Probability for Finance

Ekkehard Kopp, Jan Malczak and Tomasz Zastawniak

Solutions to Exercises

Chapter 1

1.1. First, take an infinite sequence $A_1, A_2, \ldots \in \mathcal{F}$. By de Morgan's law,

$$\bigcap_{i=1}^{\infty} A_i = \Omega \setminus \bigcup_{i=1}^{\infty} (\Omega \setminus A_i).$$

By Definition 1.10, Ω is in $\mathcal{F}$, all the $\Omega \setminus A_i$ are in $\mathcal{F}$, their union is in $\mathcal{F}$, and finally, its complement is in $\mathcal{F}$.

Now for a finite sequence $A_1, \ldots, A_n \in \mathcal{F}$, put $A_{n+i} = \Omega$ for $i = 1, 2, \ldots$. Since $\Omega \in \mathcal{F}$, we have

$$\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

1.2. Fix $0 \leq a < b \leq 1$. We show that the open interval (a, b) is in $\mathcal{F}$. For each $n = 1, 2, \ldots$ large enough so that $\frac{1}{2n} < b - a$ we can find rationals $r_n < s_n$ with $a < r_n < a + \frac{1}{n}$ and $b - \frac{1}{n} < s_n < b$. Now $(a, b) = \bigcup_{n=1}^{\infty} (r_n, s_n)$, and as $\mathcal{F}$ is a σ-field containing all (r_n, s_n), we have $(a, b) \in \mathcal{F}$.

1.3. (1) We first show that if $\{\mathcal{F}_j\}_{j \in J}$ is any collection of σ-fields defined on the same set Ω, then their intersection $\bigcap_{j \in J} \mathcal{F}_j$ is also a σ-field. We have $\Omega \in \mathcal{F}_j$ for all j, so $\Omega \in \bigcap_{j \in J} \mathcal{F}_j$. Next, take $A \in \bigcap_{j \in J} \mathcal{F}_j$. Then $A \in \mathcal{F}_j$, so $\Omega \setminus A \in \mathcal{F}_j$ for all $j \in J$; hence, $\Omega \setminus A \in \bigcap_{j \in J} \mathcal{F}_j$.

Finally, suppose that $A_1, A_2, \ldots \in \bigcap_{j \in J} \mathcal{F}_j$. Then $A_i \in \mathcal{F}_j$ for all i and j, so that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_j$ for all j. Thus, $\bigcup_{i=1}^{\infty} A_i \in \bigcap_{j \in J} \mathcal{F}_j$.

Apply this with $\Omega = \mathbb{R}$ and $\{\mathcal{F}_j\}_{j \in J} = \{\mathcal{F} : \mathcal{F}$ is a σ-field on $\mathbb{R}$ and $\mathcal{I} \subset \mathcal{F}\}$ to see that $\mathcal{B}(\mathbb{R})$ is a σ-field containing $\mathcal{I}$.

(2) Since $\mathcal{F}$ is a σ-field on $\mathbb{R}$ containing $\mathcal{I}$, it appears in the intersection defining $\mathcal{B}(\mathbb{R})$, which means that $\mathcal{B}(\mathbb{R}) \subset \mathcal{F}$.

1.4. For any singleton $\{a\}$, where $a \in \mathbb{R}$, we have $m(\{a\}) = 0$. Indeed, by definition, $m(\{a\})$ is the infimum of the sum of the lengths of any countable covering. Given any $\varepsilon > 0$ we can cover $\{a\}$ by the open intervals $I_n = \left(a - \frac{\varepsilon}{2^{n+2}}, a + \frac{\varepsilon}{2^{n+2}}\right)$ for $n = 1, 2, \ldots$. Then $\sum_{n=1}^{\infty} l(I_n) = \frac{\varepsilon}{2} < \varepsilon$. Hence $m(\{a\}) = 0$.

It follows that $m([\frac{1}{2}, 2)) = m(\{\frac{1}{2}\}) + m((\frac{1}{2}, 2)) = 0 + (2 - \frac{1}{2}) = \frac{3}{2}$ since the union is disjoint and $\{\frac{1}{2}\}$ is a singleton.

Write $[-2, 3] \cup [3, 8] = \{-2\} \cup (-2, 3) \cup \{3\} \cup (3, 8) \cup \{8\}$ to obtain a disjoint union, and use the additivity of m to obtain, similarly, that $m([-2, 3] \cup [3, 8]) = 10$.

1.5. By countable additivity, and since $m\left(\left\{\frac{1}{n}\right\}\right) = 0$ for each $n = 1, 2, \ldots$, we have

$$m\left(\bigcup_{n=2}^{\infty}\left(\frac{1}{n+1}, \frac{1}{n}\right]\right) = \sum_{n=2}^{\infty} m\left(\left(\frac{1}{n+1}, \frac{1}{n}\right]\right) = \sum_{n=2}^{\infty}\left(\frac{1}{n} - \frac{1}{n+1}\right) = \frac{1}{2}.$$

1.6. Any countable subset of $\mathbb{R}$ has Lebesgue measure 0. To see this, write $A = \{a_n : n \geq 1\}$ as a disjoint union of singletons and use countable additivity:

$$m(A) = m\left(\bigcup_{n=1}^{\infty}\{a_n\}\right) = \sum_{n=1}^{\infty} m(\{a_n\}) = 0.$$

Therefore the countable sets $\mathbb{N}, \mathbb{Q}$ and $\{x \in \mathbb{R} : \sin x = \cos x\}$ have Lebesgue measure 0. Since $m(\mathbb{R}) = \infty$, we have $m(\mathbb{R}\backslash\mathbb{Q}) = m(\mathbb{R}) - m(\mathbb{Q}) = \infty$.

1.7. First, we show that the Cantor set C is uncountable. We adapt the proof of the uncountability of $\mathbb{R}$. Each $x \in [0, 1]$ can expressed in ternary form as

$$x = \sum_{k=1}^{\infty} \frac{a_k}{3^k} = 0.a_1 a_2 \ldots$$

with coefficients $a_k = 0, 1$ or 2. We have a one-to-one correspondence between C and ternary expansions of the form $0.a_1 a_2 \ldots$ with each a_k equal 0 or 2.

Suppose that C is countable, so that it can be arranged into a sequence: $C = \{x_1, x_2, x_3, \ldots\}$. By the diagonal procedure, we define a number x in the following way: if the nth digit in the ternary expansion of x_n is equal to 0 (or 2), then we take the nth digit in the ternary expansion of x to equal to 2 (or 0, respectively). That is, we simply interchange 0 and 2 at the nth position. Hence the ternary expansion of x contains only 0 or 2 in any position, but differs from the ternary expansion of each x_n in at least one position, so $x \notin C$. The contradiction shows that C is uncountable.

All that remains is to check that $m(C) = 0$. By definition, $C = \bigcap_{n=1}^{\infty} C_n$ and C_n consists of 2^n disjoint closed intervals, each of length $\left(\frac{1}{3}\right)^n$. The total length of this sequence of intervals equals $\left(\frac{2}{3}\right)^n$. Since $C_n \supset C_{n+1}$ for each n, we have

$$m(C) = \lim_{n\to\infty} m(C_n) = \lim_{n\to\infty} \left(\frac{2}{3}\right)^n = 0.$$

1.8. First note that $l((a, b)) = l((a + x, b + x)) = b - a$. If a sequence of open intervals $(I_k)_{k=1}^{\infty}$ covers $A \in \mathcal{B}(\mathbb{R})$, that is, if

$$A \subset \bigcup_{k=1}^{\infty} I_k,$$

then the sequence of intervals $(I_k + x)_{k=1}^{\infty}$ covers $A + x$,

$$A + x \subset \bigcup_{k=1}^{\infty} (I_k + x).$$

So we have a one-one correspondence between the interval coverings of A and $A+x$. Moreover, the total length of a family of intervals does not change when we shift each by x,

$$\sum_{k=1}^{\infty} l(I_k) = \sum_{k=1}^{\infty} l(I_k + x).$$

This implies that the collections of total lengths of such coverings satisfy

$$\left\{\sum_{k=1}^{\infty} l(I_k) : A \subset \sum_{k=1}^{\infty} I_k\right\} = \left\{\sum_{k=1}^{\infty} l(\tilde{I}_k) : A + x \subset \bigcup_{k+1}^{\infty} \tilde{I}_k\right\}$$

were I_k and $\tilde{I}_k$ are open intervals. So their infima are equal, which proves that $m(A) = m(A + x)$.

1.9. We use the Riemann integral and standard calculus techniques in solving this exercise, as the integrands are continuous. Clearly, $f(x) \geq 0$ for all $x \in \mathbb{R}$. We need to verify that the integral from $-\infty$ to $+\infty$

of $f(x)$ is 1. Let us begin with $\mu = 0$ and $\sigma = 1$. Then

$$\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx\right)^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2+y^2}{2}\right) dxdy$$
$$= \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} r \exp\left(-\frac{r^2}{2}\right) d\varphi dr$$
$$= \int_0^{\infty} r \exp\left(-\frac{r^2}{2}\right) dr = -\exp\left(-\frac{r^2}{2}\right)\Big|_0^{\infty} = 1$$

To compute the above integral, we substituted the polar coordinates: $x = r\cos\varphi$, $y = r\sin\varphi$.

Now, for any μ and σ,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{z^2}{2}\right) dz = 1,$$

where the substitution $\mu + \sigma z = x$ is used. If follows that $\int_{-\infty}^{\infty} f(x)dx = 1$, as required.

1.10 As a simple consequence of Exercise 1.9 we have

$$\int_0^{\infty} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) dz = 1,$$

where the substitution $z = \ln x$ is used.

1.11. Let

$$r = \sum_{i=1}^{m} r_i \mathbf{1}_{A_i}, \quad s = \sum_{j=1}^{n} s_j \mathbf{1}_{B_j},$$

where $A_1, \ldots, A_m \in \mathcal{F}$, $A_i \cap A_j = \varnothing$ for $i \neq j$, $\bigcup_{i=1}^m A_i = \Omega$ and $B_1, \ldots, B_n \in \mathcal{F}$, $B_i \cap B_j = \varnothing$ for $i \neq j$, $\bigcup_{j=1}^n B_j = \Omega$.

For any $a, b \geq 0$ we have

$$ar + bs = \sum_{i=1}^{m} \sum_{j=1}^{n} (ar_i + bs_j) \mathbf{1}_{E_{ij}},$$

where $E_{ij} = A_i \cap B_j$, for $i = 1, \ldots, m$, $j = 1, \ldots, n$. Thus $ar + bs$ is a simple function. By Definition 1.25

$$
\begin{aligned}
\int_\Omega (ar+bs)d\mu &= \sum_{i=1}^m \sum_{j=1}^n (ar_i + bs_j)\mu(E_{ij}) \\
&= \sum_{i=1}^m \sum_{j=1}^n ar_i\mu(E_{ij}) + \sum_{i=1}^m \sum_{j=1}^n bs_j\mu(E_{ij}) \\
&= \sum_{i=1}^m \left(ar_i \sum_{j=1}^n \mu(E_{ij}) \right) + \sum_{j=1}^n \left(bs_j \sum_{i=1}^m \mu(E_{ij}) \right) = I_1 + I_2.
\end{aligned}
$$

We apply additivity of the measure μ to find

$$
\sum_{j=1}^n \mu(E_{ij}) = \mu\left(\bigcup_{j=1}^n (A_i \cap B_j) \right) = \mu\left(A_i \cup \left(\bigcup_{j=1}^n B_j \right) \right) = \mu(A_i).
$$

Similarly,

$$
\sum_{i=1}^m \mu(E_{ij}) = \mu\left(\bigcup_{i=1}^m (A_i \cap B_j) \right) = \mu\left(\left(\bigcup_{i=1}^m A_i \right) \cap B_j \right) = \mu(B_j).
$$

As a consequence,

$$
I_1 + I_2 = a\sum_{i=1}^m r_i\mu(A_i) + b\sum_{j=1}^n s_j\mu(B_j) = a\int_\Omega r d\mu + b\int_\Omega s d\mu.
$$

1.12. Let

$$
r = \sum_{i=1}^m r_i \mathbf{1}_{A_i}, \quad s = \sum_{j=1}^n s_j \mathbf{1}_{B_j},
$$

where $A_1, \ldots, A_m \in \mathcal{F}$, $A_i \cap A_j = \varnothing$ for $i \neq j$, $\bigcup_{i=1}^m A_i = \Omega$ and $B_1, \ldots, B_n \in \mathcal{F}$, $B_i \cap B_j = \varnothing$ for $i \neq j$, $\bigcup_{j=1}^n B_j = \Omega$. By the additivity of the measure μ we get

$$
\int_\Omega r d\mu = \sum_{i=1}^m r_i\mu(A_i) = \sum_{i=1}^m \sum_{j=1}^n r_i\mu(A_i \cap B_j).
$$

Similarly,

$$
\int_\Omega s d\mu = \sum_{j=1}^n s_j\mu(B_j) = \sum_{i=1}^m \sum_{j=1}^n s_j\mu(A_i \cap B_j).
$$

Since for any $\omega \in \Omega$, $r(\omega) \leq s(\omega)$, it follows that $r_i \leq s_j$ on $A_i \cap B_j$.

This implies that $r_i\mu(A_i \cap B_j) \leq s_j\mu(A_i \cap B_j)$, which yields

$$\int_\Omega r d\mu \leq \int_\Omega s d\mu.$$

1.13. The 'only if' part of the statement is obvious: $(a, \infty) \in \mathcal{B}(\mathbb{R})$, so $\{f > a\} = \{f \in (a, \infty)\} \in \mathcal{F}$ when f is measurable.

The 'if' part is based on the fact the σ-field of Borel sets is generated by intervals.

First, the family C of all sets $A \subset \mathbb{R}$ such that $\{f \in A\} \in \mathcal{F}$ is a σ-field. Let us verify the conditions of Definition 1.10. Since $\{f \in \mathbb{R}\} = \Omega \in \mathcal{F}$, it follows that $\mathbb{R} \in C$. Suppose $A \in C$. Then $\{f \in \mathbb{R} \setminus A\} = \Omega \setminus \{f \in A\}$ also lies in C because $\mathcal{F}$ is a σ-field. Now, let $A_1, A_2, \ldots \in C$, then $\{f \in \bigcup_{i=1}^\infty A_i\} = \bigcup_{i=1}^\infty \{f \in A_i\} \in \mathcal{F}$. It follows that $\bigcup_{i=1}^\infty A_i \in C$.

Second, C contains the family $\mathcal{I}$ of all open intervals: since $(a, \infty) \in C$, it follows that $(-\infty, a] = \mathbb{R} \setminus (a, \infty) \in C$, $(a, b] = (a, \infty) \cap (-\infty, b] \in C$, $[a, b] = \bigcap_{n=1}^\infty (a - \frac{1}{n}, b] \in C$ and finally $(a, b) = \bigcup_{n=1}^\infty [a + \frac{1}{n}, b - \frac{1}{n}] \in C$. By part (2) of Exercise 1.3, we obtain $\mathcal{B}(\mathbb{R}) \subset C$.

1.14. Let $s = \sum_{i=1}^n s_i \mathbf{1}_{A_i}$, where $A_1, \ldots, A_n \in \mathcal{F}$ are pairwise disjoint with $\bigcup_{i=1}^n A_i = \Omega$. For any $a \in \mathbb{R}$, by Exercise 1.13, it suffices to show that $\{s > a\} \in \mathcal{F}$. But $\{s > a\} = \bigcup A_i$, where union extends over all i such that $s_i > a$. This yields $\{s > a\} \in \mathcal{F}$.

1.15. By Exercise 1.13, it is enough to show that $\{g \circ f > a\} \in \mathcal{F}$ for any $a \in \mathbb{R}$. Let $A = (a, \infty)$. Then by the continuity of g, the inverse image $g^{-1}(A) = \{g \in A\}$ is an open subset of $\mathbb{R}$. We claim that any open set U in $\mathbb{R}$ is a union of countably many open intervals. To see this, consider all open intervals contained in U with rational endpoints. Their union is clearly contained in U. In fact, it is equal to U since for any $x \in U$, we have $(x - \varepsilon, x + \varepsilon) \subset U$ for some ε, and we can find rational numbers a and b such that $x \in (a, b) \subset (x - \varepsilon, x + \varepsilon)$. This proves that $U \in \mathcal{B}(\mathbb{R})$. As a consequence, we have $g^{-1}((a, \infty)) = \{g > a\} \in \mathcal{B}(\mathbb{R})$. By Exercise 1.13, it follows that

$$\{g \circ f > a\} = f^{-1}(g^{-1}(a, \infty)) \in \mathcal{F}.$$

This proves that $g \circ f$ is a measurable function.

1.16. We observe that the following are measurable sets for any $a \in \mathbb{R}$:

$$\{\max\{f_1, \ldots, f_n\} > a\} = \bigcup_{k=1}^{n} \{f_k > a\},$$
$$\{\min\{f_1, \ldots, f_n\} > a\} = \bigcap_{k=1}^{n} \{f_k > a\}.$$

By Exercise 1.13 we obtain the conclusion.

1.17. Analogously to Exercise 1.16, the following are measurable sets for any $a \in \mathbb{R}$:

$$\{\sup_{n \geq 1} f_n > a\} = \bigcup_{n=1}^{\infty} \{f_n > a\},$$
$$\{\inf_{n \geq 1} f_n > a\} = \bigcap_{n=1}^{\infty} \{f_n > a\}.$$

1.18. By Exercise 1.16, $\sup_{n \leq k} f_n$ and $\inf_{n \leq k} f_n$ are measurable functions. It is now immediate from Exercise 1.17 that

$$\lim_{n\to\infty} \sup f_n = \inf_{k \geq 1} (\sup_{n \leq k} f_n),$$
$$\lim_{n\to\infty} \inf f_n = \sup_{k \geq 1} (\inf_{n \leq k} f_n)$$

are measurable functions.

1.19. This is immediate since

$$\lim_{n\to\infty} f_n = \limsup_{n\to\infty} f_n = \liminf_{n\to\infty} f_n,$$

so that we can apply Exercise 1.18.

1.20. In order to prove $s_n \leq s_{n+1}$ for $n = 1, 2, \ldots$ it is sufficient to observe that

$$\begin{aligned} A_{i,n} &= \left\{ \frac{i}{2^n} \leq f < \frac{i+1}{2^n} \right\} \\ &= \left\{ \frac{2i}{2^{n+1}} \leq f < \frac{2i+1}{2^{n+1}} \right\} \cup \left\{ \frac{2i+1}{2^{n+1}} \leq f < \frac{2i+2}{2^{n+1}} \right\} \\ &= A_{2i,n+1} \cup A_{2i+1,n+1}. \end{aligned}$$

By definition, s_n is constant on $A_{i,n}$ and its value is $\frac{i}{2^n}$. Similarly

$$s_{n+1}(\omega) = \begin{cases} \frac{i}{2^n} & \text{for } \omega \in A_{2i,n+1}, \\ \frac{2i+1}{2^{n+1}} & \text{for } \omega \in A_{2i+1,n+1}. \end{cases}$$

This shows that $s_n(\omega) \leq s_{n+1}(\omega)$ and $s_n(\omega) \leq f(\omega)$ for all $\omega \in \Omega$, $n = 1, 2, \ldots$. Now we show that $f = \lim_{n\to\infty} s_n$ pointwise, which means that $f(\omega) = \lim_{n\to\infty} s_n(\omega)$ for any $\omega \in \Omega$.

Let $\omega \in \Omega$ and $\varepsilon > 0$. There exists an integer N such that $f(\omega) < N$ and $\frac{1}{2^N} < \varepsilon$. Choose $i \in \{1, 2, \ldots, N2^N\}$ such that

$$\frac{i}{2^N} \leq f(\omega) < \frac{i+1}{2^N}.$$

This implies that $f(\omega) - s_N(\omega) < \frac{1}{2^N} < \varepsilon$. Since s_n is a non-decreasing sequence, we have $f(\omega) - s_n(\omega) < \varepsilon$ for all $n \geq N$.

1.21. By Proposition 1.28, there are non-decreasing sequences $\{s_n\}$, $\{t_n\}$ of simple functions such that $f = \lim_{n\to\infty} s_n$ and $g = \lim_{n\to\infty} t_n$. This immediately gives $af + bg = \lim_{n\to\infty}(as_n + bt_n)$. Since $as_n + bt_n$ is a simple function for $n = 1, 2, \ldots$, by Exercise 1.14 it is also measurable. It follows by Exercise 1.19 that $\lim_{n\to\infty}(as_n + bt_n) = a\lim_{n\to\infty} s_n + b\lim_{n\to\infty} t_n = af + bg$ is measurable.

1.22. We can repeat the argument of Exercise 1.21. Namely, by Proposition 1.28 there are two non-decreasing sequences $\{s_n\}$, $\{t_n\}$ of simple functions such that $f = \lim_{n\to\infty} s_n$ and $g = \lim_{n\to\infty} t_n$. Obviously, $\lim_{n\to\infty} s_n t_n = \lim_{n\to\infty} s_n \lim_{n\to\infty} t_n = fg$. Since $s_n t_n$ is a simple function, by Exercise 1.14 it is measurable, and by Exercise 1.19, $\lim_{n\to\infty} s_n t_n = fg$ is measurable.

1.23. For any simple function r, Exercise 1.12 implies that Definitions 1.25 and 1.29 give the same result since

$$\sup\left\{\int_\Omega s d\mu : s \text{ is a simple function such that } s \leq r\right\} = \int_\Omega r d\mu.$$

For any measurable functions $f, g : \Omega \to [0, \infty)$ such that $f \leq g$ we have the inclusion

$$\left\{\int_\Omega s d\mu : s \text{ is a simple function such that } s \leq f\right\}$$
$$\subset \left\{\int_\Omega s d\mu : s \text{ is a simple function such that } s \leq g\right\},$$

and the supremum of the bigger set is larger. It follows that

$$\int_\Omega f d\mu \leq \int_\Omega g d\mu.$$

1.24. By Proposition 1.28 (see also Exercise 1.20), there are non-decreasing sequences $\{s_n\}$ and $\{t_n\}$ of simple measurable functions such that

$\lim_{n\to\infty} s_n = f_1$ and $\lim_{n\to\infty} t_n = f_2$. Then $\lim_{n\to\infty}(s_n + t_n) = f_1 + f_2$, and this implies that $f_1 + f_2$ is a measurable function. The monotone convergence theorem (Theorem 1.31) combined with Exercise 1.11 shows that

$$\int_\Omega (f_1 + f_2) d\mu = \int_\Omega f_1 d\mu + \int_\Omega f_2 d\mu.$$

Next, define $g_n = f_1 + \cdots + f_n$. By Exercise 1.21, g_n is a measurable function, and $\lim_{n\to\infty} g_n = \sum_{n=1}^\infty f_n$. By Exercise 1.19, $\sum_{n=1}^\infty f_n$ is a non-negative measurable function. Applying induction, we have

$$\int_\Omega g_n d\mu = \sum_{i=1}^n \int_\Omega f_i d\mu.$$

Using the monotone convergence theorem once again, we have

$$\int_\Omega \left(\sum_{n=1}^\infty f_n \right) d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.$$

1.25. We have $|f| = f^+ + f^-$. For an integrable f the functions f^+, f^- are measurable by Exercise 1.16, and both $\int_\Omega f^+ d\mu$ and $\int_\Omega f^- d\mu$ are finite. By Proposition 1.32,

$$\int_\Omega |f| d\mu = \int_\Omega f^+ d\mu + \int_\Omega f^- d\mu$$

and the integral on the left is finite, hence so is the right-hand side. If $|f|$ is integrable, then by Exercise 1.23 both $\int_\Omega f^+ d\mu$ and $\int_\Omega f^- d\mu$ are integrable since $f^+ \le |f|$ and $f^- \le |f|$. Therefore f is integrable.

1.26. Using $f = f^+ - f^-$ and $|f| = f^+ + f^-$, we have

$$\begin{aligned} \left| \int_\Omega f d\mu \right| &= \left| \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu \right| \\ &\le \int_\Omega f^+ d\mu + \int_\Omega f^- d\mu = \int_\Omega |f| d\mu. \end{aligned}$$

1.27. Let f, g be arbitrary integrable functions. Note that by Exercise 1.23 and Proposition 1.32

$$\int_\Omega |f + g| d\mu \le \int_\Omega |f| + |g| d\mu = \int_\Omega |f| d\mu + \int_\Omega |g| d\mu,$$

so the left-hand side is finite. By Exercise 1.25, $f + g$ is integrable. We can write

$$(f + g)^+ - (f + g)^- = f + g = (f^+ - f^-) + (g^+ - g^-).$$

Rearranging to have only non-negative functions on either side, we get

$$(f+g)^+ + f^- + g^- = f^+ + g^+ + (f+g)^-.$$

By Proposition 1.32,

$$\int_\Omega (f+g)^+ d\mu + \int_\Omega f^- d\mu + \int_\Omega g^- d\mu = \int_\Omega f^+ d\mu + \int_\Omega g^+ d\mu + \int_\Omega (f+g)^- d\mu,$$

hence

$$\int_\Omega (f+g)^+ d\mu - \int_\Omega (f+g)^- d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu + \int_\Omega g^+ d\mu - \int_\Omega g^- d\mu.$$

By Definition 1.33, this shows that

$$\int_\Omega (f+g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu.$$

Next, for any integrable function f and any $c \in \mathbb{R}$ we have $(cf)^+ = c^+ f^+ + c^- f^-$ and $(cf)^- = c^- f^+ + c^+ f^-$. By Definition 1.33 and Proposition 1.32, cf is integrable and

$$\begin{aligned}\int_\Omega cf d\mu &= \int_\Omega (cf)^+ \, d\mu - \int_\Omega (cf)^- \, d\mu \\ &= \int_\Omega (c^+ f^+ + c^- f^-) \, d\mu - \int_\Omega (c^- f^+ + c^+ f^-) \, d\mu \\ &= c^+ \int_\Omega f^+ d\mu + c^- \int_\Omega f^- d\mu - c^- \int_\Omega f^+ d\mu - c^+ \int_\Omega f^- d\mu \\ &= c \int_\Omega f d\mu.\end{aligned}$$

Finally, for any integrable functions f, g and for any $a, b \in \mathbb{R}$ the above results show that af, bg are integrable and therefore $af + bg$ is also integrable, and

$$\int_\Omega (af + bg) \, d\mu = \int_\Omega af d\mu + \int_\Omega bg d\mu = a \int_\Omega f d\mu + b \int_\Omega g d\mu.$$

1.28. If $f \le g$, then $f^+ \le g^+$ and $f^- \ge g^-$. These inequalities imply

$$\int_\Omega f^+ d\mu \le \int_\Omega g^+ d\mu$$

and

$$\int_\Omega g^- d\mu \le \int_\Omega f^- d\mu$$

by Exercise 1.23. Adding and rearranging gives the result.

1.29. Let $f = \lim_{n\to\infty} f_n$. We show that

$$\lim_{n\to\infty}\int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

The limit exists and satisfies

$$\lim_{n\to\infty}\int_\Omega f_n d\mu \le \int_\Omega f d\mu$$

because $\int_\Omega f_n d\mu$ is a non-decreasing sequence bounded above by $\int_\Omega f d\mu$ by Exercise 1.28. Consider $g_n = f_n - f_1 \ge 0$. It is a non-decreasing sequence of non-negative integrable functions such that $\lim_{n\to\infty} g_n = f - f_1$. Applying the monotone convergence Theorem 1.31 and Exercise 1.27, we get

$$\lim_{n\to\infty}\int_\Omega g_n d\mu = \lim_{n\to\infty}\left(\int_\Omega f_n d\mu - \int_\Omega f_1 d\mu\right) = \int_\Omega (f - f_1) d\mu.$$

This implies that

$$\lim_{n\to\infty}\int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

1.30. The integral is defined as long as at least one of $\int_\Omega f^+ d\mu$ and $\int_\Omega f^- d\mu$ is finite. So if $\int_\Omega f d\mu = 0$ we know that both are finite, hence f is integrable (else its integral is $\pm\infty$). Hence $\mu(\{f = +\infty\}) = \mu(\{f = -\infty\}) = 0$, so f is μ-a.e. finite. Now suppose that $\int_B f d\mu = 0$ for all $B \in \mathcal{F}$. As f is measurable, $\{f > 0\}, \{f < 0\} \in \mathcal{F}$, and we have

$$0 = \int_{\{f>0\}} f d\mu = \int_{\{f>0\}} f^+ d\mu - \int_{\{f>0\}} 0 d\mu,$$
$$0 = \int_{\{f<0\}} f d\mu = \int_{\{f<0\}} 0 d\mu - \int_{\{f<0\}} f^- d\mu.$$

Proposition 1.36 now shows that the non-negative measurable functions f^+, f^- are both 0 μ-a.e. Since $\{f \neq 0\} = \{f^+ > 0\} \cup (f^- > 0\}$, we have $\mu(\{f \neq 0\}) = 0$. Conversely, if $f = 0$ μ-a.e., then $0 = \mu(\{f \neq 0\}) = \mu(\{f > 0\}) + \mu(\{f < 0\})$, and μ takes only non-negative values, so $\mu(\{f^+ > 0\}) = \mu(\{f^- > 0\}) = 0$. Proposition 1.36 thus applies to both functions, hence $\int_\Omega f^+ d\mu = \int_\Omega f^- d\mu = 0$. This means that for any $B \in \mathcal{F}$ we have $\int_B f^+ d\mu = \int_B f^- d\mu = 0$, hence $\int_B f d\mu = 0$.

1.31. We have

$$f(x) = \sum_{k=1}^{\infty} k \mathbf{1}_{A_k}(x),$$

where A_k is the union of the 2^{k-1} intervals of length 3^{-k} each that are removed from $[0,1]$ at the kth stage of constructing the Cantor set. Let us define the nth partial sum of the series

$$r_n(x) = \sum_{k=1}^{n} k\mathbf{1}_{A_k}(x).$$

Since $r_n(x)$ is a non-decreasing sequence converging to $f(x)$ for each $x \in [0,1]$, by Theorem 1.31

$$\int_{[0,1]} f dm = \lim_{n\to\infty} \int_{[0,1]} r_n dm = \lim_{n\to\infty} \sum_{k=1}^{n} k\frac{2^{k-1}}{3^k} = \frac{1}{3}\sum_{k=1}^{\infty} k\left(\frac{2}{3}\right)^{k-1}.$$

Since $\sum_{k=1}^{\infty} \alpha^k = \frac{1}{1-\alpha}$ if $|\alpha| < 1$, differentiation term-by-term with respect to α shows that

$$\sum_{k=1}^{\infty} k\alpha^{k-1} = \frac{1}{(1-\alpha)^2}.$$

With $\alpha = \frac{2}{3}$ we get $\int_{[0,1]} f dm = 3$.

In order to show that the Riemann integral $\int_0^1 f(x)dx$ does not exist, we define two sequences of approximating sums with different limits. For any $n \in \mathbb{N}$ let us take a partition of $[0,1]$ given by

$$0 < \frac{1}{3^n} < \frac{2}{3^n} < \ldots < \frac{3^n - 1}{3^n} < 1.$$

For each $i = 1,\ldots,3^n$ we have either $\left(\frac{i-1}{3^n}, \frac{i}{3^n}\right) \subset A_k$ for some $k = 1,\ldots,n$ or $\left(\frac{i-1}{3^n}, \frac{i}{3^n}\right) \subset C_n$. If $\left(\frac{i-1}{3^n}, \frac{i}{3^n}\right) \subset A_k$ for some $k = 1,\ldots,n$, we take any $x_i, y_i \in \left(\frac{i-1}{3^n}, \frac{i}{3^n}\right)$, so $f(x_i) = f(y_i) = k$. If, on the other hand, $\left(\frac{i-1}{3^n}, \frac{i}{3^n}\right) \subset C_n$, then there are points $x_i, y_i \in \left(\frac{i-1}{3^n}, \frac{i}{3^n}\right)$ such that $x_i \in C$ and $y_i \in A_{3^n}$, so that $f(x_i) = 0$ and $f(y_i) = 3^n$. Now consider the approximating sums

$$\alpha_n = \sum_{i=1}^{3^n} f(x_i)\left(\frac{i}{3^n} - \frac{i-1}{3^n}\right) = \sum_{i=1}^{3^n} \frac{1}{3^n} f(x_i) = \sum_{k=1}^{n} k\frac{2^{k-1}}{3^k},$$

$$\beta_n = \sum_{i=1}^{3^n} f(y_i)\left(\frac{i}{3^n} - \frac{i-1}{3^n}\right) = \sum_{i=1}^{3^n} \frac{1}{3^n} f(y_i) = \sum_{k=1}^{n} k\frac{2^{k-1}}{3^k} + 3^n\left(\frac{2}{3}\right)^n.$$

We have already seen that $\lim_{n\to\infty} \alpha_n = 3$. But $\lim_{n\to\infty} \beta_n = \infty$. This means that the Riemann integral $\int_0^1 f(x)dx$ does not exist.

1.32. By assumption, both f^+ and f^- are integrable. By Definition 1.29,

$$\int_{\mathbb{R}} f^+ dm = \sup\left\{\int_{\mathbb{R}} s dm : \ s \text{ is a simple function such that } s \leq f^+\right\}.$$

Let $s = \sum_{i=1}^n s_i \mathbf{1}_{A_i}$, where $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$ are pairwise disjoint sets with $\bigcup_{i=1}^n A_i = \mathbb{R}$ and $s_i \geq 0$ for $i = 1, \ldots, n$ such that $s \leq f^+$. Consider the simple function $r(x) = s(x - a)$. We have

$$r(x) = \sum_{i=1}^n s_i \mathbf{1}_{A_i}(x - a) = \sum_{i=1}^n s_i \mathbf{1}_{A_i + a} \leq f^+(x - a) = g^+(x).$$

By Definition 1.25 and Exercise 1.8, we have

$$\int_{\mathbb{R}} s dm = \sum_{i=1}^n s_i m(A_i) = \sum_{i=1}^n s_i m(A_i + a) = \int_{\mathbb{R}} r dm.$$

This implies that the sets

$$\left\{\int_{\mathbb{R}} s dm : \ s \text{ is a simple function such that } s \leq f^+\right\},$$
$$\left\{\int_{\mathbb{R}} r dm : \ r \text{ is a simple function such that } r \leq g^+\right\}$$

are the same, so their suprema are equal, which means that $\int_{\mathbb{R}} f^+ dm = \int_{\mathbb{R}} g^+ dm$. For the same reason $\int_{\mathbb{R}} f^- dm = \int_{\mathbb{R}} g^- dm$.

1.33. Apply the first Fatou lemma with $f_n = \mathbf{1}_{A_n}$, so that

$$\begin{aligned} P\left(\bigcup_{n\geq 1}\bigcap_{k\geq n} A_k\right) &= \int_{\Omega} \liminf_{n\to\infty} \mathbf{1}_{A_n} \, dP \\ &\leq \liminf_{n\to\infty} \int_{\Omega} \mathbf{1}_{A_n} \, dP \\ &= \liminf_{n\to\infty} P(A_n). \end{aligned}$$

1.34. First note that for any sequence g_n of non-negative measurable functions, the partial sums $h_n = \sum_{k=1}^n g_k$ are non-decreasing and that $\lim_{n\to\infty} h_n = \sum_{k=1}^\infty g_k$, so by monotone convergence

$$\begin{aligned} \int_{\Omega} \sum_{k=1}^\infty g_k \, dP &= \int_{\Omega} \lim_{n\to\infty} h_n \, dP = \lim_{n\to\infty} \int_{\Omega} h_n \, dP \\ &= \lim_{n\to\infty} \sum_{k=1}^n \int_{\Omega} g_k \, dP = \sum_{k=1}^\infty \int_{\Omega} g_k \, dP. \end{aligned}$$

For the given sequence f_k, apply this result to $g_k = |f_k|$, which non-negative and measurable. Letting $\varphi = \sum_{k=1}^{\infty} |f_k|$, we obtain

$$\int_{\Omega} \varphi \, d\mu = \sum_{k=1}^{\infty} \int |f_k| \, d\mu.$$

The right-hand side is finite by hypothesis, so φ is integrable. Therefore φ is finite μ-a.s. So the series $\sum_{k=1}^{\infty} |f_k|$ converges μ-a.s., and therefore the series $\sum_{k=1}^{\infty} f_k$ converges (since it converges absolutely) μ-a.s. Let $f = \sum_{k=1}^{\infty} f_k$ (put $f = 0$ on the set of μ-measure 0 for which the series diverges). For all partial sums we have

$$\left| \sum_{k=1}^{n} f_k \right| \leq \varphi,$$

so we can apply the dominated convergence theorem to find

$$\begin{aligned} \int_{\Omega} f \, d\mu &= \int_{\Omega} \lim_{n\to\infty} \sum_{k=1}^{n} f_k \, d\mu = \lim_{n\to\infty} \int_{\Omega} \sum_{k=1}^{n} f_k \, d\mu \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \int_{\Omega} f_k \, d\mu = \sum_{k=1}^{\infty} \int_{\Omega} f_k \, d\mu, \end{aligned}$$

as required.

1.35. If $x > 0$, we have $e^{-x} \in (0, 1)$, so

$$\sum_{n=1}^{\infty} (e^{-x})^n = \frac{e^{-x}}{1 - e^{-x}} = \frac{1}{e^x - 1}.$$

Hence the integrand can be written as $\frac{x}{e^x-1} = \sum_{n=1}^{\infty} xe^{-nx}$. Integration by parts shows that $\int_{[0,\infty)} xe^{-nx} dm(x) = \int_0^{\infty} xe^{-nx} dx = \frac{1}{n^2}$ for each $n = 1, 2 \ldots$. By Exercise 1.34,

$$\int_0^{\infty} \frac{x}{e^x - 1} dx = \int_{[0,\infty)} \frac{x}{e^x - 1} dm(x) = \sum_{n=1}^{\infty} \int_{[0,\infty)} xe^{-nx} dm(x) = \sum_{n=1}^{\infty} \frac{1}{n^2},$$

and the last sum is well-known to be $\frac{\pi^2}{6}$.

1.36. Suppose that $\omega \mapsto f(\omega, s)$ is integrable for some $s \in [a, b]$ and define $I_s = \int_{\Omega} f(\omega, s) d\mu(\omega)$. It is obvious that $\lim_{s\to t} I_s = I_t$ if and only if $\lim_{n\to\infty} I_{s_n} = I_t$ for any sequence s_n such that $\lim_{n\to\infty} s_n = t$. For any such s_n we are given that $\lim_{n\to\infty} f(\omega, s_n) = f(\omega, t)$ for each $\omega \in \Omega$. Moreover, $|f(\omega, s_n)| \leq g(\omega)$ for each n and each $\omega \in \Omega$. By Theorem 1.43, $\omega \mapsto f(\omega, s_n)$ and $\omega \mapsto f(\omega, t)$ are integrable and $\lim_{n\to\infty} I_{s_n} = I_t$. It follows that $\lim_{s\to t} I_s = I_t$.

Chapter 2

2.1. Denote by D the set of all points of discontinuity of F. For each point $a \in D$ there exists a rational number q_a such that

$$F(a-) < q_a < F(a)$$

If a, b are discontinuity points of F and $a < b$ then $q_a \neq q_b$ because $F(a) \leq F(b-)$. So we have a one-one correspondence between a subset of rational numbers and the set D, which is therefore countable.

2.2. The definition of P shows that

$$P(\Omega) = \sum_{k=1}^{\infty} \alpha_k P_k(\Omega) = \sum_{k=1}^{\infty} \alpha_k.$$

So $\sum_{k=1}^{\infty} \alpha_k = 1$ is a necessary condition for P to be a probability measure. This condition is also sufficient. Suppose that $\sum_{k=1}^{\infty} \alpha_k = 1$ holds. We show that P is countably additive. Let $A_1, A_2, \ldots \in \mathcal{F}$ be a sequence of pairwise disjoint events. Then

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{k=1}^{\infty} \alpha_k P_k \left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{k=1}^{\infty} \alpha_k \sum_{i=1}^{\infty} P_k(A_i) \\ &= \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \alpha_k P_k(A_i) = \sum_{i=1}^{\infty} P(A_i). \end{aligned}$$

2.3. The family $\sigma(X)$ trivially contains Ω and the empty set. Take $A_n \in \sigma(X)$ for $n = 1, 2, \ldots$. . Then $A_n = \{X \in B_n\} = X^{-1}(B_n)$ for some $B_n \in \mathcal{B}(\mathbb{R})$. Using the properties of the inverse image, we get

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} X^{-1}(B_n) = X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) \in \sigma(X).$$

For $A \in \sigma(X)$ we also have $A = X^{-1}(B)$ for some $B \in \mathcal{B}(\mathbb{R})$ and $\Omega \setminus A = \Omega \setminus X^{-1}(B) = X^{-1}(\mathbb{R} \setminus B) \in \sigma(X)$.

2.4. Suppose at first Y is a simple function and let $y_1, \ldots, y_m$ be its different possible values.

Since $A_i = \{Y = y_i\} \in \sigma(X)$, we have $A_i = \{X \in B_i\}$ for some $B_i \in \mathcal{B}(\mathbb{R})$. Define $h = \sum_{i=1}^{m} y_i \mathbf{1}_{B_i}$. Thus h is a Borel function. We show that for each $\omega \in \Omega$, $X(\omega)$ belongs to only one B_i. For if $X(\omega) \in B_i \cap B_j$ for some $i \neq j$, we have $\omega \in X^{-1}(B_i \cap B_j) \subset X^{-1}(B_i) \cap X^{-1}(B_j) = A_i \cap A_j$, which is impossible since $A_i \cap A_j = \varnothing$ for $i \neq j$. For $X(\omega) \in B_i$ observe that $h(X(\omega)) = y_i = Y(\omega)$ since $\{X \in B_i\} = A_i = \{Y = y_i\}$.

For a general random variable Y, applying Proposition 1.28 to the positive and negative parts Y^+ and Y^- of Y, we find a sequence of simple random variables Y_n such that $Y_n(\omega) \to Y(\omega)$ for each $\omega \in \Omega$. By the first part, for each n there is a Borel function $h_n : \mathbb{R} \to \mathbb{R}$ such that $Y_n(\omega) = h_n(X(\omega))$, $\omega \in \Omega$.

Let B be the set of $x \in \mathbb{R}$ for which $\{h_n(x)\}$ converges. By Exercise 1.18, B is a Borel set since $B = \{x \in \mathbb{R} : \liminf_{n\to\infty} h_n(x) - \limsup_{n\to\infty} h_n(x) = 0\}$. Let $h(x) = \lim_{n\to\infty} h_n(x)$ for $x \in B$ and let $h(x) = 0$ for $x \in \mathbb{R} \setminus B$. Since $h_n \mathbf{1}_B$ is measurable, it follows by Exercise 1.19 that $h = \lim_{n\to\infty} h_n \mathbf{1}_B$ is measurable. For each ω, $Y(\omega) = \lim_{n\to\infty} h_n(X(\omega))$; this implies that $X(\omega) \in B$ and consequently $Y(\omega) = \lim_{n\to\infty} h_n(X(\omega)) = h(X(\omega))$.

The opposite implication is trivial.

2.5. If $x < -1$, then $P(\{X \le x\}) = 0$. If $-1 \le x < 1$, then $P(\{X \le x\}) = \frac{1}{2}$. If $x \ge 2$, then $P(\{X \le x\} = 1$. The result is the distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < -1, \\ \frac{1}{2} & \text{if } -1 \le x < 1, \\ 1 & \text{if } 1 \le x. \end{cases}$$

This function is shown in Figure S.1. The dots represent the values of $F_X(x)$ at $x = -1, 1$, where the distribution function has discontinuities.

2.6. For $n = 1, 2, \ldots$ the probability that the nth toss of the coin is the first to yield 'heads' is $P(\{X = n\}) = \frac{1}{2^n}$. We can see that $\sum_{n=1}^{\infty} P(\{X = n\}) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$. Thus the distribution function F_X is given by

$$\begin{aligned} F_X(x) &= P_X((-\infty, x]) = P(\{X \le x\}) = \sum_{n \le x} P(\{X = n\}) \\ &= \begin{cases} 0 & \text{if } x < 1, \\ \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^n} & \text{if } n \le x < n+1, n = 1, 2, 3, \ldots \end{cases} \\ &= \begin{cases} 0 & \text{if } x < 1, \\ 1 - \frac{1}{2^n} & \text{if } n \le x < n+1, n = 1, 2, 3, \ldots \end{cases} \end{aligned}$$

2.7. Suppose that each X_n for $n = 1, 2, 3, \ldots$ is a random variable having the binomial distribution with parameters n, p (see Example 2.2), where $p = \frac{\lambda}{n}$ for some $\lambda > 0$. Then

$$\begin{aligned} P_{X_n}(k) = P(\{X_n = k\}) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{(n-k)! n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n \frac{\lambda^k}{k!} \to e^{-\lambda} \frac{\lambda^k}{k!} \end{aligned}$$

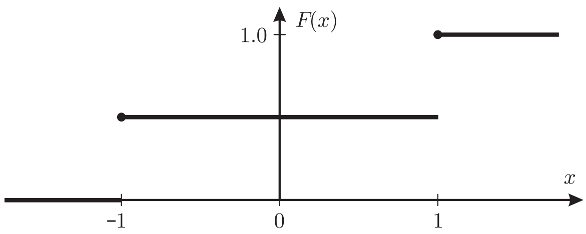

Figure S.1 Distribution function in Excercise 2.5.

since $\frac{n!}{(n-k)!n^k} = (1-\frac{1}{n})(1-\frac{2}{n})\cdot\ldots\cdot(1-\frac{k}{n}) \to 1$, $(1-\frac{\lambda}{n})^{-k} \to 1$ and $(1-\frac{\lambda}{n})^n \to e^{-\lambda}$ as $n \to \infty$.

2.8. The event $\{Y = n\}$ occurs exactly when the number of trading dates is n. It is the intersection of two independent events, requiring that the first $n-1$ trading dates record $r-1$ upward price moves and that the price also moves up on the nth date. These events have probabilities given by $\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$ and p, respectively. Multiply these probabilities (which is justified formally in Chapter 3), to find $P(\{Y=n\}) = \binom{n-1}{r-1}p^r(1-p)^{n-r}$.

The discrete random variable Y has distribution $P_Y = \sum \alpha_n\delta_{x_n}$, where $x_n = n$ and $\alpha_n = \binom{n-1}{r-1}p^r(1-p)^{n-r} \geq 0$ for $n = r, r+1, r+2, \ldots$. We verify that α_n add up to 1.

$$\sum_{n=r}^{\infty}\alpha_n = \sum_{n=r}^{\infty}\binom{n-1}{r-1}p^r(1-p)^{n-r}$$
$$= p^r\sum_{n=r}^{\infty}\binom{n-1}{n-r}(1-p)^{n-r} = p^r\frac{1}{p^r} = 1.$$

The sum is computed by using $x = 1-p$ in the Taylor expansion $\sum_{n=r}^{\infty}\binom{n-1}{r-1}x^{n-r}$ of $(1-x)^{-r}$ around 0.

2.9. Using the density of the normal distribution and making a change of variables in the integral, for any $y \in \mathbb{R}$ we have

$$P(Y \leq y) = P(\mu + \sigma X \leq y) = P\left(X \leq \frac{y-\mu}{\sigma}\right)$$
$$= \int_{-\infty}^{\frac{y-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx$$
$$= \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(z-\mu)^2}{2\sigma^2}}\,dz \quad \text{(substituting } z = \mu + \sigma x\text{)},$$

which shows that Y has the normal distribution $N(\mu, \sigma^2)$.

2.10. For any Borel set $B \in \mathcal{B}(\mathbb{R})$,

$$P(\{b + aX \in B\}) = P(\{X \in (B - b)/a\}) = \int_{(B-b)/a} f_X(x)dx.$$

Using the substitution $y = ax + b$, we can transform the last integral into $\int_B \frac{1}{|a|} f(\frac{y-b}{a})dy$.

2.11. For any Borel set $B \in \mathcal{B}(\mathbb{R})$ we define $B^* = \{y \in \mathbb{R} : \frac{1}{y} \in B\}$. Thus, $B^* \in \mathcal{B}(\mathbb{R})$, $0 \notin B^*$ and $\frac{1}{X} \in B \Leftrightarrow X \in B^*$. We calculate the density function of $Y = \frac{1}{X}$:

$$\begin{aligned} P\left(\frac{1}{X} \in B\right) &= P(X \in B^*) = \int_{B^*} \frac{1}{2} \mathbf{1}_{[-1,1]}(x)dx = \int_{B^*} \frac{1}{2} \mathbf{1}_{[-1,1]\cap B^*}(x)dx \\ &= \int_{\mathbb{R}} \frac{1}{2} \mathbf{1}_{[-1,0)\cap B^*}(x) + \int_{\mathbb{R}} \frac{1}{2} \mathbf{1}_{(0,1]\cap B^*}(x)dx. \end{aligned}$$

Substituting $y = \frac{1}{x}$ in both integrals, we get

$$\begin{aligned} \int_{\mathbb{R}} \frac{1}{2} \mathbf{1}_{[-1,0)\cap B^*}(x)dx &= \int_{\mathbb{R}} \frac{1}{2y^2} \mathbf{1}_{(-\infty,-1]\cap B}(y)dy, \\ \int_{\mathbb{R}} \frac{1}{2} \mathbf{1}_{(0,1]\cap B^*}(x)dx &= \int_{\mathbb{R}} \frac{1}{2y^2} \mathbf{1}_{[1,\infty)\cap B}(y)dy. \end{aligned}$$

Finally,

$$\begin{aligned} P\left(\frac{1}{X} \in B\right) &= \int_{\mathbb{R}} \frac{1}{2y^2} \mathbf{1}_{\{(-\infty,-1]\cup[1,\infty)\}\cap B}(y)dy \\ &= \int_B \frac{1}{2y^2} \mathbf{1}_{(-\infty,-1]\cup[1,\infty)}(y)dy. \end{aligned}$$

2.12. The assumptions on g ensure that it is an invertible function. For any Borel set $B \subset \mathbb{R}$

$$\begin{aligned} P(g(X) \in B) = P(X \in g^{-1}(B)) &= \int_{g^{-1}(B)} f_X(x)dx \\ &= \int_B \frac{f_X(g^{-1}(y))}{|g'^{-1}(y))|} \mathbf{1}_{g(\mathbb{R})}(y)dy, \end{aligned}$$

where we make the substitution $x = g^{-1}(y)$ in the integral.

2.13. For a put option written on a log-normally distributed stock, the dis-

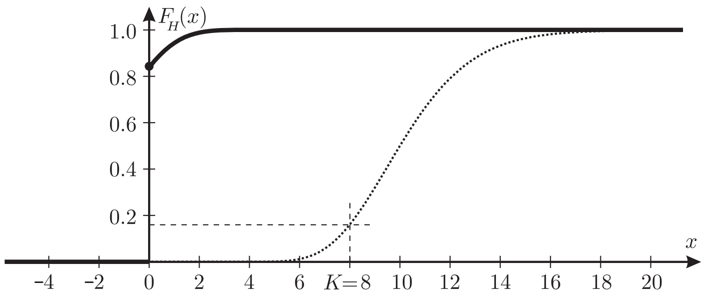

Figure S.2 Distribution function for the put option payoff in Exercise 2.13.

tribution function of the payoff $H = (K - S(T))^+$ can be written as

$$\begin{aligned} F_H(x) &= P\{(K - S(T))^+ \leq x\} \\ &= \begin{cases} 0 & \text{if} \quad x < 0, \\ P\{K - x \leq S(T)\} & \text{if} \quad x \geq 0, \end{cases} \\ &= \begin{cases} 0 & \text{if} \quad x < 0, \\ 1 - F_{S(T)}(K - x) & \text{if} \quad x \geq 0. \end{cases} \end{aligned}$$

The graph of F_H is shown in Figure S.2 in the case of a put option with expiry time $T = 1$ and strike price $K = 8$, and a log-normally distributed stock with parameters μ and σ as in Example 1.24. The dot indicates the value $F_H(0) = 1 - F_{S(T)}(K)$ at $x = 0$, where F_H has a discontinuity. For comparison, the log-normal distribution function $F_{S(1)}$ is shown as a dotted line.

2.14. Let X be a random variable with values $x_n = (-2)^n$ for $n = 1, 2, 3, \ldots$ and corresponding probabilities $p_n = P(\{X = x_n\}) = \frac{1}{2^n}$ for $n = 1, 2, 3, \ldots$. Then

$$\sum_{n=1}^{2N-1} x_n p_n = -1, \quad \sum_{n=1}^{2N} x_n p_n = 1$$

for each $N = 1, 2, \ldots$. Hence $\mathbb{E}(X) = \sum_{n=1}^{\infty} x_n p_n$ is undefined because the series does not converge.

2.15. By definition,

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{\infty} kP(\{x_k = k\}) = \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

2.16. First consider the case of $g(x) = \mathbf{1}_B(x)$ for a Borel set $B \subset \mathbb{R}$,

$$\mathbb{E}(g \circ X) = \int_\Omega \mathbf{1}_B(X(\omega))dP(\omega) = \int_\Omega \mathbf{1}_{\{X(\omega)\in B\}}(\omega)dP(\omega)$$
$$= P(\{X \in B\}) = \int_B f_X(x)dm(x) = \int_{\mathbb{R}} \mathbf{1}_B(x)f_X(x)dm(x).$$

Then by linearity we have the result for simple functions. A non-negative g can be written as the limit of some $s_n \uparrow g$ (see Exercise 1.20). By the monotone convergence theorem (Theorem 1.31) we have

$$\int_\Omega g(X(\omega))dP(\omega) = \lim \int_\Omega s_n(X(\omega))dP(\omega)$$
$$= \lim \int_{\mathbb{R}} s_n(x)f_X(x)dm(x) = \int_{\mathbb{R}} g(x)f_X(x)dm(x)$$

since $s_n \circ X \uparrow g \circ X$ and $s_n f_X \uparrow g f_X$. For a general integrable g such that $g \circ X$ is integrable we consider positive and negative parts separately.

2.17. First we prove that $\int_{\mathbb{R}} |x| f_X(x)dx$ is finite:

$$\int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} |\sigma y + \mu| \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$
$$\leq \sigma \int_{-\infty}^{\infty} |y| \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + |\mu| \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{2\sigma}{\sqrt{2\pi}} + |\mu|$$
$$= \lim_{a\to\infty} \left(-\frac{2}{\sqrt{2\pi}} e^{-\frac{a^2}{2}} + \frac{2}{\sqrt{2\pi}} \right) = \frac{2}{\sqrt{2\pi}}$$

since

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1$$

and

$$\int_{-\infty}^{\infty} |y| \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} y e^{-\frac{y^2}{2}} dy = \lim_{a\to\infty} \int_0^a \frac{2}{\sqrt{2\pi}} y e^{-\frac{y^2}{2}} dy.$$

Now

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} (\sigma y + \mu) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \mu$$

since, by Example 2.34,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} y e^{-\frac{y^2}{2}} dy = 0.$$

2.18. In the case of the Cauchy distribution we have

$$\int_{-\infty}^{\infty} |x| f_X(x) dx = \int_{-\infty}^{\infty} |x| \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} \int_0^{\infty} \frac{2x}{1+x^2} dx$$
$$= \frac{1}{\pi} \lim_{a\to\infty} \int_0^a \frac{2x}{1+x^2} dx = \frac{1}{\pi} \lim_{a\to\infty} \ln(1+a^2) = \infty,$$

which means that the expectation is undefined.

2.19. Suppose that a random variable X has distribution $P_X = \frac{1}{4}\delta_0 + \frac{3}{4}P$, where P is the exponential distribution with density (2.1). Then

$$\mathbb{E}(X) = \frac{1}{4} \cdot 0 + \frac{3}{4} \int_{\mathbb{R}} x \mathbf{1}_{[0,\infty)} \lambda e^{-\lambda x} dm(x) = \frac{3}{4} \int_0^{\infty} \lambda x 3^{-\lambda x} dx = \frac{3}{4\lambda}.$$

2.20. If X has the Poisson distribution with parameter λ, then

$$\mathbb{E}(X^2) = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} [k(k-1)+k] \frac{\lambda^k}{k!}$$
$$= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$
$$= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda.$$

Since $\mathbb{E}(X) = \lambda$ by Exercise 2.15, it follows that

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

2.21. Suppose that $X \sim N(\mu, \sigma^2)$. By Exercise 2.17, $\mathbb{E}(X) = \mu$. Let us compute $\mathrm{Var}(X)$:

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy = \sigma^2$$

since $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$. The last integral is found by integrating by parts.

2.22. (1) For a random variable X with exponential distribution we compute

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx = \int_{-\infty}^{\infty} x \lambda e^{-\lambda x} \mathbf{1}_{[0,\infty)}(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Obviously, the integral is absolutely convergent. In order to compute

$\mathrm{Var}(X)$ we first calculate

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_{-\infty}^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

Then

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

(2) For a random variable X with log-normal density (see (1.7)) we have

$$\begin{aligned}\mathbb{E}(X) &= \int_{\mathbb{R}} x f_X(x)dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \mathbf{1}_{[0,\infty)}(x)dx = \int_0^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} dx.\end{aligned}$$

By substituting $y = \ln x$, we obtain

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^y e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\ &= e^{\mu+\frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-[y-(\mu+\sigma^2)]^2}{2\sigma^2}} dy = e^{\mu+\sigma^2}\end{aligned}$$

since

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-[y-(\mu+\sigma^2)]^2}{2\sigma^2}} dy = 1.$$

Next, we compute

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \mathbf{1}_{[0,\infty)}(x)dx = \int_0^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{2y} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[y-(\mu+2\sigma^2)]^2}{2\sigma^2}} e^{2(\mu+\sigma^2)} dy \\ &= e^{2(\mu+\sigma^2)}.\end{aligned}$$

As above, we substituted $y = \ln x$. Finally,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

2.23. (1) $\phi_X(0) = \mathbb{E}(e^0) = 1$.
(2) $|\phi_X(t)| = |\mathbb{E}(e^{itX})| \le \mathbb{E}(|e^{itX}|) = \mathbb{E}(1) = 1$.

2.24. Let X have the Poisson distribution with parameter $\lambda > 0$. Then e^{itX} is a discrete random variable with values e^{itk} and corresponding probabilities $e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$. It follows that

$$\phi_X(t) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}e^{itk} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!} = e^{-\lambda}e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

2.25. If X is a random variable with the standard normal distribution $N(0, 1)$, then $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. It follows that

$$\begin{aligned}\phi_X(t) &= \int_{-\infty}^{\infty} e^{itx} f_X(x)dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}+itx}dx \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2}dx = e^{-\frac{1}{2}t^2}\end{aligned}$$

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

2.26. If $Y = aX + b$, then

$$\phi_Y(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{it(aX+b)}) = e^{itb}\mathbb{E}(e^{iatX}) = e^{itb}\phi_X(at).$$

When Y has the normal distribution $N(\mu, \sigma^2)$, we can write $Y = \sigma X + \mu$, where X has the standard normal distribution $N(0, 1)$. By Exercise 2.25, it follows that

$$\phi_Y(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}.$$

2.27. Since $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, Theorem 2.42 shows that $\mathrm{Var}(X) = -\phi_X''(0) + \phi_X'(0)^2$. But $\phi_X(0) = 1$, hence simple transformations show that $\mathrm{Var}(X) = -(\ln\phi_X)''(0)$.

2.28. Suppose that $X \sim \mathbb{N}(0, \sigma^2)$. By (2.3) (see Exercise 2.16),

$$\mathbb{E}(X^n) = \frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} x^n e^{-\frac{x^2}{2\sigma^2}}dx.$$

For odd n we have $\mathbb{E}(X^n) = 0$ since we integrate an odd function. Let $n = 2k$, $k = 1, 2, 3, \ldots$. Then

$$\mathbb{E}(X^{2k}) = \frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} x^{2k} e^{-\frac{x^2}{2\sigma^2}}dx.$$

Integrate by parts with $u = x^{2k-1}$ and $v = xe^{-\frac{x^2}{2\sigma^2}}$ to get

$$\begin{aligned}\mathbb{E}(X^{2k}) &= -\frac{\sigma}{\sqrt{2\pi}}x^{2k-1}e^{-\frac{x^2}{2\sigma^2}}\Big|_{-\infty}^{\infty} + \sigma^2(2k-1)\frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} x^{2k-2}e^{-\frac{x^2}{2\sigma^2}}dx \\ &= \sigma^2(2k-1)\frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} x^{2k-2}e^{-\frac{x^2}{2\sigma^2}}dx\end{aligned}$$

since first term vanishes. By repeating integration by parts, we can prove that $\mathbb{E}(X^{2k}) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2k-1)\sigma^{2k}$.

In the general case, when $X \sim N(\mu, \sigma^2)$, we have $\phi_X(t) = e^{-\frac{1}{2}\sigma^2 t^2 + i\mu t}$ by Exercise 2.26. The derivatives of $\phi_X(t)$ with respect to t are

$$\begin{aligned}
\phi'_X(t) &= (-\sigma^2 t + i\mu)e^{-\frac{1}{2}\sigma^2 t^2 + i\mu t},\\
\phi''_X(t) &= (\sigma^2 + (-\sigma^2 t + i\mu)^2)e^{-\frac{1}{2}\sigma^2 t^2 + i\mu t},\\
\phi'''_X(t) &= [-3\sigma^2(-\sigma^2 t + i\mu) + (-\sigma^2 t + i\mu)^3]e^{-\frac{1}{2}\sigma^2 t^2 + i\mu t},\\
\phi'^{v}_X(t) &= [3\sigma^4 - 6\sigma^2(-\sigma^2 t + i\mu)^2 + (-\sigma^2 t + i\mu)^4]e^{-\frac{1}{2}\sigma^2 t^2 + i\mu t}.
\end{aligned}$$

Substituting $t = 0$, we obtain

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{i}\phi'_X(0) = \mu,\\
\mathbb{E}(X^2) &= \frac{1}{i^2}\phi''_X(0) = \mu + \sigma^2,\\
\mathbb{E}(X^3) &= \frac{1}{i^3}\phi'''_X(0) = \mu^3 + 3\mu\sigma^2,\\
\mathbb{E}(X^4) &= \frac{1}{i^4}\phi'^{v}_X(0) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4.
\end{aligned}$$

Chapter 3

3.1. Take two subsets $A, B \subset \mathbb{R}^2$ of the form $A = A_1 \times A_2$, $B = B_1 \times B_2$, where A_1, A_2, B_1, B_2 are non-empty Borel sets and $A_1 \cap B_1 = \emptyset$, $A_2 \cap B_2 = \emptyset$. Assume further that $(A_1 \times A_2) \cup (B_1 \times B_2) = C_1 \times C_2$, where C_1, C_2 are Borel sets.

Since $A_1 \times A_2 \subset C_1 \times C_2$, $B_1 \times B_2 \subset C_1 \times C_2$, we have $A_1, B_1 \subset C_1$ and $A_2, B_2 \subset C_2$. Hence $(A_1 \cup B_1) \times (A_2 \cup B_2) \subset C_1 \times C_2$.

Now, taking $x \in A_1$ and $y \in B_2$, we have $(x, y) \in C_1 \times C_2$ but $(x, y) \notin (A_1 \times A_2) \cup (B_1 \times B_2)$, which is a contradiction.

3.2. First note that for any $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$ we have $\{Pr_1 \in A_1\} = A_1 \times \Omega_2$ and $\{Pr_2 \in A_2\} = \Omega_1 \times A_2$. Define the family of sets ('cylinders')

$$C = \{A_1 \times \Omega_2 : A_1 \in \mathcal{F}_1\} \cup \{\Omega_1 \times A_2 : A_2 \in \mathcal{F}_2\}.$$

By construction, the projections are measurable with respect $\sigma(C)$. Further, since C is contained in the family consisting of all finite unions of elements of $\mathcal{R}$, it follows that $\sigma(C) \subset \sigma(\mathcal{R})$.

We also have

$$A_1 \times A_2 = (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) \in \sigma(C).$$

Thus, $\sigma(\mathcal{R}) \subset \sigma(\sigma(C)) = \sigma(C)$, and $\sigma(C) = \sigma(\mathcal{R}) = \mathcal{F}_1 \otimes \mathcal{F}_2$.

3.3. Denote by $\mathcal{F}_{\mathcal{I}}$ the smallest σ-field on $\mathbb{R}^2$ containing the family

$$\mathcal{I} = \{I_1 \times I_2 : I_1, I_2 \text{ are open intervals in } \mathbb{R}\}.$$

Since $\mathcal{I} \subset \mathcal{B}(\mathbb{R}^2)$, it follows that $\sigma(\mathcal{I}) \subset \mathcal{B}(\mathbb{R}^2)$. In order to show that $\mathcal{B}(\mathbb{R}^2) \subset \sigma(\mathcal{I})$ we first prove that for any $A, B \in \mathcal{B}(\mathbb{R})$ we have $A \times \mathbb{R}$ and $\mathbb{R} \times B \in \sigma(\mathcal{I})$.

Let us consider two families

$$D_1 = \{A \subset \mathbb{R} : A \times \mathbb{R} \in \sigma(\mathcal{I})\}, \quad D_2 = \{B \subset \mathbb{R} : \mathbb{R} \times B \in \sigma(\mathcal{I})\}.$$

We verify that D_1 and D_2 are σ-fields. We do this for D_1 (for D_2 the proof is identical). Of course $\mathbb{R} \in D_1$. If $A \in D_1$, then $A \times \mathbb{R} \in \sigma(\mathcal{I})$ and $\mathbb{R}^2 \setminus (A \times \mathbb{R}) = (\mathbb{R} \setminus A) \times \mathbb{R} \in \sigma(\mathcal{I})$, which means that $\mathbb{R} \setminus A \in D_1$. Finally, if $A_1, A_2, \ldots \in D_1$, then $A_i \times \mathbb{R} \in \sigma(\mathcal{I})$ for $i = 1, 2, \ldots$, and

$$\bigcup_{i=1}^{\infty} (A_i \times \mathbb{R}) = \left(\bigcup_{i=1}^{\infty} A_i \right) \times \mathbb{R} \in \sigma(\mathcal{I}).$$

Hence, $\bigcup_{i=1}^{\infty} A_i \in D_1$.

It follows $D_1 \cap D_2$ is a σ-field. Moreover, since D_1 and D_2 contain all open intervals, so does $D_1 \cap D_2$. Hence, by Exercise 1.3, $\mathcal{B}(\mathbb{R}) \subset D_1 \cap D_2$. This implies that $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \subset D_1 \otimes D_2$. To show that $D_1 \otimes D_2 = \sigma(\mathcal{I})$ take any $A_1 \in D_1$ and $A_2 \in D_2$. Then $A_1 \times A_2 = (A_1 \times \mathbb{R}) \cap (\mathbb{R} \times A_2) \in \sigma(\mathcal{I})$, and so $D_1 \otimes D_2 \subset \sigma(\mathcal{I})$.

3.4. For a measurable rectangle $B = B_1 \times B_2$ with $B_1 \in \mathcal{F}_1$ and $B_2 \in \mathcal{F}_2$, we obtain $B_{\omega_1} = B_2$ if $\omega_1 \in B_1$ and $B_{\omega_1} = \varnothing$ otherwise, hence

$$B_1 \times B_2 \in \mathcal{H} = \{H \in \mathcal{F}_1 \otimes \mathcal{F}_2 : H_{\omega_1} \in \mathcal{F}_2 \text{ for all } \omega_1 \in \Omega_1\}.$$

We need to show that $\mathcal{H}$ is a σ-field. To verify that take $H \in \mathcal{H}$. Then

$$\begin{aligned}(\Omega \setminus H)_{\omega_1} &= \{\omega_2 : (\omega_1, \omega_2) \in (\Omega \setminus H)\} \\ &= \Omega_2 \setminus \{\omega_2 : (\omega_1, \omega_2) \in H\} = \Omega_2 \setminus H_{\omega_1}\end{aligned}$$

and $\Omega \setminus H \in \mathcal{H}$. Now take $H_1, H_2, \ldots$ such that $H_i \in \mathcal{H}$. We have

$$\begin{aligned}\left(\bigcup_{i=1}^{\infty} H_i\right)_{\omega_1} &= \left\{\omega_2 : (\omega_1, \omega_2) \in \left(\bigcup_{i=1}^{\infty} H_i\right)\right\} \\ &= \bigcup_{i=1}^{\infty} \{\omega_2 : (\omega_1, \omega_2) \in H_i\} = \bigcup_{i=1}^{\infty} (H_i)_{\omega_1}.\end{aligned}$$

Since $\mathcal{H}$ is a σ-field containing any measurable rectangles $B = B_1 \times B_2$, we have $\mathcal{F}_1 \otimes \mathcal{F}_2 \subset \mathcal{H}$. By the definition of $\mathcal{H}$, this implies that $A_{\omega_1} \in \mathcal{F}_2$ for any $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ and $\omega_1 \in \Omega_1$. The proof for A_{ω_2} is identical.

3.5. First we show that the limit in Definition 3.9 (iii) exists. By (3.4) (Theorem 3.5 (iii)) and Definition 3.9, we have

$$\begin{aligned}\mu_1^{(n)} \otimes \mu_2^{(n)}(C \cap (A_n \times B_n)) &= \int_{\Omega_1} \mu_2^{(n)}((C \times (A_n \times B_n))_{\omega_1}) d\mu_1^{(n)}(\omega_1) \\ &= \int_{\Omega_1} \mu_2(C_{\omega_1} \cap B_n) \mathbf{1}_{A_n}(\omega_1) d\mu_1(\omega_1) \\ &\leq \int_{\Omega_1} \mu_2(C_{\omega_1} \cap B_{n+1}) \mathbf{1}_{A_{n+1}}(\omega_1) d\mu_1(\omega_1) \\ &= \mu_1^{(n+1)} \otimes \mu_2^{(n+1)}(C \cap (A_{n+1} \times B_{n+1}))\end{aligned}$$

since

$$(C \cap (A_n \times B_n))_{\omega_1} = \begin{cases} C_{\omega_1} \cap B_n & \text{if } \omega_1 \in A_n \\ \varnothing & \text{otherwise.} \end{cases}$$

Hence $\mu_1^{(n)} \otimes \mu_2^{(n)}(C \cap (A_n \times B_n))$ is a non-decreasing sequence of non-negative numbers. This implies that the limit exists.

In order to prove that this limit does not depend on the choice of the sequences A_n, B_n take two other sequences of events $\tilde{A}_n \in \mathcal{F}_1$ with $\mu_1(\tilde{A}_n) < \infty$ and $\tilde{A}_n \subset \tilde{A}_{n+1}$, and $\tilde{B}_n \in \mathcal{F}_2$ with $\mu_2(\tilde{B}_n) < \infty$ and $\tilde{B}_n \subset \tilde{B}_{n+1}$ for $n = 1, 2, 3, \ldots$ such that $\Omega_1 = \bigcup_{n=1}^{\infty} \tilde{A}_n$ and $\Omega_2 = \bigcup_{n=1}^{\infty} \tilde{B}_n$. Denote by $\tilde{\mu}_1^{(n)}$ the restriction of μ_1 to $\tilde{A}_n$ and by $\tilde{\mu}_2^{(n)}$ the restriction of μ_2 to $\tilde{B}_n$. For the same reason as above,

$$\tilde{\mu}_1^{(n)} \otimes \tilde{\mu}_2^{(n)}(C \cap (\tilde{A}_n \times \tilde{B}_n)) = \int_{\Omega_1} \mu_2(C_{\omega_1} \cap \tilde{B}_n) \mathbf{1}_{\tilde{A}_n}(\omega_1) d\mu_1(\omega_1).$$

By Theorem 3.5 (i), for any $C \in \mathcal{F}_1 \otimes \mathcal{F}_2$ the functions $\omega_1 \mapsto \mu_2(C_{\omega_1} \cap B_n) \mathbf{1}_{A_n}(\omega_1)$, $\omega_1 \mapsto \mu_2(C_{\omega_1} \cap \tilde{B}_n) \mathbf{1}_{\tilde{A}_n}(\omega_1)$ are measurable with respect to $\mathcal{F}_1$ for all $n = 1, 2, 3, \ldots$. Moreover, for any $\omega_1 \in \Omega_1$ we have

$$\lim_{n\to\infty} \mu_2(C_{\omega_1} \cap B_n)\mathbf{1}_{A_n}(\omega_1) = \lim_{n\to\infty} \mu_2(C_{\omega_1} \cap \tilde{B}_n)\mathbf{1}_{\tilde{A}_n}(\omega_1) = \mu_2(C_{\omega_1}).$$

Then, by Exercise 1.19, the function: $\omega_1 \mapsto \mu_2(C_{\omega_1})$ is measurable with respect to $\mathcal{F}_1$. Using the monotone convergence theorem, we have

$$\begin{aligned}
\lim_{n\to\infty} \mu_1^{(n)} \otimes \mu_2^{(n)}(C \cap (A_n \times B_n)) &= \lim_{n\to\infty} \int_{\Omega_1} \mu_2(C_{\omega_1} \cap B_n) d\mu_1(\omega_1) \\
&= \int_{\Omega_1} \mu_2(C_{\omega_1})\mathbf{1}_{A_n}(\omega_1) d\mu_1(\omega_1) \\
&= \lim_{n\to\infty} \int_{\Omega_1} \mu_2(C_{\omega_1} \cap \tilde{B}_n)\mathbf{1}_{\tilde{A}_n}(\omega_1) d\mu_1(\omega_1) \\
&= \lim_{n\to\infty} \tilde{\mu}_1^{(n)} \otimes \tilde{\mu}_2^{(n)}(C \cap (\tilde{A}_n \times \tilde{B}_n)).
\end{aligned}$$

3.6. Take A_n, B_n as in Definition 3.9 (i). We have $A_n \times B_n \subset A_{n+1} \times B_{n+1}$ for all $n = 1, 2, 3, \ldots$ and

$$\Omega_1 \times \Omega_2 = \left(\bigcup_{n=1}^{\infty} A_n\right) \times \left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} A_n \times B_n.$$

Fix any $n = n_0$. By Theorem 3.5 (iii), we have

$$\begin{aligned}
(\mu_1 \otimes \mu_2)(A_{n_0} \times B_{n_0}) &= \lim_{n\to\infty} \mu_1^{(n)} \otimes \mu_2^{(n)}((A_{n_0} \times B_{n_0}) \cap (A_n \times B_n)) \\
&= \mu_1^{(n_0)} \otimes \mu_2^{(n_0)}(A_{n_0} \times B_{n_0})
\end{aligned}$$

since $(A_{n_0} \times B_{n_0}) \cap (A_n \times B_n) = A_{n_0} \times B_{n_0}$ for $n \geq n_0$. On the other hand,

$$\begin{aligned}
\mu_1^{(n_0)} \otimes \mu_2^{(n_0)}((A_{n_0} \times B_{n_0}) &= \int_{\Omega_1} \mu_2^{(n_0)}(A_{n_0} \times B_{n_0})_{\omega_1}) d\mu_1^{(n_0)}(\omega_1) \\
&= \int_{\Omega_1} \mu_2^{(n_0)}(B_{n_0})\mathbf{1}_{A_{n_0}}(\omega_1) d\mu_1^{(n_0)}(\omega_1) \\
&= \mu_1^{(n_0)}(A_{n_0})\mu_2^{(n_0)}(B_{n_0}) \\
&= \mu_1(A_{n_0})\mu_2(B_{n_0}) < \infty.
\end{aligned}$$

3.7. By $m^{(n)}$ we denote the restriction of the Lebesgue measure m to $(-n, n)$. In order to apply Definition 3.9 (iii) to the Lebesgue measure m on $\mathbb{R}$, we compute

$$\begin{aligned}
&m^{(n)} \otimes m^{(n)}(((a,b)\times(c,d)) \cap ((-n,n)\times(-n,n)))\\
&= \int_{\mathbb{R}} m^{(n)}((((a,b)\times(c,d)) \cap ((-n,n)\times(-n,n))_x) dm^{(n)}(x)\\
&= \int_{\mathbb{R}} m^{(n)}((c,d)\cap(-n,n))\mathbf{1}_{(a,b)\cap(-n,n)}(x) dm^{(n)}(x)\\
&= \int_{\mathbb{R}} m((c,d)\cap(-n,n))\mathbf{1}_{(a,b)\cap(-n,n)}(x) dm(x)\\
&= m((c,d)\cap(-n,n))m((a,b)\cap(-n,n)) = (b-a)(c-d)
\end{aligned}$$

for all n such that (a,b) and (c,d) are contained in $(-n,n)$. We have applied the simple observation that

$$\begin{aligned}
&(((a,b)\times(c,d)) \cap ((-n,n)\times(-n,n)))_x\\
&= \begin{cases} (c,d)\cap(-n,n) & \text{if } x \in (a,b)\cap(-n,n),\\ \varnothing & \text{otherwise.}\end{cases}
\end{aligned}$$

This gives the conclusion.

3.8. Suppose that $(X,Y) : \Omega \to \mathbb{R}^2$ is a random vector on $(\Omega,\mathcal{F})$. Let B be a Borel set in $\mathbb{R}$. Then $B\times\mathbb{R} \in \mathcal{B}(\mathbb{R}^2)$ and $\{X \in B\} = \{(X,Y) \in B\times\mathbb{R}\} \in \mathcal{F}$. This implies that X is a random variable. Similarly for Y.

To prove the converse consider the family

$$\mathcal{H} = \{B \in \mathcal{B}(\mathbb{R}^2) : \{(X,Y) \in B\} \in \mathcal{F}\}.$$

Any measurable rectangle $B_1 \times B_2$ with $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ belongs to $\mathcal{H}$ since $\{(X,Y) \in B_1 \times B_2\} = \{X \in B_1\} \cap \{Y \in B_2\} \in \mathcal{F}$. If $B \in \mathcal{H}$, then the complement $\mathbb{R}^2 \setminus B$ belongs to $\mathcal{H}$ since $\{(X,Y) \in \mathbb{R}^2 \setminus B\} = \Omega \setminus \{(X,Y) \in B\} \in \mathcal{F}$. Now, suppose that $B_n \in \mathcal{H}$ for $n = 1,2,3,\ldots$. Since $\{(X,Y) \in \bigcup_{n=1}^\infty B_n\} = \bigcup_{n=1}^\infty \{(X,Y) \in B_n\} \in \mathcal{F}$, the union $\bigcup_{n=1}^\infty B_n$ is also in $\mathcal{H}$. Hence $\mathcal{H}$ is a σ-field and $\mathcal{B}(\mathbb{R}^2) \subset \mathcal{H}$. This implies that (X,Y) is a random vector.

3.9. The random vector (X_1,Y_1) takes values $(X_1,Y_1)(\omega_1) = (110,60)$ and $(X_1,Y_1)(\omega_2) = (90,40)$ with probabilities $\frac{1}{2}$ and $\frac{1}{2}$. Then for any $B \in \mathcal{B}(\mathbb{R}^2)$ we have

$$P_{X_1,Y_1}(B) = \begin{cases} 0 & \text{if } \{(90,40),(110,60)\} \cap B = \varnothing,\\ 1 & \text{if } \{(90,40),(110,60)\} \subset B,\\ \frac{1}{2} & \text{otherwise.}\end{cases}$$

If we denote by $\delta_{x,y}$ the Dirac measure on $\mathcal{B}(\mathbb{R}^2)$ concentrated at a point (x, y), then the joint distribution of (X_1, Y_1) can be written in the form

$$P_{X_1,Y_1} = \frac{1}{2}\delta_{110,60} + \frac{1}{2}\delta_{90,40}.$$

Similarly, for (X_2, Y_2)

$$P_{X_2,Y_2} = \frac{1}{2}\delta_{110,40} + \frac{1}{2}\delta_{90,60}.$$

Hence, $P_{X_1,Y_1} \neq P_{X_2,Y_2}$. For the marginal distributions we have

$$\begin{aligned} P_{X_1} &= P_{X_2} = \frac{1}{2}\delta_{90} + \frac{1}{2}\delta_{110}, \\ P_{Y_1} &= P_{Y_2} = \frac{1}{2}\delta_{40} + \frac{1}{2}\delta_{60}. \end{aligned}$$

3.10. If $t < s$, then $(-\infty, t] \subset (-\infty, s]$ and

$$F_{X,Y}(t, y) = P(X \leq t, Y \leq y) \leq P(X \leq s, Y \leq y) = F_{X,Y}(s, y).$$

This means that $F_{X,Y}(x, y)$ is non-decreasing in x. Similarly for y. Now, by Theorem 1.11 (v), we have

$$\begin{aligned} \lim_{y\to\infty} F_{X,Y}(a, y) = \lim_{n\to\infty} F_{X,Y}(a, n) &\\ &= \lim_{n\to\infty} P_{X,Y}((-\infty, a] \times (-\infty, n]) \\ &= P_{X,Y}((-\infty, a] \times \bigcup_{n=1}^{\infty}(-\infty, n]) \\ &= P_{X,Y}((-\infty, a] \times \mathbb{R}) = P(X \in (-\infty, a]) = F_X(a). \end{aligned}$$

The proof that $\lim_{y\to\infty} F_{X,Y}(x, b) = F_Y(b)$ is similar.

3.11. Since

$$\begin{aligned} F_X(a) &= P(X \leq a, Y > b) + F_{X,Y}(a, b), \\ F_Y(b) &= P(X > a, Y \leq b) + F_{X,Y}(a, b), \end{aligned}$$

and so

$$\begin{aligned} 1 &= P(X > a, Y > b) + P(X \leq a, Y > b) \\ &\qquad + P(X > a, Y \leq b) + F_{X,Y}(a, b) \\ &= P(X > a, Y > b) + F_X(a) + F_Y(b) - F_{X,Y}(a, b), \end{aligned}$$

it follows that

$$P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F_{X,Y}(a, b).$$

3.12. First we show that $f_{X,Y}$ given by (3.5) is a density. Observe that

$$\frac{1}{1-\rho^2}(x_1^2 - 2\rho x_1 x_2 + x_2^2) = x_1^2 + \frac{1}{1-\rho^2}(x_2 - \rho x_1)^2.$$

By Fubini's theorem and since the bivariate density is a continuous function, we have

$$\int_{\mathbb{R}^2} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right) dm_2(x_1, x_2)$$
$$= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{1}{2}\left(\frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}\right)^2\right) dx_2\right) dx_1$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} dx_1 = 1$$

since integrating by substitution with $y = \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}$ gives

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{1}{2}\left(\frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}\right)^2\right) dx_2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

Now, by (3.6), we have

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dm(y) = \int_{\mathbb{R}} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{1-\rho^2}\right) dy$$
$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \exp\left(-\frac{1}{2(1-\rho^2)}(y - \rho x)^2\right) dy$$
$$= \frac{1}{2\pi} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) dy$$
$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

The proof that $f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$ is similar.

3.13. Let B be a Borel set in $\mathbb{R}$ and let $A = \{(x,y) \in \mathbb{R}^2 : x + y \in B\}$. Then

$$\begin{aligned}
P_{X+Y}(B) &= P(X+Y \in B) = P((X,Y) \in A) \\
&= \int\int_{\mathbb{R}^2} \mathbf{1}_A(x,y) f_{X,Y}(x,y)\,dm_2(x,y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(x,y) f_{X,Y}(x,y)\,dm(y) \right) dm(x) \\
&\quad \text{by Fubini's theorem} \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_B(z) f_{X,Y}(x,z-x)\,dm(z) \right) dm(x) \\
&\quad \text{by the substitution } z = x+y \text{ and by Exercise 1.32} \\
&= \int_{\mathbb{R}} \left(\int_{B} f_{X,Y}(x,z-x)\,dm(z) \right) dm(x) \\
&= \int_{B} \left(\int_{\mathbb{R}} f_{X,Y}(x,z-x)\,dm(x) \right) dm(z) \\
&\quad \text{by Fubini's theorem.}
\end{aligned}$$

Hence $\int_{\mathbb{R}} f_{X,Y}(x,z-x)\,dm(x)$ is the density for $X+Y$.

3.14. Suppose that random variables X, Y have joint density

$$f_{X,Y}(x,y) = e^{-(x+y)} \mathbf{1}_{(0,\infty)\times(0,\infty)}.$$

To find the density of X/Y we compute the distribution function $F_{X/Y}$. Let $z > 0$. Then

$$\begin{aligned}
F_{X/Y}(z) &= P(X/Y \leq z) = P_{X,Y}(\{(x,y) : \frac{x}{y} \leq z, x,y > 0\} \\
&= \int_{\{(x,y):\, \frac{x}{y} \leq z\}} e^{-(x+y)} \mathbf{1}_{(0,\infty)\times(0,\infty)}(x,y)\,dm_2(x,y) \\
&\quad \text{by Fubini's theorem and since } e^{-(x+y)} \text{ is continuous} \\
&= \int_0^\infty \left(\int_{\frac{x}{z}}^\infty e^{-x} e^{-y}\,dy \right) dx = \int_0^\infty \left(e^{-x} \int_{\frac{x}{z}}^\infty e^{-y}\,dy \right) dx \\
&= \int_0^\infty e^{-x} e^{-\frac{x}{z}}\,dx = \frac{z}{1+z}.
\end{aligned}$$

Since $F_{X/Y}(z)$ is a differentiable function for all $z \in (0,\infty)$, we have $F'_{X/Y}(z) = f_{X/Y}(z) = \frac{1}{(1+z)^2}$.

3.15. The proof is similar to that for $n = 2$ (Exercise 3.3). Since $\mathcal{I}_n \subset \mathcal{R}_n$, we have $\sigma(\mathcal{I}_n) \subset \mathcal{B}(\mathbb{R}^n)$. To show the reverse inclusion define

$$D_i = \{A \subset \mathbb{R} : \mathbb{R}^{i-1} \times A \times \mathbb{R}^{n-i} \in \sigma(\mathcal{I}_n)\} \text{ for } i = 1,2,3,\ldots,n.$$

We verify in the same way as in Exercise 3.3 that D_i, $i = 1, \ldots, n$ are σ-fields. Hence, by Exercise 1.3 we have $\mathcal{B}(\mathbb{R}) \subset D_i$ for $i = 1, 2, 3, \ldots, n$. This implies that $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R}) \subset D_1 \otimes \ldots \otimes D_n$. In order to show that $D_1 \otimes \ldots \otimes D_n = \sigma(\mathcal{I}_n)$, take any $A_i \in D_i$ for $i = 1, 2, 3, \ldots, n$. By the definition of D_i, we have

$$A_1 \times A_2 \times \cdots \times A_n = \bigcap_{i=1}^{n} \{\mathbb{R}^{i-1} \times A_i \times \mathbb{R}^{n-i}\} \in \sigma(\mathcal{I}_n).$$

It follows that

$$D_1 \otimes \ldots \otimes D_n \subset \sigma(\mathcal{I}_n).$$

3.16. The proof is similar as in the case $n = 2$ (Exercise 3.8). Suppose that $X = (X_1, \ldots, X_n)$ is a random vector on $(\Omega, \mathcal{F})$. Let B be a Borel set in $\mathbb{R}$. Then $\mathbb{R}^{i-1} \times B \times \mathbb{R}^{n-i} \in \mathcal{B}(\mathbb{R}^n)$ for $i = 1, \ldots, n$, and $\{X_i \in B\} = \{X \in \mathbb{R}^{i-1} \times B \times \mathbb{R}^{n-i}\} \in \mathcal{F}$. This implies that X_i is a random variable for $i = 1, \ldots, n$.

To prove the converse, consider the family $\mathcal{H} = \{B \in \mathcal{B}(\mathbb{R}^n) : \{X \in B\} \in \mathcal{F}\}$. Any measurable rectangle $B_1 \times B_2 \times \cdots \times B_n$ belongs to $\mathcal{H}$ for $B_i \in \mathcal{B}(\mathbb{R})$ and $i = 1, \ldots, n$ since $\{X \in B_1 \times B_2 \times \cdots \times B_n\} = \bigcap_{i=1}^n \{X_i \in B_i\} \in \mathcal{F}$. If $B \in \mathcal{H}$, then $\mathbb{R}^n \setminus B \in \mathcal{H}$ since $\{X \in \mathbb{R}^n \setminus B\} = \Omega \setminus \{X \in B\} \in \mathcal{F}$. Finally, if $B_i \in \mathcal{H}$ for $i = 1, 2, 3, \ldots$, then $\{X \in \bigcup_{i=1}^\infty B_i\} = \bigcup_{i=1}^\infty \{X \in B_i\} \in \mathcal{F}$. This implies that $\bigcup_{i=1}^\infty B_i \in \mathcal{H}$. Hence $\mathcal{H}$ is a σ-field and $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{H}$. This proves that X is a random vector.

3.17. To verify that f_X given by (3.7) is a density, we need the following well-known result in linear algebra: For any non-singular positive definite symmetric $n \times n$ matrix Σ there is an $n \times n$ matrix A which is orthogonal (i.e. $AA^T = I$, so $A^{-1} = A^T$ and $|\det A| = 1$) and such that $B = A^{-1}\Sigma A$, where B is a diagonal matrix with $b_{ii} > 0$, $i = 1, 2, \ldots, n$ being the eigenvalues of Σ. Clearly, $\Sigma = ABA^{-1}$, $\Sigma^{-1} = AB^{-1}A^{-1}$ and $\det\Sigma = \det B$.

Now, making the substitution $Ay = x - \mu$ and remembering that $A^T = A^{-1}$, we have

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = y^T A^T A B^{-1} A^{-1} A y = y^T B^{-1} y.$$

If we put $b_{ii} = \sigma_i^2$, then $\det\Sigma = \det B = \sigma_1^2 \cdots \sigma_n^2$ and

$$B^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n^2} \end{bmatrix}.$$

We are now ready to calculate

$$\int_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n \det\Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) dm_n(x)$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^n}\sigma_1\cdots\sigma_n} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) dx_1\cdots dx_n$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^n}\sigma_1\cdots\sigma_n} \exp\left(\frac{1}{2}y^T B^{-1} y\right) dy_1\cdots dy_n$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{y_1^2}{2\sigma_1^2}} \cdots \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{y_n^2}{2\sigma_n^2}} dy_1\cdots dy_n$$
$$= \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{y_1^2}{2\sigma_1^2}} dy_1\right)\cdots\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{y_n^2}{2\sigma_n^2}} dy_n\right) = 1.$$

3.18. Let $X : \Omega \to \mathbb{R}$ be a continuous random variable with density f_X. Put $Y = X$. Suppose that the vector (X, Y) has a density $f_{X,Y}(x, y)$, i.e. that X, Y are jointly continuous. By Fubini's theorem

$$P(\{X < Y\}) = \int\int_{\mathbb{R}^2} \mathbf{1}_{\{(x,y):x<y\}} f(x,y) dm_2(x,y)$$
$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} \mathbf{1}_{\{(x,y):x<y\}} f(x,y) dm(x)\right) dm(y)$$
$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{y} f(x,y) dm(x)\right) dm(y).$$

Similarly,

$$P(\{X > Y\}) = \int\int_{\mathbb{R}^2} \mathbf{1}_{\{(x,y):x>y\}} f(x,y) dm_2(x,y)$$
$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} \mathbf{1}_{\{(x,y):x>y\}} f(x,y) dm(x)\right) dm(y)$$
$$= \int_{-\infty}^{\infty}\left(\int_{y}^{\infty} f(x,y) dm(x)\right) dm(y).$$

Now, define $B = \{(x, x) : x \in \mathbb{R}\}$. We have

$$0 = P(\{X \neq Y\}) = P(\{(X,Y) \notin B\}) = P(\{X < Y\}) + P(\{X > Y\})$$
$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{y} f(x,y) dm(x)\right) dm(y) + \int_{-\infty}^{\infty}\left(\int_{y}^{\infty} f(x,y) dm(x)\right) dm(y)$$
$$= \int\int_{\mathbb{R}^2} f(x,y) dm_2(x,y) = 1,$$

which gives a contradiction.

3.19. If $\rho = 0$ in (3.5), then by Proposition 3.20

$$f_X(x_1) = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-\frac{x_1^2+x_2^2}{2}} dm(x_2)$$
$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x_1^2}{2}} e^{-\frac{x_2^2}{2}} dx_2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}},$$

since $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} dx_2 = 1$. Similarly,

$$f_Y(x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}}.$$

Then $f_{X,Y}(x_1, x_2) = f_X(x_1) f_Y(x_2)$. For any $B_1, B_2 \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} P_{X,Y}(B_1 \times B_2) &= \int_{B_1 \times B_2} f_{X,Y}(x,y) dm_2(x,y) \\ &= \int_{B_1 \times B_2} f_X(x) f_Y(y) dm_2(x,y) \\ &= \int_{B_1} \left(\int_{B_2} f_X(x) f_Y(y) dm(y) \right) dm(x) \\ &= \left(\int_{B_1} f_X(x) dm(x) \right) \left(\int_{B_2} f_Y(y) dm(y) \right) \\ &= P(X_1 \in B_1) P(Y \in B_2) \end{aligned}$$

by Fubini's theorem. Hence X and Y are independent.

3.20. Let B be a Borel set in $\mathbb{R}$. Then

$$\begin{aligned} P_{X+Y}(B) &= P(X + Y \in B) = P((X,Y) \in \{(x,y) \in \mathbb{R}^2 : x + y \in B\}) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_{\{x+y \in B\}}(x,y) f_{X,Y}(x,y) dm_2(x,y) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_{\{x+y \in B\}}(x,y) f_X(x) f_Y(y) dm_2(x,y), \text{ by (3.9)} \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{\{x+y \in B\}}(x,y) f_X(x) f_Y(y) dm(y) \right) dm(x) \end{aligned}$$

by Fubini's theorem. Next, by the substitution $z = x + y$ and by

Exercise 1.32,

$$\begin{aligned}P_{X+Y}(B) &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} \mathbf{1}_{\{z\in B\}}(x, z-x)f_X(x)f_Y(z-x)dm(y)\right)dm(x)\\ &= \int_{\mathbb{R}}\left(\int_{B} f_X(x)f_Y(z-x)dm(z)\right)dm(x)\\ &= \int_{B}\left(\int_{\mathbb{R}} f_X(x)f_Y(z-x)dm(x)\right)dm(z)\end{aligned}$$

again by Fubini's theorem.

3.21. First we prove Theorem 3.31. We follow the arguments used in the proof of Theorem 3.27. Since intervals are Borel sets, the necessity is obvious. For sufficiency, we use induction on $n = 2, 3, \ldots$. For $n = 2$ this has been proved in Theorem 3.27. The induction hypothesis states that if

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_n)$$

for each $x_1, \ldots, x_n \in \mathbb{R}$, then $X_1, \ldots, X_n$ are independent. Suppose that for any $x_1, \ldots, x_{n+1} \in \mathbb{R}$

$$F_{X_1,\ldots,X_{n+1}}(x_1, \ldots, x_{n+1}) = F_{X_1}(x_1)\cdots F_{X_{n+1}}(x_{n+1}). \tag{S.1}$$

Let $X = (X_1, \ldots, X_n)$, $Y = X_{n+1}$. We show that for any $B \in \mathcal{B}(\mathbb{R}^n)$ and $H \in \mathcal{B}(\mathbb{R})$

$$P_{X,Y}(B \times H) = P_X(B)P_Y(H). \tag{S.2}$$

In order to prove this consider the class C of all Borel sets $A \in \mathcal{B}(\mathbb{R}^n)$ such that for each $y \in \mathbb{R}$

$$P(X \in A, Y \leq y) = P(X \in A)P(Y \leq y),$$

and the class $\mathcal{D}$ of all Borel sets $H \in \mathcal{B}(\mathbb{R})$ such that for each $A \in C$

$$P(X \in A, Y \in H) = P(X \in A)P(Y \in H).$$

Our aim is to show that C and $\mathcal{D}$ are equal to the σ-fields $\mathcal{B}(\mathbb{R}^n)$ and $\mathcal{B}(\mathbb{R})$, respectively. By the induction hypothesis, C contains the collection of all sets $(-\infty, x_1]\times\cdots\times(-\infty, x_n]$ with $x_1, \ldots, x_n \in \mathbb{R}$, and this collection is closed under intersection; we only need to check that C is a d-system, which will mean that it contains the σ-field $\mathcal{B}(\mathbb{R}^n)$. This in turn will mean that $\mathcal{D}$ contains all intervals $(-\infty, y]$, hence to show that it contains $\mathcal{B}(\mathbb{R})$ we need to show that $\mathcal{D}$ is a d-system.

Now we check that C satisfies the conditions for a d-system. The proof for $\mathcal{D}$ is almost identical. We have $\Omega \times \cdots \times \Omega \in C$ since for any $y \in \mathbb{R}$

$$P(X \in \Omega\times\cdots\times\Omega, Y \leq y) = P(Y \leq y) = P(X \in \Omega\times\cdots\times\Omega)P(Y \leq y).$$

If $A_1, A_2 \in C$ and $A_1 \subset A_2$, then

$$\begin{aligned}
&P(X \in (A_2 \setminus A_1), Y \leq y) = P(X \in A_2, Y \leq y) - P(X \in A_1, Y \leq y)\\
&= P(X \in A_2)P(Y \leq y) - P(X \in A_1)P(Y \leq y)\\
&= P(X \in (A_2 \setminus A_1))P(Y \leq y).
\end{aligned}$$

Finally, if $A_k \subset A_{k+1}$ with $A_k \in C$ for all $k = 1, 2, \ldots$ and $\bigcup_{k=1}^{\infty} A_k = A$, then

$$\begin{aligned}
&P(X \in A, Y \leq y) = P\left(\bigcup_{k=1}^{\infty}\{X \in A_k, Y \leq y\}\right)\\
&= \lim_{k\to\infty} P(X \in A_k, Y \leq y) = \lim_{k\to\infty} P(X \in A_k)P(Y \leq y)\\
&= P\left(\bigcup_{k=1}^{\infty} A_k\right) P(Y \leq y).
\end{aligned}$$

Thus C is a d-system. By Proposition 3.56, C is a σ-field containing all sets of the form $(-\infty, x_1]\times\cdots\times(-\infty, x_n]$, and so it contains $\mathcal{B}(\mathbb{R}^n)$. This ends the proof of (S.2).

Further, observe that from (S.1) it follows that

$$\begin{aligned}
&F_X(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n, X_{n+1} \in \mathbb{R})\\
&= P\left(X_1 \leq x_1, \ldots, X_n \leq x_n, X_{n+1} \in \bigcup_{k=1}^{\infty}(-\infty, k]\right)\\
&= \lim_{k\to\infty} P(X_1 \leq x_1, \ldots, X_n \leq x_n, X_{n+1} \leq k)\\
&= \lim_{k\to\infty} F_{X_1}(x_1)\cdots F_{X_n}(x_n)F_{X_{n+1}}(k)\\
&= F_{X_1}(x_1)\cdots F_{X_n}(x_n) \lim_{k\to\infty} F_{X_{n+1}}(k)\\
&= F_{X_1}(x_1)\cdots F_{X_n}(x_n).
\end{aligned}$$

Substituting $B = B_1 \times \cdots \times B_n$ with $B_i \in \mathcal{B}(\mathbb{R})$, $i = 1, 2 \ldots, n$ into (S.2), we have

$$P_{X,Y}(B_1 \times \cdots \times B_n \times H) = P_X(B_1 \times \cdots \times B_n)P_Y(H).$$

By the induction hypothesis, we obtain

$$P_{X_1,\ldots,X_{n+1}}(B_1 \times \cdots \times B_n \times H) = P_{X_1}(B_1)\cdots P_{X_n}(B_n)P_{X_{n+1}}(H),$$

so $X_1, \dots, X_{n+1}$ are independent.

We turn to the proof of Theorem 3.32. If the random vector $X = (X_1, \dots, X_n)$ has joint density f_X, then each X_i has a density of the form

$$f_{X_i}(x) = \int_{\mathbb{R}^{n-1}} f_X(x', x, x'')dm_{n-1}(x', x'')$$

for all $x \in \mathbb{R}$. Here $(x', x'') \in \mathbb{R}^{i-1} \times \mathbb{R}^{n-i}$ has been identified with a point in $\mathbb{R}^{n-1}$, and $(x', x, x'') \in \mathbb{R}^{i-1} \times \mathbb{R} \times \mathbb{R}^{n-i}$ with a point in $\mathbb{R}^n$. For any $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ we have

$$P_X(B_1 \times \cdots \times B_n) = \int_{B_1 \times \cdots \times B_n} f_X(x)dm_n(x),$$

while

$$\begin{aligned}
P_{X_1}(B_1) \cdots P_{X_n}(B_n) &= \left(\int_{B_1} f_{X_1}(x_1)dm(x_1)\right) \cdots \left(\int_{B_n} f_{X_n}(x_n)dm(x_n)\right) \\
&= \left(\int_{B_1} f_{X_1}(x_1)dm(x_1)\right) \cdots \left(\int_{B_{n-2}} f_{X_{n-2}}(x_{n-2})dm(x_{n-2})\right) \\
&\qquad \left(\int_{B_{n-1}} \left(\int_{B_n} f_{X_{n-1}}(x_{n-1}) f_{X_n}(x_n)dm(x_n)\right) dm(x_{n-1})\right) \\
&= \left(\int_{B_1} f_{X_1}(x_1)dm(x_1)\right) \cdots \left(\int_{B_{n-2}} f_{X_{n-2}}(x_{n-2})dm(x_{n-2})\right) \\
&\qquad \int_{B_{n-2} \times B_n} f_{X_{n-1}}(x_{n-1}) f_{X_n}(x_n)dm_2(x_{n-1}, x_n) \\
&= \int_{B_1 \times \cdots \times B_n} f_{X_1}(x_1) \cdots f_{X_n}(x_n)dm_n(x_1, \dots, x_n)
\end{aligned}$$

by Fubini's theorem. If $f_X(x) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$, then

$$P_X(B_1 \times \cdots \times B_n) = P_{X_1}(B_1) \cdots P_{X_n}(B_n)$$

Conversely, if

$$P_X(B_1 \times \cdots \times B_n) = P_{X_1}(B_1) \cdots P_{X_n}(B_n)$$

then, proceeding as above, we have

$$\int_{B_1 \times \cdots \times B_n} f_X(x)dm_n(x) = \int_{B_1 \times \cdots \times B_n} f_{X_1}(x_1) \cdots f_{X_n}(x_n)dm_n(x_1, \dots, x_n)$$

for any Borel sets $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$. It follows from Lemma 3.58

that

$$\int_B f_X(x)dm_n(x) = \int_B f_{X_1}(x_1)\cdots f_{X_n}(x_n)dm_n(x_1,\ldots,x_n)$$

for any Borel set $B \in \mathcal{B}(\mathbb{R}^n)$ because, by Theorem 1.35, the integrals on both sides of the last equality are measures when regarded as functions of $B \in \mathcal{B}(\mathbb{R}^n)$. By Exercise 1.30, this implies $f_X(x) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$, m_n-a.e.

3.22. Suppose that the random vector $X = (X_1,\ldots,X_n)$ has joint density (3.7). Then each X_i has density

$$f_{X_i}(x) = \int_{\mathbb{R}^{n-1}} f_X(x',x,x'')dm_{n-1}(x',x'')$$

for any $x \in \mathbb{R}$, where $(x',x'') \in \mathbb{R}^{i-1}\times\mathbb{R}^{n-i}$ is identified with a point in $\mathbb{R}^{n-1}$, and $(x',x,x'') \in \mathbb{R}^{i-1}\times\mathbb{R}\times\mathbb{R}^{n-i}$ with a point in $\mathbb{R}^n$. By Exercise 2.16, the expectation of X_i can be computed as

$$\mathbb{E}(X_i) = \int_{\mathbb{R}} x_i f_{X_i}(x_i)dm(x_i).$$

Then, by Fubini's theorem we have

$$\begin{aligned}
\mathbb{E}(X_i) &= \int_{\mathbb{R}}\left(x_i\int_{\mathbb{R}^{n-1}} f_X(x',x_i,x'')dm_{n-1}(x',x'')\right)dm(x_i)\\
&= \int_{\mathbb{R}^n} x_i\frac{1}{(\sqrt{2\pi})^n\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dm_n(x)\\
&= \int_{\mathbb{R}^n}(x_i-\mu_i)\frac{1}{(\sqrt{2\pi})^n\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right)dm_n(x).
\end{aligned}$$

We recall (see the solution to Exercise 3.17) that there is an orthogonal $n\times n$ matrix $A = [a_{ij}]$ such that $B = A^{-1}\Sigma A$ is a diagonal matrix, B^{-1} has the form

$$B^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n^2}\end{bmatrix},$$

and $\det\Sigma = \det B = \sigma_1\cdots\sigma_n$. Now, making the substitution $Ay+\mu =$

x, we have

$$\mathbb{E}(X_i) = \int_{\mathbb{R}^n} \left(\sum_{j=1}^{n} a_{ij} y_j + \mu_i \right) \frac{1}{\sqrt{(2\pi)^n} \sigma_1 \cdots \sigma_n} \exp\left(-\frac{1}{2} y^T B^{-1} y \right) dm_n(y)$$

$$= \sum_{j=1}^{n} a_{ij} \int_{\mathbb{R}^n} y_i \frac{1}{\sqrt{(2\pi)^n} \sigma_1 \cdots \sigma_n} \exp\left(-\frac{1}{2} y^T B^{-1} y \right) dm_n(y)$$

$$+\mu_i \int_{\mathbb{R}^n} \frac{1}{(\sqrt{(2\pi)^n} \sigma_1 \cdots \sigma_n} \exp\left(-\frac{1}{2} y^T B^{-1} y \right) dm_n(y) = \mu_i$$

since by Exercise 3.17

$$\int_{\mathbb{R}^n} \frac{1}{(\sqrt{2\pi})^n \sigma_1 \cdots \sigma_n} \exp\left(-\frac{1}{2} y^T B^{-1} y \right) dm_n(y) = 1$$

and for any $i = 1, 2, \ldots, n$

$$\int_{\mathbb{R}^n} y_i \frac{1}{(\sqrt{2\pi})^n \sigma_1 \cdots \sigma_n} \exp\left(-\frac{1}{2} y^T B^{-1} y \right) dm_n(y)$$

$$= \int_{\mathbb{R}} y_i \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{y_i^2}{2\sigma_i^2}} dm(y_i) \prod_{k=1, k \neq i}^{n} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi} \sigma_k} e^{-\frac{y_k^2}{2\sigma_k^2}} dm(y_k) = 0.$$

Here we have used Fubini's theorem and the two equalities

$$\int_{\mathbb{R}} y_i \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{y_i^2}{2\sigma_i^2}} dm(y_i) = 0,$$

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi} \sigma_k} e^{-\frac{y_k^2}{2\sigma_k^2}} dm(y_k) = 1.$$

If Σ is a diagonal matrix, then $\Sigma = B$, and it follows that

$$f_X(x) = \frac{1}{(\sqrt{(2\pi)^n}) \sigma_1 \cdots \sigma_n} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \right)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

By Exercise 3.21, this means that $X_1, \ldots, X_n$ are independent.

3.23. Suppose that $A_1, A_2 \in \mathcal{F}$ are independent. Then

$$P(A_1 \cap (\Omega \setminus A_2)) = P(A_1 \setminus (A_1 \cap A_2))$$

$$= P(A_1) - P(A_1 \cap A_2) = P(A_1) - P(A_1)P(A_2)$$

$$= P(A_1)(1 - P(A_2)) = P(A_1)P(\Omega \setminus A_2).$$

So A_1 and $\Omega \setminus A_2$ are independent. Conversely, if A_1 and $\Omega \setminus A_2$ are independent, then

$$\begin{aligned}P(A_1 \cap A_2) &= P(A_1 \setminus (A_1 \cap (\Omega \setminus A_2)))\\ &= P(A_1) - P(A_1 \cap (\Omega \setminus A_2)) = P(A_1) - P(A_1)P(\Omega \setminus A_2)\\ &= P(A_1)(1 - P(\Omega \setminus A_2)) = P(A_1)P(A_2).\end{aligned}$$

By symmetry, $\Omega \setminus A_1$, A_2 are independent and $\Omega \setminus A_1$, $\Omega \setminus A_2$ are independent.

3.24. For any events $A_1, A_2 \in \mathcal{F}$ and for every choice of Borel sets $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ we have

$$\{\mathbf{1}_{A_i} \in B_i\} = \begin{cases} A_i & \text{if } 0 \notin B_i \text{ and } 1 \in B_i,\\ \Omega \setminus A_i & \text{if } 0 \in B_i \text{ and } 1 \notin B_i,\\ \Omega & \text{if } 0 \in B_i \text{ and } 1 \in B_i,\\ \varnothing & \text{otherwise,}\end{cases}$$

so $\{\mathbf{1}_{A_i} \in B_1\} \in \{A_i, \Omega \setminus A_i, \Omega, \varnothing\}$ for $i = 1, 2$.

If A_1, A_2 are independent events, from the solution of Exercise 3.23 we therefore know that $\{\mathbf{1}_{A_1} \in B_1\}, \{\mathbf{1}_{A_2} \in B_2\}$ are independent events, and so

$$\begin{aligned}P_{\mathbf{1}_{A_1},\mathbf{1}_{A_2}}(B_1 \times B_2) &= P(\mathbf{1}_{A_1} \in B_1, \mathbf{1}_{A_2} \in B_2)\\ &= P(\mathbf{1}_{A_1} \in B_1)P(\mathbf{1}_{A_2} \in B_2) = P_{\mathbf{1}_{A_1}}(B_1)P_{\mathbf{1}_{A_2}}(B_2)\end{aligned}$$

for any Borel sets $B_1, B_2 \in \mathcal{B}(\mathbb{R})$. This means that $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}$ are independent random variables.

Conversely, suppose that $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}$ are independent random variables. Taking $B_1 = \{1\}$ and $B_2 = \{1\}$, we get

$$\begin{aligned}P(A_1 \cap A_2) &= P(\mathbf{1}_{A_1} \in B_1, \mathbf{1}_{A_2} \in B_2)\\ &= P_{\mathbf{1}_{A_1},\mathbf{1}_{A_2}}(B_1 \times B_2) = P_{\mathbf{1}_{A_1}}(B_1)P_{\mathbf{1}_{A_2}}(B_2)\\ &= P(\mathbf{1}_{A_1} \in B_1)P(\mathbf{1}_{A_2} \in B_2) = P(A_1)P(A_2),\end{aligned}$$

so A_1, A_2 are independent events.

3.25. Consider the following closed intervals: $A = [\frac{1}{8}, \frac{5}{8}]$, $B = [\frac{1}{4}, \frac{3}{4}]$, $C = [\frac{1}{2}, 1]$. We have

$$P(A \cap B \cap C) = \frac{1}{8} = P(A)P(B)P(C),$$

but A, B are not independent since $P(A \cap B) = \frac{3}{8} \neq P(A)P(B) = \frac{1}{4}$.

Let us now consider the following subsets of $[0, 1]$:

$$C = \left[0, \frac{1}{8}\right] \cup \left[\frac{1}{4}, \frac{3}{8}\right] \cup \left[\frac{1}{2}, \frac{5}{8}\right] \cup \left[\frac{3}{4}, \frac{7}{8}\right],$$
$$D = \left[\frac{1}{8}, \frac{3}{8}\right] \cup \left[\frac{1}{2}, \frac{5}{8}\right] \cup \left[\frac{7}{8}, 1\right],$$
$$E = \left[0, \frac{1}{4}\right] \cup \left[\frac{3}{4}, 1\right].$$

Then C, D, E are pairwise independent, but (3.12) fails.

3.26. Consider the events A, B in Example 3.34 and a third event C defined as follows:

$$\begin{aligned} C = \{&(1,2),(2,1),(3,2),(4,1),(5,2),(6,1),\\ &(1,4),(2,3),(3,4),(4,3),(5,4),(6,3),\\ &(1,6),(2,5),(3,6),(4,5),(5,6),(6,5)\}. \end{aligned}$$

Then each pair of these events is independent, but (3.12) fails since $A \cap B \cap C = \varnothing$.

3.27. Since $A \cup B = (A \setminus (A \cap B)) \cup B$ and $(A \setminus (A \cap B)) \cap B = \varnothing$, we have

$$P(A \cup B) = P(A \setminus (A \cap B)) + P(B) = P(A) - P(A \cap B) + P(B)$$

for any events A, B.

When A, B, C are independent events, we use this property to obtain

$$\begin{aligned} &P(A \cup B) \cap C) = P((A \cap C) \cup (B \cap C)) \\ &= P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) \\ &= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C) \\ &= P(C)(P(A) + P(B) - P(A)P(B)) \\ &= P(C)(P(A) + P(B) - P(A \cap B)) = P(C)P(A \cup B). \end{aligned}$$

Thus, $A \cup B$ and C are independent.

3.28. For any events A, B we know from Exercise 3.23 that A, B are independent if and only if $A, \Omega \setminus B$ are independent. Moreover, if A is an event, then $A, \varnothing$ are independent. It follows that if $A_1, \ldots, A_n$ are independent events and $C_i \in \{\varnothing, A_i, \Omega \setminus A_i, \Omega\}$ for $i = 1, \ldots, n$, then $C_1, \ldots, C_n$ are independent.

Now suppose that $A_1, \ldots, A_n$ are independent events. Then for any Borel sets $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ the events $\{\mathbf{1}_{A_1} \in B_1\}, \ldots, \{\mathbf{1}_{A_n} \in B_n\}$

are also independent because $\{\mathbf{1}_{A_i} \in B_i\} \in \{\varnothing, A_i, \Omega \setminus A_i, \Omega\}$ for each $i = 1, \ldots, n$. This, in turn, means that

$$\begin{aligned} P_{\mathbf{1}_{A_1},\ldots,\mathbf{1}_{A_n}}(B_1 \times \cdots \times B_n) &= P(\mathbf{1}_{A_1} \in B_1, \ldots, \mathbf{1}_{A_n} \in B_n) \\ &= P(\mathbf{1}_{A_1} \in B_1) \cdots P(\mathbf{1}_{A_n} \in B_n) \\ &= P_{\mathbf{1}_{A_1}}(B_1) \cdots P_{\mathbf{1}_{A_n}}(B_n). \end{aligned}$$

We have shown that $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n}$ are independent random variables.

Conversely, suppose that $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n}$ are independent random variables. For any subsequence $i_1, \ldots, i_k$ of the sequence $1, \ldots, n$ we take

$$B_i = \begin{cases} \{1\} & \text{if } i \in \{i_1, \ldots, i_k\} \\ \mathbb{R} & \text{if } i \notin \{i_1, \ldots, i_k\} \end{cases}$$

so that

$$\{\mathbf{1}_{A_i} \in B_i\} = \begin{cases} A_i & \text{if } i \in \{i_1, \ldots, i_k\} \\ \Omega & \text{if } i \notin \{i_1, \ldots, i_k\} \end{cases}$$

for each $i = 1, \ldots, n$. Then

$$\begin{aligned} P(A_{i_1} \cap \cdots \cap A_{i_k}) &= P(\mathbf{1}_{A_1} \in B_1, \ldots, \mathbf{1}_{A_n} \in B_n) \\ &= P_{\mathbf{1}_{A_1},\ldots,\mathbf{1}_{A_n}}(B_1 \times \cdots \times B_n) \\ &= P_{\mathbf{1}_{A_1}}(B_1) \cdots P_{\mathbf{1}_{A_n}}(B_n) \\ &= P(\mathbf{1}_{A_1} \in B_1) \cdots P(\mathbf{1}_{A_n} \in B_n) \\ &= P(A_{i_1}) \cdots P(A_{i_k}). \end{aligned}$$

This proves that $A_1, \ldots, A_n$ are independent events.

3.29. By Exercise 3.23, A, B are independent events if and only if C, D are independent for any $C \in \{\varnothing, A, \Omega \setminus A, \Omega\}$ and $D \in \{\varnothing, B, \Omega \setminus B, \Omega\}$, which in turn means that $\{\varnothing, A, \Omega \setminus A, \Omega\}$ and $\{\varnothing, B, \Omega \setminus B, \Omega\}$ are independent σ-fields.

3.30. If a σ-field $\mathcal{G}$ is independent of itself, it means that any event $A \in \mathcal{G}$ is independent of itself, so $P(A) = P(A \cap A) = P(A)P(A)$. It follows that $P(A)(1 - P(A)) = 0$, which means that $P(A) = 1$ or 0.

3.31. First suppose that the random variables $X_1, \ldots, X_n$ are independent. For any choice $A_1 \in \sigma(X_1), \ldots, A_n \in \sigma(X_n)$ we have $A_1 = \{X_1 \in B_1\}, \ldots, A_n = \{X_n \in B_n\}$ for some Borel sets $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$. For any subsequence $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ we define $C_i = B_i$ for

$i \in \{i_1, \ldots, i_k\}$ and $C_i = \mathbb{R}$ for $i \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$. Then

$$\begin{aligned} &P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(X_{i_1} \in B_{i_1}, \ldots, X_{i_k} \in B_{i_k}) \\ &= P(X_1 \in C_1, \ldots, X_n \in C_n) = P(X_1 \in C_1) \cdots P(X_n \in C_n) \\ &= P(X_{i_1} \in B_{i_1}) \cdots P(X_{i_k} \in B_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}), \end{aligned}$$

so that $A_1, \ldots, A_n$ are independent events. This means that the σ-fields $\sigma(X_1), \ldots, \sigma(X_n)$ are independent.

Conversely, suppose that the σ-fields $\sigma(X_1), \ldots, \sigma(X_n)$ are independent. For any choice of Borel sets $B_1, B_2, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned} P(X_1 \in B_1, \ldots, X_n \in B_n) &= P(X_1 \in B_1, \ldots, X_n \in B_n) \\ &= P(X_1 \in B_1) \cdots P(X_n \in B_n) \end{aligned}$$

since $\{X_1 \in B_1\} \in \sigma(X_1), \ldots, \{X_n \in B_1\} \in \sigma(X_n)$. It follows that $X_1, \ldots, X_n$ are independent random variables.

3.32. Note that (see the solution of Exercise 3.28) $A_1, \ldots, A_n$ are independent events if and only if $B_1, \ldots, B_n$ are independent for any choice of $B_i \in \mathcal{G}_i$, where $\mathcal{G}_i = \{\emptyset, A_i, \Omega \setminus A_i, \Omega\}$ for $i = 1, \ldots, n$, which in turn is equivalent to $\mathcal{G}_1, \ldots, \mathcal{G}_n$ being independent σ-fields.

3.33. Define the family of sets

$$\mathcal{G} = \{B \in \mathcal{B}(\mathbb{R}^n) : \{X \in B\}, D \text{ are independent for each } D \in \sigma(Y)\}.$$

We have $\mathbb{R}^n \in \mathcal{G}$ because $\{X \in \mathbb{R}^n\} = \Omega$ and Ω, D are independent for each $D \in \sigma(Y)$. Moreover, if $B \in \mathcal{G}$, then $\mathbb{R}^n \setminus B \in \mathcal{G}$ by Exercise 3.23. Finally, if $B_i \in \mathcal{G}$ for $i = 1, 2, \ldots$ and the B_i are pairwise disjoint, then

$$\begin{aligned} P\left(\left(\bigcup_{i=1}^{\infty} \{X \in B_i\}\right) \cap D\right) &= P\left(\bigcup_{i=1}^{\infty} (\{X \in B_i\} \cap D)\right) \\ &= \sum_{i=1}^{\infty} P(\{X \in B_i\} \cap D) \\ &= \sum_{i=1}^{\infty} P(X \in B_i)P(D) = P\left(\bigcup_{i=1}^{\infty} \{X \in B_i\}\right) P(D). \end{aligned}$$

This means that $\mathcal{G}$ is a σ-field on $\mathbb{R}^n$. In addition, since $X_1, \ldots, X_n, Y$ are independent, and so $\{X \in B_1 \times \cdots \times B_n\}$ and D are independent for any $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ and for any $D \in \sigma(Y)$, it follows that $B_1 \times \cdots \times B_n \in \mathcal{G}$ for any $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$. This proves that $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{G}$, which, in turn, implies that Y is independent of the σ-field $\sigma(X)$ generated by the random vector $X = (X_1, \ldots, X_n)$.

3.34. We prove the assertion by induction on n. By Theorem 3.42, for $n = 2$ we know that X_1X_2 is integrable and $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$. Suppose that for some $n = 2, 3, \ldots$ the product $\prod_{i=1}^n X_i$ is integrable and

$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i).$$

Moreover, suppose that $X_1, \ldots, X_{n+1}$ are independent random variables. By Exercise 3.33, X_{n+1} is independent of the σ-field $\sigma(X)$ generated by the random vector $X = (X_1, \ldots, X_n)$, which by definition consists of all events of the form $\{X \in B\}$ with $B \in \mathcal{B}(\mathbb{R}^n)$.

Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x_1, \ldots, x_n) = x_1 \cdots x_n$. Since f is continuous, it is a Borel function. The function $f(X_1, \ldots, X_n) = X_1 \cdots X_n$ is measurable with respect to the σ-field $\sigma(X)$ since $\{f \in B\} \in \mathcal{B}(\mathbb{R}^n)$ for any $B \in \mathcal{B}(\mathbb{R})$ and

$$\{f(X_1, \ldots, X_n) \in B\} = \{(X_1, \ldots, X_n) \in \{f \in B\}\} \in \sigma(X).$$

It follows that the random variables $f(X_1, \ldots, X_n) = X_1 \cdots X_n$ and X_{n+1} are independent. Applying Theorem 3.42 and the induction hypothesis, we can see that $f(X_1, \ldots, X_n)X_{n+1} = X_1 \cdots X_n X_{n+1}$ is integrable and

$$\begin{aligned}
&\mathbb{E}(f(X_1, \ldots, X_n)X_{n+1}) = \mathbb{E}(X_1 \cdots X_n X_{n+1}) \\
&= \mathbb{E}\left(\prod_{i=1}^n X_i\right)\mathbb{E}(X_{n+1}) = \prod_{i=1}^{n+1} \mathbb{E}(X_i).
\end{aligned}$$

3.35. Suppose that $a, b \neq 0$. If X, Y are independent random variables, then so are aX, bY. Suppose that $B_1, B_2 \in \mathcal{B}(\mathbb{R})$. Then

$$\begin{aligned}
&P((aX, bY) \in B_1 \times B_2) = P(aX \in B_1, bY \in B_2) \\
&= P(X \in \frac{1}{a}B_1, Y \in \frac{1}{b}B_2) = P(X \in \frac{1}{a}B_1)P(Y \in \frac{1}{b}B_2) \\
&= P(aX \in B_1)P(bY \in B_2).
\end{aligned}$$

It follows that aX, bY are independent. By Corollary 3.45 and Exercise 2.25, we have

$$\begin{aligned}
&\phi_{aX+bY}(t) = \phi_{aX}(t)\phi_{bY}(t) = \phi_X(at)\phi_Y(bt) \\
&= e^{-\frac{1}{2}a^2t^2}e^{-\frac{1}{2}b^2t^2} = e^{-\frac{1}{2}(a^2+b^2)t^2}.
\end{aligned}$$

The case when $a = 0$ or $b = 0$ is trivial.

3.36. We prove the statement by induction. By Proposition 3.46, for $n = 2$

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

if X_1, X_2 are independent integrable random variables.

Suppose that for some $n = 2, 3, \ldots$

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$$

if $X_1, \ldots, X_n$ are independent integrable random variables. Also suppose that $X_1, \ldots X_n, X_{n+1}$ are independent integrable random variables. By Exercise 3.33, X_{n+1} is independent of the σ-field $\sigma(X)$ generated by the random vector $X = (X_1, \ldots, X_n)$. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x_1, \ldots, x_n) = x_1 + \cdots + x_n$. Since the function is continuous, it is a Borel function. Moreover, the function $f(X_1, \ldots, X_n) = X_1 + \ldots + X_n$ is measurable with respect to σ-field $\sigma(X)$ and integrable. It follows that $X_1 + \cdots + X_n$ and X_{n+1} are independent random variables. By Proposition 3.46 and the induction hypothesis,

$$\begin{aligned}&\mathrm{Var}((X_1 + \cdots + X_n) + X_{n+1}) = \mathrm{Var}(X_1 + \cdots + X_n) + \mathrm{Var}(X_{n+1})\\&= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_{n+1}).\end{aligned}$$

3.37. Suppose that X, Y have the bivariate normal distribution with density (3.5). To compute $\mathrm{Cov}(X, Y)$ it suffices to compute $\mathbb{E}(XY)$ since, by Exercise 3.12, $\mathbb{E}(X) = \mathbb{E}(Y) = 0$.

In order to compute $\mathbb{E}(XY)$ we extend (2.3) to random vectors. Suppose that $X = (X_1, \ldots, X_n)$ is a random vector whose joint distribution P_X has density f_X, and $g : \mathbb{R}^n \to \mathbb{R}$ is an integrable function with respect to P_X. Then

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^n} g(x) f_X(x) dm_n(x). \tag{S.3}$$

For $g = \mathbf{1}_B(x)$ with $B \in \mathcal{B}(\mathbb{R}^n)$ we have

$$\mathbb{E}(\mathbf{1}_B(X)) = P(X \in B) = \int_B f_X(x) dm_n(x) = \int_{\mathbb{R}^n} \mathbf{1}_B(x) f_X(x) dm_n(x).$$

If $g = \sum_{i=1}^n a_i \mathbf{1}_{B_i}$ with $B_i \in \mathcal{B}(\mathbb{R}^n)$ and $a_i \in \mathbb{R}$ for $i = 1, \ldots, n$, then (S.3) follows by linearity. When g is a non-negative measurable function, by Proposition 1.28 there is a non-decreasing sequence of non-negative simple functions s_n, $n = 1, 2, \ldots$, such that $\lim_{n\to\infty} s_n =$

g. By the monotone convergence theorem it follows that

$$\int_\Omega g(X(\omega))dP(\omega) = \lim_{n\to\infty}\int_\Omega s_n(X(\omega))dP(\omega)$$
$$= \lim_{n\to\infty}\int_{\mathbb{R}^n} s_n(x)f_X(x)dm_n(x) = \int_{\mathbb{R}^n} g(x)f_X(x)dm_n(x).$$

If g is an integrable function, then (S.3) follows from the result for g^+ and g^-.

Now put $g(x_1, x_2) = x_1x_2$. Then

$$\mathbb{E}(XY) = \int_{\mathbb{R}^n} x_1x_2 f_{X,Y}(x_1, x_2)dm_2(x_1, x_2)$$
$$= \int_{\mathbb{R}^n} x_1x_2 \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1x_2 + x_2^2}{2(1-\rho^2)}\right)dx_1dx_2$$
$$= \int_{\mathbb{R}^n} x_1x_2 \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2}{2} - \frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right)dx_1dx_2$$
$$= \int_{\mathbb{R}}\left(\frac{x_1}{\sqrt{2\pi}}x_1 \exp\left(-\frac{x_1^2}{2}\right)\int_{\mathbb{R}} \frac{x_2}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right)dx_2\right)dx_1$$

by Fubini's theorem. To compute

$$I = \int_{\mathbb{R}} \frac{x_2}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right)dx_2$$

make the substitution $z = x_2 - \rho x_1$. Then

$$I = \frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{\mathbb{R}}(z+\rho x_1)\exp\left(-\frac{z^2}{2(1-\rho^2)}\right)dz$$
$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{\mathbb{R}} z\exp\left(-\frac{z^2}{2(1-\rho^2)}\right)dz$$
$$+\frac{\rho x_1}{\sqrt{2\pi(1-\rho^2)}}\int_{\mathbb{R}} \exp\left(-\frac{z^2}{2(1-\rho^2)}\right)dz = \rho x_1$$

since

$$\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{\mathbb{R}} z\exp\left(-\frac{z^2}{2(1-\rho^2)}\right)dz = 0,$$
$$\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{\mathbb{R}} \exp\left(-\frac{z^2}{2(1-\rho^2)}\right)dz = 1.$$

Finally,

$$\mathbb{E}(XY) = \rho \int_{\mathbb{R}} x_1^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right) dx_1 = \rho.$$

Hence $\mathrm{Cov}(X, Y) = \rho$

3.38. By Exercise 3.37, $\mathrm{Cov}(X, Y) = \rho$ for random variables X, Y with density (3.5). By Exercise 3.12, f_X and f_Y are standard normal densities. Hence

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \rho.$$

If X and Y are uncorrelated, then $\rho_{X,Y} = \rho = 0$. By Exercise 3.19, they are independent.

3.39. Consider a quadratic expression

$$\eta(t) = \mathbb{E}(X_0 + tY_0)^2 = \mathbb{E}(X_0^2) + 2t\mathbb{E}(X_0 Y_0) + t^2 \mathbb{E}(Y_0^2)$$

for $t \in \mathbb{R}$, where $X_0 = X - \mathbb{E}(X)$ and $Y_0 = Y - \mathbb{E}(Y)$. Since $\eta(t) \geq 0$ for each $t \in \mathbb{R}$, the discriminant satisfies

$$\Delta = 4(\mathbb{E}(X_0 Y_0))^2 - 4\mathbb{E}(X_0^2)\mathbb{E}(Y_0^2) \leq 0.$$

Moreover, $\Delta = 0$ if and only if there is $t_0 \in \mathbb{R}$ such that $0 = \eta(t_0) = \mathbb{E}((X_0 + tY_0)^2)$. By Proposition 1.36, this is equivalent to $P(X_0 + t_0 Y_0 = 0) = 1$. On the other hand $|\rho_{X,Y}| = 1$ if and only if $\Delta = 0$. It follows that

$$P(X - \mathbb{E}(X) + t_0(Y - \mathbb{E}(Y)) = 0) = 1.$$

In other words, $X = aY + b$, P-a.e. for some $a, b \in \mathbb{R}$.

3.40. When $\mathcal{F}_1, \mathcal{F}_2$ are σ-fields on Ω_1, Ω_2, respectively, the family of measurable rectangles is defined as

$$\mathcal{R} = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}.$$

Suppose that $A_1 \times A_2, B_1 \times B_2 \in \mathcal{R}$ for some $A_1, B_1 \in \mathcal{F}_1$ and $A_2, B_2 \in \mathcal{F}_2$. Then

$$(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2),$$

where $A_1 \cap B_1 \in \mathcal{F}_1$ and $A_2 \cap B_2 \in \mathcal{F}_2$, so $(A_1 \times A_2) \cap (B_1 \times B_2) \in \mathcal{R}$. This means that $\mathcal{R}$ is closed under intersection.

3.41. This is a very simple consequence of Proposition 3.56 since the family I of open intervals in $\mathbb{R}$ is closed under intersection.

3.42. Let $\mathcal{D}$ be the family of all Borel sets $A \subset [0,1]$ such that

$$\int_A X\,dm = 0.$$

Observe that $\mathcal{D}$ is a d-system. Also denote by $\mathcal{C}$ the family of intervals $\left(\frac{i}{2^n}, \frac{j}{2^n}\right]$ such that $n = 0, 1, \ldots$ and $i, j = 0, 1, \ldots, 2^n$ with $i \leq j$, and observe that $\mathcal{C}$ is closed under intersections. It follows by Proposition 3.56 that $d(\mathcal{C}) = \sigma(\mathcal{C})$. Moreover, since $\mathcal{C} \subset \mathcal{D}$ and $\sigma(\mathcal{C})$ is the σ-field of Borel sets on $[0,1]$, we have $\mathcal{D} \subset \sigma(\mathcal{C}) = d(\mathcal{C}) \subset \mathcal{D}$, so $\mathcal{D}$ contains all Borel subsets in $[0,1]$.

We have shown that $\int_A X\,dm = 0$ for each Borel set $A \subset [0,1]$. For each $n = 1, 2, \ldots$ let $A_n = \left\{X \geq \frac{1}{n}\right\}$. Then

$$0 = \int_{A_n} X\,dm \geq \frac{1}{n} m(A_n),$$

which means that $m(A_n) = 0$. Since $\{X > 0\} = \bigcup_{n=1}^{\infty} A_n$ and $A_n \subset A_{n+1}$ for each n, it follows that $m(\{X > 0\}) = \lim_{n\to\infty} m(A_n) = 0$. In the same manner we can show that $m(\{X < 0\}) = 0$, and therefore deduce that $X = 0$, m-a.s.

Chapter 4

4.1. Suppose that $\mathcal{F}$ is the family of all possible countable unions of sets belonging to $\mathcal{P} = \{B_1, B_2, \ldots\}$. (Recall that 'countable' means finite or countably infinite.) Clearly, $\Omega \in \mathcal{F}$. If $A \in \mathcal{F}$, then $A = \bigcup_{i\in I} B_i$ for some $I \subset \mathbb{N}$ and $\Omega \setminus A = \bigcup_{i\in\mathbb{N}\setminus I} B_i$, so $\Omega \setminus A \in \mathcal{F}$. Further, if $A_k \in \mathcal{F}$ for $k = 1, 2, \ldots$, then $A_k = \bigcup_{i\in I_k} B_i$, for some $I_k \subset \mathbb{N}$ and $A = \bigcup_{k=1}^{\infty} A_k = \bigcup_{i\in\bigcup_{k=1}^{\infty} I_k} B_i \in \mathcal{F}$. Hence $\mathcal{F}$ is a σ-field. This implies that $\sigma(\mathcal{P}) \subset \mathcal{F}$.

Now suppose that $A \in \mathcal{F}$, but $A \notin \sigma(\mathcal{P})$. Then $\Omega \setminus A \notin \sigma(\mathcal{P})$. This implies that $\Omega \notin \sigma(\mathcal{P})$, a contradiction. So $\mathcal{F} \subset \sigma(\mathcal{P})$.

4.2. Suppose that the family $\mathcal{A} = \{A_1, A_2, \ldots\}$ of all atoms in $\mathcal{F}$ is a partition of Ω. Of course, $\sigma(\mathcal{A}) \subset \mathcal{F}$. To prove the inverse inclusion take any $A \in \mathcal{F}$. For every $A_i \in \mathcal{A}$ such that $A_i \cap A \neq \emptyset$ we have $A_i \subset A$. Otherwise $A_i = (A_i \cap A) \cup (A_i \cap (\Omega \setminus A))$ and $A_i \cap A \neq \emptyset$, $A_i \cap (\Omega \setminus A) \neq \emptyset$, which is impossible by the definition of an atom. Since $A \subset \bigcup_{i=1}^{\infty} A_i = \Omega$, it follows that A is the union of some atoms, $A = \bigcup_{i\in I} A_i$, where $I \subset \{1, 2, \ldots\}$ and $A_i \subset A$ for each $i \in I$. This implies that $A \in \sigma(\mathcal{A})$ and $\mathcal{F} \subset \sigma(\mathcal{A})$.

4.3. Suppose that $P(B) > 0$ and let $\mathcal{F}_B$ be the σ-field consisting of all events $A \in \mathcal{F}$ such that $A \subset B$. We show that P_B is a probability on $\mathcal{F}_B$. First, $P_B(B) = \frac{P(B \cap B)}{P(B)} = 1$. Let $A_i \in \mathcal{F}_B$ be a sequence of pairwise disjoint events. Then

$$P_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{P(\bigcup_{i=1}^{\infty} A_i)}{P(B)}$$

$$= \frac{1}{P(B)} \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} \frac{P(A_i)}{P(B)} = \sum_{i=1}^{\infty} P_B(A_i).$$

4.4. Suppose that X has the Poisson distribution with parameter λ. First we calculate

$$P(X \text{ is odd}) = 1 - P(X \text{ is even})\}) = 1 - P\left(\bigcup_{k=0}^{\infty} \{X = 2k\}\right)$$

$$= 1 - \sum_{k=0}^{\infty} P(\{X = 2k\}) = 1 - \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} e^{-\lambda} = 1 - e^{-\lambda} \cosh \lambda.$$

By (4.1), we have

$$\mathbb{E}(X | \{X \text{ is odd}\}) = \frac{1}{P(X \text{ is odd})} \mathbb{E}(\mathbf{1}_{\{X \text{ is odd}\}} X)$$

$$= \frac{1}{(1 - e^{-\lambda} \cosh \lambda)} \sum_{k=0}^{\infty} (2k+1) \frac{\lambda^{2k+1} e^{-\lambda}}{(2k+1)!}$$

$$= \frac{\lambda}{e^{\lambda} - \cosh \lambda} \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} = \frac{\lambda \cosh \lambda}{e^{\lambda} - \cosh \lambda}.$$

4.5. By Definition 4.10, if Z is a discrete random variable, then the conditional expectation $\mathbb{E}(X|Z)$ is the conditional expectation of X with respect to the partition $\mathcal{P}$ generated by Z. The partition generated by Z consists of two subsets of the interval $[0, 1]$, namely $B_1 = [0, \frac{1}{2})$ and $B_2 = [\frac{1}{2}, 1]$. Hence $\mathbb{E}(X|Z)$ is constant on B_1 and on B_2. By (4.1), we have

$$\mathbb{E}(X|B_1) = \frac{\mathbb{E}(\mathbf{1}_{B_1} X)}{P(B_1)} = \frac{1}{\frac{1}{2}} \int_0^{\frac{1}{2}} x dx = \frac{1}{4},$$

$$\mathbb{E}(X|B_2) = \frac{\mathbb{E}(\mathbf{1}_{B_2} X)}{P(B_2)} = \frac{1}{\frac{1}{2}} \int_{\frac{1}{2}}^{1} x dx = \frac{3}{4}.$$

It follows that

$$\mathbb{E}(X|Z)(\omega) = \begin{cases} \frac{1}{4} & \text{if } \omega \in [0, \frac{1}{2}), \\ \frac{3}{4} & \text{if } \omega \in [\frac{1}{2}, 1]. \end{cases}$$

4.6. Similarly as in Exercise 4.5, we have

$$\mathbb{E}(X|B_1) = \frac{\mathbb{E}(\mathbf{1}_{B_1}X)}{P(B_1)} = \frac{1}{\frac{1}{2}} \int_0^{\frac{1}{2}} (1-x)dx = \frac{3}{4},$$
$$\mathbb{E}(X|B_2) = \frac{\mathbb{E}(\mathbf{1}_{B_2}X)}{P(B_2)} = \frac{1}{\frac{1}{2}} \int_{\frac{1}{2}}^{1} (1-x)dx = \frac{1}{4}.$$

Hence

$$\mathbb{E}(X|Z)(\omega) = \begin{cases} \frac{3}{4} & \text{if } \omega \in [0, \frac{1}{2}), \\ \frac{1}{4} & \text{if } \omega \in [\frac{1}{2}, 1]. \end{cases}$$

4.7. Take $\Omega = [0, 1]$ with its Borel subsets and Lebesgue measure. Let V be the random variable equal to 0 on $[0, \frac{3}{4})$ and 1 on $[\frac{3}{4}, 1]$. Furthermore, let W be the random variable with three values, -1 on $[0, \frac{1}{3})$, 0 on $[\frac{1}{3}, \frac{1}{2})$ and 1 on $[\frac{1}{2}, 1]$, and let X be the random variable defined as $X(\omega) = 2\omega - 1$ for $\omega \in [0, 1]$.

First we calculate $\mathbb{E}(X|V)$. Put $A_1 = \{V = 0\} = [0, \frac{3}{4})$ and $A_2 = \{V = 1\} = [\frac{3}{4}, 1]$. The conditional expectation $\mathbb{E}(X|V)$ is constant on A_1 and A_2, and

$$\mathbb{E}(X|A_1) = \frac{\mathbb{E}(\mathbf{1}_{A_1}X)}{P(A_1)} = \frac{1}{\frac{3}{4}} \int_0^{\frac{3}{4}} (2x-1)dx = -\frac{1}{4},$$
$$\mathbb{E}(X|A_2) = \frac{\mathbb{E}(\mathbf{1}_{A_2}X)}{P(A_2)} = \frac{1}{\frac{1}{4}} \int_{\frac{3}{4}}^{1} (2x-1)dx = \frac{3}{4}.$$

It follows that

$$\mathbb{E}(X|V)(\omega) = \begin{cases} -\frac{1}{4} & \text{if } \omega \in [0, \frac{3}{4})\ , \\ \frac{3}{4} & \text{if } \omega \in [\frac{3}{4}, 1]. \end{cases}$$

Next we compute $\mathbb{E}(X|W)$. The random variable W generates the partition consisting of three sets: $B_1 = \{W = -1\} = [0, \frac{1}{3})$, $B_2 = \{W =$

$0\} = [\frac{1}{3}, \frac{1}{2})$ and $B_3 = \{W = 1\} = [\frac{1}{2}, 1]$. Moreover,

$$\mathbb{E}(X|B_1) = \frac{\mathbb{E}(\mathbf{1}_{B_1}X)}{P(A_1)} = \frac{1}{\frac{1}{3}}\int_0^{\frac{1}{3}}(2x-1)dx = -\frac{2}{3},$$
$$\mathbb{E}(X|B_2) = \frac{\mathbb{E}(\mathbf{1}_{B_2}X)}{P(B_2)} = \frac{1}{\frac{1}{6}}\int_{\frac{1}{3}}^{\frac{1}{2}}(2x-1)dx = -\frac{1}{6},$$
$$\mathbb{E}(X|B_3) = \frac{\mathbb{E}(\mathbf{1}_{B_3}X)}{P(B_3)} = \frac{1}{\frac{1}{2}}\int_{\frac{1}{2}}^{1}(2x-1)dx = \frac{1}{2}.$$

It follows that

$$\mathbb{E}(X|W)(\omega) = \begin{cases} -\frac{2}{3} & \text{if } \omega \in [0, \frac{1}{3}) , \\ -\frac{1}{6} & \text{if } \omega \in [\frac{1}{3}, \frac{1}{2}) , \\ \frac{1}{2} & \text{if } \omega \in [\frac{1}{2}, 1], \end{cases}$$

and $\mathbb{E}(X|V) \neq \mathbb{E}(X|W)$.

4.8. Since $Y(n) = (-1)^n$ takes only two values -1 and 1, it generates the partition $\mathcal{P} = \{A, B\}$ consisting of two sets, $A = \{Y = -1\} = \{\text{odd numbers}\}$ and $B = \{Y = 1\} = \{\text{even numbers}\}$. We calculate $P(A)$ and $P(B)$ as follows:

$$P(A) = P(\{Y = -1\}) = \sum_{k=1}^{\infty} P(\{2k-1\}) = \sum_{k=1}^{\infty} 2 \cdot 3^{-(2k-1)} = \frac{3}{4},$$
$$P(B) = P(\{Y = 1\}) = \sum_{k=1}^{\infty} P(\{2k\}) = \sum_{k=1}^{\infty} 2 \cdot 3^{-2k} = \frac{1}{4}.$$

It follows that $\mathbb{E}(X|Y)$ has only two values

$$\mathbb{E}(X|A) = \frac{\mathbb{E}(\mathbf{1}_A X)}{P(A)} = \frac{1}{\frac{3}{4}}\sum_{k=1}^{\infty} 2^{2k-1} \cdot 2 \cdot 3^{-(2k-1)} = \frac{16}{5},$$
$$\mathbb{E}(X|B) = \frac{\mathbb{E}(\mathbf{1}_B X)}{P(B)} = \frac{1}{\frac{1}{4}}\sum_{k=1}^{\infty} 2^{2k} \cdot 2 \cdot 3^{-2k} = \frac{32}{5},$$

and so

$$\mathbb{E}(X|Y)(n) = \begin{cases} \frac{16}{5} & \text{if } n \text{ is odd,} \\ \frac{35}{5} & \text{if } n \text{ is even.} \end{cases}$$

4.9. Suppose that the discrete random variable Y has pairwise distinct values $y_1, y_2, \ldots$, and let $B_i = \{Y = y_i\}$ for each $i = 1, 2, \ldots$. If

$P(B_i) \neq 0$, then

$$\mathbb{E}(\mathbf{1}_{B_i}\mathbb{E}(X|Y)) = \mathbb{E}(\mathbf{1}_{B_i}\mathbb{E}(X|B_i)) = \mathbb{E}(X|B_i)\mathbb{E}(\mathbf{1}_{B_i})$$
$$= \mathbb{E}(X|B_i)P(B_i) = \mathbb{E}(\mathbf{1}_{B_i}X).$$

If $P(B_i) = 0$, then we also have $\mathbb{E}(\mathbf{1}_{B_i}\mathbb{E}(X|Y)) = \mathbb{E}(\mathbf{1}_{B_i}X)$ since both sides of the equality are equal to 0. Every set $B \in \sigma(Y)$ can be written as $B = \bigcup_{i\in I} B_i$ for some $I \subset \{1, 2, \ldots\}$. Let $I_n = I \cap \{1, 2, \ldots\}$. Since the B_i are disjoint,

$$\mathbf{1}_B = \mathbf{1}_{\bigcup_{i\in I} B_i} = \lim_{n\to\infty} \mathbf{1}_{\bigcup_{i\in I_n} B_i} = \lim_{n\to\infty} \sum_{i\in I_n} \mathbf{1}_{B_i}$$

Because X is integrable, so is $\mathbb{E}(X|Y)$. By the dominated convergence theorem, we therefore get

$$\mathbb{E}(\mathbf{1}_B X) = \lim_{n\to\infty} \mathbb{E}(\mathbf{1}_{\bigcup_{i\in I_n} B_i} X) = \lim_{n\to\infty} \sum_{i\in I_n} \mathbb{E}(\mathbf{1}_{B_i} X),$$
$$\mathbb{E}(\mathbf{1}_B \mathbb{E}(X|Y)) = \lim_{n\to\infty} \mathbb{E}(\mathbf{1}_{\bigcup_{i\in I_n} B_i}\mathbb{E}(X|Y))$$
$$= \lim_{n\to\infty} \sum_{i\in I_n} \mathbb{E}(\mathbf{1}_{B_i}\mathbb{E}(X|Y)) = \lim_{n\to\infty} \sum_{i\in I_n} \mathbb{E}(\mathbf{1}_{B_i} X),$$

so $\mathbb{E}(\mathbf{1}_B\mathbb{E}(X|Y)) = \mathbb{E}(\mathbf{1}_B\mathbb{E}(X|Y))$.

4.10. Suppose $\mathcal{P}_1$ and $\mathcal{P}_2$ are partitions of some set Ω. Consider the partition

$$\mathcal{P} = \{C = A \cap B : A \in \mathcal{P}_1 \text{ and } B \in \mathcal{P}_2\}.$$

We show that $\mathcal{P}$ refines $\mathcal{P}_1$ and $\mathcal{P}_2$. Take any $A \in \mathcal{P}_1$. Since $\bigcup_{B\in\mathcal{P}_2} B = \Omega$,

$$A = A \cap \left(\bigcup_{B\in\mathcal{P}_2} B \right) = \bigcup_{B\in\mathcal{P}_2, A\cap B\neq\emptyset} A \cap B.$$

Similarly, for any $B \in \mathcal{P}_2$ we have

$$B = B \cap \left(\bigcup_{A\in\mathcal{P}_1} B \right) = \bigcup_{A\in\mathcal{P}_1, A\cap B\neq\emptyset} B \cap A.$$

Now consider any partition $\mathcal{P}'$ which refines $\mathcal{P}_1$ and $\mathcal{P}_2$. Take any $C \in \mathcal{P}$. Then there are sets $A \in \mathcal{P}_1$ and $B \in \mathcal{P}_2$ such that $C = A \cap B$. Because $\mathcal{P}'$ refines $\mathcal{P}_1$, we have $A = \bigcup_{i\in I} A_i$ for a countable family of sets $A_i \in \mathcal{P}_1$, $i \in I$. Moreover, Because $\mathcal{P}'$ refines $\mathcal{P}_2$ and $B \in \mathcal{P}_2$ we have $A_i \subset B$ or $A_i \subset \Omega \setminus B$ for each $i \in I$. Similarly, $B = \bigcup_{j\in J} B_j$ for a countable family of sets $B_j \in \mathcal{P}_2$, $j \in J$, and $B_j \subset A$ or $B_j \subset \Omega \setminus A$

for each $j \in J$. It follows that $C = A \cap B = \bigcup_{(i,j)\in K} A_i \cap B_j$, where $K = \{(i, j) \in I \times J : A_i \subset B, B_j \subset A\}$. This means that $\mathcal{P}'$ refines $\mathcal{P}$.

We have shown that $\mathcal{P}$ is the coarsest partition that refines $\mathcal{P}_1$ and $\mathcal{P}_2$.

4.11. By Exercise 3.31, X, Y are independent if and only if their generated σ-fields $\sigma(X), \sigma(Y)$ are independent. For Y discrete with values $y_1, y_2, \ldots$, the σ-field $\sigma(Y)$ is generated by the partition $\mathcal{P} = \{B_i : i = 1, 2, \ldots\}$, where $B_i = \{Y = y_i\}$. Since $\sigma(\mathbf{1}_{B_i}) = \{\Omega, B_i, \Omega \setminus B_i, \varnothing\}$, it follows that $\sigma(\mathbf{1}_{B_i}) \subset \sigma(Y)$. This implies that X and $\mathbf{1}_{B_i}$ are independent for every $i = 1, 2, \ldots$. By Definition 4.10, $\mathbb{E}(X|Y)$ is constant on each B_i, $i = 1, 2, \ldots$, so the equality $\mathbb{E}(X|Y) = \mathbb{E}(X)$ suffices on B_i. By Theorem 3.42, we have

$$\mathbb{E}(\mathbf{1}_{B_i} X) = \mathbb{E}(\mathbf{1}_{B_i})\mathbb{E}(X) = P(B_i)\mathbb{E}(X)$$

and

$$\mathbb{E}(X|B_i) = \frac{\mathbb{E}(\mathbf{1}_{B_i} X)}{P(B_i)} = \frac{P(B_i)\mathbb{E}(X)}{P(B_i)} = \mathbb{E}(X).$$

4.12. Since Y is symmetric with respect to the line $x = \frac{1}{2}$ and $Y([0, 1]) = [0, \frac{1}{2}]$, we claim that $\sigma(Y) = \{B \cup (1 - B) : B \subset [0, \frac{1}{2}]$ is a Borel set$\}$. To verify this, first take any $A \in \mathcal{B}(\mathbb{R})$. Then $\{Y \in A\} \cap [0, \frac{1}{2}] = B$ is a Borel set in $[0, \frac{1}{2}]$ and $\{Y \in A\} = B \cup (1 - B)$. This implies that

$$\sigma(Y) \subset \left\{ B \cup (1 - B) : B \subset \left[0, \frac{1}{2}\right] \text{ is a Borel set} \right\}.$$

Now, if $B \subset [0, \frac{1}{2}]$ is a Borel set, then $A = \frac{1}{2} - B$ is a Borel set and $\{Y \in A\} = B \cup (1 - B)$. This gives the converse inclusion.

In order to calculate $\mathbb{E}(X|Y)$, take $B \cup (1 - B) \in \sigma(Y)$, where $B \subset [0, \frac{1}{2}]$ is a Borel set. If we put $X_1(\omega) = X(1 - \omega)$, then $\mathbb{E}(\mathbf{1}_{B\cup(1-B)} X) = \mathbb{E}(\mathbf{1}_{B\cup(1-B)} X_1)$ and

$$\begin{aligned}\mathbb{E}(\mathbf{1}_{B\cup(1-B)} X) &= \frac{1}{2}\left[\mathbb{E}(\mathbf{1}_{B\cup(1-B)} X) + \mathbb{E}(\mathbf{1}_{B\cup(1-B)} X_1)\right] \\ &= \mathbb{E}\left(\mathbf{1}_{B\cup(1-B)} \frac{1}{2}(X + X_1)\right).\end{aligned}$$

Since $\frac{1}{2}(X + X_1)$ is $\sigma(Y)$-measurable,

$$\mathbb{E}(X|Y)(\omega) = \frac{1}{2}(X(\omega) + X_1(\omega)) = \frac{1}{2}\left(\left|\omega - \frac{1}{3}\right| + \left|\omega - \frac{2}{3}\right|\right).$$

4.13. First we prove the condition (1) (linearity). By Definition 4.20 (i), the conditional expectations $\mathbb{E}(X|\mathcal{G})$ and $\mathbb{E}(Y|\mathcal{G})$ are $\mathcal{G}$-measurable. By Exercise 1.21, for any $a, b \in \mathbb{R}$, $a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ is $\mathcal{G}$-measurable. Take any $B \in \mathcal{G}$. By Definition 4.20 (ii), we have

$$\begin{aligned}
&\mathbb{E}(\mathbf{1}_B(a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G}))) = \mathbb{E}(a\mathbf{1}_B\mathbb{E}(X|\mathcal{G}) + b\mathbf{1}_B\mathbb{E}(Y|\mathcal{G}))\\
&= a\mathbb{E}(\mathbf{1}_B\mathbb{E}(X|\mathcal{G})) + b\mathbb{E}(\mathbf{1}_B\mathbb{E}(Y|\mathcal{G}))\\
&= a\mathbb{E}(\mathbf{1}_B X) + b\mathbb{E}(\mathbf{1}_B Y) = \mathbb{E}(\mathbf{1}_B(aX + bY)).
\end{aligned}$$

We have verified conditions (i) and (ii) of Definition 4.20. To prove (2) (positivity) put $B = \{\mathbb{E}(X|\mathcal{G}) < 0\}$. We have $B = \bigcup_{n=1}^{\infty} B_n$, where $B_n = \{\mathbb{E}(X|\mathcal{G}) < -\frac{1}{n}\}$ and $B_n, B \in \mathcal{G}$. The B_n form an increasing sequence of sets in $\mathcal{G} \subset \mathcal{F}$, so $P(B) = \lim_{n\to\infty} P(B_n)$. The simple function $s_n = -\frac{1}{n}\mathbf{1}_{B_n}$ is $\mathcal{G}$-measurable and satisfies

$$\mathbf{1}_{B_n}\mathbb{E}(X|\mathcal{G}) < s_n,$$

so

$$0 \le \mathbb{E}(\mathbf{1}_B X) = \mathbb{E}(\mathbf{1}_{B_n}\mathbb{E}(X|\mathcal{G})) \le \mathbb{E}(\mathbf{1}_{B_n} s_n) = -\frac{1}{n}P(B_n) \le 0,$$

which means that $P(B_n) = 0$ for all n, hence $P(B) = 0$. Therefore $\mathbb{E}(X|\mathcal{G}) \ge 0$, P-a.e.

4.14. By Exercise 4.13, the conditional expectations $\mathbb{E}(X_n|\mathcal{G})$ form a non-decreasing sequence of integrable random variables: since $X_{n+1} - X_n \ge 0$, by linearity and positivity we obtain $\mathbb{E}(X_{n+1} - X_n|\mathcal{G}) \ge 0$ and

$$\mathbb{E}(X_{n+1} - X_n|\mathcal{G}) = \mathbb{E}(X_{n+1}|\mathcal{G}) - \mathbb{E}(X_n|\mathcal{G}) \ge 0.$$

So $\lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{G}) = Y$ exists P-a.s., and by Exercise 1.19, Y is measurable with respect to $\mathcal{G}$.

We show that $Y = \mathbb{E}(X|\mathcal{G})$. For any $B \in \mathcal{G}$ we apply the monotone convergence theorem and the definition of conditional expectation to get

$$\lim_{n\to\infty}(\mathbf{1}_B\mathbb{E}(X_n|\mathcal{G})) = \lim_{n\to\infty}(\mathbb{E}(\mathbf{1}_B X_n) = \mathbb{E}(\mathbf{1}_B X) = \mathbb{E}(\mathbf{1}_B\mathbb{E}(X|\mathcal{G})).$$

Moreover

$$\lim_{n\to\infty}(\mathbf{1}_B\mathbb{E}(X_n|\mathcal{G})) = \mathbb{E}(\mathbf{1}_B Y).$$

This implies that $\mathbb{E}(\mathbf{1}_B\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(\mathbf{1}_B Y)$. It follows that $\mathbb{E}(X|\mathcal{G}) = Y$, P-a.s.

4.15. Theorem 4.27 can be extended as follows:

Let (Ω, F, P) be a probability space and let $G \subset F$ be a σ-field. Suppose that $X : \Omega \to \mathbb{R}^m$ is a G-measurable random vector and $Y : \Omega \to \mathbb{R}^n$ is a random vector independent of G. If $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ is a bounded Borel measurable function, then $g_f : \mathbb{R}^m \to \mathbb{R}$ defined for any $x \in \mathbb{R}^m$ by

$$g_f(x) = \mathbb{E}(f(x, Y)) = \int_{\mathbb{R}^n} f(x, y) dP_Y(y)$$

is a bounded Borel measurable function, and we have

$$\mathbb{E}(f(X, Y)|\mathcal{G}) = g_f(X), \ \ P\text{-a.s.}$$

Proof In fact we can repeat the proof of Theorem 4.27 without any significant changes.

By Proposition 3.17, g_f is a Borel measurable function. It follows that $g_f(X)$ is $\sigma(X)$-measurable. By the definition of conditional expectation it suffices to show that

$$\mathbb{E}(\mathbf{1}_G f(X, Y)) = \mathbb{E}(g_f(X)\mathbf{1}_G)$$

for each $G \in \mathcal{G}$.

By hypothesis, $\sigma(Y)$ and $\mathcal{G}$ are independent σ-fields. For any bounded $\mathcal{G}$-measurable random variable Z we have $\sigma(X, Z) \subset \mathcal{G}$ since for any $B_1 \times \cdots \times B_m \times B_{m+1}$, $B_i \in \mathcal{B}(\mathbb{R})$, $i = 1, \ldots, m+1$ we have

$$\{(X, Z) \in B_1 \times \cdots \times B_m \times B_{m+1}\} = \{X \in B_1 \times \cdots \times B_m\} \cap \{Z \in B_{m+1}\} \in \mathcal{G}.$$

Hence Y and (X, Z) are independent. This means (see Remark 3.40) that their joint distribution is the product measure $P_{X,Z} \otimes P_Y$.

In order to compute $\mathbb{E}(f(X, Y)Z)$ we apply Proposition 1.37, in which we take

$$\begin{aligned}
&(\Omega, \mathcal{F}, \mu) = (\Omega, \mathcal{F}, P), \\
&(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu}) = (\mathbb{R}^{m+n+1}, \mathcal{B}(\mathbb{R}^{m+n+1}), P_{X,Z} \otimes P_Y),
\end{aligned}$$

$\varphi = (X, Z, Y)$ and $\tilde{g} : \mathbb{R}^{m+n+1} \to \mathbb{R}$ such that $\tilde{g}(x, z, y) = f(x, y)z$ for $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, $z \in \mathbb{R}$. Hence

$$\mathbb{E}(f(X, Y)Z) = \int_{\Omega} f(X, Y)ZdP = \int_{\mathbb{R}^{m+n+1}} f(x, y)zd(P_{X,Z} \otimes P_Y)(x, z, y).$$

Applying Fubini's theorem, we obtain

$$
\begin{aligned}
\mathbb{E}(f(X,Y)Z) &= \int_{\mathbb{R}^{m+1}} \left(\int_{\mathbb{R}^n} f(x,y) z \, dP_Y(y) \right) dP_{X,Z}(x,z) \\
&= \int_{\mathbb{R}^{m+1}} g_f(x) z \, dP_{X,Z}(x,z) \\
&\quad \text{by Proposition 1.37 once again} \\
&= \mathbb{E}(g_f(X)Z).
\end{aligned}
$$

If we put $Z = \mathbf{1}_G$, we get

$$
\mathbb{E}(\mathbf{1}_G f(X,Y)) = \mathbb{E}(g_f(X)\mathbf{1}_G).
$$

This completes the proof.

4.16. Let $f_{X,Y}(x,y)$ be the bivariate normal density given in Example 3.16. Similarly as in Exercise 3.12, $f_{X,Y}(x,y)$ can be written in the form

$$
f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right),
$$

and $f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$. Then, by Definition 4.30, the conditional density of X given Y is

$$
h(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right).
$$

Furthermore, by Proposition 4.31, we have

$$
\mathbb{E}(X|Y) = \int_{\mathbb{R}} x h(x,Y) \, dm(x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{\mathbb{R}} x \exp\left(-\frac{(x-\rho Y)^2}{2(1-\rho^2)}\right) dx.
$$

Substituting $z = x - \rho Y$, we get

$$
\mathbb{E}(X|Y) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{\mathbb{R}} (z + \rho Y) \exp\left(-\frac{z^2}{2(1-\rho^2)}\right) dz.
$$

Since

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{\mathbb{R}} z \exp\left(-\frac{z^2}{2(1-\rho^2)}\right) dz &= 0, \\
\frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{\mathbb{R}} \exp\left(-\frac{z^2}{2(1-\rho^2)}\right) dz &= 1,
\end{aligned}
$$

we have $\mathbb{E}(X|Y) = \rho Y$.

4.17. Suppose that $x, y, z \in (a, b)$. We claim that the function $y \mapsto \frac{\phi(y)-\phi(x)}{y-x}$ is non-decreasing. Consider three cases: (i) $y < x < z$, (ii) $x < y < z$, and (iii) $y < z < x$. Suppose (i) holds. Taking $\lambda = \frac{z-x}{z-y}$, we have $x = \lambda y + (1-\lambda)z$, so the convexity of ϕ gives

$$\phi(x) \le \frac{z-x}{z-y}\phi(y) + \frac{x-y}{z-y}\phi(z).$$

Rearranging, we get

$$\frac{\phi(y)-\phi(x)}{y-x} \le \frac{\phi(z)-\phi(x)}{z-x}.$$

In cases (ii) and (iii) we verify the claim similarly. If $h > 0$ and $x-h, x+h \in (a, b)$, then we have

$$\frac{\phi(x-h)-\phi(x)}{-h} = \frac{\phi(x)-\phi(x-h)}{h} \le \frac{\phi(x+h)-\phi(x)}{h}.$$

By the claim, for $h, t > 0$, $h < t$ and $x+h, x+t, x-h, x-t \in (a, b)$ we also get

$$\frac{\phi(x+h)-\phi(x)}{h} \le \frac{\phi(x+t)-\phi(x)}{t},$$
$$\frac{\phi(x)-\phi(x-t)}{t} \le \frac{\phi(x)-\phi(x-h)}{h}.$$

It follows that the ratio $\frac{1}{h}[\phi(x+h)-\phi(x)]$ decreases as $h \searrow 0$ and $\frac{1}{h}[\phi(x)-\phi(x-h)]$ increases as $h \searrow 0$.

Finally, fix $t_0 > 0$ such that $x-t_0, x+t_0 \in (a, b)$. Applying the claim again, we obtain for any $h > 0$ such that $x+h \in (a, b)$

$$\frac{\phi(x-t_0)-\phi(x)}{-t_0} \le \frac{1}{h}[\phi(x+h)-\phi(x)]$$
$$\frac{1}{h}[\phi(x)-\phi(x-h)] \le \frac{\phi(x+t_0)-\phi(x)}{t_0}.$$

This means that the ratio $\frac{1}{h}[\phi(x+h)-\phi(x)]$ is bounded below by a constant and $\frac{1}{h}[\phi(x)-\phi(x-h)]$ is bounded above by a constant.

4.18. Take any $X \in L^2(P)$ and a sequence $X_1, X_2, \ldots \in L^2(P)$ such that $\lim_{n\to\infty} \|X_n - X\|_2 = 0$. For given $Y \in L^2(P)$, the Schwarz inequality implies that

$$|\langle X_n, Y\rangle - \langle X, Y\rangle| = |\langle X_n - X, Y\rangle| \le \|X_n - X\|_2\|Y\|_2.$$

By the hypothesis, $\lim_{n\to\infty} |\langle X_n, Y\rangle - \langle X, Y\rangle| = 0$. This means that the map $X \mapsto \langle X, Y\rangle$ is norm continuous.

Next, using the triangle inequality, we have

$$\|X_n\|_2 = \|(X_n - X) + X\|_2 \le \|X_n - X\|_2 + \|X\|_2,$$
$$\|X\|_2 = \|(X - X_n) + X_n\|_2 \le \|X_n - X\|_2 + \|X_n\|_2.$$

So we get

$$|\|X_n\|_2 - \|X\|_2| \le \|X_n - X\|_2 \to 0 \text{ as } n \to \infty.$$

This implies norm continuity of the L^2-norm.

4.19. Let $X, X_1, X_2, \ldots \in L^2(P)$. Suppose that $\lim_{n\to\infty} \|X_n - X\|_2 = 0$, where $X \in L^2(P)$. Given $\varepsilon > 0$, there is $k \in \mathbb{N}$ such that $\|X_n - X\|_2 < \frac{\varepsilon}{2}$ for each $n \ge k$. By the triangle inequality, for any $m, n \ge k$ we have

$$\|X_m - X_n\|_2 = \|(X_m - X) + (X - X_n)\|_2 \le \|X_m - X\|_2 + \|X_n - X\|_2 < \varepsilon.$$

This implies that

$$\sup_{m,n\ge k} \|X_m - X_n\|_2 < \varepsilon.$$

We have proved that $X_1, X_2, \ldots$ is a Cauchy sequence.

4.20. Let $X, Y \in L^2(\mathcal{F}, P)$. Using the definition of the norm, we have

$$\begin{aligned}\|X + Y\|_2^2 &= \mathbb{E}((X + Y)^2) = \mathbb{E}(X^2) + 2\mathbb{E}(X, Y) + \mathbb{E}(Y^2)\\ &= \|X\|_2^2 + 2\langle X, Y\rangle + \|Y\|_2^2.\end{aligned}$$

If $\langle X, Y\rangle = 0$, then $\|X + Y\|_2^2 = \|X\|_2^2 + \|Y\|_2^2$.

4.21. Let $X, Y \in L^2(\mathcal{F}, P)$. Using the definition of the norm, we have

$$\begin{aligned}&\|X + Y\|_2^2 + \|X - Y\|_2^2 = \mathbb{E}((X + Y)^2) + \mathbb{E}((X - Y)^2)\\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) + \mathbb{E}(X^2) - 2\mathbb{E}(XY) + \mathbb{E}(Y^2)\\ &= 2\|X\|_2^2 + 2\|Y\|_2^2.\end{aligned}$$

4.22. Let $X_n(\omega) = \sin n\omega$ and $Y_m(\omega) = \cos m\omega$. We show that $X_n(\omega)$, $Y_m(\omega)$ are orthogonal in $L^2[-\pi, \pi]$ for any $m, n = 1, 2, \ldots$. Observe that

$$\sin n\omega \cos m\omega = \frac{1}{2}[\sin(n + m)\omega + \sin(n - m)\omega].$$

By the definition of the inner product in $L^2[-\pi, \pi]$, we can verify that

$$\begin{aligned}\langle X_n, Y_m\rangle &= \int_{-\pi}^{\pi} \sin nx \cos mx dx\\ &= \frac{1}{2}\int_{-\pi}^{\pi} [\sin(n + m)x + \sin(n - m)x]dx\\ &= \frac{1}{2}\left[-\frac{\cos(n + m)x}{n + m} - \frac{\cos(n - m)x}{n - m}\right]\Big|_{-\pi}^{\pi} = 0.\end{aligned}$$

4.23. Let $X \in L^2(\mathcal{F}, P)$. Suppose that $Y \in M$ satisfies $\langle X - Y, Z\rangle = 0$ for any $Z \in M$. It follows that $\langle X - Y, Y - Z\rangle = 0$ for any $Z \in M$. By Pythagoras' theorem, we get

$$\|X - Z\|_2^2 = \|(X - Y) + (Y - Z)\|_2^2 = \|X - Y\|_2^2 + \|Y - Z\|_2^2.$$

It follows that $\|X - Y\|_2 \leq \|X - Z\|_2$ for any $Z \in M$. Since $Y \in M$, we have $\|X - Y\|_2 = \inf\{\|X - Z\|_2 : Z \in M\}$.

4.24. Consider the random variables $X(\omega) = \omega$ and $Y(\omega) = 1 - \omega$ defined on the probability space $[0, 1]$ with Borel sets and Lebesgue measure.

We compute $\|X\|_1$, $\|Y\|_1$, $\|X+Y\|_1$ and $\|X-Y\|_1$. By Definition 4.49,

$$\|X\|_1 = \mathbb{E}(|X|) = \frac{1}{2}, \quad \|Y\|_1 = \mathbb{E}(|Y|) = \frac{1}{2},$$
$$\|X + Y\|_1 = \mathbb{E}(|X + Y|) = 1, \quad \|X - Y\|_1 = \mathbb{E}(|X - Y|) = \frac{1}{2}.$$

The parallelogram law stated in Exercise 4.21 fails in the L^1-norm since in our case $\|X + Y\|_1^2 + \|X - Y\|_1^2 = \frac{3}{2}$, but $2\|X\|_1^2 + 2\|Y\|_1^2 = 2$.

The parallelogram law holds for any norm induced by an inner product, that is, if H is a real or complex normed vector space with an inner product such that $\|h\|^2 = \langle h, h\rangle$. The proof is the same as for L^2. The above example shows that the L^1-norm is not induced by an inner product.

4.25. Let the assumptions of Theorem 4.54 be satisfied, so that there exists a random variable $Z \in L^1(P)$ such that for each $A \in \mathcal{F}$

$$Q(A) = \int_A Z dP.$$

Since Q is a probability measure, we know that $Z \geq 0$ on Ω and $\mathbb{E}_P(Z) = 1$. For any $B \in \mathcal{F}$ we have

$$\mathbb{E}_Q(\mathbf{1}_B) = Q(B) = \mathbb{E}_P(\mathbf{1}_B Z).$$

By linearity, this extends to $\mathbb{E}_Q(s) = \mathbb{E}_P(sZ)$ for any simple function s. Approximating any non-negative random variable X by a non-decreasing sequence of simple functions, we obtain by monotone convergence that

$$\mathbb{E}_Q(X) = \mathbb{E}_P(XZ).$$

Finally, we can extend the last identity to any random variable X integrable under Q by considering X^+ and X^- and using linearity once again.

4.26. To verify (1) take any $A \in \mathcal{F}$ with $P(A) = 0$. For $\lambda \in (0,1)$, we have

$$(\lambda Q + (1-\lambda)R)(A) = \lambda Q(A) + (1-\lambda)R(A) = 0$$

since $Q \ll P$ and $R \ll P$. Write $\frac{dQ}{dP} = Z$, $\frac{dR}{dP} = U$ and $\frac{d(\lambda Q+(1-\lambda)R)}{dP} = W$. For each $A \in \mathcal{F}$ we get

$$\begin{aligned}\int_A W dP &= (\lambda Q + (1-\lambda)R)(A) = \lambda Q(A) + (1-\lambda)R(A) \\ &= \int_A \lambda Z dP + \int_A (1-\lambda) U dP = \int_A (\lambda Z + (1-\lambda)U) dP.\end{aligned}$$

By Exercise 1.30, $W = \lambda Z + (1-\lambda)U$, P-a.e.

(2) Suppose that $R \ll Q$ and $Q \ll P$. For any $A \in \mathcal{F}$, $P(A) = 0$ implies $Q(A) = 0$, and this in turn implies $R(A) = 0$. It follows that $R \ll P$. Writing $\frac{dR}{dP} = U$, $\frac{dQ}{dP} = Z$ and $\frac{dR}{dQ} = V$, we have $R(A) = \int_A U dP = \int_A V dQ$ and $Q(A) = \int_A Z dP$ for any $A \in \mathcal{F}$. By Exercise 4.25, we obtain

$$R(A) = \mathbb{E}_Q(\mathbf{1}_A V) = \mathbb{E}_P(\mathbf{1}_A VZ) = \mathbb{E}_P(\mathbf{1}_A U).$$

By Exercise 1.30, it follows that $VZ = U$, P-a.e. Thus $\frac{dR}{dP} = \frac{dR}{dQ}\frac{dQ}{dP}$.

(3) Suppose that $P \sim Q$. Write $\frac{dQ}{dP} = Z$ and $\frac{dP}{dQ} = S$. Since $Q(A) = \int_A Z dP$ and $P(A) = \int_A S dQ$ for any $A \in \mathcal{F}$, using Exercise 4.25 once again, we get

$$\mathbb{E}_P(\mathbf{1}_A) = \mathbb{E}_Q(\mathbf{1}_A S) = \mathbb{E}_P(\mathbf{1}_A SZ).$$

By Exercise 1.30, we have $SZ = 1$, P-a.e. This implies

$$\frac{dP}{dQ} = \left(\frac{dQ}{dP}\right)^{-1}.$$

Chapter 5

5.1. If $X, Y \in L^2(P)$ and $E_1, E_2, \ldots \in L^2(P)$ is a complete orthonormal sequence, then (5.5) holds for X and Y, so

$$\langle X, Y\rangle = \left\langle \sum_{i=1}^{\infty} \langle X, E_i\rangle E_i, \sum_{j=1}^{\infty} \langle Y, E_j\rangle E_j \right\rangle = \sum_{i=1}^{\infty} \langle X, E_i\rangle\langle Y, E_i\rangle$$

since $\langle E_i, E_j\rangle = 1$ if $i = j$, and 0 otherwise.

5.2. We have

$$\|H_0\|_2^2 = \int_{[0,1]} (H_0)^2 dm = \int_0^1 dx = 1,$$

and any $j = 0, 1, \dots$ and $k = 0, \dots, 2^j - 1$

$$\begin{aligned}\|H_{2^j+k}\|_2^2 &= \int_{[0,1]} (H_{2^j+k})^2 dm = 2^j \int_0^1 \mathbf{1}_{(\frac{2k}{2^{j+1}}, \frac{2k+2}{2^{j+1}}]} dx \\ &= 2^j \left(\frac{2k+2}{2^{j+1}} - \frac{2k}{2^{j+1}} \right) = 1.\end{aligned}$$

This proves that each Haar function H_n has L^2-norm 1.

Next we show that H_0 is orthogonal to H_{2^j+k} for any $j = 0, 1, \dots$ and $k = 0, 1, \dots, 2^j - 1$:

$$\begin{aligned}\langle H_0, H_{2^j+k} \rangle &= \int_{[0,1]} H_0 H_{2^j+k} dm = 2^{\frac{j}{2}} \int_{[0,1]} \left(\mathbf{1}_{\left(\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}}\right]} - \mathbf{1}_{\left(\frac{2k+1}{2^{j+1}}, \frac{2k+2}{2^{j+1}}\right]} \right) dm \\ &= 2^{\frac{j}{2}} \left(\frac{2k+1}{2^{j+1}} - \frac{2k}{2^{j+1}} \right) - 2^{\frac{j}{2}} \left(\frac{2k+2}{2^{j+1}} - \frac{2k+1}{2^{j+1}} \right) = 0.\end{aligned}$$

Now for any $i, j = 0, 1, \dots$ such that $i \leq j$ and for any $k = 0, 1, \dots, 2^i - 1$ and $l = 0, 1, \dots, 2^j - 1$

$$\begin{aligned}\langle H_{2^i+k}, H_{2^j+l} \rangle &= \int_{[0,1]} H_{2^i+k} H_{2^j+l} dm \\ &= 2^{\frac{i}{2}} 2^{\frac{j}{2}} \int_{[0,1]} \left(\mathbf{1}_{\left(\frac{2k}{2^{i+1}}, \frac{2k+1}{2^{i+1}}\right]} - \mathbf{1}_{\left(\frac{2k+1}{2^{i+1}}, \frac{2k+2}{2^{i+1}}\right]} \right) \left(\mathbf{1}_{\left(\frac{2l}{2^{j+1}}, \frac{2l+1}{2^{j+1}}\right]} - \mathbf{1}_{\left(\frac{2l+1}{2^{j+1}}, \frac{2l+2}{2^{j+1}}\right]} \right) dm.\end{aligned}$$

Observe that

$$\begin{array}{ll} \left(\frac{2l}{2^{j+1}}, \frac{2l+2}{2^{j+1}}\right] \subset \left(\frac{2k}{2^{i+1}}, \frac{2k+1}{2^{i+1}}\right] & \text{if } k2^{j-i} \leq l \leq k2^{j-i} + 2^{j-i-1} - 1, \\ \left(\frac{2l}{2^{j+1}}, \frac{2l+2}{2^{j+1}}\right] \subset \left(\frac{2k+1}{2^{i+1}}, \frac{2k+2}{2^{i+1}}\right] & \text{if } k2^{j-i} + 2^{j-i-1} \leq l \leq k2^{j-i} + 2^{j-i} - 1, \\ \left(\frac{2l}{2^{j+1}}, \frac{2l+2}{2^{j+1}}\right] \cap \left(\frac{2k}{2^{i+1}}, \frac{2k+2}{2^{i+1}}\right] = \varnothing & \text{otherwise.} \end{array}$$

Therefore,

$$\begin{aligned}\langle H_{2^i+k}, H_{2^j+l} \rangle &= 2^{\frac{i}{2}} 2^{\frac{j}{2}} \int_{\left(\frac{2k}{2^{i+1}}, \frac{2k+1}{2^{i+1}}\right]} \left(\mathbf{1}_{\left(\frac{2l}{2^{j+1}}, \frac{2l+1}{2^{j+1}}\right]} - \mathbf{1}_{\left(\frac{2l+1}{2^{j+1}}, \frac{2l+2}{2^{j+1}}\right]} \right) dm \\ &= 2^{\frac{i}{2}} 2^{\frac{j}{2}} \left(\left(\frac{2l+1}{2^{j+1}} - \frac{2l}{2^{j+1}} \right) - \left(\frac{2l+1}{2^{j+1}} - \frac{2l+1}{2^{j+1}} \right) \right) = 0\end{aligned}$$

if $k2^{j-i} \leq l \leq k2^{j-i} + 2^{j-i-1} - 1$,

$$\begin{aligned}\langle H_{2^i+k}, H_{2^j+l}\rangle &= -2^{\frac{i}{2}}2^{\frac{j}{2}} \int_{\left(\frac{2k+1}{2^{i+1}}, \frac{2k+2}{2^{i+1}}\right]} \left(\mathbf{1}_{\left(\frac{2l}{2^{j+1}}, \frac{2l+1}{2^{j+1}}\right]} - \mathbf{1}_{\left(\frac{2l+1}{2^{j+1}}, \frac{2l+2}{2^{j+1}}\right]}\right) dm \\ &= -2^{\frac{i}{2}}2^{\frac{j}{2}} \left(\left(\frac{2l+1}{2^{j+1}} - \frac{2l}{2^{j+1}}\right) - \left(\frac{2l+1}{2^{j+1}} - \frac{2l+1}{2^{j+1}}\right)\right) = 0\end{aligned}$$

if $k2^{j-i} + 2^{j-i-1} \leq l \leq k2^{j-i} + 2^{j-i} - 1$, and

$$\langle H_{2^i+k}, H_{2^j+l}\rangle = 0$$

in all other cases. This proves that the H_n are orthogonal to one another.

5.3. First we prove the following claim:

Suppose that $f \in L^1(\Omega, \mathcal{F}, \mu)$. Then for each $\varepsilon > 0$ there is $\delta > 0$ such that $\int_A |f| d\mu < \varepsilon$ whenever $A \in \mathcal{F}$ and $\mu(A) < \delta$.

Proof Let $f_n = \min(n, |f|)$ for $n = 1, 2, \ldots$. Since f_n is non-decreasing and f_n converges to $|f|$, μ-a.e., it follows by the monotone convergence theorem that

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega |f| d\mu.$$

For any given $\varepsilon > 0$ there is an integer k such that

$$\int_\Omega (|f| - f_k) d\mu < \frac{\varepsilon}{2}.$$

Put $\delta = \frac{\varepsilon}{2k}$. Since $\int_A (|f| - f_k) d\mu \leq \int_\Omega (|f| - f_k) d\mu$, it follows that

$$\int_A |f| d\mu = \int_A (|f| - f_k) d\mu + \int_A f_k d\mu < \frac{\varepsilon}{2} + k\mu(A) < \varepsilon$$

whenever $\mu(A) < \delta$. The claim has been proved.

Now we prove the assertion stated in Exercise 5.3. It suffices to show that $\|f - f_h\|_2 \to 0$ as $h \to 0$, where $f_h(x) = f(x + h)$. Take any $\varepsilon > 0$. By Lemma 5.8, there is a continuous function g defined on $[0, 1]$ such that $\|f - g\|_2 < \frac{\varepsilon}{4}$. Let us define an extension $\tilde{g}$ of g to the interval $[a, b]$ as follows

$$\tilde{g}(x) = \begin{cases} g(0) & \text{if } x \in [a, 0], \\ g(x) & \text{if } x \in [0, 1], \\ g(1) & \text{if } x \in [1, b]. \end{cases}$$

So $\tilde{g}$ is continuous on $[a, b]$ and $g = \mathbf{1}_{[0,1]}\tilde{g}$. In the sequel we consider h small enough so that $[h, 1 + h] \subset [a, b]$, that is, such that

$|h| < \min\{|a|, b-1\}$. Applying the triangle inequality, we have

$$\|f - f_h\|_2 \leq \|f - g\|_2 + \|g - \tilde{g}_h\|_2 + \|\tilde{g}_h - f_h\|_2.$$

Now, we have to estimate the second and third term of the right hand side of this inequality. Any continuous function on $[0,1]$ is uniformly continuous, so for $\frac{\varepsilon}{4} > 0$ there is $\delta_1 > 0$ such that $|g(x) - \tilde{g}(x+h)| < \frac{\varepsilon}{4}$, whenever $|h| < \delta_1$ and $x \in [0,1]$. This implies that $\|g - \tilde{g}_h\|_2 < \frac{\varepsilon}{4}$. Further, applying Exercise 1.32 for $\mathbf{1}_{[0,1]}(x)(g(x+h) - f(x+h))^2$, we have

$$\|\tilde{g}_h - f_h\|_2^2 = \int_{[0,1]} (\tilde{g}(x+h) - f(x+h))^2 dm(x)$$
$$= \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(\tilde{g}(x+h) - f(x+h))^2 dm(x)$$
$$= \int_{\mathbb{R}} \mathbf{1}_{[h,1+h]}(y)(\tilde{g}(y) - f(y))^2 dm(y)$$
$$= \int_{[h,1+h]} (\tilde{g}(y) - f(y))^2 dm(y) \leq \int_{[-|h|,1+|h|]} (\tilde{g}(y) - f(y))^2 dm(y)$$
$$= \int_{[0,1]} (\tilde{g}(y) - f(y))^2 dm(y) + \int_{[-|h|,0]} (\tilde{g}(y) - f(y))^2 dm(y)$$
$$+ \int_{[1,1+|h|]} (\tilde{g}(y) - f(y))^2 dm(y).$$

From the claim proved at the outset, given $\frac{\varepsilon^2}{32}$ there is $\delta_2 > 0$ such that $\int_{[-|h|,0]}(\tilde{g}(y) - f(y))^2 dm(y)$ and $\int_{[1,1+|h|]}(\tilde{g}(y) - f(y))^2 dm(y)$ are less then $\frac{\varepsilon^2}{32}$ if $|h| < \delta_2$. Finally, $\|\tilde{g}_h - f_h\|_2^2 < \frac{\varepsilon^2}{8}$ and $\|f - f_h\|_2 < \frac{(2+\sqrt{2})\varepsilon}{4} < \varepsilon$ for $|h| < \delta = \min(\delta_1, \delta_2)$.

5.4. From the obvious inequalities

$$|x_i^{(k)} - x_i| \leq \|x^{(k)} - x\|_1 < n\|x^{(k)} - x\|_2$$

we get the following implications: (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1). So, it suffices to show that (1) implies (2). Suppose that (1) holds. Then for any $\varepsilon > 0$ there is an integer K such that for any $k \geq K$ we have

$$|x_i^{(k)} - x_i| < \frac{\varepsilon}{\sqrt{n}} \quad \text{for } i = 1, \ldots, n.$$

Hence

$$\|x^{(k)} - x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i^{(k)} - x_i|^2} < \varepsilon.$$

5.5. Because

$$\limsup_{n\to\infty} \mathbf{1}_{A_n} = \mathbf{1}_{\limsup_{n\to\infty} A_n}$$

and all indicator functions are bounded above by $\mathbf{1}_\Omega$, which is integrable, we can apply the second Fatou lemma (Lemma 1.41 (ii)) with $f_n = \mathbf{1}_{A_n}$. This gives

$$\begin{aligned} P\left(\limsup_{n\to\infty} A_n\right) &= \int_\Omega \mathbf{1}_{\limsup_{n\to\infty} A_n}\, dP = \int_\Omega \limsup_{n\to\infty} \mathbf{1}_{A_n}\, dP \\ &\geq \limsup_{n\to\infty} \int_\Omega \mathbf{1}_{A_n}\, dP = \limsup_{n\to\infty} P(A_n). \end{aligned}$$

5.6. The 'if' claim follows immediately since for $\varepsilon = 1$ we can find $K > 0$ such that $\int_{\{|X|\geq K\}} |X| dP < 1$. Then

$$\int_\Omega |X| dP = \int_{\{|X|<K\}} |X| dP + \int_{\{|X|\geq K\}} |X| dP \leq K + 1 < \infty.$$

The opposite implication ('only if') is proved as follows: Suppose that $X \in L^1(P)$. Then $X_n = |X|\mathbf{1}_{\{|X|\geq n\}} \to 0$, P-a.e., with $X_n \leq |X|$. Thus by the dominated convergence theorem

$$\lim_{n\to\infty} \int_{\{|X|\geq n\}} |X|\, dP = \lim_{n\to\infty} \int_\Omega X_n\, dP = \int_\Omega \lim_{n\to\infty} X_n\, dP = 0.$$

So for any given $\varepsilon > 0$ there is $K > 0$ such that

$$\int_{\{|X|\geq K\}} |X| dP < \varepsilon.$$

5.7. If $X_1, X_2, \ldots$ is a sequence of random variable such that $|X_n| \leq Y$, P-a.e. for all n, then the sequence is uniformly integrable.

This is a simple consequence of Exercise 5.6. Since Y is integrable, for any given $\varepsilon > 0$ there is a $K > 0$ such that $\int_{\{|Y|\geq K\}} |Y| < \varepsilon$. Because $\{|X_n| \geq K\} \subset \{|Y| \geq K\}$ for $n = 1, 2, \ldots$, we obtain

$$\int_{\{|X_n|\geq K\}} |X_n| dP \leq \int_{\{|Y|\geq K\}} |Y| dP < \varepsilon.$$

5.8. We have the following inequalities for any $x \in \mathbb{R}$ and $\varepsilon > 0$:

$$\begin{aligned} P(Y_n \leq x - \varepsilon) - P(|X_n - Y_n| > \varepsilon) &\leq P(X_n \leq x) \\ &\leq P(Y_n \leq x + \varepsilon) + P(|X_n - Y_n| > \varepsilon). \end{aligned}$$

From Exercise 2.1 we know that there are at most countably many

$\varepsilon > 0$ such that F_Y has a discontinuity at $x+\varepsilon$ or $x-\varepsilon$. For any other $\varepsilon > 0$ we get

$$F_Y(x-\varepsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_Y(x+\varepsilon)$$

by letting $n \to \infty$. Because x can be approached from the left by $x-\varepsilon$ and from the right by $x+\varepsilon$ such that $\varepsilon > 0$ does not belong to that countable set, if x is a continuity point of F_Y, then we obtain

$$\lim_{n\to\infty} F_{X_n}(x) = F_Y(x),$$

showing that $X_n \Longrightarrow Y$.

5.9. Observe that for any $x \in \mathbb{R}$

$$F_{-X_n}(x) = P(-X_n \leq x) \geq P(-X_n < x) = 1-P(X_n \leq -x) = 1-F_{X_n}(-x)$$

Suppose that x is a continuity point of F_{-X}. It follows that $P\{-X = x\} = P\{X = -x\} = 0$, which means that $-x$ is a continuity point of F_X, and therefore

$$\liminf_{n\to\infty} F_{-X_n}(x) \geq 1 - F_X(-x) = F_{-X}(x).$$

On the other hand, for any $x \in \mathbb{R}$ and $\varepsilon > 0$

$$\begin{aligned} F_{-X_n}(x) &= P(-X_n \leq x) \leq P(-X_n < x+\varepsilon) \\ &= 1 - P(X_n \leq -x-\varepsilon) = 1 - F_{X_n}(-x-\varepsilon). \end{aligned}$$

From Exercise 2.1 we know that the set C consisting of all $\varepsilon > 0$ such that F_{-X} has a discontinuity at $x+\varepsilon$ (equivalently, F_X has a discontinuity at $-x-\varepsilon$) is at most countable. For any $\varepsilon > 0$ such that $\varepsilon \notin C$ we therefore have

$$\limsup_{n\to\infty} F_{-X_n}(x) \leq 1 - F_X(-x-\varepsilon) = F_{-X}(x+\varepsilon).$$

Since $\varepsilon > 0$ such that $\varepsilon \notin C$ can be arbitrarily small, from the right-continuity of the distribution function F_{-X} we get

$$\limsup_{n\to\infty} F_{-X_n}(x) \leq F_{-X}(x).$$

We have shown that

$$\lim_{n\to\infty} F_{-X_n}(x) = F_{-X}(x)$$

whenever x is a continuity point of F_{-X}, that is, $-X_n \Longrightarrow -X$.

5.10. For any $x \in \mathbb{R}$ and $\varepsilon > 0$

$$\{X_n + c \leq x - \varepsilon\} \subset \{X_n + Y_n \leq x\} \cup \{|Y_n - c| > \varepsilon\},$$
$$\{X_n + Y_n \leq x\} \subset \{X_n + c \leq x + \varepsilon\} \cup \{|Y_n - c| > \varepsilon\}.$$

It follows that

$$P(X_n + c \leq x - \varepsilon) \leq P(X_n + Y_n \leq x) + P(|Y_n - c| > \varepsilon),$$
$$P(X_n + Y_n \leq x) \leq P(X_n + c \leq x + \varepsilon) + P(|Y_n - c| > \varepsilon).$$

From Exercise 2.1 we know that the set C consisting of all $\varepsilon > 0$ such that F_X has a discontinuity at $x - c + \varepsilon$ or $x + c - \varepsilon$ is at most countable. For any $\varepsilon > 0$ such that $\varepsilon \notin C$ we get

$$F_X(x - c - \varepsilon) \leq \liminf_{n\to\infty} F_{X_n+Y_n}(x) \leq \limsup_{n\to\infty} F_{X_n+Y_n}(x) \leq F_X(x - c + \varepsilon)$$

since $P(|Y_n - c| > \varepsilon) \to 0$ as $n \to \infty$. Observe that $F_{X+c}(x) = F_X(x - c)$ for any $x \in \mathbb{R}$. Hence, if x is a continuity point of F_{X+c}, then $x - c$ is a continuity point of F_X, and because $x - c$ can be approached from the left by $x - c - \varepsilon$ and from the right by $x - c + \varepsilon$ such that $0 < \varepsilon \notin C$, we obtain

$$\lim_{n\to\infty} F_{X_n+Y_n}(x) = F_X(x - c) = F_{X+c}(x)$$

when x is a continuity point of F_{X+c}, which proves that $X_n + Y_n \Longrightarrow X + c$.

5.11. We show that $\lim_{T\to\infty} \int_0^T \frac{\sin x}{x} dx = \frac{\pi}{2}$. Consider the function $f : (0, T) \times (0, \infty) \to \mathbb{R}$ given $f(x, y) = e^{-xy} \sin x$. First we show that $f \in L^1((0, T] \times (0, \infty), \mathcal{B}((0, T] \times (0, \infty)), m_2)$. We calculate the iterated integral

$$\int_0^T \left(\int_0^\infty |e^{-xy} \sin x| dy \right) dx = \int_0^T \left(|\sin x| \frac{1}{x} \right) dx \leq T < \infty$$

since $\int_0^\infty e^{-xy} dy = \frac{1}{x}$. Then, reasoning as in the proof of Fubini's theorem (Theorem 3.18), we obtain that

$$\int\int_{(0,T)\times(0,\infty)} |e^{-xy} \sin x| dx dy = \int_0^T \left(\int_0^\infty |e^{-xy} \sin x| dy \right) dx \leq T.$$

We can apply Fubini's theorem to have

$$\int_0^T \frac{\sin x}{x} dx = \int_0^T \sin x \left(\int_0^\infty e^{-xy} dy \right) dx$$
$$= \int_0^T \left(\int_0^\infty e^{-xy} \sin x dy \right) dx = \int_0^\infty \left(\int_0^T e^{-xy} \sin x dx \right) dy.$$

Differentiating both sides with respect to T, we can verify that

$$\int_0^T e^{-xy}\sin x dx = \frac{1}{1+y^2}(1 - e^{Ty}(y\sin T + \cos T)).$$

Then

$$\int_0^T \frac{\sin x}{x}dx = \int_0^\infty \frac{1}{1+y^2}dy - \int_0^\infty \frac{e^{-Ty}}{1+y^2}(y\sin T + \cos T)dy.$$

Since $\int_0^\infty \frac{dy}{1+y^2} = \arctan y|_0^\infty = \frac{\pi}{2}$ and

$$\left|\int_0^\infty \frac{e^{-Ty}}{1+y^2}(y\sin T + \cos T)dy\right| \le \int_0^\infty \frac{e^{-Ty}}{1+y^2}|y\sin T + \cos T|dy$$
$$\le \int_0^\infty \frac{1+y}{1+y^2}e^{-Ty}dy = \int_0^1 \frac{1+y}{1+y^2}e^{-Ty}dy + \int_1^\infty \frac{1+y}{1+y^2}e^{-Ty}dy$$
$$\le \max_{y\in[0,1]}\left(\frac{1+y}{1+y^2}\right)\left(\frac{1}{T}(1-e^{-T})\right) + \frac{1}{T} \to 0,$$

if $T \to \infty$, we have

$$\lim_{T\to\infty}\int_0^T \frac{\sin x}{x}dx = \frac{\pi}{2}.$$

5.12. Let $a \le b$ be continuity points of F_X. With

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

we have

$$\begin{aligned}
\int_a^b f_X(x)\,dx &= \int_{[a,b]} f_X(x)\,dm(x)\\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\left(\int_{[a,b]} e^{-itx}\,dm(x)\right)\phi_X(t)\,dm(t)\\
&= \lim_{T\to\infty}\frac{1}{2\pi}\int_{[-T,T]}\left(\int_a^b e^{-itx}\,dx\right)\phi_X(t)\,dm(t)\\
&= \lim_{T\to\infty}\frac{1}{2\pi}\int_{[-T,T]}\frac{e^{-ita}-e^{-itb}}{it}\phi_X(t)\,dm(t)\\
&= F_X(b) - F_X(a)
\end{aligned}$$

by Fubini's theorem and the inversion formula. In fact $\int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$ holds for all $a, b \in \mathbb{R}$ since F_X is right-continuous and the integral on the left is continuous with respect to a and b. It follows that f_X is indeed the density of X.

5.13. Suppose that X is an integer-valued random variable. Then e^{itX} is a discrete random variable with values e^{itn} for $n \in \mathbb{Z}$. Put $p_n = P(X = n)$. By the definition of $\phi_X(t)$, we get

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \sum_{n\in\mathbb{Z}} e^{itn} p_n.$$

Fix $k \in \mathbb{Z}$. Then

$$\int_0^{2\pi} e^{-itk}\phi_X(t)dt = \int_0^{2\pi} \sum_{n\in\mathbb{Z}} e^{it(n-k)} p_n dt.$$

In order to calculate the last integral we show that

$$\int_0^{2\pi} \sum_{n\in\mathbb{Z}} e^{it(n-k)} p_n dt = \sum_{n\in\mathbb{Z}} p_n \int_0^{2\pi} e^{it(n-k)}dt.$$

Put

$$\begin{aligned} s &= \int_0^{2\pi} \sum_{n\in\mathbb{Z}} e^{it(n-k)} p_n dt \\ s_m &= \sum_{n=-m}^{m} p_n \int_0^{2\pi} e^{it(n-k)}dt = \int_0^{2\pi} \sum_{n=-m}^{m} e^{it(n-k)} p_n dt. \end{aligned}$$

It suffices to verify that $s_m \to s$ as $m \to \infty$. The follows because

$$\begin{aligned} |s_m - s| &= \left| \int_0^{2\pi} \sum_{|n|>m} e^{it(n-k)} p_n dt \right| \le \int_0^{2\pi} \left| \sum_{|n|>m} e^{it(n-k)} p_n \right| dt \\ &\le \int_0^{2\pi} \sum_{|n|>m} \left| e^{it(n-k)} \right| p_n dt = 2\pi \sum_{|n|>m} p_n \to 0 \end{aligned}$$

as $n \to \infty$ since $\sum_{n\in\mathbb{Z}} p_n = 1$.

Finally, we have

$$\int_0^{2\pi} e^{-itk}\phi_X(t)dt = \sum_{n\in\mathbb{Z}} p_n \int_0^{2\pi} e^{it(n-k)}dt = 2\pi p_k$$

since

$$\begin{aligned} \int_0^{2\pi} e^{it(n-k)}dt &= \int_0^{2\pi} \cos(n-k)tdt + i\int_0^{2\pi} \sin(n-k)tdt \\ &= \begin{cases} 2\pi & \text{if } n = k \\ 0 & \text{if } n \neq k. \end{cases} \end{aligned}$$

It follows that

$$P(X = k) = \frac{1}{2\pi} \int_0^{2\pi} e^{-itk} \phi_X(t) dt.$$

5.14. Let S be the number of 'heads' obtained in $n = 10\,000$ tosses of a fair coin. Then $S = \sum_{i=1}^{10\,000} X_i$, where X_i, $i = 1, \ldots, 10\,000$ are i.i.d. random variables with $P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$. Furthermore, $\mathrm{Var}(S) = \sqrt{np(1-p)} = \sqrt{10\,000/4} = 50$. By Corollary 5.53, we have

$$P(4900 < S < 5100) = P(|S - 500| < 100) = P\left(\frac{|S - 500|}{50} < \frac{1}{2}\right)$$
$$\approx \phi(\frac{1}{2}) - \phi(-\frac{1}{2}) = 2\phi(\frac{1}{2}) - 1,$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$. It follows that $P(4900 < S < 5100) \approx 2 \cdot 0.6915 - 1 = 0.383$.