, \quad \text{for } i = 1, \dots, n \text{ and } j = 1, \dots, m.$$

A version of the conditional expectation $\mathbb{E}[X \mid \sigma(Z)]$ of X given $\sigma(Z)$ is given by

$$\begin{aligned}
\mathbb{E}[X \mid Z](\omega) &= \sum_{j=1}^{m} \mathbb{E}[X \mid Z = z_j] \mathbf{1}_{\{Z = z_j\}}(\omega) \\
&= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} x_i \, p_{X|Z}(x_i|z_j) \right) \mathbf{1}_{\{Z = z_j\}}(\omega).
\end{aligned}$$

Here, we adopt the usual habit of writing $\mathbb{E}[X \mid Z]$ instead of $\mathbb{E}[X \mid \sigma(Z)]$, which simplifies the notation.

Proof. First, we observe that the σ-algebra $\sigma(Z)$ is easy to describe: it consists of all possible unions of sets in the family $\big\{\{Z = z_1\}, \dots, \{Z = z_m\}\big\}$, namely,

$$\sigma(Z) = \left\{ \bigcup_{k \in J} \{Z = z_k\} \;\middle|\; J \subseteq \{1, \dots, m\} \right\}, \tag{52}$$

with the convention that

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

Now, we can verify that the random variable Y defined by

$$Y = \sum_{j=1}^{m}\sum_{i=1}^{n} x_i\, p_{X|Z}(x_i|z_j)\mathbf{1}_{\{Z=z_j\}}$$

has the properties (i)–(iii) in Definition 8 above as follows.

(i) In view of (52), 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

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

(iii) Let C be any event in $\sigma(Z)$. In view of (52), 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

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

13. *Lemma*. 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. Now, we define the *conditional probability density function* of X given Z by

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

If we set

$$I(z) = \int_{-\infty}^{\infty} x\, f_{X|Z}(x|z)\, dx,$$

then $I(Z)$ is a version of the conditional expectation $\mathbb{E}[X \mid Z]$ of X given Z.

Properties of conditional expectation

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

(i) $\mathbb{E}[X \mid \{\Omega, \emptyset\}] = \mathbb{E}[X]$.

(ii) If Y is a version of $\mathbb{E}[X \mid \mathcal{G}]$, then $\mathbb{E}[Y] = \mathbb{E}[X]$.

(iii) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$, $\mathbb{P}$-a.s..

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

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

Precisely, if Y_1 is a version of $\mathbb{E}[X_1 \mid \mathcal{G}]$ and Y_2 is a version of $\mathbb{E}[X_2 \mid \mathcal{G}]$, then $a_1 Y_1 + a_2 Y_2$ is a version of $\mathbb{E}[a_1 X_1 + a_2 X_2 \mid \mathcal{G}]$.

(v) (*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

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

(vi) (*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

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

(vii) (*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

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

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

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

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

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

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

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

(xi) (*Independence*) If $\mathcal{H}$ is independent of $\sigma(\sigma(X) \cup \mathcal{G})$, then

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

In particular, if $\sigma(X)$ and $\mathcal{H}$ are independent σ-algebras,

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

P6. Stochastic processes

1. In what follows, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables considered are defined.

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.

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

$$\mathcal{F}_t \subseteq \mathcal{F} \text{ for all } t \in \mathcal{T}, \quad \text{and} \quad \mathcal{F}_s \subseteq \mathcal{F}_t \text{ for all } s, t \in \mathcal{T} \text{ such that } s \leq t. \tag{53}$$

   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 (53) 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

$$\tau^{-1}\big([0, t]\big) = \{\omega \in \Omega \mid \ \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \text{for all } t \in \mathcal{T}. \tag{54}$$

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 information $\mathcal{F}_t$ that is available to us at time 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 $\tau = \min\big\{\tau_1, \tau_2\big\}$ is an $(\mathcal{F}_t)$-stopping time.

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

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

    Therefore

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

    which proves the claim.

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

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

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

Proof. Since $|X_1 + X_2 + \cdots + X_n| \le |X_1| + |X_2| + \cdots + |X_n|$, the assumption that $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ for all $n \ge 1$, implies that $\mathbb{E}[|S_n|] < \infty$ for all $n \ge 1$. Moreover, the assumption that $X_1, X_2, \ldots$ are independent implies that

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

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 $t \in \mathcal{T}$. Furthermore, given any times $s < t$, the tower property of conditional expectation implies

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

Brownian motion

20. *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., $W_t - W_s \sim N(0, t-s)$.

21. In this definition, it is important to observe that we must require *any finite number* of increments in non-overlapping time intervals to be independent. Indeed, there exists a process (X_t) which satisfies (i), (ii), (iv) of Definition 20, and

   (iii') any two increments of (X_t) are independent, i.e., given any times $a < b < c$, the increments $X_b - X_a$ and $X_c - X_b$ are independent random variables,

   that is *not* a Brownian motion.

22. 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,

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

23. *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:

   - $B_0 = W_T - W_{T-0} = 0$.
   - Plainly, (B_t) has continuous sample paths because this is true for (W_t).

- Given $0 \le t_1 < t_2 < \cdots < t_k \le T$, 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 $B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}$ are independent random variables because this is true for the random variables $W_{T-t_1} - W_{T-t_2}, \ldots, W_{T-t_{k-1}} - W_{T-t_k}$ which are increments of the Brownian motion (W_t) in non-overlapping time intervals.
- Given any times $s < t$, since $B_t - B_s = -(W_{T-t} - W_{t-s})$, $B_t - B_s \sim N(0, t-s)$ because (W_t) is a Brownian motion, and so, $W_{T-t} - W_{t-s} \sim N(0, t-s)$.

24. *Example*. Given $s < t$, the properties of the normal distribution imply

$$\mathbb{E}[(W_{t+2} - W_t)^3] = 0 \quad \text{and} \quad \mathbb{E}[(W_t - W_s)^4] = 3(t-s)^2.$$

25. *Lemma*. The sample paths of the standard Brownian motion are nowhere differentiable functions, $\mathbb{P}$-a.s..

26. *Definition*. An n-dimensional standard Brownian motion (W_t) is a vector $(W_t^1, \ldots, W_t^n)$ of independent standard one-dimensional Brownian motions $(W_t^1), \ldots, (W_t^n)$.

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

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

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

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

28. *Lemma*. Every $(\mathcal{G}_t)$-Brownian motion (W_t) is a $(\mathcal{G}_t)$-martingale.

Proof. First, we note that the inequalities

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

imply that $W_t \in \mathcal{L}^1$ for all $t \ge 0$. Next, we observe that, given any times $s < t$,

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

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

P7. Itô calculus

1. Throughout this 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). We also denote by $(\mathcal{F}_t^W)$ the natural filtration of W.

   For technical reasons, we assume that every filtration $(\mathcal{G}_t)$ we consider is right-continuous, i.e., $\mathcal{G}_t = \bigcap_{\varepsilon>0} \mathcal{G}_{t+\varepsilon}$ for all $t \geq 0$, as well as augmented by the $\mathbb{P}$-negligible sets in $\mathcal{G}_\infty = \sigma\left(\bigcup_{t\geq 0} \mathcal{G}_t\right)$; we do not expand on such issues here.

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 $(\mathcal{F}_t)$-*progressively measurable*. This measurability assumption is slightly stronger than assuming that (K_t) is $(\mathcal{F}_t)$-adapted.

   Note that every $(\mathcal{F}_t)$-adapted process with continuous (or, more generally, with right-continuous and left-limited) sample paths is $(\mathcal{F}_t)$-progressively measurable.

   We also assume that the sample paths of (K_t) are "reasonable" in the sense that
$$\int_0^t K_s^2\, ds < \infty \quad \text{for all } t \geq 0, \ \mathbb{P}\text{-a.s.}.$$

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 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 2 above, then its stochastic integral satisfies the *Itô isometry*:
$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T K_s\, dW_s\right)^2\right] &= \mathbb{E}\left[\int_0^T K_s^2\, ds\right] \\ &= \int_0^T \mathbb{E}\left[K_s^2\right]\, ds. \qquad (55)\end{aligned}$$

6. *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, $a \wedge b = \min(a, b)$.

   Note that a local martingale is *not* necessarily a martingale.

7. Every Itô integral is a local martingale.

   An Itô integral that satisfies

$$\mathbb{E}\left[\left(\int_0^T K_s\, dW_s\right)^2\right] < \infty \quad \text{for all } T \geq 0,$$

   is a *square integrable* martingale.

   Note that *not every* martingale is a square integrable one.

Itô's formula

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

$$dX_t = a_t\,dt + b_t\,dW_t \tag{56}$$

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.

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

$$(dW_t)^2 = dt, \quad dW_t\,dt = 0, \quad (dt)^2 = 0, \tag{57}$$

which imply that, if X is the Itô process given by (56), then

$$\begin{aligned}(dX_t)^2 &= a_t^2\,(dt)^2 + 2a_tb_t\,dW_t\,dt + b_t^2\,(dW_t)^2\\ &= b_t^2\,dt.\end{aligned} \tag{58}$$

At this point, it should be stressed that the expressions (57) and (58) are not rigorous mathematics: one should use them as a mnemonic rule only!

10. Given a $C^{1,2}$ function $(t,x) \mapsto f(t,x)$ and the Itô process (X_t) given by (56), *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

$$\begin{aligned}df(t,X_t) &= f_t(t,X_t)\,dt + f_x(t,X_t)\,dX_t + \frac{1}{2}f_{xx}(t,X_t)\,(dX_t)^2\\ &= \left[f_t(t,X_t) + a_tf_x(t,X_t) + \frac{1}{2}b_t^2f_{xx}(t,X_t)\right]dt + b_tf_x(t,X_t)\,dW_t,\end{aligned} \tag{59}$$

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:

$$\begin{aligned}df(t,W_t) &= f_t(t,W_t)\,dt + f_x(t,W_t)\,dW_t + \frac{1}{2}f_{xx}(t,W_t)\,(dW_t)^2\\ &= \left[f_t(t,W_t) + \frac{1}{2}f_{xx}(t,W_t)\right]dt + f_x(t,W_t)\,dW_t.\end{aligned} \tag{60}$$

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

$$\begin{aligned} df(X_t) &= f'(X_t)\,dX_t + \frac{1}{2} f''(X_t)\,(dX_t)^2 \\ &= \left[a_t f'(X_t) + \frac{1}{2} b_t^2 f''(X_t) \right] dt + b_t f'(X_t)\,dW_t, \end{aligned} \tag{61}$$

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

$$\begin{aligned} df(W_t) &= f'(W_t)\,dW_t + \frac{1}{2} f''(W_t)\,(dW_t)^2 \\ &= \frac{1}{2} f''(W_t)\,dt + f'(W_t)\,dW_t. \end{aligned} \tag{62}$$

12. *Example*. The solution of the stochastic equation

$$dS_t = \mu_t S_t\,dt + \sigma_t S_t\,dW_t \tag{63}$$

is given by

$$S_t = S_0 \exp\left(\int_0^t \left(\mu_u - \tfrac{1}{2}\sigma_u^2\right) du + \int_0^t \sigma_u\,dW_u \right). \tag{64}$$

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 (61) to calculate

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

which implies that

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

It follows that

$$
\begin{aligned}
S_t &= e^{\ln S_t} \\
&= \exp\left(\ln S_0 + \int_0^t \left(\mu_u - \frac{1}{2}\sigma_u^2\right) du + \int_0^t \sigma_u \, dW_u\right),
\end{aligned}
$$

which establishes that the solution of (63) is given by (64).

Way 2. We consider the Itô process

$$
dX_t = \left(\mu_t - \tfrac{1}{2}\sigma_t\right) dt + \sigma_t \, dW_t,
$$

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

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

Using Itô's formula (61), we can see that the process (S_t) defined by (64) satisfies

$$
\begin{aligned}
dS_t &= df(X_t) \\
&= f'(X_t) \, dX_t + \frac{1}{2} f''(X_t) \, (dX_t)^2 \\
&= \left(\mu_t - \tfrac{1}{2}\sigma_t^2\right) f(X_t) \, dt + \sigma_t f(X_t) \, dW_t + \tfrac{1}{2}\sigma_t^2 f(X_t) \, dt \\
&= \mu_t S_t \, dt + \sigma_t S_t \, dW_t, \qquad (65)
\end{aligned}
$$

which proves that (S_t) satisfies (63).

13. The results above can be generalised in a straightforward way to account for multi-dimensional Itô processes.

    To fix ideas, let (W_t) be an n-dimensional Brownian motion, and consider the Itô processes (X_t^1), ..., (X_t^m) given by

$$
dX_t^i = a_t^i \, dt + \sum_{j=1}^{n} b_t^{ij} \, dW_t^j, \quad \text{for } i = 1, \ldots, m,
$$

    where (a_t^i) and $\left(b_t^{ij}\right)$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, are appropriate stochastic processes.

    To simplify the notation, we can introduce a "vector formalism" and write

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

    where (b_t^i) is the vector process given by $b_t^i = (b_t^1, \ldots, b_t^m)'$.

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

$$\begin{aligned}
df(t, X_t^1, \ldots, X_t^m) &= f_t(t, X_t^1, \ldots, X_t^m)\, dt + \sum_{i=1}^m f_{x_i}(t, X_t^1, \ldots, X_t^m)\, dX_t^i \\
&\quad + \frac{1}{2} \sum_{i,k=1}^m f_{x_i x_k}(t, X_t^1, \ldots, X_t^m) \left(dX_t^i\right) \left(dX_t^k\right) \\
&= \Bigg[f_t(t, X_t^1, \ldots, X_t^m) + \sum_{i=1}^m a_t^i f_{x_i}(t, X_t^1, \ldots, X_t^m) \\
&\quad + \frac{1}{2} \sum_{i,k=1}^m \left(\sum_{\ell=1}^n b_t^{i\ell} b_t^{k\ell} \right) f_{x_i x_k}(t, X_t^1, \ldots, X_t^m) \Bigg] dt \\
&\quad + \sum_{i=1}^m f_{x_i}(t, X_t^1, \ldots, X_t^m) b_t^i \cdot dW_t. \qquad (66)
\end{aligned}$$

It is worth noting that the second expression here follows immediately from the first one if we consider the *formal* expressions

$$(dt)^2 = 0, \quad dW_t^i\, dt = 0 \quad \text{and} \quad dW_t^i\, dW_t^j = \begin{cases} dt, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \qquad (67)$$

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

$$\begin{aligned}
dX_t &= a_t\, dt + b_t \cdot dW_t, \quad t \geq 0, \\
dY_t &= c_t\, dt + d_t \cdot dW_t, \quad t \geq 0,
\end{aligned}$$

the product process $(X_t Y_t)$ is again an Itô process, and

$$\begin{aligned}
d\,(X_t Y_t) &= X_t\, dY_t + Y_t\, dX_t + (dX_t)(dY_t) \\
&= [Y_t a_t + X_t c_t + b_t' d_t]\, dt + [Y_t b_t + X_t d_t] \cdot dW_t, \qquad (68)
\end{aligned}$$

where we have used the formal expressions (67).

Martingale representation theorem

15. Stochastic integrals are local martingales. Is it true that every local martingale can be written as a stochastic integral? The answer is yes if information coincides with the natural filtration $(\mathcal{F}_t^W)$ of the Brownian motion (W_t).

16. Suppose that (W_t) is an n-dimensional Brownian motion. The *martingale representation theorem* states that, given any $(\mathcal{F}_t^W)$-local martingale (M_t), there exist a constant M_0 and an appropriate process (K_t) such that

$$M_t = M_0 + \int_0^t K_s \cdot dW_s.$$

Changes of probability measures

17. 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

$$Y \geq 0, \quad \mathbb{P}\text{-a.s.}, \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}\big[Y\big] = 1. \tag{69}$$

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

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}\left[Y \mathbf{1}_A\right] \quad \text{for all } A \in \mathcal{F}. \tag{70}$$

18. The probability measure $\mathbb{Q}$ defined by (70) has the property that

$$\text{given any } A \in \mathcal{F}, \quad \mathbb{P}(A) = 0 \quad \Rightarrow \quad \mathbb{Q}(A) = 0. \tag{71}$$

The construction in Paragraph 17 has a converse:

Let $\mathbb{Q}$ be any probability measure on $(\Omega, \mathcal{F})$ such that (71) is true. Then, there exists a random variable Y satisfying (69) such that (70) is true. Moreover, this random variable is unique $\mathbb{P}$-a.s..

In this case, we say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$, and we write $\mathbb{Q} \ll \mathbb{P}$. The random variable Y is the associated *Radon-Nikodym derivative*, and we write

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = Y \quad \text{or} \quad d\mathbb{Q} = Y \, d\mathbb{P}. \tag{72}$$

If $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, then we say that $\mathbb{Q}$ and $\mathbb{P}$ are *equivalent*, and we write $\mathbb{Q} \sim \mathbb{P}$.

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 can define a probability measure $\mathbb{Q}_T$ on the measurable space $(\Omega, \mathcal{F}_T)$ by

$$\mathbb{Q}_T(A) = \mathbb{E}^{\mathbb{P}}\left[L_T \mathbf{1}_A\right] \quad \text{for all } A \in \mathcal{F}_T. \tag{73}$$

The family of probability measures $\{\mathbb{Q}_T, \ T \geq 0\}$ thus arising is consistent in the following sense. Given any times $t < T$, we can use the properties of conditional expectation to calculate

$$\begin{aligned}
\text{for all } A \in \mathcal{F}_t, \quad \mathbb{Q}_T(A) &= \mathbb{E}^{\mathbb{P}}\left[L_T \mathbf{1}_A\right] \\
&= \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}}\left[L_T \mathbf{1}_A \mid \mathcal{F}_t\right]\Big] \\
&= \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}}\left[L_T \mid \mathcal{F}_t\right] \mathbf{1}_A\Big] \\
&= \mathbb{E}^{\mathbb{P}}\left[L_t \mathbf{1}_A\right] \\
&= \mathbb{Q}_t(A).
\end{aligned}$$

In other words, if we consider the *restriction* of the probability measures $\mathbb{P}$ and $\mathbb{Q}_T$ on the σ-algebra $\mathcal{F}_t$, then $d\mathbb{Q}_T = L_t \, d\mathbb{P}$ for all $t \in [0, T]$.

This observation shows that martingales are intimately related to changes of probability measures.

20. In the context of the previous paragraph, if Z is an $\mathcal{F}_T$-measurable random variable satisfying appropriate integrability conditions, then

$$\mathbb{E}^{\mathbb{Q}_T}[Z \mid \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]}{L_s} \quad \text{for all } s \in [0, T]. \tag{74}$$

This is a very useful result!

To see (74), 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 $A \in \mathcal{F}_s$,

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

Girsanov's theorem

21. Suppose that the $(\mathcal{F}_t)$-Brownian motion is n-dimensional, and let (X_t) be an n-dimensional, $(\mathcal{F}_t)$-progressively measurable 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

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

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

$$L_t = 1 + \int_0^t L_s X_s \cdot dW_s \quad \text{for all } t \geq 0, \tag{75}$$

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

22. Under appropriate conditions, the process (L_t) defined by (75) 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*:

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

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

$$\mathbb{Q}_T(A) = \mathbb{E}\left[L_T \mathbf{1}_A\right] \quad \text{for all } A \in \mathcal{F}_T. \tag{76}$$

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

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

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

24. Given a constant ϑ, the process

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

where (W_t) is a one-dimensional Brownian motion, is an $(\mathcal{F}_t)$-martingale. It follows that, if $\mathbb{Q}_T$ is the probability measure defined on the measurable space $(\Omega, \mathcal{F}_T)$ by (76), then the process

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

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

In this context,

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

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

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

and

$$\mathbb{E}^{\mathbb{Q}_T}\left[\int_0^t K_u\, dW_u^{\vartheta}\right] = 0.$$

Exercises 1

1. Suppose that $\Omega = \mathbb{R}$. Which of the following families of sets are σ-algebras on Ω?

$$\begin{aligned}
&\text{(i) } \mathcal{F} = \big\{\Omega, \emptyset, (-\infty, a], (b, \infty) \mid \ a, b \in \mathbb{R}\big\},\\
&\text{(ii) } \mathcal{F} = \big\{A \mid \ \text{either } A \text{ or } A^c \text{ is countable}\big\},\\
&\text{(iii) } \mathcal{F} = \big\{\mathbb{R}, \emptyset, (-\infty, 5], (5, \infty), (-\infty, 3), [3, \infty), [3, 5], (-\infty, 3) \cup (5, \infty)\big\}.
\end{aligned}$$

2. Find the σ-algebra on Ω generated by $\mathcal{C}$ if

$$\begin{aligned}
&\text{(i) } \Omega = \mathbb{R} \quad \text{and} \quad \mathcal{C} = \big\{(-20, \sqrt{2}), (-15, \infty)\big\},\\
&\text{(ii) } \Omega = \mathbb{R} \quad \text{and} \quad \mathcal{C} = \big\{(1, 2], \{2\}\big\},\\
&\text{(iii) } \Omega = \{1, 2, 3, 4\} \quad \text{and} \quad \mathcal{C} = \big\{\emptyset, \{2, 3\}\big\},\\
&\text{(iv) } \Omega = \{1, 2, 3, 4\} \quad \text{and} \quad \mathcal{C} = \big\{\{3\}, \{2, 3, 4\}\big\}.
\end{aligned}$$

3. Consider a measurable space $(\Omega, \mathcal{F})$ and any set $\Omega' \subseteq \Omega$. Prove that the family of sets

$$\mathcal{H} = \big\{A \cap \Omega' \mid \ A \in \mathcal{F}\big\}$$

is a σ-algebra on Ω'.

4. (*First Borel-Cantelli lemma*) Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and any sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$ such that

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty. \tag{77}$$

Prove that

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0. \tag{78}$$

Note: The event $\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ is also called "A_n infinitely often (i.o.)" and is denoted by "$\limsup_{n\to\infty} A_n$".

Hint: Observe that the sequence of events $B_n := \bigcup_{m=n}^{\infty} A_m$ is decreasing, and use the "continuity" of a probability measure.

5. Consider two sets Ω, S and a function $f : \Omega \to S$. Given a set $A \subseteq S$, the inverse image $f^{-1}(A)$ of A under f is defined by

$$\begin{aligned}
f^{-1}(A) &= \{\omega \in \Omega \mid \ f(\omega) \in A\}\\
&= \{\omega \in \Omega \mid \ \text{there exists } x \in A \text{ such that } f(\omega) = x\}.
\end{aligned} \tag{79}$$

Given a collection $(A_i,\ i \in I)$ of subsets of S, where $I \neq \emptyset$ is an index set, we can see that

$$\begin{aligned}
\omega \in f^{-1}\left(\bigcup_{i\in I} A_i\right) \quad &\Leftrightarrow \quad \text{there exists } x \in \bigcup_{i\in I} A_i \text{ such that } f(\omega) = x \\
&\Leftrightarrow \quad \text{for some } i \in I,\ \text{there exists } x \in A_i \text{ such that } f(\omega) = x \\
&\Leftrightarrow \quad \text{there exists } i \in I \text{ such that } \omega \in f^{-1}(A_i) \\
&\Leftrightarrow \quad \omega \in \bigcup_{i\in I} f^{-1}(A_i). \qquad (80)
\end{aligned}$$

These equivalences prove that

$$f^{-1}\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f^{-1}(A_i). \qquad (81)$$

Use a similar reasoning to show that

$$f^{-1}\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f^{-1}(A_i). \qquad (82)$$

Also, show that

$$f^{-1}(S \setminus A) = \Omega \setminus f^{-1}(A) \quad \text{for all } A \subseteq S. \qquad (83)$$

Exercises 2

1. Consider a real-valued random variable X and a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$f^{-1}(C) = \big\{ a \in \mathbb{R} \mid \ f(a) \in C \big\} \in \mathcal{B}(\mathbb{R}) \quad \text{for all } C \in \mathcal{B}(\mathbb{R}).$$

Show that $\sigma\big(f(X)\big) \subseteq \sigma(X)$, and conclude that $f(X)$ is a random variable.

2. Suppose that a random variable X can take only four possible values, i.e., suppose that there exist distinct $x_1, x_2, x_3, x_4 \in \mathbb{R}$ such that

$$X(\omega) \in \{x_1, x_2, x_3, x_4\} \quad \text{for all } \omega \in \Omega.$$

Describe explicitly the σ-algebra $\sigma(X)$ generated by X.

3. Given a random variable X, prove that

(i) if $\mathcal{H} = \{\emptyset, \Omega\}$, then X is $\mathcal{H}$-measurable if and only if X is constant, and

(ii) if $\mathcal{H}$ is a σ-algebra such that $\mathbb{P}(A) = 0$ or 1 for every $A \in \mathcal{H}$, then $\mathbb{P}(X = c) = 1$, for some constant c.

4. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable space $(S, \mathcal{S})$, and let X be an $(S, \mathcal{S})$-valued random variable defined on $(\Omega, \mathcal{F})$, i.e., let X be a function mapping Ω into S such that

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

Also, define the function $\overline{\mathbb{P}} : \mathcal{S} \to [0, 1]$ by

$$\overline{\mathbb{P}}(A) = \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) \in A\}\big) \quad \text{for } A \in \mathcal{S}.$$

Prove that $(S, \mathcal{S}, \overline{\mathbb{P}})$ is a probability space.

Remark. Suppose that $S = \mathbb{R}$ and $\mathcal{S} = \mathcal{B}(\mathbb{R})$, so that X is a real-valued random variable. In this case, compare the relevant definitions to conclude that

$$\overline{\mathbb{P}}\big((-\infty, a]\big) = F(a) \quad \text{for all } a \in \mathbb{R},$$

where F is the distribution function of X.

5. Consider a random variable X with distribution function F. Prove the following results:

$$\begin{aligned} &\text{(i) } \mathbb{P}(a < X \le b) = F(b) - F(a), \\ &\text{(ii) } \mathbb{P}(a \le X \le b) = F(b) - F(a-), \\ &\text{(iii) } \mathbb{P}(X = a) = F(a) - F(a-). \end{aligned}$$

In (ii) and (iii), $F(a-)$ is the left-hand limit of F at a, i.e., $F(a-) = \lim_{c \uparrow a} F(c)$.

6. Consider tossing a coin that lands heads with probability $p \in (0,1)$ three times, and let X be the number of heads observed. Determine the distribution function of X.

7. Which of the following functions are probability distribution functions?

$$\text{(i) } F(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ 1 - 0.3e^{-x}, & \text{if } x > 0, \end{cases}$$

$$\text{(ii) } F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 0.5, & \text{if } 0 \leq x < 2, \\ 0.3, & \text{if } 2 \leq x < 4, \\ 1, & \text{if } 4 \leq x, \end{cases}$$

$$\text{(iii) } F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 0.3(1 - e^{-x}), & \text{if } x \geq 0. \end{cases}$$

8. **(i)** Give an example of a probability distribution function F that has infinite discontinuities.

(ii) Prove that a probability distribution function F has at most countably many discontinuities.

Hint: Recalling that F is an increasing function with values in $[0,1]$, how many points x such that $F(x) - F(x-) \in \left(\frac{1}{n+1}, \frac{1}{n}\right]$ can we have for each $n \geq 1$?

Exercises 3

1. Suppose that $\Omega = \{1,2,3,4,5,6,7,8\}$, $\mathcal{F}$ is the collection of all subsets of Ω, and the probability measure $\mathbb{P}$ assigns mass $\frac{1}{8}$ on each point of Ω.

(i) Are the following events independent?

$$A_1 = \{1,2,3,4\}, \quad A_2 = \{5,6,7,8\}.$$

(ii) Are the following events independent?

$$B_1 = \{1,2,3,4\}, \quad B_2 = \{3,4,5,6\}, \quad B_3 = \{2,4,6,8\}.$$

(iii) Are the following events independent?

$$C_1 = \{1,2,3,4\}, \quad C_2 = \{3,4,5,6\}, \quad C_3 = \{3,4,7,8\}.$$

(iv) Are the following events independent?

$$D_1 = \{1,2,3,4\}, \quad D_2 = \{4,5,6,7\}, \quad D_3 = \{4,6,7,8\}.$$

2. Prove that if $A \cap B = \emptyset$, then A and B *cannot* be independent unless $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$.

3. Suppose that the events $A, B, C \in \mathcal{F}$ form a partition of Ω 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

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

(i) Is it true that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$?

(ii) Are X and Y independent?

4. Show that, if X and Y are independent random variables, then

$$\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y).$$

5. Suppose that X has the geometric distribution with parameter $p \in (0,1)$, i.e.,

$$\mathbb{P}(X = n) = (1-p)p^{n-1}, \quad \text{for } n = 1, 2, \ldots.$$

Calculate the expectation of the random variable $Y = \left(\frac{1}{2}\right)^X$.

6. The *moment generating function* M_X (or simply M if there is no ambiguity) of a random variable X is defined by

$$M_X(t) = \mathbb{E}\left[e^{tX}\right].$$

Assuming that $M_X(t)$ is finite for all $t \in [-\varepsilon, \varepsilon]$, for some $\varepsilon > 0$, show formally that

$$\mathbb{E}\big[X^n\big] = M^{(n)}(0), \quad \text{for } n = 1, 2, \ldots,$$

where $M^{(n)}$ is the n-th derivative of M.

7. **(i)** Suppose that Z is a Bernoulli random variable with parameter p, i.e.,

$$\mathbb{P}(X = 0) = 1 - p \quad \text{and} \quad \mathbb{P}(X = 1) = p.$$

What is the moment generating function of Z?

(ii) Suppose that X is a binomial random variable with parameters n, p. Find the moment generating function of X.

Hint. There are two ways to derive this result:

(1) Start from the definition of moment generating functions and use the binomial expansion formula

$$(a + b)^n = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} a^i b^{n-i}.$$

(2) Use the fact that X has the same distribution as $Z_1 + Z_2 + \cdots + Z_n$, where $Z_1, Z_2, \ldots, Z_n$ are independent Bernoulli random variables, each with parameter p.

(iii) Suppose that X is a binomial random variable with parameters n, p. Calculate its mean and its variance from its moment generating function.

(iv) Suppose that X and Y are independent binomial random variables with parameters n, p and m, p, respectively. What is the distribution of the random variable $X + Y$?

8. Suppose that a random variable X has the uniform distribution over the interval (a, b), i.e., X has the probability density function given by

$$f(x) = \begin{cases} 1/(b-a), & \text{if } a < x < b, \\ 0, & \text{otherwise.} \end{cases}$$

Find the moment generating function of X.

9. Suppose that X is a normal random variable with parameters m, σ. Show that the moment generating function of X is given by

$$M(t) = e^{mt + \frac{\sigma^2 t^2}{2}}.$$

Calculate: **(i)** the mean of X, **(ii)** the variance of X, and **(iii)** $\mathbb{E}[X^4]$.

Hint: You may use the result of Exercise 6 above.

10. Suppose that U is an *exponential* random variable with parameter $\mu > 0$, so that the probability density function of U is given by

$$f(u) = \mu e^{-\mu u}.$$

Calculate: **(i)** the moment generating function M_U, and **(ii)** the mean and the variance of U.

11. Prove that, if Z is a normal random variable with mean 0 and variance σ^2, then

$$\mathbb{E}\left[\exp\left(-\frac{\sigma^2}{2} - Z\right)\right] = 1.$$

Exercises 4

1. Prove the following statements:

(i) If (Z_k) is a sequence of positive random variables (i.e., $Z_k \geq 0$ for all k), then

$$\mathbb{E}\left[\sum_{k=1}^{\infty} Z_k\right] = \sum_{k=1}^{\infty} \mathbb{E}\big[Z_k\big] \leq \infty.$$

Hint: You may use the monotone convergence theorem.

(ii) If (Z_k) is a sequence of positive random variables such that $\sum_{k=1}^{\infty} \mathbb{E}\big[Z_k\big] < \infty$, then

$$\sum_{k=1}^{\infty} Z_k < \infty, \ \mathbb{P}\text{-a.s.}, \quad \text{which implies that} \quad \lim_{k\to\infty} Z_k = 0, \ \mathbb{P}\text{-a.s.}.$$

2. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an event $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. Prove that the function $\mathbb{P}(\cdot \mid B) : \mathcal{F} \to \mathbb{R}$ defined by

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{for } A \in \mathcal{F},$$

is a probability measure on $(\Omega, \mathcal{F})$.

3. 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?

4. An insurance company classifies drivers as class X, Y or Z. Experience indicates that the probability that a class X driver has at least one accident in any given year is 0.01, while the corresponding probabilities for classes Y and Z are 0.05 and 0.10, respectively. The company has also found that, of the drivers who apply for cover, 30% are class X, 60% class Y and 10% class Z.

i) A certain new client had an accident within one year. What is the probability that he is a class Z risk?

ii) Another client goes for n years without an accident. Assuming the incidence of accidents in different years to be independent, how large must n be before the company decides that she is more likely to belong to class X than to class Y?

5. Suppose that a random variable X has the geometric distribution with parameter p, so that

$$\mathbb{P}(X = j) = p(1-p)^{j-1}, \quad \text{for } j = 1, 2, \ldots.$$

Show that, given any $n, k = 1, 2, \ldots,$

$$\mathbb{P}(X = n + k \mid X > n) = \mathbb{P}(X = k).$$

6. If X and Y are independent Poisson random variables with parameters λ and μ, respectively, show that

(i) $X + Y$ is a Poisson random variable with parameter $\lambda + \mu$,

(ii) The conditional distribution of X given that $X + Y = n$ is binomial with parameters n, $\frac{\lambda}{\lambda+\mu}$.

7. A certain region is inhabited by two types of insect. Each insect caught will be of type 1 with probability p and type 2 with probability $1 - p$, independently of previous catches. Suppose that a random number N of catches are made, and the number of type 1 insects caught is X.

i) For $n = 0, 1, 2, \ldots$, find $\mathbb{E}[X \mid N = n]$.

ii) If $\mathbb{E}[N] = \mu$, find $\mathbb{E}[X]$.

8. Let X be a simple random variable. Given an event A, describe explicitly a version of the conditional probability $\mathbb{P}\big(A \mid \sigma(X)\big)$.

9. Suppose that a random variable X is equal to a constant c, $\mathbb{P}$-a.s.. Show that, given any σ-algebra $\mathcal{G}$, $\mathbb{E}[X \mid \mathcal{G}] = c$.

Hint. You may use the following property of the expectation operator that you are not required to prove here: if Z_1, Z_2 are random variables such that $Z_1 = Z_2$, $\mathbb{P}$-a.s., then $\mathbb{E}[Z_1] = \mathbb{E}[Z_2]$.

10. Suppose that X is a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X \geq 0$, $\mathbb{P}$-a.s., and $\mathbb{E}\big[X\big] < \infty$. Given a σ-algebra $\mathcal{H} \subseteq \mathcal{F}$, prove that

$$\mathbb{E}\big[X \mid \mathcal{H}\big] \geq 0, \quad \mathbb{P}\text{-a.s.}.$$

Hint. You may use the following property of the expectation operator that you are not required to prove here: if Z is a random variable such that $Z \geq 0$, $\mathbb{P}$-a.s., then $\mathbb{E}[Z] \geq 0$.

11. Suppose that X is a random variable in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Prove that

$$\mathbb{E}\big[X \mid \{\Omega, \emptyset\}\big] = \mathbb{E}\big[X\big].$$

12. Consider random variables $X, X_1, X_2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and two σ-algebras $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$. Use the definition of conditional expectation to prove the following properties:

(i) If Y is a version of $\mathbb{E}[X \mid \mathcal{G}]$, then $\mathbb{E}[Y] = \mathbb{E}[X]$.

(ii) (*Linearity*) Given any constants $a_1, a_2 \in \mathbb{R}$,

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

Hint. To answer this question, you can use the linearity of expectation, which you are not required to prove here.

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

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

Exercises 5

1. Let $X_0 = 1$, and let $X_1, X_2, \ldots$ be a sequence of independent positive random variables with

$$\mathbb{E}\big[X_k\big] = 1 \quad \text{for all } k.$$

Define

$$M_n = X_0 X_1 \cdots X_n, \quad \text{for } n \geq 0,$$

and let $\mathcal{F}_n$ be the σ-algebra generated by the the random variables $X_0, X_1, \ldots, X_n$. Prove that the process (M_n) is an $(\mathcal{F}_n)$-martingale.

2. Let $X_0 = 0$, and let $X_1, X_2, \ldots$ be a sequence of random variables such that $\mathbb{E}[|X_k|] < \infty$ for all $k \geq 1$. Also, let $\mathcal{F}_n$ be the σ-algebra generated by the the random variables $X_0, X_1, \ldots, X_n$, and define

$$M_0 = X_0 \quad \text{and} \quad M_n = \sum_{i=1}^{n} \big(X_i - \mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big), \quad \text{for } n \geq 1.$$

Prove that the process (M_n) is an $(\mathcal{F}_n)$-martingale.

3. Consider a filtration $(\mathcal{F}_n)$ and an $(\mathcal{F}_n)$-adapted stochastic process (X_n) such that $X_0 = 0$ and $\mathbb{E}\big[|X_n|\big] < \infty$ for all $n \geq 0$. Also, let (c_n) be a sequence of constants. Define $M_0 = 0$ and

$$M_n = c_n X_n - \sum_{j=1}^{n} c_j \mathbb{E}\big[X_j - X_{j-1} \mid \mathcal{F}_{j-1}\big] - \sum_{j=1}^{n} \big(c_j - c_{j-1}\big) X_{j-1}, \quad \text{for } n \geq 1.$$

Prove that (M_n) is an $(\mathcal{F}_n)$-martingale.

4. Let (W_t) be a standard one-dimensional Brownian motion. Given times $r < s < t < u$, calculate the expectations

(i) $\mathbb{E}\left[(W_t - W_s)(W_s - W_r)\right]$,

(ii) $\mathbb{E}\left[(W_u - W_t)^2 (W_s - W_r)^2\right]$,

(iii) $\mathbb{E}\left[(W_u - W_s)(W_t - W_r)\right]$,

(iv) $\mathbb{E}\left[(W_t - W_r)(W_s - W_r)^2\right]$, and

(v) $\mathbb{E}\left[W_r W_s W_t\right]$.

5. *Scaling of the standard Brownian motion.* Let (W_t) be a standard Brownian motion. Given a constant $c > 0$, show that the stochastic process (X_t) defined by

$$X_t = \frac{1}{\sqrt{c}} W_{ct}, \quad \text{for } t \geq 0,$$

is a standard Brownian motion.

6. Suppose that the process (W_t) is a standard one-dimensional $(\mathcal{F}_t)$-Brownian motion.

(I) Prove that the process (X_t) defined by

$$X_t = W_t^2 - t, \quad \text{for } t \geq 0,$$

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

Hint. Observe that the $(\mathcal{F}_t)$-martingale property of (X_t) is equivalent to

$$\mathbb{E}[W_t^2 \mid \mathcal{F}_s] - W_s^2 = t - s \quad \text{for all } s < t. \tag{84}$$

Then consider $\mathbb{E}[(W_t - W_s)^2 \mid \mathcal{F}_s]$ and prove (84).

(II) Prove that the process (Y_t) defined by

$$Y_t = \exp\left(-\frac{1}{2}\theta^2 t - \theta W_t\right), \quad \text{for } t \geq 0,$$

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

Hint. Given times $s < t$, you can use Exercise 3.11 to calculate $\mathbb{E}[Y_t Y_s^{-1} \mid \mathcal{F}_s]$.

Exercises 6

1. Consider a standard one-dimensional Brownian motion (W_t). Use Itô's formula to calculate

$$W_t^2 = t + 2\int_0^t W_s \, dW_s,$$

and

$$W_t^{27} = 351\int_0^t W_s^{25}\, ds + 27\int_0^t W_s^{26}\, dW_s.$$

2. Consider a standard one-dimensional Brownian motion (W_t). Given $k \geq 2$ and $t \geq 0$, use Itô's formula to prove that

$$\mathbb{E}\left[W_t^k\right] = \frac{1}{2}k(k-1)\int_0^t \mathbb{E}\left[W_u^{k-2}\right]\, du.$$

Use this expression to calculate $\mathbb{E}\left[W_t^4\right]$ and $\mathbb{E}\left[W_t^6\right]$.

Hint: You may assume that all stochastic integrals with respect to a Brownian motion that you encounter in this exercise are martingales, so they have expectation 0.

3. Consider the following stochastic differential equation

$$Z_t = -\int_0^t Z_u \, du + \int_0^t e^{-u}\, dW_u.$$

Prove that its solution is given by

$$Z_t = e^{-t} W_t.$$

4. In Vasicek's interest rate model, the dynamics of the short rate process (r_t) are given by the stochastic differential equation

$$dr_t = k(\vartheta - r_t)\, dt + \sigma\, dW_t, \qquad (85)$$

where k, ϑ and σ are strictly positive constants

(a) Show that the solution of (85) is given by

$$r_t = \vartheta + (r_0 - \vartheta)e^{-kt} + \sigma e^{-kt}\int_0^t e^{ks}\, dW_s.$$

Hint: Consider the Itô processes (X_t) and (Y_t) defined by

$$X_t = e^{kt} \quad \text{and} \quad Y_t = r_t,$$

and use the integration by parts formula.

(b) Calculate the mean $\mathbb{E}[r_t]$ and the variance $\text{var}(r_t)$ of the random variable r_t.

Hint. To calculate the variance of r_t, you may use Itô's isometry. Also, you may assume that all stochastic integrals with respect to a Brownian motion that you encounter in this exercise are martingales, so they have expectation 0.

5. In the Cox-Ingersoll-Ross interest rate model, the dynamics of the short rate process (r_t) are given by the stochastic differential equation

$$dr_t = k(\vartheta - r_t)\, dt + \sigma\sqrt{r_t}\, dW_t,$$

where k, ϑ and σ are strictly positive constants. Prove that the stochastic process (r_t) satisfies

$$r_t = \vartheta + (r_0 - \vartheta)e^{-kt} + \sigma e^{-kt}\int_0^t e^{ks}\sqrt{r_s}\, dW_s,$$

and then calculate the mean $\mathbb{E}[r_t]$ and the variance $\text{var}(r_t)$ of the random variable r_t.

Hint. You may assume that all stochastic integrals with respect to a Brownian motion that you encounter in this exercise are martingales, so they have expectation 0.

6. Consider a standard one-dimensional Brownian motion (W_t), and the Itô process given by

$$dX_t = te^{W_t}\, dt + \cos(t^2 W_t)\, dW_t, \quad X_0 = \sqrt{\pi}.$$

Also, let (Z_t) be the Itô process defined by

(i) $Z_t = \sin(tX_t)$, or

(ii) $Z_t = X_t \exp(t^2 X_t)$, or

(iii) $Z_t = X_t^3 + t\cos(X_t)$.

In each of these cases, use Itô's formula to provide expressions for the constant Z_0, and the processes (A_t) and (C_t) such that

$$Z_t = Z_0 + \int_0^t A_s\, ds + \int_0^t C_s\, dW_s.$$

7. Consider the exponential martingale (L_t) defined by the stochastic differential equation

$$dL_t = \vartheta L_t\, dW_t, \quad L_0 = 1,$$

where ϑ is a constant, and let (π_t) be the process defined by

$$\pi_t = \frac{L_t}{1 + L_t}.$$

Prove that (π_t) satisfies the stochastic differential equation

$$d\pi_t = -\vartheta^2 \pi_t^2\, (1 - \pi_t)\, dt + \vartheta\pi_t\, (1 - \pi_t)\, dW_t.$$

Exercises 7

1. Suppose that (X_t) is a continuous $(\mathcal{F}_t)$-local martingale such that $X_t \geq 0$ for all $t \geq 0$, $\mathbb{P}$-a.s.. Prove that (X_t) is an $(\mathcal{F}_t)$-supermartingale.

Hint. You may use the definition of a local martingale and Fatou's lemma.

2. Let Y be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$Y \geq 0, \quad \mathbb{P}\text{-a.s.}, \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}\big[Y\big] = 1. \tag{86}$$

Also, define the function $\mathbb{Q} : \mathcal{F} \to \mathbb{R}$ by

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}\left[Y \mathbf{1}_A\right] \quad \text{for all } A \in \mathcal{F}. \tag{87}$$

Prove that the function $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{F})$.

3. Consider the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ defined by (86)–(87) in the previous exercise. Prove that if Z is a random variable such that either $Z \geq 0$, $\mathbb{P}$-a.s., or $\mathbb{E}^{\mathbb{Q}}[|Z|] < \infty$, then

$$\mathbb{E}^{\mathbb{Q}}[Z] = \mathbb{E}^{\mathbb{P}}[YZ].$$

4. Consider the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ defined by (86)–(87). Prove that

$$\mathbb{Q}\big(\{Y = 0\}\big) = 0.$$

5. Consider the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ defined by (86)–(87), and suppose that $\mathbb{Q} \sim \mathbb{P}$. Show that if $d\mathbb{Q} = Y\, d\mathbb{P}$, then $Y > 0$, $\mathbb{P}$-a.s., and $d\mathbb{P} = Y^{-1}\, d\mathbb{Q}$.

6. 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 $T > 0$, define the probability measure $\mathbb{Q}_T$ on the measurable space $(\Omega, \mathcal{F}_T)$ by

$$\mathbb{Q}_T(A) = \mathbb{E}^{\mathbb{P}}\left[L_T \mathbf{1}_A\right], \quad \text{for } A \in \mathcal{F}_T. \tag{88}$$

Given times $0 \leq s \leq t \leq T$, show that, if Z is an $\mathcal{F}_t$-measurable random variable, then

$$\mathbb{E}^{\mathbb{Q}_T}[Z \mid \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[L_t Z \mid \mathcal{F}_s]}{L_s}. \tag{89}$$

7. Suppose that (L_t) is an $(\mathcal{F}_t)$-martingale as in Exercise 6 above. Also, suppose that (M_t) is an $(\mathcal{F}_t)$-martingale with respect to the probability measure $\mathbb{P}$. Prove that the process $(L_t^{-1} M_t,\ t \in [0, T])$, is an $(\mathcal{F}_t)$-martingale with respect to the probability measure $\mathbb{Q}_T$ defined by (88).

8. Given an n-dimensional $(\mathcal{F}_t)$-Brownian motion W and an n-dimensional, $(\mathcal{F}_t)$-progressively measurable process (X_t) satisfying

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

show that the process (L_t) given by

$$L_t = \exp\left(-\frac{1}{2}\int_0^t |X_s|^2 \, ds + \int_0^t X_s \cdot dW_s\right) \tag{90}$$

satisfies the SDE

$$L_t = 1 + \int_0^t L_s X_s \cdot dW_s. \tag{91}$$

9. Suppose that (X_t) is an n-dimensional, $(\mathcal{F}_t)$-progressively measurable process that is bounded in the sense that

$$\left|X_t^i\right| \leq K \quad \text{for all } t \geq 0 \text{ and } i = 1, \ldots, n, \quad \mathbb{P}\text{-a.s.},$$

where $K > 0$ is a constant. Prove that the process (L_t) defined by (90) or (91) in the previous exercise is an $(\mathcal{F}_t)$-martingale.

Solutions 1

1. Suppose that $\Omega = \mathbb{R}$.

(i) The family of events

$$\mathcal{F} = \big\{\Omega, \emptyset, (-\infty, a], (b, \infty) \mid \ a, b \in \mathbb{R}\big\}$$

is not a σ-algebra. For example, given any $a \in \mathbb{R}$,

$$(-\infty, -\sqrt{2}] \cup (1, \infty) \notin \mathcal{F}.$$

(ii) The family of events

$$\mathcal{F} = \big\{A \mid \ \text{either } A \text{ or } A^c \text{ is countable}\big\}$$

is a σ-algebra on $\mathbb{R}$. To verify this claim, we check the three properties that characterise a σ-algebra:

- $\Omega \equiv \mathbb{R} \in \mathcal{F}$ because $\Omega^c = \emptyset$ and the empty set is "countable" (it has 0 elements).
- Given any event $A \in \mathcal{F}$, plainly $A^c \in \mathcal{F}$.
- Fix any sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$. We have two possibilities:
  1) If $\bigcup_{n=1}^{\infty} A_n$ is countable, then clearly $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.
  2) If $\bigcup_{n=1}^{\infty} A_n$ is uncountable, then at least one of the sets A_n, $n \geq 1$, is uncountable, because

$$\text{if } C_n \text{ is countable for all } n \geq 1, \text{ then } \bigcup_{n=1}^{\infty} C_n \text{ is countable.}$$

  So, suppose that, e.g., A_1 is uncountable. Since $A_1 \in \mathcal{F}$, that means that A_1^c is countable. Furthermore, noting that

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c \subseteq A_1^c,$$

  we can see that $\left(\bigcup_{n=1}^{\infty} A_n\right)^c$ is countable (because any subset of a countable set is countable), and therefore, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

(iii) It is straightforward to check that the family of events

$$\mathcal{F} = \big\{\mathbb{R},\ \emptyset,\ (-\infty, 5],\ (5, \infty),\ (-\infty, 3),\ [3, \infty),\ [3, 5],\ (-\infty, 3) \cup (5, \infty)\big\}$$

is a σ-algebra.

2. **(i)** Suppose that $\Omega = \mathbb{R}$ and $\mathcal{C} = \big\{(-20, \sqrt{2}), (-15, \infty)\big\}$. If we define

$$A = (-\infty, -20], \quad B = (-20, -15] \quad C = (-15, \sqrt{2}), \quad D = [\sqrt{2}, \infty),$$

then we can check that

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

(ii) Suppose that $\Omega = \mathbb{R}$ and $\mathcal{C} = \big\{(1,2], \{2\}\big\}$. If we define

$$A = (1,2), \quad B = \{2\}, \quad C = (-\infty, 1] \cup (2, \infty),$$

then we can check that

$$\begin{aligned}\sigma(\mathcal{C}) &= \big\{\Omega, \emptyset, A, B, C, A \cup B, A \cup C, B \cup C\big\} \\ &= \big\{\mathbb{R},\ \emptyset,\ (1,2),\ \{2\},\ (-\infty,1] \cup (2,\infty),\ (1,2],\ (-\infty,2) \cup (2,\infty),\ (-\infty,1] \cup [2,\infty)\big\}.\end{aligned}$$

(iii) Suppose that $\Omega = \{1,2,3,4\}$ and $\mathcal{C} = \big\{\emptyset, \{2,3\}\big\}$. If we define

$$A = \{2,3\}, \quad B = \{1,4\},$$

then we can check that

$$\begin{aligned}\sigma(\mathcal{C}) &= \big\{\Omega, \emptyset, A, B\big\} \\ &= \big\{\{1,2,3,4\},\ \emptyset,\ \{2,3\},\ \{1,4\}\big\}.\end{aligned}$$

(iv) Suppose that $\Omega = \{1,2,3,4\}$ and $\mathcal{C} = \big\{\{3\}, \{2,3,4\}\big\}$. If we define

$$A = \{1\}, \qquad B = \{2,4\}, \qquad C = \{3\},$$

then we can check that

$$\begin{aligned}\sigma(\mathcal{C}) &= \big\{\Omega, \emptyset, A, B, C, A \cup B, A \cup C, B \cup C\big\} \\ &= \big\{\{1,2,3,4\},\ \emptyset,\ \{1\},\ \{2,4\},\ \{3\},\ \{1,2,4\},\ \{1,3\},\ \{2,3,4\}\big\}.\end{aligned}$$

3. Given a measurable space $(\Omega, \mathcal{F})$ and a set $\Omega' \subseteq \Omega$, we can show that the family of sets

$$\mathcal{H} = \big\{A \cap \Omega' \mid\ A \in \mathcal{F}\big\}$$

is a σ-algebra on Ω' by checking that it satisfies the defining properties of a σ-algebra as follows:

- $\Omega' \in \mathcal{H}$ because $\Omega' = \Omega \cap \Omega'$ and $\Omega \in \mathcal{F}$ (since $\mathcal{F}$ is a σ-algebra on Ω).
- Let any $C \in \mathcal{H}$. By the definition of $\mathcal{H}$, there exists $A \in \mathcal{F}$ such that $C = A \cap \Omega'$. Now since $\Omega \setminus A \in \mathcal{F}$ (because $\mathcal{F}$ is a σ-algebra on Ω), it follows that

$$\begin{aligned}\Omega' \setminus C &= \Omega' \setminus \big(A \cap \Omega'\big) \\ &= \Omega' \setminus A \\ &= \big(\Omega \setminus A\big) \cap \Omega' \in \mathcal{H},\end{aligned}$$

which proves that the complement relative to Ω' of any set in $\mathcal{H}$ belongs to $\mathcal{H}$.

- Let any sequence of sets $C_1, C_2, \ldots, C_n, \ldots \in \mathcal{H}$. By the definition of $\mathcal{H}$, there exists a sequence of sets $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$ such that $C_n = A_n \cap \Omega'$ for all $n = 1, 2, \ldots$. Now, since $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (because $\mathcal{F}$ is a σ-algebra),

$$\begin{aligned}\bigcup_{n=1}^{\infty} C_n &= \bigcup_{n=1}^{\infty} (A_n \cap \Omega') \\ &= \left(\bigcup_{n=1}^{\infty} A_n \right) \cap \Omega' \in \mathcal{H},\end{aligned}$$

which proves that $\mathcal{H}$ is closed under countable unions.

4. First, we note that the assumption $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ implies

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) = 0.$$

Also, we observe that the sequence of events $B_1, \ldots, B_n, \ldots \in \mathcal{F}$ defined by

$$B_n = \bigcup_{m=n}^{\infty} A_m$$

is decreasing, i.e., $B_1 \supseteq \cdots \supseteq B_n \supseteq \cdots$.

In view of these observations and the continuity of a probability measure, we calculate

$$\begin{aligned}\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \right) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n \right) \\ &= \mathbb{P}\left(\lim_{n\to\infty} B_n \right) \\ &= \lim_{n\to\infty} \mathbb{P}(B_n) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m \right) \\ &\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) \\ &= 0,\end{aligned}$$

the inequality following by the sub-additivity property of a probability measure, and the result follows.

5. Given a collection $(A_i,\ i \in I)$ of subsets of S, where I is an index set, we can see that

$$\begin{aligned}
\omega \in f^{-1}\left(\bigcap_{i\in I} A_i\right) \quad &\Leftrightarrow \quad \text{there exists } x \in \bigcap_{i\in I} A_i \text{ such that } f(\omega) = x \\
&\Leftrightarrow \quad \text{for all } i \in I,\ \text{there exists } x \in A_i \text{ such that } f(\omega) = x \\
&\Leftrightarrow \quad \text{for all } i \in I,\ \omega \in f^{-1}(A_i) \\
&\Leftrightarrow \quad \omega \in \bigcap_{i\in I} f^{-1}(A_i). \qquad (92)
\end{aligned}$$

These equivalences prove that

$$f^{-1}\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f^{-1}(A_i). \qquad (93)$$

Similarly, we can see that, given any $A \subseteq S$,

$$\begin{aligned}
\omega \in f^{-1}(S \setminus A) \quad &\Leftrightarrow \quad \text{there exists } x \in S \setminus A \text{ such that } f(\omega) = x \\
&\Leftrightarrow \quad \text{for all } x \in A,\ f(\omega) \neq x \\
&\Leftrightarrow \quad \omega \in \Omega \setminus f^{-1}(A), \qquad (94)
\end{aligned}$$

which prove that

$$f^{-1}(S \setminus A) = \Omega \setminus f^{-1}(A) \quad \text{for all } A \subseteq S. \qquad (95)$$

Solutions 2

1. First, we recall the definition: a real-valued random variable is any function $Y : \Omega \to \mathbb{R}$ such that

$$Y^{-1}(A) = \big\{\omega \in \Omega \mid \; Y(\omega) \in A\big\} \in \mathcal{F} \quad \text{for all } A \in \mathcal{B}(\mathbb{R}).$$

In view of this, we can verify that the function $\omega \mapsto f(X(\omega)) \equiv (f \circ X)(\omega)$ is a random variable by showing that

$$(f \circ X)^{-1}(A) = \big\{\omega \in \Omega \mid \; f(X(\omega)) \in A\big\} \in \mathcal{F} \quad \text{for all } A \in \mathcal{B}(\mathbb{R}).$$

Now, we observe that, given any $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}(f \circ X)^{-1}(A) &= \big\{\omega \in \Omega \mid \; f(X(\omega)) \in A\big\} \\ &= \big\{\omega \in \Omega \mid \; X(\omega) \in \{x \in \mathbb{R} \mid \; f(x) \in A\}\big\}.\end{aligned}$$

It follows that, since X is a random variable, this set will belong to $\mathcal{F}$ if

$$\{x \in \mathbb{R} \mid \; f(x) \in A\} \in \mathcal{B}(\mathbb{R}),$$

which is true by the assumptions on f.

2. For $i = 1, 2, 3, 4$, we define

$$A_i = \big\{\omega \in \Omega \mid \; X(\omega) = x_i\big\}.$$

It is straightforward to verify that

$$\begin{aligned}\sigma(X) = \big\{&\emptyset, \Omega, A_1, A_2, A_3, A_4, \; A_1 \cup A_2, \; A_1 \cup A_3, \; A_1 \cup A_4, \\ &A_2 \cup A_3, \; A_2 \cup A_4, \; A_3 \cup A_4, \; A_1 \cup A_2 \cup A_3, \; A_1 \cup A_3 \cup A_4, \; A_2 \cup A_3 \cup A_4\big\}.\end{aligned}$$

3. (i) First, suppose that X is $\{\emptyset, \Omega\}$-measurable and that X is not constant. Then, there exist at least two real numbers $c_1 \neq c_2$ such that the sets

$$\big\{\omega \in \Omega \mid \; X(\omega) = c_1\big\} \quad \text{and} \quad \big\{\omega \in \Omega \mid \; X(\omega) = c_2\big\}$$

are not empty. However, none of these two events belongs to $\{\emptyset, \Omega\}$, which is a contradiction.

Conversely, suppose that X is constant, i.e., that there exists $c \in \mathbb{R}$ such that $X(\omega) = c$ for all $\omega \in \Omega$. In this case, given $A \in \mathcal{B}(\mathbb{R})$,

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \; X(\omega) \in A\big\} = \begin{cases}\Omega, & \text{if } c \in A, \\ \emptyset, & \text{if } c \notin A,\end{cases}$$

which proves that X is $\{\emptyset, \Omega\}$-measurable.

(ii) We can solve this exercise in two ways.

1st way. Let $\mathcal{H}$ be a σ-algebra such that $\mathbb{P}(A) \in \{0,1\}$ for every $A \in \mathcal{H}$, and let X be an $\mathcal{H}$-measurable random variable. In this context,

$$\text{either} \quad \mathbb{P}(X = x) = 0 \quad \text{or} \quad \mathbb{P}(X = x) = 1, \quad \text{for all } x \in \mathbb{R}. \tag{96}$$

Now, we argue by contradiction, and we assume that there exists no $c \in \mathbb{R}$ such that $\mathbb{P}(X = c) = 1$. Such an assumption and (96) imply that

$$\mathbb{P}(X = x) = 0 \quad \text{for all } x \in \mathbb{R},$$

which implies that the distribution function F of X is continuous (see also Exercise 2.5.(iii)). Now, the possibility that X has a continuous distribution function implies that, for every $b \in (0,1)$, there exists $a \in \mathbb{R}$ such that

$$F(a) = b.$$

Given any such reals a and b, the event $\big\{\omega \in \Omega \mid \; X(\omega) \le a\big\}$ has probability $F(b)$, which is different from 0 or 1. However, this contradicts (96).

2nd way. Let $\mathcal{H}$ be a σ-algebra such that $\mathbb{P}(A) \in \{0,1\}$ for every $A \in \mathcal{H}$. If X is an $\mathcal{H}$-measurable random variable, then

$$\text{either} \quad F(a) = \mathbb{P}(X \le a) = 0 \quad \text{or} \quad F(a) = \mathbb{P}(X \le a) = 1 \quad \text{for all } a \in \mathbb{R}.$$

Since the distribution function F of X is non-decreasing, this implies that there exists $c \in \mathbb{R}$ such that

$$F(a) = \begin{cases} 0 & \text{for all } a < c, \\ 1 & \text{for all } a \ge c, \end{cases}$$

which establishes the fact that there exists $c \in \mathbb{R}$ such that

$$\mathbb{P}(X = c) = F(c) - F(c-) = 1.$$

4. Since $(S, \mathcal{S})$ is a measurable space, we will prove that $(S, \mathcal{S}, \overline{\mathbb{P}})$ is a probability space if we show that $\overline{\mathbb{P}}$ is a probability measure on $(S, \mathcal{S})$. To achieve this, we have to show that the function $\overline{\mathbb{P}} : \mathcal{S} \to [0,1]$, satisfies properties (i) and (ii) of Definition C1.29.

(i) Since $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space,

$$\overline{\mathbb{P}}(S) = \mathbb{P}\big(\{\omega \in \Omega \mid \; X(\omega) \in S\}\big) = \mathbb{P}(\Omega) = 1.$$

(ii) Let $A_1, \ldots, A_n, \ldots \in \mathcal{S}$ be any sequence of pairwise disjoint subsets of S. Noting that the events

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \; X(\omega) \in A_n\big\}, \quad n \ge 1,$$

are pairwise disjoint, we use the fact that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space to calculate

$$\begin{aligned}
\overline{\mathbb{P}}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}\left(\left\{\omega \in \Omega \mid\ X(\omega) \in \bigcup_{n=1}^{\infty} A_n\right\}\right) \\
&= \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{\omega \in \Omega \mid\ X(\omega) \in A_n\}\right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}\big(\{\omega \in \Omega \mid\ X(\omega) \in A_n\}\big) \\
&= \sum_{n=1}^{\infty} \overline{\mathbb{P}}(A_n).
\end{aligned}$$

5. (i) The required identity follows immediately from the calculations

$$\begin{aligned}
F(b) &= \mathbb{P}\big(X \le b\big) \\
&= \mathbb{P}\big(\{\omega \in \Omega \mid\ X(\omega) \le a\} \cup \{\omega \in \Omega \mid\ a < X(\omega) \le b\}\big) \\
&= \mathbb{P}\big(X \le a\big) + \mathbb{P}\big(a < X \le b\big). \\
&= F(a) + \mathbb{P}\big(a < X \le b\big).
\end{aligned}$$

(ii) Using the previous result and the "continuity" of a probability measure, we can see that

$$\begin{aligned}
\mathbb{P}\big(a \le X \le b\big) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \{\omega \in \Omega \mid\ X(\omega) \in (a - \tfrac{1}{n}, b]\}\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(a - \tfrac{1}{n} < X \le b\right) \\
&= \lim_{n\to\infty} \left[F(b) - F\left(a - \tfrac{1}{n}\right)\right] \\
&= F(b) - F(a-).
\end{aligned}$$

(iii) Using the previous two results, we calculate

$$\begin{aligned}
\mathbb{P}\big(X = a\big) &= \mathbb{P}\left(X \in [a, a+\sqrt{2}] \setminus (a, a+\sqrt{2}]\right) \\
&= \mathbb{P}\left(a \le X \le a+\sqrt{2}\right) - \mathbb{P}\left(a < X \le a+\sqrt{2}\right) \\
&= F(a) - F(a-).
\end{aligned}$$

6. First, we note that the possible values of X are 0, 1, 2 or 3.

Now, $X = 2$ if we observe one of the following sequences:

HHT, HTH or THH.

Each of these sequences has probability $p^2(1-p)$, so $\mathbb{P}(X=2) = 3p^2(1-p)$.

Similar arguments holding for the other possibilities, we conclude that the distribution of X is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ (1-p)^3, & \text{if } 0 \le x < 1, \\ (1-p)^3 + 3p(1-p)^2, & \text{if } 1 \le x < 2, \\ (1-p)^3 + 3p(1-p)^2 + 3p^2(1-p), & \text{if } 2 \le x < 3, \\ 1, & \text{if } 3 \le x. \end{cases}$$

7. (i) F is not a distribution function because it is not right-continuous (at $x = 0$).

(ii) F is not a distribution function because it is not non-decreasing. For example,

$$F(1) = 0.5 > 0.3 = F(3).$$

(iii) F is not a distribution function because

$$\lim_{x\to\infty} = 0.3 \neq 1.$$

8. (i) If F is the distribution function of a Poisson random variable with parameter λ, then

$$F(n) - F(n-) = e^{-\lambda}\frac{\lambda^n}{n!} > 0 \quad \text{for all } n = 0, 1, 2, \dots.$$

(ii) Let F be a distribution function. The facts that F is non-decreasing and takes values in $[0,1]$, imply that F can have at most n discontinuities with size that is greater than $\frac{1}{n+1}$ and less than or equal to $\frac{1}{n}$, i.e., the set

$$\left\{a \in \mathbb{R} \mid \ \frac{1}{n+1} < F(a) - F(a-) \le \frac{1}{n}\right\}$$

can have at most n elements. It follows that, since

$$\big\{a \in \mathbb{R} \mid \ F(a) - F(a-) > 0\big\} = \bigcup_{n=1}^{\infty}\left\{a \in \mathbb{R} \mid \ \frac{1}{n+1} < F(a) - F(a-) \le \frac{1}{n}\right\},$$

the set of all discontinuities of F is countable because it can be expressed as the countable union of finite sets.

Solutions 3

1. (i) *NO*: $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(\emptyset) = 0 \neq \frac{1}{4} = \mathbb{P}(A_1)\mathbb{P}(A_2)$.

(ii) *YES*:

$$\begin{aligned}
\mathbb{P}(B_1 \cap B_2) = \mathbb{P}(\{3,4\}) =& \frac{1}{4} = \mathbb{P}(B_1)\mathbb{P}(B_2),\\
\mathbb{P}(B_1 \cap B_3) = \mathbb{P}(\{2,4\}) =& \frac{1}{4} = \mathbb{P}(B_1)\mathbb{P}(B_3),\\
\mathbb{P}(B_2 \cap B_3) = \mathbb{P}(\{4,6\}) =& \frac{1}{4} = \mathbb{P}(B_2)\mathbb{P}(B_3),\\
\mathbb{P}(B_1 \cap B_2 \cap B_3) = \mathbb{P}(\{4\}) =& \frac{1}{8} = \mathbb{P}(B_1)\mathbb{P}(B_2)\mathbb{P}(B_3).
\end{aligned}$$

(iii) *NO*: Although

$$\begin{aligned}
\mathbb{P}(C_1 \cap C_2) = \mathbb{P}(\{3,4\}) =& \frac{1}{4} = \mathbb{P}(C_1)\mathbb{P}(C_2),\\
\mathbb{P}(C_1 \cap C_3) = \mathbb{P}(\{3,4\}) =& \frac{1}{4} = \mathbb{P}(C_1)\mathbb{P}(C_3),\\
\mathbb{P}(C_2 \cap C_3) = \mathbb{P}(\{3,4\}) =& \frac{1}{4} = \mathbb{P}(C_2)\mathbb{P}(C_3),
\end{aligned}$$

we have

$$\mathbb{P}(C_1 \cap C_2 \cap C_3) = \mathbb{P}(\{3,4\}) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(C_1)\mathbb{P}(C_2)\mathbb{P}(C_3).$$

(iv) *NO*: Although

$$\mathbb{P}(D_1 \cap D_2 \cap D_3) = \mathbb{P}(\{4\}) = \frac{1}{8} = \mathbb{P}(D_1)\mathbb{P}(D_2)\mathbb{P}(D_3),$$

we have

$$\begin{aligned}
\mathbb{P}(D_1 \cap D_2) = \mathbb{P}(\{4\}) = \frac{1}{8} \neq \frac{1}{4} =& \mathbb{P}(D_1)\mathbb{P}(D_2),\\
\mathbb{P}(D_1 \cap D_3) \neq& \mathbb{P}(D_1)\mathbb{P}(D_3),\\
\mathbb{P}(D_2 \cap D_3) \neq& \mathbb{P}(D_2)\mathbb{P}(D_3).
\end{aligned}$$

2. If $A \cap B = \emptyset$, then

$$\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0. \tag{97}$$

On the other hand, A and B are independent if and only if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{98}$$

The claim now follows immediately from equating the right-hand sides of (97) and (98).

3. **(i)** We can easily verify that $XY = X$. Combining this observation 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 $\mathbb{E}[XY] = \mathbb{E}[X] = \mathbb{E}[X]\mathbb{E}[Y]$.

(ii) The random variables X and Y are *not* independent. For instance,

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

4. Using the assumption that X and Y are independent, we calculate

$$\begin{aligned}\operatorname{var}(X+Y) &= \mathbb{E}\big[(X+Y)^2\big] - \big(\mathbb{E}\,[X+Y]\big)^2 \\ &= \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] + 2\mathbb{E}\,[XY] - \big(\mathbb{E}\,[X]\big)^2 - \big(\mathbb{E}\,[Y]\big)^2 - 2\mathbb{E}\,[X]\,\mathbb{E}\,[Y] \\ &= \mathbb{E}\left[X^2\right] - (\mathbb{E}\,[X])^2 + \mathbb{E}\left[Y^2\right] - (\mathbb{E}\,[Y])^2 \\ &= \operatorname{var}(X) + \operatorname{var}(Y)\end{aligned}$$

5. The expectation of the random variable $\left(\frac{1}{2}\right)^X$ is given by

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{2}\right)^X\right] &= \sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n (1-p)p^{n-1} \\ &= \frac{1-p}{p}\sum_{n=1}^{\infty}\left(\frac{p}{2}\right)^n \\ &= \frac{1-p}{p}\frac{p/2}{1-p/2} \\ &= \frac{1-p}{2-p}.\end{aligned}$$

6. We calculate

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

Evaluating this expression at $t = 0$ yields immediately the required formula.

7. **(i)** If Z is a Bernoulli random variable with parameter p, its moment generating function is given by

$$M_Z(t) = \mathbb{E}\left[e^{tZ}\right] = 1 - p + pe^t.$$

(ii) Suppose that X is a binomial random variable with parameters n, p. We can calculate its moment generating function in two ways as follows:

(1) Using the binomial expansion formula, we can calculate

$$\begin{aligned} M_X(t) &= \mathbb{E}\left[e^{tX}\right] \\ &= \sum_{i=0}^{n} e^{ti} \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\ &= \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} \left(pe^t\right)^i (1-p)^{n-i} \\ &= \left(1-p+pe^t\right)^n . \end{aligned}$$

(2) Since X has the same distribution as $Z_1 + Z_2 + \cdots + Z_n$, where $Z_1, Z_2, \ldots, Z_n$ are independent Bernoulli random variables, each with parameter p,

$$\begin{aligned} M_X(t) &= M_{Z_1+Z_2+\cdots+Z_n}(t) \\ &= M_{Z_1}(t) M_{Z_2}(t) \cdots M_{Z_n}(t) \\ &= \left(1-p+pe^t\right)^n , \end{aligned}$$

the last equality following by part (i) of this exercise.

(iii) The mean of a random variable X with the binomial distribution with parameters n, p is

$$\mathbb{E}[X] = M_X'(0) = npe^t \left(1-p+pe^t\right)^{n-1} \Big|_{t=0} = np.$$

The second moment of X is

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

and its variance is

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

(iv) Since X and Y are independent, the moment generating function of $X + Y$ is

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) \\ &= \left(1-p+pe^t\right)^n \left(1-p+pe^t\right)^m \\ &= \left(1-p+pe^t\right)^{n+m} , \end{aligned}$$

the second inequality following by using the result in (ii) above. However, this shows that the distribution of $X + Y$ is binomial with parameters $n + m$, p.

8. We calculate

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \int_a^b e^{tx} \frac{1}{b-a}\, dx = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

9. The moment generating function of a normal random variable with parameters m and σ is given by

$$\begin{aligned} M(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx \\ &= e^{mt+\frac{\sigma^2 t^2}{2}} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-m-\sigma^2 t)^2}{2\sigma^2}}\, dx \\ &= e^{mt+\frac{\sigma^2 t^2}{2}}. \end{aligned}$$

The expectation of X is given by

$$\mathbb{E}\big[X\big] = M'(0) = (m+\sigma^2 t) e^{mt+\frac{\sigma^2 t^2}{2}}\Big|_{t=0} = m,$$

its second moment is given by

$$\mathbb{E}\big[X^2\big] = M''(0) = \left[\sigma^2 + (m+\sigma^2 t)^2\right] e^{mt+\frac{\sigma^2 t^2}{2}}\Big|_{t=0} = \sigma^2 + m^2,$$

and its variance is

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

Also, its fourth moment is

$$\begin{aligned} \mathbb{E}\big[X^4\big] &= M^{(4)}(0) \\ &= \left[3\sigma^4 + 6\sigma^2(m+\sigma^2 t)^2 + (m+\sigma^2 t)^4\right] e^{mt+\frac{\sigma^2 t^2}{2}}\Big|_{t=0} \\ &= 3\sigma^4 + 6\sigma^2 m^2 + m^4. \end{aligned}$$

10. The moment generating function of an exponentially distributed random variable U with parameter $\mu > 0$ is

$$M_U(t) = \int_0^{\infty} e^{tx} \mu e^{-\mu x}\, dx = \mu \int_0^{\infty} e^{-(\mu - t)x}\, dx = \frac{\mu}{\mu - t}, \quad \text{if } t < \mu.$$

Using this expression, we can calculate

$$\begin{aligned} \mathbb{E}\big[X\big] &= \frac{\mu}{(\mu-t)^2}\Big|_{t=0} = \frac{1}{\mu}, \\ \mathbb{E}\big[X^2\big] &= \frac{2\mu}{(\mu-t)^3}\Big|_{t=0} = \frac{2}{\mu^2}, \end{aligned}$$

and

$$\mathrm{var}(X) = \frac{1}{\mu^2}. \tag{99}$$

11. If Z is a Gaussian random variable with mean 0 and variance σ^2, then

$$\begin{aligned}\mathbb{E}\left[\exp\left(-\frac{\sigma^2}{2}-Z\right)\right] &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{\sigma^2}{2}-a}e^{-\frac{a^2}{2\sigma^2}}\,da \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(a+\sigma^2)^2}{2\sigma^2}}\,da \\ &= 1,\end{aligned}$$

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

Solutions 4

1. (i) Consider the sequence of random variables

$$X_n = \sum_{k=1}^{n} Z_k, \quad n = 1, 2, \ldots .$$

Since $X_{n+1} = X_n + Z_{n+1}$ and $Z_{n+1} \geq 0$, $\mathbb{P}$-a.s.,

$$X_1 \leq X_2 \leq \cdots \leq X_n \leq \cdots \quad \text{and} \quad \lim_{n\to\infty} X_n = \sum_{k=1}^{\infty} Z_k, \quad \mathbb{P}\text{-a.s..}$$

In view of these observation, we can use the monotone convergence theorem and the linearity of expectation, to calculate

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{\infty} Z_k\right] &= \mathbb{E}\left[\lim_{n\to\infty} X_n\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[X_n\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sum_{k=1}^{n} Z_k\right] \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[Z_k\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[Z_k\right].
\end{aligned}$$

(ii) Suppose that (Z_k) is a sequence of positive random variables such that

$$\sum_{k=1}^{\infty} \mathbb{E}[Z_k] < \infty.$$

Part (i) of this exercise implies that this sequence satisfies

$$\mathbb{E}\left[\sum_{k=1}^{\infty} Z_k\right] < \infty.$$

The required result now follows from the fact that, if X is a random variable such that $X \geq 0$, $\mathbb{P}$-a.s., and $\mathbb{E}[X] < \infty$, then $X < \infty$, $\mathbb{P}$-a.s..
To see this last claim, we argue by contradiction, and we assume that there exists a random variable $X \geq 0$ such that $\mathbb{E}[X] < \infty$ and $\mathbb{P}(X = \infty) > 0$. In such a case,

$$\mathbb{E}[X] = \mathbb{E}\left[X\left(\mathbf{1}_{\{X=\infty\}} + \mathbf{1}_{\{X<\infty\}}\right)\right] \geq \mathbb{E}\left[X\mathbf{1}_{\{X=\infty\}}\right] = \infty \times \mathbb{P}(X = \infty) = \infty,$$

which contradicts the assumption that $\mathbb{E}[X] < \infty$.

2. We use the fact that $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$ to verify that, given any fixed event $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, the function $\mathbb{P}(\cdot \mid B) : \mathcal{F} \to \mathbb{R}$ defined by

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{for } A \in \mathcal{F}, \tag{100}$$

is a probability measure on $(\Omega, \mathcal{F})$.

First, we observe that $\mathbb{P}(A \mid B) \geq 0$ for all $A \in \mathcal{F}$, because $\mathbb{P}(A \cap B), \mathbb{P}(B) \geq 0$. Also, $\mathbb{P}(A \mid B) \leq 1$ for all $A \in \mathcal{F}$, because $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$. Therefore, (100) defines a function $\mathbb{P}(\cdot \mid B) : \mathcal{F} \to [0, 1]$. Furthermore, this function is a probability measure because

(i) $\mathbb{P}(\Omega \mid B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = 1$.

(ii) Given any sequence of pairwise disjoint events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$, the events $A_1 \cap B, A_2 \cap B, \ldots, A_n \cap B, \ldots$ are pairwise elements of $\mathcal{F}$. Therefore, using the countable additivity property of $\mathbb{P}$, we can see that

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \,\middle|\, B\right) &= \frac{\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \cap B\right)}{\mathbb{P}(B)} \\
&= \frac{\sum_{n=1}^{\infty} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} \\
&= \sum_{n=1}^{\infty} \mathbb{P}(A_n \mid B).
\end{aligned}$$

3. Consider the events

$$H = \{\text{the person does not have the disease}\},$$

and

$$T = \{\text{the test result is positive}\}.$$

We are given the probabilities

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

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

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

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.

4. We consider the events

$$
\begin{aligned}
A &= \{\text{the client has an accident within a year}\},\\
X &= \{\text{the client is of class } X \text{ risk}\},\\
Y &= \{\text{the client is of class } Y \text{ risk}\},
\end{aligned}
$$

and

$$
Z = \{\text{the client is of class } Z \text{ risk}\}.
$$

(i) Applying Bayes' lemma, we calculate

$$
\begin{aligned}
\mathbb{P}(Z \mid A) &= \frac{\mathbb{P}(A \mid Z)\mathbb{P}(Z)}{\mathbb{P}(A \mid X)\mathbb{P}(X) + \mathbb{P}(A \mid Y)\mathbb{P}(Y) + \mathbb{P}(A \mid Z)\mathbb{P}(Z)}\\
&= \frac{0.1 \times 0.1}{0.01 \times 0.3 + 0.05 \times 0.6 + 0.1 \times 0.1} \simeq 0.233.
\end{aligned}
$$

(ii) Let B_n be the event that the client goes for n years without an accident. We want to find how large n must be so that $\mathbb{P}(X \mid B_n) > \mathbb{P}(Y \mid B_n)$. In view of Bayes' lemma, this inequality is equivalent to

$$
\frac{\mathbb{P}(B_n \mid X)\mathbb{P}(X)}{\mathbb{P}(B_n)} > \frac{\mathbb{P}(B_n \mid Y)\mathbb{P}(Y)}{\mathbb{P}(B_n)},
$$

which is equivalent to

$$
\mathbb{P}(B_n \mid X)\mathbb{P}(X) > \mathbb{P}(B_n \mid Y)\mathbb{P}(Y). \tag{101}
$$

Since *different years are independent of each other*,

$$
\begin{aligned}
\mathbb{P}(B_n \mid X) &= [\mathbb{P}(B_1 \mid X)]^n\\
&= (1 - 0.01)^n = 0.99^n.
\end{aligned}
$$

Similarly,

$$
\mathbb{P}(B_n \mid Y) = 0.95^n.
$$

Therefore, (101) is equivalent to

$$
\begin{aligned}
0.99^n \times 0.3 > 0.95^n \times 0.6 \quad &\Leftrightarrow \quad n \log \frac{0.99}{0.95} > \log 2\\
&\Leftrightarrow \quad n > \frac{\log 2}{\log 0.99/0.95} \simeq 16.8.
\end{aligned}
$$

We conclude that the client must go for 17 years without accident before the company decides that he/she is more likely to belong to class X than to class Y.

5. Given any $n, k = 1, 2, \ldots$,

$$\begin{aligned}\mathbb{P}(X = n+k \mid X > n) &= \frac{\mathbb{P}(X = n+k\,,\ X > n)}{\mathbb{P}(X > n)} \\ &= \frac{\mathbb{P}(X = n+k)}{\sum_{j=1}^{\infty} \mathbb{P}(X = n+j)}.\end{aligned}$$

Combining this expression with the calculation

$$\begin{aligned}\sum_{j=1}^{\infty} \mathbb{P}(X = n+j) &= \sum_{j=1}^{\infty} p(1-p)^{n+j-1} \\ &= p(1-p)^n \sum_{j=0}^{\infty} (1-p)^j \\ &= p(1-p)^n \frac{1}{1-(1-p)} \\ &= (1-p)^n,\end{aligned}$$

we can see that

$$\begin{aligned}\mathbb{P}(X = n+k \mid X > n) &= \frac{p(1-p)^{n+k-1}}{(1-p)^n} \\ &= p(1-p)^{k-1} \\ &= \mathbb{P}(X = k).\end{aligned}$$

6. (i) We can derive the result in two ways:

1st way. Given $n = 0, 1, 2, \ldots$, we calculate

$$\begin{aligned}\mathbb{P}(X + Y = n) &= \sum_{k=0}^{n} \mathbb{P}(X = k, Y = n-k) \\ &= \sum_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n-k) \quad \text{(because } X, Y \text{ are independent)} \\ &= \sum_{k=0}^{n} e^{-\lambda} \frac{\lambda^k}{k!}\, e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda^k \mu^{n-k} \\ &= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!},\end{aligned}$$

which proves that $X + Y$ is a Poisson random variable with parameter $\lambda + \mu$.

2nd way. The moment generating functions of the random variables X and Y are given by

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

and

$$M_Y(t) = e^{-\mu+\mu e^t}. \tag{102}$$

Now, since X and Y are independent,

$$M_{X+Y}(t) \quad = M_X(t)M_Y(t) = e^{-(\lambda+\mu)+(\lambda+\mu)e^t},$$

which proves that $X+Y$ is a Poisson random variable with parameter $\lambda+\mu$.

(ii) Given $m = 0, 1, 2, \ldots, n$,

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

which proves that the conditional distribution of X given that $X+Y = n$ is binomial with parameters n, $\frac{\lambda}{\lambda+\mu}$.

7. i) Given that $N = n$, X is binomial with parameters n, p, and so,

$$\mathbb{E}\left[X \mid N = n\right] = pn.$$

ii) From the previous step, we can see that

$$\mathbb{E}\left[X \mid \sigma(N)\right] \equiv \mathbb{E}\left[X \mid N\right] = pN.$$

In view of this observation and the tower property of conditional expectation, we calculate

$$\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}\left[X \mid N\right]\right] = \mathbb{E}[pN] = p\mu.$$

8. Suppose that X is a simple random variable, so that

$$X = \sum_{i=1}^{n} x_i \mathbf{1}_{B_i}, \tag{103}$$

for some *distinct* $x_1, \ldots, x_n \in \mathbb{R}$ and some pairwise disjoint events $B_1, \ldots, B_n \in \mathcal{F}$ such that $\bigcup_{i=1}^n B_i = \Omega$. Since the numbers x_i in the representation (103) of X are distinct, the σ-algebra $\sigma(X)$ is easy to describe: it consists of all possible unions of sets in the family $\{B_1, \ldots, B_n\}$, i.e.,

$$\sigma(X) = \left\{ \bigcup_{j \in J} B_j \mid \ J \subseteq \{1, \ldots, n\} \right\},$$

with the usual convention that

$$\bigcup_{j \in \emptyset} B_j = \emptyset.$$

For example, if $n = 4$,

$$\sigma(X) = \big\{\emptyset, B_1, B_2, B_3, B_4, B_1 \cup B_2,\ B_1 \cup B_3,\ B_1 \cup B_4,\ B_2 \cup B_3,\ B_2 \cup B_4,\ B_3 \cup B_4,\\ B_1 \cup B_2 \cup B_3,\ B_1 \cup B_2 \cup B_4,\ B_2 \cup B_3 \cup B_4, \Omega\big\}.$$

Now, let any event $A \in \mathcal{F}$. A version of the conditional probability $\mathbb{P}\big(A \mid \sigma(X)\big) \equiv \mathbb{E}\big[1_A \mid \sigma(X)\big]$ is given by the random variable Y defined by

$$Y = \sum_{i=1}^n \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \mathbf{1}_{B_i}.$$

To prove this claim, we check the defining properties of conditional expectation:

(i) Noting that

$$Y = \sum_{i=1}^n c_i \mathbf{1}_{B_i},$$

where the *constants* c_i are given by

$$c_i = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)}, \quad i = 1, \ldots, n,$$

we can see that Y is $\sigma(X)$-measurable.

(ii) We calculate

$$\mathbb{E}\big[|Y|\big] = \mathbb{E}\big[Y\big] = \sum_{i=1}^n \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \mathbb{P}(B_i) = \mathbb{P}(A) < \infty,$$

the last equality following because $B_1, \ldots, B_n$ are pairwise disjoint.

(iii) Let B be any event in $\sigma(X)$. There exists a set $J \subseteq \{1, \dots, n\}$ such that

$$B = \bigcup_{j \in J} B_j.$$

Since $B_1, \dots, B_n$ are pairwise disjoint,

$$\mathbf{1}_B = \sum_{j \in J} \mathbf{1}_{B_j} \quad \text{and} \quad \mathbf{1}_{B_i} \mathbf{1}_{B_j} = 0 \quad \text{for all } i \neq j.$$

In view of these observations, we calculate

$$\begin{aligned}
\mathbb{E}\big[Y \mathbf{1}_B\big] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \mathbf{1}_{B_i}\right)\left(\sum_{j \in J} \mathbf{1}_{B_j}\right)\right] \\
&= \mathbb{E}\left[\sum_{j \in J} \sum_{i=1}^{n} \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \mathbf{1}_{B_i} \mathbf{1}_{B_j}\right] \\
&= \mathbb{E}\left[\sum_{j \in J} \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(B_j)} \mathbf{1}_{B_j}\right] \\
&= \sum_{j \in J} \mathbb{P}(A \cap B_j) \\
&= \sum_{j \in J} \mathbb{E}\left[\mathbf{1}_{A \cap B_j}\right] \\
&= \sum_{j \in J} \mathbb{E}\left[\mathbf{1}_A \mathbf{1}_{B_j}\right] \\
&= \mathbb{E}\left[\mathbf{1}_A \sum_{j \in J} \mathbf{1}_{B_j}\right] \\
&= \mathbb{E}\left[\mathbf{1}_A \mathbf{1}_B\right].
\end{aligned}$$

9. Suppose that a random variable X satisfies

$$X = c, \quad \mathbb{P}\text{-a.s.}, \tag{104}$$

for some constant $c \in \mathbb{R}$. We can prove that the random variable $Y = c$ is a version of $\mathbb{E}[X \mid \mathcal{G}]$, for any σ-algebra $\mathcal{G}$, by checking the defining properties of conditional expectation:

(i) We can easily check that Y is $\mathcal{G}$-measurable (do convince yourselves about this!).

(ii) $\mathbb{E}\big[|Y|\big] = |c| < \infty$.

(iii) Given any event $A \in \mathcal{G}$,

$$\mathbb{E}\left[Y \mathbf{1}_A\right] = \mathbb{E}\left[c \mathbf{1}_A\right] = \mathbb{E}\left[X \mathbf{1}_A\right].$$

The second equality here follows because $\mathbb{E}[Z_1] = \mathbb{E}[Z_2]$ for all random variables Z_1, Z_2 such that $Z_1 = Z_2$, $\mathbb{P}$-a.s., and (104) implies that $X \mathbf{1}_A = c \mathbf{1}_A$, $\mathbb{P}$-a.s..

10. To prove this result, we argue by contradiction and we assume that X is a random variable such that $X \geq 0$, $\mathbb{P}$-a.s., and

$$\mathbb{P}\big(\mathbb{E}[X \mid \mathcal{G}] < 0\big) > 0. \tag{105}$$

We consider the sequence of events $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ defined by

$$A_n = \left\{\omega \in \Omega \mid \ \mathbb{E}[X \mid \mathcal{G}](\omega) \leq -\frac{1}{n}\right\},$$

and we observe that

$$\big\{\omega \in \Omega \mid \ \mathbb{E}[X \mid \mathcal{G}](\omega) < 0\big\} = \bigcup_{n=1}^{\infty} A_n.$$

Since $\mathbb{E}[X \mid \mathcal{G}]$ is a $\mathcal{G}$-measurable random variable, by definition, $A_n \in \mathcal{G}$ for all $n \geq 1$. Furthermore, the "continuity" of a probability measure and (105) imply that there exists $N \geq 1$ such that $\mathbb{P}(A_N) > 0$. In view of these observations, we calculate

$$\begin{aligned}
\mathbb{E}\left[X\mathbf{1}_{A_N}\right] &= \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]\mathbf{1}_{A_N}\right] \\
&\leq \mathbb{E}\left[-\frac{1}{N}\mathbf{1}_{A_N}\right] \\
&= -\frac{1}{N}\mathbb{P}(A_N) \\
&< 0.
\end{aligned}$$

These inequalities contradict the assumption that $X \geq 0$, $\mathbb{P}$-a.s., because the inequality $X \geq 0$, $\mathbb{P}$-a.s., implies that $X\mathbf{1}_{A_N} \geq 0$, $\mathbb{P}$-a.s., which, in view of the fact that $\mathbb{E}[Z] \geq 0$ for every random variable Z such that $Z \geq 0$, $\mathbb{P}$-a.s., implies that $\mathbb{E}\left[X\mathbf{1}_{A_N}\right] \geq 0$. As a consequence, (105) cannot be true, which proves that $\mathbb{E}[X \mid \mathcal{G}] \geq 0$, $\mathbb{P}$-a.s..

11. We can show that the random variable $Y = \mathbb{E}[X]$ is a version of the conditional expectation $\mathbb{E}\left[X \mid \{\Omega, \emptyset\}\right]$ of the random variable X given the trivial σ-algebra $\{\Omega, \emptyset\}$ by checking the three defining properties of conditional expectation:

(i) Y is $\{\Omega, \emptyset\}$-measurable because $Y = \mathbb{E}[X]$ is a constant.

(ii) Plainly, $\mathbb{E}[|Y|] = |\mathbb{E}[X]| < \infty$.

(iii) Given $C \in \{\Omega, \emptyset\}$, we can see that

$$\begin{aligned}
\mathbb{E}\left[Y\mathbf{1}_C\right] &= \mathbb{E}[X]\mathbb{E}\left[\mathbf{1}_C\right] = \mathbb{E}[X]\mathbb{P}(C) \\
&= \begin{cases} \mathbb{E}[X], & \text{if } C = \Omega, \\ 0, & \text{if } C = \emptyset, \end{cases} = \begin{cases} \mathbb{E}[X\mathbf{1}_\Omega], & \text{if } C = \Omega, \\ \mathbb{E}[X\mathbf{1}_\emptyset], & \text{if } C = \emptyset, \end{cases} = \mathbb{E}\left[X\mathbf{1}_C\right].
\end{aligned}$$

12. **(i)** Since $\Omega \in \mathcal{G}$, we can appeal to part (iii) of the definition of conditional expectation to obtain

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

(ii) For $i = 1, 2$, let Y_i be a version of $\mathbb{E}[X_i \mid \mathcal{G}]$. Then, $a_1Y_1 + a_2Y_2$ is integrable because $\mathbb{E}\big[|a_1Y_1 + a_2Y_2|\big] \le |a_1|\mathbb{E}\big[|Y_1|\big] + |a_2|\mathbb{E}\big[|Y_2|\big] < \infty$, and $a_1Y_1 + a_2Y_2$ is $\mathcal{G}$-measurable. Also, given an event $A \in \mathcal{G}$, we can use the linearity of expectation to obtain

$$\begin{aligned}
\mathbb{E}\big[(a_1Y_1 + a_2Y_2)\mathbf{1}_A\big] &= a_1\mathbb{E}\big[Y_1\mathbf{1}_A\big] + a_2\mathbb{E}\big[Y_2\mathbf{1}_A\big] \\
&= a_1\mathbb{E}\big[X_1\mathbf{1}_A\big] + a_2\mathbb{E}\big[X_2\mathbf{1}_A\big] \\
&= \mathbb{E}\big[(a_1X_1 + a_2X_2)\mathbf{1}_A\big].
\end{aligned}$$

These considerations prove that $a_1Y_1 + a_2Y_2$ is a version of conditional expectation $\mathbb{E}[a_1X_1 + a_2X_2 \mid \mathcal{G}]$, and the result follows.

(iii) $\mathbb{E}[X \mid \mathcal{H}]$ is a version of the conditional expectation of $\mathbb{E}[X \mid \mathcal{G}]$ given $\mathcal{H}$ because it is integrable and $\mathcal{H}$-measurable, and, given an event $A \in \mathcal{H} \subseteq \mathcal{G}$,

$$\begin{aligned}
\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{H}]\mathbf{1}_A\big] &= \mathbb{E}\big[X\mathbf{1}_A\big] \\
&= \mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}]\mathbf{1}_A\big].
\end{aligned}$$

Solutions 5

1. First, we observe that, since $X_k \in \mathcal{L}^1$ for all $k \geq 1$, and $X_1, X_2, \ldots, X_n$ are independent, positive random variables, $M_n = X_1 X_2 \cdots X_n \in \mathcal{L}^1$ for all $n \geq 1$.

Now, we can check the martingale property as follows. Given $m < n$,

$$\begin{aligned}
\mathbb{E}\big[M_n \mid \mathcal{F}_m\big] &= \mathbb{E}\big[X_1 X_2 \cdots X_m X_{m+1} \cdots X_n \mid \mathcal{F}_m\big] \\
&= X_1 X_2 \cdots X_m \mathbb{E}\big[X_{m+1} \cdots X_n \mid \mathcal{F}_m\big] \\
&= X_1 X_2 \cdots X_m \\
&= M_m.
\end{aligned}$$

The second equality follows because $X_1, X_2, \ldots, X_m$ are $\mathcal{F}_m$ measurable, and the third one because $X_{m+1}, \ldots, X_n$ are independent of $\mathcal{F}_m = \sigma(X_1, X_2, \ldots, X_m)$.

2. Since $X_k \in \mathcal{L}^1$ for all $k \geq 1$, the definition of conditional expectation implies that $\mathbb{E}\big[X_k \mid \mathcal{F}_{k-1}\big] \in \mathcal{L}^1$ for all $k \geq 1$. Therefore,

$$\begin{aligned}
\mathbb{E}\big[|M_n|\big] &= \left| \sum_{i=1}^{n} \big(X_i - \mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big) \right| \\
&\leq \sum_{i=1}^{n} \big(|X_i| + \big|\mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big|\big) \\
&< \infty.
\end{aligned}$$

Now, we can check the martingale property as follows. Given $m < n$,

$$\begin{aligned}
\mathbb{E}\big[M_n \mid \mathcal{F}_m\big] &= \mathbb{E}\left[\sum_{i=1}^{n} \big(X_i - \mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big) \mid \mathcal{F}_m \right] \\
&= \sum_{i=1}^{m} \big(X_i - \mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big) + \mathbb{E}\left[\sum_{i=m+1}^{n} \big(X_i - \mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big) \mid \mathcal{F}_m \right] \\
&= \sum_{i=1}^{m} \big(X_i - \mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big) + \sum_{i=m+1}^{n} \big(\mathbb{E}\big[X_i \mid \mathcal{F}_m\big] - \mathbb{E}\big[\mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big] \mid \mathcal{F}_m\big] \\
&= \sum_{i=1}^{m} \big(X_i - \mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big]\big).
\end{aligned}$$

Note that the fourth equality here follows by the tower property of conditional expectation. Indeed, since $\mathcal{F}_{i-1} \supseteq \mathcal{F}_m$ for all $i \geq m+1$,

$$\mathbb{E}\big[\mathbb{E}\big[X_i \mid \mathcal{F}_{i-1}\big] \mid \mathcal{F}_m\big] = \mathbb{E}\big[X_i \mid \mathcal{F}_m\big] \quad \text{for all } i \geq m+1.$$

3. Using the fact that $\mathbb{E}\big[X_j \mid \mathcal{F}_j\big] = X_j$, which follows from the assumption that (X_n) is $(\mathcal{F}_n)$-adapted, and the assumption that $X_0 = 0$, we first observe that

$$M_n = \sum_{j=1}^{n} c_j X_j - \sum_{j=1}^{n} c_j \mathbb{E}\big[X_j \mid \mathcal{F}_{j-1}\big].$$

In view of this expression, we can use the properties of conditional expectation to check the martingale property as follows.

$$\begin{aligned}
\mathbb{E}\big[M_{n+1} \mid \mathcal{F}_n\big] &= \mathbb{E}\left[\sum_{j=1}^{n+1} c_j X_j - \sum_{j=1}^{n+1} c_j \mathbb{E}\big[X_j \mid \mathcal{F}_{j-1}\big] \,\middle|\, \mathcal{F}_n\right] \\
&= \sum_{j=1}^{n+1} c_j \mathbb{E}\big[X_j \mid \mathcal{F}_n\big] - \sum_{j=1}^{n+1} c_j \mathbb{E}\big[\mathbb{E}\big[X_j \mid \mathcal{F}_{j-1}\big] \mid \mathcal{F}_n\big] \\
&= c_{n+1}\mathbb{E}\big[X_{n+1} \mid \mathcal{F}_n\big] + \sum_{j=1}^{n} c_j X_j - c_{n+1}\mathbb{E}\big[X_{n+1} \mid \mathcal{F}_n\big] - \sum_{j=1}^{n} c_j \mathbb{E}\big[X_j \mid \mathcal{F}_{j-1}\big] \\
&= M_n.
\end{aligned}$$

4. Using the properties of the standard Brownian motion, we can calculate

(i)

$$\begin{aligned}
\mathbb{E}\left[(W_t - W_s)(W_s - W_r)\right] &= \mathbb{E}\left[W_t - W_s\right]\,\mathbb{E}\left[W_s - W_r\right] \quad \text{(independence)} \\
&= 0 \times 0 \quad \text{(normality)} \\
&= 0.
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathbb{E}\left[(W_u - W_t)^2(W_s - W_r)^2\right] &= \mathbb{E}\left[(W_u - W_t)^2\right]\,\mathbb{E}\left[(W_s - W_r)^2\right] \quad \text{(independence)} \\
&= (u - t)(s - r). \quad \text{(normality)}
\end{aligned}$$

(iii)

$$\begin{aligned}
&\mathbb{E}\big[(W_u - W_s)(W_t - W_r)\big] \\
&\quad = \mathbb{E}\left[\Big\{(W_u - W_t) + (W_t - W_s)\Big\}\Big\{(W_t - W_s) + (W_s - W_r)\Big\}\right] \\
&\quad = \mathbb{E}\left[(W_u - W_t)(W_t - W_s)\right] + \mathbb{E}\left[(W_u - W_t)(W_s - W_r)\right] \\
&\qquad + \mathbb{E}\left[(W_t - W_s)^2\right] + \mathbb{E}\left[(W_t - W_s)(W_s - W_r)\right] \\
&\quad = \mathbb{E}\left[W_u - W_t\right]\,\mathbb{E}\left[W_t - W_s\right] + \mathbb{E}\left[W_u - W_t\right]\,\mathbb{E}\left[W_s - W_r\right] \\
&\qquad + \mathbb{E}\left[(W_t - W_s)^2\right] + \mathbb{E}\left[W_t - W_s\right]\,\mathbb{E}\left[W_s - W_r\right] \quad \text{(independence)} \\
&\quad = 0 \times 0 + 0 \times 0 + (t - s) + 0 \times 0 \quad \text{(normality)} \\
&\quad = t - s.
\end{aligned}$$

(iv)

$$
\begin{aligned}
&\mathbb{E}\big[(W_t - W_r)(W_s - W_r)^2\big] \\
&\qquad = \mathbb{E}\left[\left\{(W_t - W_s) + (W_s - W_r)\right\}(W_s - W_r)^2\right] \\
&\qquad = \mathbb{E}\left[(W_t - W_s)(W_s - W_r)^2\right] + \mathbb{E}\left[(W_s - W_r)^3\right] \\
&\qquad = \mathbb{E}\left[W_t - W_s\right]\mathbb{E}\left[(W_s - W_r)^2\right] + \mathbb{E}\left[(W_s - W_r)^3\right] \quad \text{(independence)} \\
&\qquad = 0 \times (s - r) + 0 \quad \text{(normality)} \\
&\qquad = 0.
\end{aligned}
$$

(v)

$$
\begin{aligned}
\mathbb{E}\left[W_r W_s W_t\right] = \mathbb{E}\Big[&(W_r - W_0)\left\{(W_s - W_r) + (W_r - W_0)\right\} \\
&\times\left\{(W_t - W_s) + (W_s - W_r) + (W_r - W_0)\right\}\Big] \quad (W_0 = 0) \\
= \mathbb{E}&\left[(W_r - W_0)(W_s - W_r)(W_t - W_s)\right] + \mathbb{E}\left[(W_r - W_0)(W_s - W_r)^2\right] \\
&+ \mathbb{E}\left[(W_r - W_0)^2(W_s - W_r)\right] + \mathbb{E}\left[(W_r - W_0)^2(W_t - W_s)\right] \\
&+ \mathbb{E}\left[(W_r - W_0)^2(W_s - W_r)\right] + \mathbb{E}\left[(W_r - W_0)^3\right] \\
= \mathbb{E}&\left[W_r - W_0\right]\mathbb{E}\left[W_s - W_r\right]\mathbb{E}\left[W_t - W_s\right] + \mathbb{E}\left[W_r - W_0\right]\mathbb{E}\left[(W_s - W_r)^2\right] \\
&+ \mathbb{E}\left[(W_r - W_0)^2\right]\mathbb{E}\left[W_s - W_r\right] + \mathbb{E}\left[(W_r - W_0)^2\right]\mathbb{E}\left[W_t - W_s\right] \\
&+ \mathbb{E}\left[(W_r - W_0)^2\right]\mathbb{E}\left[W_s - W_r\right] + \mathbb{E}\left[(W_r - W_0)^3\right] \quad \text{(independence)} \\
= 0 &\times 0 \times 0 + 0 \times (s - r) + r \times 0 + r \times 0 + r \times 0 + 0 \quad \text{(normality)} \\
= 0&.
\end{aligned} \tag{106}
$$

5. We have to verify that (X_t) satisfies the properties that define a standard Brownian motion. To do so, we exploit the fact that (W_t) itself is a standard Brownian motion.

(i) Since $W_0 = 0$,

$$X_0 = \frac{1}{\sqrt{c}} W_0 = 0.$$

(ii) (X_t) has continuous sample paths because this is true for (W_t).

(iii) Given any times $t_1 < t_2 < \cdots < t_k$, the random variables $X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ are independent because

$$X_{t_i} - X_{t_{i-1}} = \frac{1}{\sqrt{c}}\left(W_{ct_i} - W_{ct_{i-1}}\right), \quad i = 2, 3, \ldots, k,$$

and the random variables $W_{ct_2} - W_{ct_1}, \ldots, W_{ct_k} - W_{ct_{k-1}}$ are independent (because they are increments of the Brownian motion (W_t) in non-overlapping time intervals).

(iv) Given times $s < t$, the increment

$$X_t - X_s = \frac{1}{\sqrt{c}} \left[W_{ct} - W_{cs}\right]$$

is normally distributed because this is true for the increment $W_{ct} - W_{cs}$ of the Brownian motion (W_t). Furthermore, $X_t - X_s$ has mean

$$\mathbb{E}\left[X_t - X_s\right] = \frac{1}{\sqrt{c}} \mathbb{E}\left[W_{ct} - W_{cs}\right] = 0$$

and variance

$$\begin{aligned}
\mathbb{E}\left[(X_t - X_s)^2\right] &= \frac{1}{c} \mathbb{E}\left[(W_{ct} - W_{cs})^2\right] \\
&= \frac{1}{c}\left(ct - cs\right) \\
&= t - s.
\end{aligned} \tag{107}$$

These observations imply that $X_t - X_s \sim N(0, t-s)$.

6. **(I)** First, we note that (X_t) is $(\mathcal{F}_t)$-adapted and

$$\mathbb{E}\big[|X_t|\big] \leq 2\mathbb{E}\big[X_t^2\big] + 2t = 4t < \infty \quad \text{for all } t \geq 0.$$

Next, we consider any times $s < t$. Using the fact that the random variable $W_t - W_s$ is independent of $\mathcal{F}_s$ and the fact that (W_t) is a Brownian motion as well as an $(\mathcal{F}_t)$-martingale, we calculate

$$\begin{aligned}
\mathbb{E}[(W_t - W_s)^2 \mid \mathcal{F}_s] &= \mathbb{E}[(W_t - W_s)^2] \\
&= t - s,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[(W_t - W_s)^2 \mid \mathcal{F}_s] &= \mathbb{E}[W_t^2 \mid \mathcal{F}_s] + \mathbb{E}[W_s^2 \mid \mathcal{F}_s] - 2\mathbb{E}[W_t W_s \mid \mathcal{F}_s] \\
&= \mathbb{E}[W_t^2 \mid \mathcal{F}_s] + W_s^2 - 2W_s \mathbb{E}[W_t \mid \mathcal{F}_s] \\
&= \mathbb{E}[W_t^2 \mid \mathcal{F}_s] - W_s^2.
\end{aligned}$$

However, equating the right hand sides of these two expressions, we obtain

$$\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s.$$

(II) First, we note that, since θW_t is a normal random variable with mean 0 and variance $\theta^2 t$, the identities

$$\mathbb{E}\big[|Y_t|\big] = \mathbb{E}\big[Y_t\big] = 1.$$

follow immediately from Exercise 3.11. Next, we consider any times $s < t$, and we use the facts that the random variable $\theta(W_t - W_s)$ is normal with mean 0 and variance $\theta^2(t-s)$, and is independent of $\mathcal{F}_s$, to calculate

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

which proves that

$$\mathbb{E}[Y_t \mid \mathcal{F}_s] = Y_s.$$

Solutions 6

1. If f is the function given by $f(x) = x^2$, then $f'(x) = 2x$ and $f''(x) = 2$, and Itô's formula implies

$$\begin{aligned}
W_t^2 &\equiv f(W_t) \\
&= f(W_0) + \int_0^t f'(W_s)\, dW_s + \tfrac{1}{2} \int_0^t f''(W_s)\,(dW_s)^2 \\
&= \int_0^t 2W_s\, dW_s + \tfrac{1}{2} \int_0^t 2\, ds \\
&= t + 2 \int_0^t W_s\, dW_s.
\end{aligned}$$

Similarly, if g is the function given by $g(y) = y^{27}$, then $g'(y) = 27y^{26}$ and $g''(y) = 702y^{25}$, and Itô's formula implies

$$\begin{aligned}
W_t^{27} &\equiv g(W_t) \\
&= g(W_0) + \int_0^t g'(W_s)\, dW_s + \tfrac{1}{2} \int_0^t g''(W_s)\,(dW_s)^2 \\
&= 351 \int_0^t W_s^{25}\, ds + 27 \int_0^t W_s^{26}\, dW_s.
\end{aligned}$$

2. We consider the function F given by $F(z) = z^k$, so that

$$F'(z) = kz^{k-1} \quad \text{and} \quad F''(z) = k(k-1)z^{k-2}.$$

An application of Itô's formula yields

$$\begin{aligned}
W_t^k &\equiv F(W_t) \\
&= F(W_0) + \int_0^t F'(W_s)\, dW_s + \tfrac{1}{2} \int_0^t F''(W_s)\,(dW_s)^2 \\
&= \frac{k(k-1)}{2} \int_0^t W_s^{k-2}\, ds + k \int_0^t W_s^{k-1}\, dW_s.
\end{aligned}$$

Taking expectations and noting that the stochastic integral has expectation 0, we obtain

$$\begin{aligned}
\mathbb{E}\left[W_t^k\right] &= \frac{k(k-1)}{2} \mathbb{E}\left[\int_0^t W_s^{k-2}\, ds\right] \\
&= \frac{k(k-1)}{2} \int_0^t \mathbb{E}\left[W_s^{k-2}\right] ds.
\end{aligned}$$

We can use this identity to calculate

$$\mathbb{E}\left[W_t^4\right] = 6 \int_0^t \mathbb{E}\left[W_s^2\right] ds = 6 \int_0^t s\, ds = 3t^2$$

and

$$\mathbb{E}\left[W_t^6\right] = 15\int_0^t \mathbb{E}\left[W_s^4\right] ds = 45\int_0^t s^2\, ds = 15t^3.$$

3. We can solve this exercise in two ways:

Way 1. We consider the function H given by $H(t,x) = e^{-t}x$, and we calculate

$$H_t(t,x) = -e^{-t}x, \quad H_x(t,x) = e^{-t} \quad \text{and} \quad H_{xx}(t,x) = 0.$$

Using Itô's formula, we can see that the stochastic process (Z_t) given by $Z_t = e^{-t}W_t$ satisfies

$$\begin{aligned} Z_t &\equiv H(t,W_t) \\ &= H(0,W_0) + \int_0^t H_t(s,W_s)\, ds + \int_0^t H_x(s,W_s)\, dW_s + \tfrac{1}{2}\int_0^t H_{xx}(s,W_s)\,(dW_s)^2 \\ &= H(0,W_0) + \int_0^t \left[H_t(s,W_s) + \tfrac{1}{2}H_{xx}(s,W_s)\right] ds + \int_0^t H_x(s,W_s)\, dW_s \\ &= -\int_0^t e^{-s}W_s\, ds + \int_0^t e^{-s}\, dW_s \\ &= -\int_0^t Z_s\, ds + \int_0^t e^{-s}\, dW_s. \end{aligned}$$

Way 2. We consider the Itô processes (X_t) and (Y_t) defined by $X_t = e^{-t}$ and $Y_t = W_t$, and we note that

$$\begin{aligned} dX_t &= -e^{-t}\, dt + 0\, dW_t, \quad X_0 = 1, \\ dY_t &= 0\, dt + dW_t, \quad Y_0 = 0. \end{aligned}$$

Using the integration by parts formula, we can see that the stochastic process (Z_t) defined by $Z_t = e^{-t}W_t \equiv X_tY_t$ satisfies

$$\begin{aligned} Z_t &\equiv X_tY_t \\ &= X_0Y_0 + \int_0^t Y_s\, dX_s + \int_0^t X_s\, dY_s + \int_0^t (dX_s)(dY_s) \\ &= X_0Y_0 + \int_0^t Y_s\left[-e^{-s}\right] ds + \int_0^t X_s dW_s \\ &= -\int_0^t e^{-s}W_s\, ds + \int_0^t e^{-s}\, dW_s \\ &= -\int_0^t Z_s\, ds + \int_0^t e^{-s}\, dW_s. \end{aligned}$$

4. (a) We consider the Itô processes (X_t) and (Y_t) defined by $X_t = e^{kt}$ and $Y_t = r_t$, and we note that

$$dX_t = ke^{kt}\,dt + 0\,dW_t, \quad X_0 = 1,$$
$$dY_t = k(\vartheta - r_t)\,dt + \sigma\,dW_t, \quad Y_0 = r_0.$$

Using the integration by parts formula, we calculate

$$\begin{aligned}
e^{kt}r_t &\equiv X_tY_t \\
&= X_0Y_0 + \int_0^t Y_s\,dX_s + \int_0^t X_s\,dY_s + \int_0^t (dX_s)(dY_s) \\
&= X_0Y_0 + \int_0^t \left[ke^{ks}Y_s + k(\vartheta - r_s)X_s\right] ds + \int_0^t \sigma X_s\,dW_s \\
&= r_0 + \int_0^t \left[ke^{ks}r_s + k(\vartheta - r_s)e^{ks}\right] ds + \int_0^t \sigma e^{ks}\,dW_s \\
&= r_0 + \vartheta e^{kt} - \vartheta + \sigma \int_0^t e^{ks}\,dW_s.
\end{aligned}$$

This implies that

$$r_t = \vartheta + (r_0 - \vartheta)e^{-kt} + \sigma e^{-kt} \int_0^t e^{ks}\,dW_s.$$

(b) Noting that the stochastic integral with respect to the Brownian motion has expectation 0, we can see that the expectation of the random variable r_t is given by

$$\mathbb{E}\left[r_t\right] = \vartheta + (r_0 - \vartheta)e^{-kt}.$$

Using Itô's isometry, we calculate

$$\begin{aligned}
\operatorname{var}(r_t) \equiv \mathbb{E}\left[\left(r_t - \mathbb{E}[r_t]\right)^2\right] &= \sigma^2 e^{-2kt}\mathbb{E}\left[\left(\int_0^t e^{ks}\,dW_s\right)^2\right] \\
&= \frac{\sigma^2}{2k}\left(1 - e^{-2kt}\right).
\end{aligned}$$

5. We consider the Itô processes (X_t) and (Y_t) defined by $X_t = e^{kt}$ and $Y_t = r_t$, and we note that

$$dX_t = ke^{kt}\,dt + 0\,dW_t, \quad X_0 = 1,$$
$$dY_t = k(\vartheta - r_t)\,dt + \sigma\sqrt{r_t}\,dW_t, \quad Y_0 = r_0.$$

Using the integration by parts formula, we calculate

$$
\begin{aligned}
e^{kt} r_t &\equiv X_t Y_t \\
&= X_0 Y_0 + \int_0^t Y_s \, dX_s + \int_0^t X_s \, dY_s + \int_0^t (dX_s)(dY_s) \\
&= X_0 Y_0 + \int_0^t \left[k e^{ks} Y_s + k(\vartheta - r_s) X_s \right] ds + \int_0^t \sigma \sqrt{r_s} X_s \, dW_s \\
&= r_0 + \int_0^t \left[k e^{ks} r_s + k(\vartheta - r_s) e^{ks} \right] ds + \int_0^t e^{ks} \sigma \sqrt{r_s} \, dW_s \\
&= r_0 + \vartheta e^{kt} - \vartheta + \sigma \int_0^t e^{ks} \sqrt{r_s} \, dW_s.
\end{aligned}
$$

It follows that

$$
r_t = \vartheta + (r_0 - \vartheta) e^{-kt} + \sigma e^{-kt} \int_0^t e^{ks} \sqrt{r_s} \, dW_s.
$$

Assuming that stochastic integrals with respect to the Brownian motion have expectation 0, we can see that the mean of r_t is given by

$$
\mathbb{E}\left[r_t\right] = \vartheta + (r_0 - \vartheta) e^{-kt}. \tag{108}
$$

Furthermore, using Itô's isometry, we calculate

$$
\begin{aligned}
\operatorname{var}(r_t) &= \mathbb{E}\left[\left(r_t - \mathbb{E}\left[r_t\right] \right)^2 \right] \\
&= \sigma^2 e^{-2kt} \mathbb{E}\left[\left(\int_0^t e^{ks} \sqrt{r_s} \, dW_s \right)^2 \right] \\
&= \sigma^2 e^{-2kt} \mathbb{E}\left[\int_0^t e^{2ks} r_s \, ds \right] \\
&= \sigma^2 e^{-2kt} \int_0^t e^{2ks} \mathbb{E}\left[r_s\right] \, ds.
\end{aligned}
$$

Substituting the expression for $\mathbb{E}\left[r_s\right]$ given by (108) into this and integrating, we obtain

$$
\operatorname{var}(r_t) = \frac{\sigma^2 \vartheta}{2k} \left(1 - e^{-kt} \right)^2 + \frac{\sigma^2}{k} r_0 e^{-kt} \left(1 - e^{-kt} \right).
$$

6. **(i)** We consider the function $F(t,x) = \sin(tx)$, and we calculate

$$
F_t(t,x) = x \cos(tx), \quad F_x(t,x) = t \cos(tx) \quad \text{and} \quad F_{xx}(t,x) = -t^2 \sin(tx).
$$

An application of Itô's formula yields

$$\begin{aligned}
Z_t &\equiv F(t, X_t) \\
&= F(0,x) + \int_0^t F_t(s, X_s)\, ds + \int_0^t F_x(s, X_s)\, dX_s + \tfrac{1}{2} \int_0^t F_{xx}(s, X_s)\, (dX_s)^2 \\
&= F(0,x) + \int_0^t \Big[F_t(s, X_s) + \tfrac{1}{2} \left[\cos(s^2 W_s)\right]^2 F_{xx}(s, X_s) + s e^{W_s} F_x(s, X_s) \Big]\, ds \\
&\quad + \int_0^t \cos(s^2 W_s) F_x(s, X_s)\, dW_s.
\end{aligned}$$

In view of these calculations, we can see that the required quantities are given by

$$\begin{aligned}
&Z_0 = 0, \\
&A_t = X_t \cos(tX_t) - \tfrac{1}{2} t^2 \cos^2(t^2 W_t) \sin(tX_t) + t^2 e^{W_t} \cos(tX_t), \\
&C_t = t \cos(t^2 W_t) \cos(tX_t).
\end{aligned}$$

(ii) We consider the function $G(t,x) = x \exp(t^2 x)$, and we calculate

$$G_t(t,x) = 2tx^2 e^{t^2 x}, \quad G_x(t,x) = \left[1 + t^2 x\right] e^{t^2 x} \quad \text{and} \quad G_{xx}(t,x) = \left[2 + t^2 x\right] t^2 e^{t^2 x}.$$

An application of Itô's formula yields

$$\begin{aligned}
Z_t &\equiv G(t, X_t) \\
&= G(0,x) + \int_0^t G_t(s, X_s)\, ds + \int_0^t G_x(s, X_s)\, dX_s + \tfrac{1}{2} \int_0^t G_{xx}(s, X_s)\, (dX_s)^2 \\
&= G(0,x) + \int_0^t \Big[G_t(s, X_s) + \tfrac{1}{2} \left[\cos(s^2 W_s)\right]^2 G_{xx}(s, X_s) + s e^{W_s} G_x(s, X_s) \Big]\, ds \\
&\quad + \int_0^t \cos(s^2 W_s) G_x(s, X_s)\, dW_s.
\end{aligned}$$

In view of these calculations, we can see that the required quantities are given by

$$\begin{aligned}
&Z_0 = \sqrt{\pi}, \\
&A_t = \left(2tX_t^2 + \tfrac{1}{2} \cos^2(t^2 W_t) \left[2 + t^2 X_t\right] t^2 + t e^{W_t} \left[1 + t^2 X_t\right]\right) e^{t^2 X_t}, \\
&C_t = \cos(t^2 W_t) \left[1 + t^2 X_t\right] e^{t^2 X_t}.
\end{aligned}$$

(iii) We consider the function $H(t,x) = x^3 + t\cos(x)$, and we calculate

$$H_t(t,x) = \cos(x), \quad H_x(t,x) = 3x^2 - t\sin(x) \quad \text{and} \quad H_{xx}(t,x) = 6x - t\cos(x).$$

An application of Itô's formula yields

$$
\begin{aligned}
Z_t &\equiv H(t, X_t) \\
&= H(0,x) + \int_0^t H_t(s, X_s)\, ds + \int_0^t H_x(s, X_s)\, dX_s + \tfrac{1}{2} \int_0^t H_{xx}(s, X_s)\, (dX_s)^2 \\
&= H(0,x) + \int_0^t \Big[H_t(s, X_s) + \tfrac{1}{2} \left[\cos(s^2 W_s)\right]^2 H_{xx}(s, X_s) + s e^{W_s} H_x(s, X_s) \Big]\, ds \\
&\quad + \int_0^t \cos(s^2 W_s) H_x(s, X_s)\, dW_s.
\end{aligned}
$$

In view of these calculations, we can see that the required quantities are given by

$$
\begin{aligned}
&Z_0 = \pi^{3/2}, \\
&A_t = \cos(X_t) + \tfrac{1}{2}\cos^2(t^2 W_t)\left[6X_t - t\cos(X_t)\right] + t e^{W_t}\left[3X_t^2 - t\sin(X_t)\right], \\
&C_t = \cos(t^2 W_t)\left[3X_t^2 - t\sin(X_t)\right].
\end{aligned}
$$

7. We consider the function F defined by $F(x) = x/(1+x)$, and we calculate

$$
F'(x) = \frac{1}{1+x} - \frac{x}{(1+x)^2} \quad \text{and} \quad F''(x) = \frac{2x}{(1+x)^3} - \frac{2}{(1+x)^2}.
$$

Using Itô's formula, we calculate

$$
\begin{aligned}
d\pi_t &= dF(L_t) \\
&= F'(L_t)\, dL_t + \tfrac{1}{2} F''(L_t)\, (dL_t)^2 \\
&= \tfrac{1}{2} (\vartheta L_t)^2\, F''(L_t)\, dt + \vartheta L_t F'(L_t)\, dW_t \\
&= -\vartheta^2 \left(\frac{L_t}{1+L_t}\right)^3 \frac{1}{L_t}\, dt + \vartheta \frac{L_t}{1+L_t}\left(1 - \frac{L_t}{1+L_t}\right) dW_t.
\end{aligned}
$$

Combining this calculation with the definition

$$
\pi_t = \frac{L_t}{1+L_t} \quad \Leftrightarrow \quad \frac{1}{L_t} = \frac{1-\pi_t}{\pi_t},
$$

we can see that the process (π_t) satisfies the stochastic differential equation

$$
d\pi_t = -\vartheta^2 \pi_t^2 \,(1-\pi_t)\, dt + \vartheta \pi_t \,(1-\pi_t)\, dW_t.
$$

Solutions 7

1. The assumption that (X_t) is an $(\mathcal{F}_t)$-local martingale implies that there exists a sequence (τ_n) of $(\mathcal{F}_t)$-stopping times such that

$$\lim_{n\to\infty} \tau_n = \infty, \quad \mathbb{P}\text{-a.s.}, \tag{109}$$

and

$$\text{the process } (X_t^{\tau_n}) \text{ defined by } X_t^{\tau_n} = X_{t\wedge\tau_n} \text{ is an } (\mathcal{F}_t)\text{-martingale.} \tag{110}$$

Now, we consider any times $s \le t$, and we note that (109) implies that

$$\lim_{n\to\infty} X_t^{\tau_n} \equiv \lim_{n\to\infty} X_{t\wedge\tau_n} = X_t \quad \text{and} \quad \lim_{n\to\infty} X_s^{\tau_n} \equiv \lim_{n\to\infty} X_{s\wedge\tau_n} = X_s.$$

Also, we observe that (110) implies that

$$\mathbb{E}\left[X_t^{\tau_n} \mid \mathcal{F}_s\right] = X_s^{\tau_n} \quad \text{for all } n \ge 1.$$

In view of these observations and Fatou's lemma, we can see that

$$\mathbb{E}\left[X_t \mid \mathcal{F}_s\right] = \mathbb{E}\left[\lim_{n\to\infty} X_t^{\tau_n} \mid \mathcal{F}_s\right] \le \lim_{n\to\infty} \mathbb{E}\left[X_t^{\tau_n} \mid \mathcal{F}_s\right] = \lim_{n\to\infty} X_s^{\tau_n} = X_s,$$

and the supermartingale inequality follows. (Note that it is for the application of Fatou's lemma that we have assumed that (X_t) is positive.)

2. First, we note that (1) in the exercises and the inequalities $0 \le \mathbf{1}_A \le 1$ imply that

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

so (2) in the exercises defines a function $\mathbb{Q} : \mathcal{F} \to [0,1]$. In particular,

$$\mathbb{Q}(\emptyset) = \mathbb{E}^{\mathbb{P}}\left[Y\mathbf{1}_\emptyset\right] = \mathbb{E}^{\mathbb{P}}\left[\mathbf{1}_\emptyset\right] = \mathbb{P}(\emptyset) = 0$$

and

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

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

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

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

3. First, we assume that Z is a *simple random variable*, so that there exist distinct constants $z_1, \ldots, z_n$ such that

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

In this case, we can use the linearity of the expectation operator and the definition of

$\mathbb{Q}$ to calculate

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

It follows that the result is true whenever Z is a simple random variable.

Next, we suppose that Z is a positive random variable, and consider any *increasing* sequence of positive simple random variables (Z_n) such that $\lim_{n\to\infty} Z_n = Z$, $\mathbb{P}$-a.s.. 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

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

which establishes the required result whenever Z is a positive random variable.

Finally, suppose that Z is a random variable such that $\mathbb{E}^{\mathbb{Q}}\big[|Z|\big] < \infty$, and let Z^+, Z^- be the positive random variables 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 proof is complete.

4. This result follows trivially from Exercise 3 once we observe that $Y\mathbf{1}_{\{Y=0\}} = 0$.

5. The assumption $\mathbb{Q} \sim \mathbb{P}$ implies that, given any $A \in \mathcal{F}$, $\mathbb{Q}(A) = 0 \Leftrightarrow \mathbb{P}(A) = 0$. Combining this observation with the assumption that $Y \geq 0$, $\mathbb{P}$-a.s. (see (1) in the exercises) and fact that $\mathbb{Q}\big(\{Y=0\}\big) = 0$ (see the previous exercise), we can see that $\mathbb{P}\big(\{Y=0\}\big) = 0$, i.e., $Y > 0$, $\mathbb{P}$-a.s..

To prove that $d\mathbb{P} = Y^{-1}\, d\mathbb{Q}$, we have to show that $\mathbb{P}(A) = \mathbb{E}^{\mathbb{Q}}\left[Y^{-1}\mathbf{1}_A\right]$ for all $A \in \mathcal{F}$. In view of Exercise 3, this claim follows immediately from the calculations

$$\mathbb{P}(A) = \mathbb{E}^{\mathbb{P}}\left[\mathbf{1}_A\right] = \mathbb{E}^{\mathbb{P}}\left[YY^{-1}\mathbf{1}_A\right] = \mathbb{E}^{\mathbb{Q}}\left[Y^{-1}\mathbf{1}_A\right].$$

6. We know that

$$\mathbb{E}^{\mathbb{Q}_T}[Z \mid \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s]}{L_s}. \tag{111}$$

Using the tower property of conditional expectation, the assumption that Z is $\mathcal{F}_t$-measurable, and the martingale property of (L_t), we can see that

$$\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}\big[\mathbb{E}^{\mathbb{P}}[L_T Z \mid \mathcal{F}_t] \mid \mathcal{F}_s\big] = \mathbb{E}^{\mathbb{P}}\big[\mathbb{E}^{\mathbb{P}}[L_T \mid \mathcal{F}_t] Z \mid \mathcal{F}_s\big] = \mathbb{E}^{\mathbb{P}}[L_t Z \mid \mathcal{F}_s].$$

Combining this observation with (111), we obtain the required result.

7. The assumption that (M_t) is an $(\mathcal{F}_t)$-martingale with respect to the probability measure $\mathbb{P}$ implies that (M_t) is $(\mathcal{F}_t)$-adapted,

$$\mathbb{E}^{\mathbb{P}}[|M_t|] < \infty \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}[M_t \mid \mathcal{F}_s] = M_s \quad \text{for all } 0 \le s \le t \le T.$$

In view of these observations, the definition of $\mathbb{Q}_T$ and Exercise 3, we can see that

$$\mathbb{E}^{\mathbb{Q}_T}\left[|L_t^{-1} M_t|\right] = \mathbb{E}^{\mathbb{P}}\left[L_t |L_t^{-1} M_t|\right] = \mathbb{E}^{\mathbb{P}}[|M_t|] < \infty \quad \text{for all } t \in [0, T].$$

Furthermore, we can use (4) from Exercise 6 and the martingale property of (M_t) to calculate

$$\mathbb{E}^{\mathbb{Q}_T}\left[L_t^{-1} M_t \mid \mathcal{F}_t\right] = \frac{\mathbb{E}^{\mathbb{P}}\left[L_t L_t^{-1} M_t \mid \mathcal{F}_s\right]}{L_s} = L_s^{-1}\mathbb{E}^{\mathbb{P}}[M_t \mid \mathcal{F}_s] = L_s^{-1} M_s$$

for all $0 \le s \le t \le T$, and the result follows.

8. Consider the Itô process (U_t) defined by

$$U_t = -\frac{1}{2}\int_0^t |X_s|^2\, ds + \int_0^t X_s \cdot dW_s,$$

so that

$$dU_t = -\frac{1}{2}|X_t|^2\, dt + X_t \cdot dW_t,$$

and the function F given by $F(u) = e^u$. Noting that

$$F''(u) = F'(u) = F(u),$$

and that the process (L_t) defined by (5) in the exercises admits the expression $L_t = F(U_t)$, we can use Itô's formula to calculate

$$\begin{aligned} dL_t &= dF(U_t) \\ &= \left[\frac{1}{2}|X_t|^2 F''(U_t) - \frac{1}{2}|X_t|^2 F'(U_t)\right] dt + F'(U_t) X_t \cdot dW_t \\ &= F(U_t) X_t \cdot dW_t \\ &= L_t X_t \cdot dW_t. \end{aligned}$$

These calculation and the initial condition $L_0 = 1$ imply that (L_t) satisfies (6) in the exercises, as required.

9. The required result follows immediately from Novikov's condition and the observation that

$$\begin{aligned} \mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t |X_s|^2\, ds\right)\right] &\le \mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t nK^2\, ds\right)\right] \\ &= \exp\left(\frac{1}{2} nK^2 t\right) \\ &< \infty \qquad \text{for all } t \ge 0. \end{aligned}$$