Exercises and solutions: Chapter 2 only

Sébastien Ott

October 27, 2025

Exercise 2.1. In each of the following cases, give a realisation set Ω and a probability mass function $p : \Omega \to [0,1]$ modelling the situation.

1. Simultaneous tossing of a 1chf coin, a 2chf coin and a 5chf coin.
2. We do the following procedure: first toss a biased coin that does head twice as often as it does tail. If the coin shows "head", roll a fair 4 faces dice. If the coin shows "tail", roll a fair 6 faces dice.
3. Toss two fair dice and look at the total number of points obtained.

Solution 2.1.

1. One can take as realisation space
$$\Omega = \{H,T\} \times \{H,T\} \times \{H,T\},$$
the first coordinate representing the result of the 1chf coin, the second coordinate the result of the 2chf coin, and the last coordinate gives the result of the 5chf coin. The probability mass function is the the uniform probability measure on Ω as all coins are fair:
$$p(\omega) = \frac{1}{|\Omega|} = \frac{1}{2^3} = \frac{1}{8}, \ \forall \omega \in \Omega.$$

2. The result of our experiment is a pair "result of the coin" and "number of points on the face of a dice". We have several options, we illustrate two here.
   - We could take $\Omega = \{H,T\} \times \{1,2,3,4,5,6\}$, and p given by
$$p(T,i) = \frac{1}{3} \cdot \frac{1}{6}, \ i = 1, \dots, 6,$$
$$p(H,i) = \frac{2}{3} \cdot \frac{1}{4}, \ i = 1, \dots, 4, \quad p(H,5) = p(H,6) = 0.$$
   - We could take
$$\Omega = \{(H,1),(H,2),(H,3),(H,4),(T,1),(T,2),(T,3),(T,4),(T,5),(T,6)\},$$
and p given by
$$p(T,i) = \frac{1}{2} \cdot \frac{1}{6}, \ i = 1, \dots, 6,$$
$$p(H,i) = \frac{1}{2} \cdot \frac{1}{4}, \ i = 1, \dots, 4.$$

3. We only look at the total number of points, which ranges from $2 = 1+1$ to $12 = 6+6$, so we can take $\Omega = \{2,3,\dots,12\}$ and p given by
$$p(2) = p(12) = \frac{1}{36}, \ p(3) = p(11) = \frac{2}{36}, \ p(4) = p(10) = \frac{3}{36},$$
$$p(5) = p(9) = \frac{4}{36}, \ p(6) = p(8) = \frac{5}{36}, \ p(7) = \frac{6}{36},$$
where the 36 is the total number of cases for the result of the two dices, and the numerators correspond to the number of cases which give the wanted number of points.

Exercise 2.2. Let Ω be finite and $p : \Omega \to [0, 1]$ be a probability mass function.

1. Show that $P_p(\Omega) = 1$, $P_p(\varnothing) = 0$.
2. Show that if $A, B \subset \Omega$ are such that $A \cap B = \varnothing$, $P_p(A \cup B) = P_p(A) + P_p(B)$.
3. Let $n \geq 2$. Show that if $A_1, \ldots, A_n \subset \Omega$ are such that $A_i \cap A_j = \varnothing$ for all $i \neq j$, then $P_p(A_1 \cup A_2 \cup \cdots \cup A_n) = P_p(A_1) + P(A_2) + \cdots + P_p(A_n)$. *Hint: remember what is a proof by induction.*
4. Show that for any $A \subset \Omega$, $P_p(A) = 1 - P_p(\Omega \setminus A)$.
5. Show that for any $A, B \subset \Omega$, $P_p(A \cup B) = P_p(A) + P_p(B) - P_p(A \cap B)$.

Solution 2.2.

1. We have
$$P_p(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1,$$
by definition of a mass function. On the other hand,
$$\sum_{\omega \in \varnothing} p(\omega) = 0$$
as it is an empty sum.

2. As $A \cap B = \varnothing$, for any $\omega \in \Omega$, $\mathbb{1}_{A \cup B}(\omega) = \mathbb{1}_A(\omega) + \mathbb{1}_B(\omega)$. Thus,
$$\begin{aligned} P_p(A \cup B) &= \sum_{\omega \in A \cup B} p(\omega) = \sum_{\omega \in \Omega} \mathbb{1}_{A \cup B}(\omega) p(\omega) = \sum_{\omega \in \Omega} \big(\mathbb{1}_A(\omega) + \mathbb{1}_B(\omega)\big) p(\omega) \\ &= \sum_{\omega \in \Omega} \mathbb{1}_A(\omega) p(\omega) + \sum_{\omega \in \Omega} \mathbb{1}_B(\omega) p(\omega) = \sum_{\omega \in A} p(\omega) + \sum_{\omega \in B} p(\omega) = P_p(A) + P_p(B) \end{aligned}$$
where we used that sums are linear.

3. The previous point show the claim when $n = 2$. We suppose that the equality is true for $n \geq 2$, and let us show that is it true for $n + 1$. By assumption, $B = A_1 \cup \cdots \cup A_n$ and A_{n+1} are such that $B \cap A_{n+1} = \varnothing$. By the previous point, we have
$$\begin{aligned} P_p(B \cup A_{n+1}) = P_p(B) + P_p(A_{n+1}) &= P_p(A_1 \cup \cdots \cup A_n) + P_p(A_{n+1}) \\ &= P_p(A_1) + \cdots + P(A_n) + P(A_{n+1}) \end{aligned}$$
as we supposed the equality true for n.

4. For any $\omega \in \Omega$, $\mathbb{1}_A(\omega) = 1 - \mathbb{1}_{\Omega \setminus A}$. So,
$$\begin{aligned} P_p(A) = \sum_{\omega \in \Omega} \mathbb{1}_A(\omega) p(\omega) &= \sum_{\omega \in \Omega} \big(1 - \mathbb{1}_{\Omega \setminus A}(\omega)\big) p(\omega) \\ &= 1 - \sum_{\omega \in \Omega} \mathbb{1}_{\Omega \setminus A}(\omega) p(\omega) = 1 - P_p(\Omega \setminus A). \end{aligned}$$

5. Proceed as in points 2. and 4. using that for any $\omega \in \Omega$, we have that

$$\mathbb{1}_{A\cup B}(\omega) = \mathbb{1}_A(\omega) + \mathbb{1}_B(\omega) - \mathbb{1}_{A\cap B}(\omega).$$

Exercise 2.3. Which of the following p are probability mass functions[a]?

1. $\Omega = \{0, 1, \ldots, L\}$, $p(n) = \frac{1}{L}$ for $n \in \Omega$.
2. $\Omega = \mathbb{N}^*$, $p(n) = \frac{6}{\pi^2 n^2}$ for $n \in \Omega$.
3. $\Omega = \mathbb{Z}^*$, $p(n) = \frac{1}{n^3}$ for $n \in \Omega$.
4. $\Omega = \mathbb{N}$, $p(n) = 2^{-n-1}$ for $n \in \Omega$.

[a] Recall that $\sum_{n=1}^{\infty} n^{-2} = \frac{\pi^2}{6}$.

Solution 2.3. 1. This is not a probability mass function as it does not sum to 1:

$$\sum_{k=0}^{L} \frac{1}{L} = \frac{L+1}{L} > 1.$$

2. This is a probability mass function: $p(n) \geq 0$ for ever $n \in \Omega$, and

$$\sum_{n\in\Omega} p(n) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} k^{-2} = 1.$$

3. This is not a probability mass function as $p(-1) = -1 < 0$.

4. This is a probability mass function: $p(n) \geq 0$ and

$$\sum_{n\in\Omega} p(n) = \frac{1}{2} \sum_{k=0}^{\infty} 2^{-k} = \frac{1}{2} \frac{1}{1-\frac{1}{2}} = 1.$$

Exercise 2.4. We have a bag containing 10 blue balls and 15 red ones. We pick four of them out of the bag. All the balls are assumed to have the same shape and weight (uniform sampling).

1. What is the probability that we picked two blue and two red balls?
2. What is the probability that we picked **at most** one red ball?
3. What is the probability that we picked **at least** one blue ball?

Solution 2.4. We start by computing the total number of possible draws. We can suppose that all the balls are numbered from 1 to 25 with the blue going from 1 to 10 and the red from 11 to 25. In that case, there are $\binom{25}{4}$ possible draws.

1. We need to pick two balls out of the 10 blues, which gives $\binom{10}{2}$ possibilities, and two out of the 15 red, which gives $\binom{15}{2}$. The probability is thus

$$\frac{\binom{10}{2} \cdot \binom{15}{2}}{\binom{25}{4}} = \frac{10!15!21!4!}{2!8!2!13!25!} = \frac{10 \cdot 9 \cdot 15 \cdot 14 \cdot 6}{25 \cdot 24 \cdot 23 \cdot 22} = \frac{189}{506} \approx 0.37.$$

2. We need to pick either 0 red ball or 1 red ball. The first case has $\binom{10}{4}$ favourables cases (pick 4 balls amongst the 10 blue), the second has $\binom{10}{3} \cdot \binom{15}{1}$ favourable cases. This gives a probability of

$$\frac{\binom{10}{4} + \binom{10}{3} \cdot \binom{15}{1}}{\binom{25}{4}} = \frac{201}{1265} \approx 0.16.$$

3. The wanted probability is 1 minus the probability of picking 0 blue ball. The later has $\binom{15}{4}$ favourable cases (4 red balls amongst 15), so the wanted probability is

$$1 - \frac{\binom{15}{4}}{\binom{25}{4}} = \frac{2257}{2530} \approx 0.89.$$

Exercise 2.5. In which cases is the function f a probability density function?

1. $\Omega = \mathbb{R}$, $f(x) = \mathbb{1}_{[0,3]}(x)$.
2. $\Omega = \mathbb{R}$, $f(x) = \frac{x}{4}\mathbb{1}_{[-1,3]}$.
3. $\Omega = \mathbb{R}$, $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
4. $\Omega = \mathbb{R}^2$, $f(x, y) = e^{-\sqrt{x^2+y^2}}$.
5. $\Omega = \mathbb{R}^2$, $f(x, y) = \mathbb{1}_{x\geq 0}\mathbb{1}_{y\geq 0}e^{-x-y}$.

Solution 2.5.

1. This is not a probability density function as $\int_{-\infty}^{+\infty} f(x)dx = \int_0^3 dx = 3 \neq 1$.
2. This is not a probability density function as $f(-1/2) = \frac{-1}{8} < 0$.
3. This is a probability density function: $f(x) \geq 0$ for all x, and

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)dx &= \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{1}{1+x^2}dx \\ &= \frac{1}{\pi}\lim_{R_-,R_+\to\infty}\big(\arctan(R_+) - \arctan(-R_-)\big) = \frac{1}{\pi}\Big(\frac{\pi}{2}+\frac{\pi}{2}\Big) = 1.\end{aligned}$$

4. This is not a probability density function: as it does not integrate to 1:

$$\begin{aligned}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\sqrt{x^2+y^2}}dxdy &= \int_0^{2\pi}\int_0^{+\infty} re^{-r}drd\theta \\ &= 2\pi\int_0^{+\infty} re^{-r}dr = 2\pi\int_0^{+\infty} e^{-r}dr = 2\pi\end{aligned}$$

where we changed to polar coordinates: $x = r\cos(\theta)$, $y = r\sin(\theta)$, $dxdy = rd\theta dr$, and integrated by parts in the fourth equality.

5. This is a probability density function: $f(x, y) \geq 0$ for all x, y, and

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = \int_0^{+\infty}\int_0^{+\infty} e^{-x-y}dxdy = \Big(\int_0^{+\infty} e^{-x}dx\Big)^2 = 1.$$

Exercise 2.6. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \frac{1}{2\pi}\frac{1}{1+x^2}e^{-|y|}.$$

1. Show that f is a probability density function on $\Omega = \mathbb{R}^2$.
2. Show that for any events $A, B \subset \mathbb{R}$, we have

$$P_f(A \times B) = P_f(A \times \mathbb{R})P_f(\mathbb{R} \times B).$$

In particular, "the first and second coordinates are independent".

Solution 2.6.

$$f(x,y) = \frac{1}{2\pi}\frac{1}{1+x^2}e^{-|y|}.$$

1. We have that $f(x,y) \geq 0$ for all x, y, so we need to check that it integrates to 1. We have

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{1+x^2}e^{-|y|}dxdy$$
$$= \frac{1}{2\pi}\Big(\int_{-\infty}^{+\infty}\frac{1}{1+x^2}dx\Big)\Big(\int_{-\infty}^{+\infty}e^{-|y|}dy\Big)$$
$$= \frac{1}{2\pi}\Big(\lim_{L,R\to\infty}\big(\arctan(R)-\arctan(-L)\big)\Big)\Big(2\int_{0}^{+\infty}e^{-y}dy\Big) = \frac{1}{2\pi}\Big(\frac{\pi}{2}+\frac{\pi}{2}\Big)\cdot 2 = 1.$$

2. For any events $A, B \subset \mathbb{R}$, we have

$$P_f(A \times B) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mathbb{1}_A(x)\mathbb{1}_B(y)f(x,y)dxdy$$
$$= \Big(\int_{-\infty}^{+\infty}\mathbb{1}_A(x)\frac{dx}{\pi(1+x^2)}\Big)\Big(\int_{-\infty}^{+\infty}\mathbb{1}_B(y)\frac{e^{-|y|}dy}{2}\Big).$$

But now, as (see the first point) $\int_{-\infty}^{+\infty}\frac{e^{-|y|}dy}{2} = 1$, we have

$$\int_{-\infty}^{+\infty}\mathbb{1}_A(x)\frac{dx}{\pi(1+x^2)} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mathbb{1}_A(x)\frac{dx}{\pi(1+x^2)}\frac{e^{-|y|}dy}{2}$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mathbb{1}_{A\times\mathbb{R}}(x,y)f(x,y)dxdy = P_f(A \times \mathbb{R}),$$

and proceed similarly to obtain $\int_{-\infty}^{+\infty}\mathbb{1}_B(y)\frac{e^{-|y|}dy}{2} = P_f(\mathbb{R} \times B)$.

***Exercise 2.7.** Consider the set $\{1,\dots,n\}$. Let $\mathcal{S}_n$ be the set of permutation of $1,\dots,n$ (i.e.: the set of bijections from $\{1,\dots,n\}$ to $\{1,\dots,n\}$). Call $\phi \in \mathcal{S}_n$ a *derangement* if ϕ has no fixed point[a]. Denote

$$D_{n,k} = \{\phi \in \mathcal{S}_n : \ \phi \text{ has exactly } k \text{ fixed points}\},$$

so that the set of derangements is $D_{n,0}$. Let P_n be the uniform probability measure on $\mathcal{S}_n$.

1. Enumerate all elements of $\mathcal{S}_3$. What are the values of $P_3(D_{3,0})$, $P_3(D_{3,1})$, $P_3(D_{3,2})$, $P_3(D_{3,3})$?
2. What is $|\mathcal{S}_n|$?

For $1 \le k \le n$ and $1 \le i_1 < i_2 < \cdots < i_k \le n$, denote

$$A_{n;i_1,\dots,i_k} = \{\phi \in \mathcal{S}_n : \ \phi(i_j) = i_j, j = 1,\dots,k\},$$

the set of permutations which fix $i_1,\dots,i_k$.

3. Show that for any fixed $i_1,\dots,i_k$ as above,

$$P_n(A_{n;i_1,\dots,i_k}) = \tfrac{(n-k)!}{n!}.$$

4. Show that $\mathcal{S}_n \setminus D_{n,0} = \cup_{i=1}^n A_{n;i}$.
5. Show that

$$P_n(D_{n,0}) = 1 + \sum_{k=1}^{n} (-1)^k \frac{1}{k!},$$

and deduce that $\lim_{n\to\infty} P_n(D_{n,0}) = e^{-1}$.

[a] A fixed point of ϕ is a number k such that $\phi(k) = k$.

Solution 2.7.
1. We can represent a permutation by a list of n elements: the first is the image of 1, the second is the image of 2 and so on and so forth. We have

$$\mathcal{S}_3 = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,2,1),(3,1,2)\}.$$

We then get

$$P_3(D_{3,0}) = \frac{2}{6} = \frac{1}{3},\ P_3(D_{3,1}) = \frac{3}{6} = \frac{1}{2},\ P_3(D_{3,2}) = \frac{0}{6} = 0,\ P_3(D_{3,3}) = \frac{1}{6}.$$

2. $n!$: there are n possible images for 1, $n-1$ for 2 (n minus the one that was picked for 1), $n-2$ for 3, and so one and so forth.

3. For $1 \le i_1 < \cdots < i_k \le n$, we have that permutations fixing $I = \{i_1,\dots,i_k\}$ are in bijection with permutations of $I^c = \{1,\dots,n\} \setminus I$: one simply restricts the first sort of permutation to I^c to obtain a permutation of the second sort, and complete a permutation of the second sort by setting it to be the identity on I to

obtain a permutation of the first sort. As there are $|I^c|! = (n-k)!$ permutations of the elements of I^c, we obtain

$$P_n(A_{n;i_1,\dots,i_k}) = \tfrac{(n-k)!}{n!}.$$

4. Every permutation that is not a derangement possess at least one fixed point, so $\mathcal{S}_n \setminus D_{n,0} \subset \cup_{i=1}^n A_{n;i}$. Moreover, if $\phi \in \cup_{i=1}^n A_{n;i}$, then there is i_* such that $\phi \in A_{n;i_*}$, so ϕ fixes i_* and is therefore not a derangement, so $\phi \in \mathcal{S}_n \setminus D_{n,0}$, which gives $\mathcal{S}_n \setminus D_{n,0} \supset \cup_{i=1}^n A_{n;i}$.

5. By the complement formula and the previous point,

$$P_n(D_{n,0}) = 1 - P\big(\mathcal{S}_n \setminus D_{n,0}\big) = 1 - P\big(\cup_{i=1}^n A_{n;i}\big).$$

Now, from the Inclusion-Exclusion principle (Lemma **??**), we obtain

$$\begin{aligned} P\big(\cup_{i=1}^n A_{n;i}\big) &= \sum_{k=1}^n (-1)^{k+1} \sum_{1\le i_1<\dots<i_k\le n} P\big(\cap_{j=1}^k A_{n;i_j}\big) \\ &= \sum_{k=1}^n (-1)^{k+1} \sum_{1\le i_1<\dots<i_k\le n} P\big(A_{n;i_1,\dots,i_k}\big) \\ &= \sum_{k=1}^n (-1)^{k+1} \sum_{1\le i_1<\dots<i_k\le n} \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^n (-1)^{k+1} \frac{(n-k)!}{n!} \binom{n}{k} = -\sum_{k=1}^n (-1)^k \frac{1}{k!} \end{aligned}$$

where we used the third point and the fact that $\sum_{1\le i_1<\dots<i_k\le n} 1 = \binom{n}{k}$ (the sum counts the number of ways to pick k elements in $\{1,\dots,n\}$). Plugging this in the previous display, we obtain

$$P_n(D_{n,0}) = 1 + \sum_{k=1}^n (-1)^k \frac{1}{k!}.$$

Notice that this is the Taylor series of the exponential function truncated at rank n and evaluated at -1. By convergence of the Taylor series, we obtain the wanted limit.

Exercise 2.8. Suppose that $X : \Omega \to \mathbb{R}$ is a random variable such that $\text{Image}(X) = \{2, 5, 7\}$. Let $p_2 = P(X = 2)$, $p_5 = P(X = 5)$, $p_7 = P(X = 7)$.

1. Show that X is a discrete random variable (i.e.: what set $\mathcal{D}_X$ can we take?).

2. What is the value of $p_2 + p_5 + p_7$?

3. Check that the function $P_X : \mathcal{P}(\mathbb{R}) \to [0, 1]$ defined in Definition **??** is a probability measure on $\mathbb{R}$.

As functions of p_2, p_5, p_7, what are the values of

4. $P(X \in [0, 6])$?

5. $P(X \le 3)$?

Solution 2.8. Suppose that $X : \Omega \to \mathbb{R}$ is a random variable such that $\text{Image}(X) = \{2, 5, 7\}$. Let $p_2 = P(X = 2)$, $p_5 = P(X = 5)$, $p_7 = P(X = 7)$.

1. We can take the set $\mathcal{D}_X = \{2, 5, 7\}$ as $P(X \in \text{Image}(X)) = 1$.

2. We have

$$p_2 + p_5 + p_7 = P(X = 2) + P(X = 5) + P(X = 7) = P(X \in \{2, 5, 7\}) = 1.$$

3. We have that $P_X(\varnothing) = P(X \in \varnothing) = 0$ and $P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1$. We only have to check additivity. Let $A_1, A_2, \dots \subset \mathbb{R}$ be disjoint. Then, the events $\{X \in A_1\}, \{X \in A_2\}, \dots$ are disjoint events (as $X : \Omega \to \mathbb{R}$ is a function). Thus, as P is a probability measure,

$$P_X\Big(\cup_{i\geq 1} A_i\Big) = P\Big(X \in \cup_{i\geq 1} A_i\Big) = P\Big(\cup_{i\geq 1}\{X \in A_i\}\Big)$$
$$= \sum_{i\geq 1} P(X \in A_i) = \sum_{i\geq 1} P_X(A_i)$$

where we used that an element belong to the union of disjoint sets if and only if it belongs to exactly one of them.

4. $P(X \in [0, 6]) = P(X \in [0, 6] \cap \{2, 5, 7\})$ as $P(X \in \{2, 5, 7\}) = 1$. So,

$$P(X \in [0, 6]) = P(X \in \{2, 5\}) = p_2 + p_5.$$

5. $P(X \leq 3) = P(X \in (-\infty, 3]) = P(X \in (-\infty, 3] \cap \{2, 5, 7\})$ as in the previous point, so $P(X \leq 3) = p_2$.

Exercise 2.9. Let $\Omega = \mathbb{R}$, $f(x) = \frac{1}{4}\mathbb{1}_{[0,4]}(x)$. Let $X : \mathbb{R} \to \mathbb{R}$ be given by $X(x) = x^2$.

1. What is the value of $P_f(X \leq 4)$? and of $P_f(X \leq 9)$?

2. Find a density function f_X such that $P_f(X \in A) = \int_{\mathbb{R}} \mathbb{1}_A(x) f_X(x)dx$.

Solution 2.9.

1. We use the definition:

$$P_f(X \leq 4) = P_f\big(\{x :\ X(x) \leq 4\}\big) = P_f\big(\{x :\ x^2 \leq 4\}\big) = P_f([-2, 2]) = \int_{-2}^{2} f(x)dx = \frac{1}{2}$$

as $x^2 \leq 4$ if and only if $x \in [-2, 2]$. We proceed similarly to obtain $P_f(X \leq 9) = \frac{3}{4}$.

2. Writing the definitions, we want to find f_X such that for any set A:

$$P_f(X \in A) = P_f\big(\{x :\ x^2 \in A\}\big) = \int_{\mathbb{R}} \mathbb{1}_A(x^2) f(x)dx$$
$$= \frac{1}{4}\int_0^4 \mathbb{1}_A(x^2)dx = \int_{\mathbb{R}} \mathbb{1}_A(x) f_X(x)dx.$$

We can then make the change of variable $y = x^2$ to obtain

$$\frac{1}{4}\int_0^4 \mathbb{1}_A(x^2)dx = \frac{1}{4}\int_0^{16} \mathbb{1}_A(y)\frac{1}{2\sqrt{y}}dy = \frac{1}{4}\int_{\mathbb{R}} \mathbb{1}_A(y)\mathbb{1}_{[0,16]}(y)\frac{1}{2\sqrt{y}}dy.$$

We thus find that we can take $f_X(x) = \mathbb{1}_{[0,16]}(x)\frac{1}{8\sqrt{x}}$.

Exercise 2.10. For each of the following random variable, say whether it is discrete of continuous.

1. The number of rainy days in Lausanne during March.
2. The volume of rain in Lausanne during March.
3. The number of problems that you will correctly solve during the exam.
4. The number of points that you will get at the exam.
5. The time you will need to complete the exam.

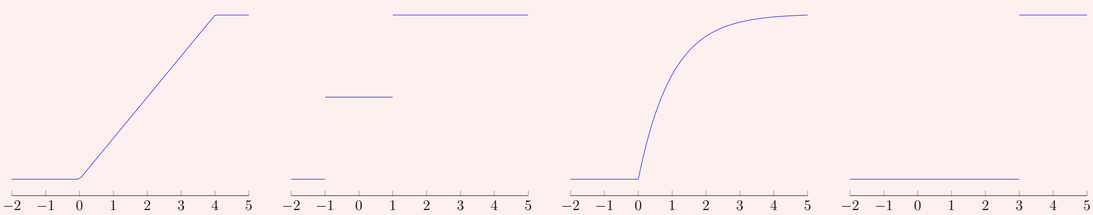
Solution 2.10.

1. A number of days is an integer, in particular, the variable takes value in (a subset of) a countable set, it is therefore discrete.
2. The volume of rain is a continuous non-negative quantity, it thus takes value in the whole of $[0, +\infty)]$ and is thus a continuous random variable.
3. This is again an integer, so discrete RV as in the first point.
4. This is again an integer, so discrete RV as in the first point.
5. This is a continuous random variable (you can take anywhere between 0 and the max duration of the exam), so continuous random variable.

Exercise 2.11. Let $X : \Omega \to \mathbb{R}$ be a random variable. Compute F_X in each of the following cases.

1. X is such that $P(X = -1) = P(X = 1) = \frac{1}{2}$.
2. X is such that $P(X = 3) = 1$.
3. X is a continuous random variable with density
$$f_X(x) = \frac{1}{4}\mathbb{1}_{[0,4]}(x).$$
4. X is a continuous random variable with density
$$f_X(x) = \mathbb{1}_{[0,+\infty)}(x)e^{-x}.$$

Associate each of the above with one of the following graphs.



(a) (b) (c) (d)

Solution 2.11.

1. We have that $P(X \in \{-1, 1\}) = 1$, so,

- for $t < -1$, $P(X \leq t) = 0$,
- for $-1 \leq t < 1$, $P(X \leq t) = P(X = -1) = \frac{1}{2}$,
- for $t \geq 1$, $P(X \leq t) = P(X \in \{-1, 1\}) = 1$.

The associated graph is the (b).

2. This is a constant variable, so

- for $t < 3$, $P(X \leq t) = 0$,
- for $t \geq 3$, $P(X \leq t) = P(X = 3) = 1$.

The associated graph is (d).

3. We can compute:

$$P(X \leq t) = \int_{-\infty}^{t} f_X(x)dx = \frac{1}{4}\int_{-\infty}^{t} \mathbb{1}_{[0,4]}(x)dx = \begin{cases} 0 & \text{if } t < 0, \\ \frac{t}{4} & \text{if } t \in [0, 4], \\ 1 & \text{if } t > 4. \end{cases}$$

The associated graph is (a).

4. We can compute:

$$\begin{aligned} P(X \leq t) = \int_{-\infty}^{t} f_X(x)dx = \int_{-\infty}^{t} \mathbb{1}_{[0,+\infty)}(x)e^{-x}dx \\ = \begin{cases} 0 & \text{if } t < 0, \\ \int_0^t e^{-x}dx = 1 - e^{-t} & \text{if } t \geq 0 \end{cases} \end{aligned}$$

The associated graph is (c).

***Exercise 2.12.** Let $X : \Omega \to \mathbb{R}$ be a random variable. Suppose that $|\text{Image}(X)| < \infty$, and that for any $x \in \text{Image}(X)$, $P(X = x) > 0$.

1. Show that there is $\epsilon > 0$ such that for any $x \in \text{Image}(X)$,

$$P(X = x) = P\big(X \in (x - \epsilon, x + \epsilon)\big).$$

2. What are the discontinuity points of F_X, the cumulative distribution function of X?

3. For every $x \in \text{Image}(X)$, express $P(X = x)$ in terms of F_X.

4. Let $Y : \Omega \to \mathbb{R}$ be a random variable. Suppose that $F_Y(t) = F_X(t)$ for every $t \in \mathbb{R}$. Show that for every event $A \subset \mathbb{R}$,

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

Solution 2.12. 1. For $x, y \in \text{Image}(X)$, consider the distance between them $|x-y|$. If $x \neq y$, $|x-y| > 0$. We can take

$$\epsilon = \frac{1}{3}\min\{|x-y| : \ x, y \in \text{Image}(X), \ x \neq y\} > 0$$

as it is the minimum of finitely many positive numbers. Indeed, for any $x \in \text{Image}(X)$, we have that $(x-\epsilon, x+\epsilon) \cap \text{Image}(X) = \{x\}$ as the distance between x and any other points in $\text{Image}(x)$ is at least 3ϵ. So,

$$P\big(X \in (x-\epsilon, x+\epsilon)\big) = P\big(X \in (x-\epsilon, x+\epsilon) \cap \text{Image}(X)\big) = P(X \in \{x\}) = P(X = x).$$

2. They are precisely $\text{Image}(X)$ as for any $t \notin \text{Image}(X)$, $F_X(t) = P(X \leq x(t))$ where $x(t)$ is the smallest point in $\text{Image}(X)$ which is larger than t. In particular, F_X is constant between two successive points in $\text{Image}(X)$, if is therefore continuous on $\mathbb{R} \setminus \text{Image}(X)$. On the other side, if $t \in \text{Image}(X)$, we have that for any $\delta > 0$,

$$\begin{aligned}|F_X(t) - F_X(t-\delta)| = F_X(t) - F_X(t-\delta) &= P(X \leq t) - P(X \leq t-\delta) \\ &= P(X \in (t-\delta, t]) \geq P(X = t) > 0\end{aligned}$$

as we assume the X takes any of its values with strictly positive probability. In particular, this shows that F_X is not left continuous at t.

3. Using the same ϵ as in the first point, we have that

$$P(X = x) = P(X \in (x-\epsilon, x+\epsilon]) = F_X(x+\epsilon) - F_X(x-\epsilon).$$

4. Use the same ϵ as in the first point. For any $\epsilon' \in (0, \epsilon]$, we have (recall that by assumption, $F_Y = F_X$)

$$\begin{aligned}P(Y \in (x-\epsilon', x+\epsilon']) = F_Y(x+\epsilon') - F_Y(x-\epsilon') &= F_X(x+\epsilon') - F_X(x+\epsilon') \\ &= P(X \in (x-\epsilon', x+\epsilon']).\end{aligned}$$

Letting $\epsilon' \to 0$, we obtain that for any $x \in \text{Image}(X)$, $P(Y = x) = P(X = x)$. In particular,

$$P(Y \in \text{Image}(X)) = \sum_{x \in \text{Image}(X)} P(Y = x) = \sum_{x \in \text{Image}(X)} P(X = x) = 1.$$

So, Y is a discrete random variable (we can take $\mathcal{D}_Y = \text{Image}(X)$), and for every event $A \subset \mathbb{R}$,

$$P(Y \in A) = \sum_{x \in A \cap \text{Image}(X)} P(Y = x) = \sum_{x \in A \cap \text{Image}(X)} P(X = x) = P(X \in A).$$

Exercise 2.13. In each of the following situations, give a probability space associated to the situation (the probability measure does not need to be an explicit one), and random variables giving the quantities of interest.

1. We test a new model of graphic cards. We have 1000 cards of this model to perform our test on, and we want to know how likely it is for a card to reach 100 degrees during the test.

2. We pick 100 humans uniformly at random in the population and give them an IQ test. We wonder how what is the best and worst IQ amongst them.

Solution 2.13. 1. We can take $\Omega = \mathbb{R}^{1000}$, each coordinate representing the temperature reached by the corresponding card. The probability measure can be taken to be a continuous probability measure

$$P(A) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_A(x_1, \ldots, x_{1000}) \prod_{i=1}^{1000} f(x_i) dx_i$$

where f is some density function on $\mathbb{R}$ which is 0 on the negative numbers. The product structure reflects the fact that the results of the tests are independent, and the fact that we use the same f for different cards reflects the fact that they are all of the same model (and so their temperatures share the same law). We can then take the random variable "average number of cards reaching 100 degrees":

$$X : \Omega \to \mathbb{R}, \quad X(x_1, \ldots, x_{1000}) = \frac{1}{1000} \sum_{i=1}^{1000} \mathbb{1}_{[100,+\infty)}(x_i)$$

2. Denote by Pop the world population (elements of Pop are living individuals). For $i \in \text{Pop}$, let IQ_i denote the IQ of i. Let then $\text{IQPop} = \{(i, IQ_i) : i \in \text{Pop}\}$. We can take Ω to be the set of subsets of IQPop with size 100:

$$\Omega = \{A : A \subset \text{IQPop}, |A| = 100\}.$$

The probability measure can then be taken to be the uniform probability measure on Ω: $P(A) = |A|/|\Omega|$ for any $A \subset \Omega$. The random variables of interest are then

- the maximal IQ of an individual in our sample: for $\omega \in \Omega$,

$$\text{MaxIQ}(\omega) = \max\{IQ_i : (i, IQ_i) \in \omega\}.$$

- the minimal IQ of an individual in our sample: for $\omega \in \Omega$,

$$\text{MinIQ}(\omega) = \min\{IQ_i : (i, IQ_i) \in \omega\}.$$

Exercise 2.14. Let Ω be a **finite** realisation set, $p : \Omega \to [0, 1]$ a probability mass function, and P_p the associated probability measure on Ω. Let $X, Y : \Omega \to \mathbb{R}$ be two random variables.

1. Show that for $a, b \in \mathbb{R}$, $E_{P_p}(aX + bY) = aE_{P_p}(X) + bE_{P_p}(Y)$.
2. Show that if $X \leq Y$, then $E_{P_p}(X) \leq E_{P_p}(Y)$.

Solution 2.14. Recall that in the present case,

$$E_{P_p}(X) = \sum_{\omega \in \Omega} p(\omega) X(\omega),$$

and that the sum is finite.

1. For $a, b \in \mathbb{R}$,

$$\begin{aligned} E_{P_p}(aX + bY) &= \sum_{\omega\in\Omega} p(\omega)(aX(\omega) + bY(\omega)) \\ &= a\sum_{\omega\in\Omega} p(\omega)X(\omega) + b\sum_{\omega\in\Omega} p(\omega)Y(\omega) = aE_{P_p}(X) + bE_{P_p}(Y) \end{aligned}$$

where we used linearity of finite sums.

2. For $X \leq Y$,

$$E_{P_p}(X) = \sum_{\omega\in\Omega} p(\omega)X(\omega) \leq \sum_{\omega\in\Omega} p(\omega)Y(\omega) = E_{P_p}(Y),$$

as $a \leq a'$ and $b \leq b'$ implies $a + b \leq a' + b'$, and $p(\omega)X(\omega) \leq p(\omega)Y(\omega)$ as $p(\omega) \geq 0$ and $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$.

Exercise 2.15. Let X the the random variable which gives the life duration of a television (in years). Assume that X is a continuous random variable with density given by

$$f_X(x) = \mathbb{1}_{[0,+\infty)}(x)\tfrac{1}{10}e^{-x/10}$$

1. What is the probability that the television lasts at least 10 years?
2. What is the average time you will be able to keep your television without repairing it?

Solution 2.15. 1. This is the probability that $X \geq 10$, which is given by

$$P(X \geq 10) = \int_{-\infty}^{\infty} f_X(x)\mathbb{1}_{[10,+\infty)}(x)dx = \tfrac{1}{10}\int_{10}^{+\infty} e^{-x/10}dx = [-e^{-x/10}]_{10}^{+\infty} = e^{-1}.$$

2. This is given by the expected value of X:

$$\begin{aligned} E(X) = \int_{-\infty}^{\infty} xf_X(x)dx = \tfrac{1}{10}\int_0^{\infty} xe^{-x/10}dx &= [xe^{-x/10}]_0^{\infty} + \int_0^{\infty} e^{-x/10}dx \\ &= [-10e^{-x/10}]_0^{\infty} = 10, \end{aligned}$$

where we used integration by part.

Exercise 2.16. A casino wants to offer the following game. The player pays a sum of money, x CHF, to start playing. The player starts with a $G_0 = 2$ CHF initial gain. Then, at every step, the player can chose to either take the current gain, or to play a fair coin toss to double his gain or lose everything. The game stops when the player choses to exit or loses. Imagine the casino has infinite money, so that the game can carry on for as long as needed. What is the amount of money x that the player should pay in order for the casino to not be at lose on average?

Solution 2.16. The casino has to ask for the expected gain of the player in order not to be at lose on average. The player can chose when to stop: so there is a number N

(chosen by the player) such that he will stop as soon as he will have won N times (if he wins that many times at all). For a fixed N, the average gain of the player is

$$2^N G_0 2^{-N} + 0(1 - 2^{-N}) = 2$$

as there is a probability 2^{-N} that he will win N consecutive coin flips and $1 - 2^{-N}$ that he loses before that. So, independently of the value of N, the expected gain is 2, so the casino should ask to pay at least $x = 2$ to play the game.

Exercise 2.17. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = \frac{1}{2\pi} e^{-(x-1.7)^2/0.08} e^{-(y-100(x-1))^2/50}.$$

1. Check that f is a probability density function.[a]

Let $X = (X_1, X_2)$ be a continuous random vector with density f. One can think of X as modelling the height (in meters) and weight (in kilograms) of an individual in a given population (which "fact" about height and weight is not respected by this modelling?).

2. What is the density of the random variable X_1?
3. What is the average height in that population?
4. What is the average weight in that population?
5. What is the probability that an individual taken uniformly at random in the population is taller that 1.8m. whilst weighting less than 70kg.? (You can answer with a definite integral, your computer can estimate it later).

[a] You can use $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$, it will be computed later in the notes.

Solution 2.17. 1. One has that $f(x, y) \geq 0$ for all $x, y \in \mathbb{R}$. We then need to check that f integrates to 1.

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-(x-1.7)^2/0.08} e^{-(y-100(x-1))^2/50} dy dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x-1.7)^2/0.08} \int_{-\infty}^{\infty} e^{-(y')^2/2} 5 \cdot dy' dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x')^2/2} 0.2 \cdot dx' \int_{-\infty}^{\infty} e^{-(y')^2/2} 5 \cdot dy' \\
&= \frac{1}{2\pi} \Big(\int_{-\infty}^{\infty} e^{-x^2/2} dx \Big)^2 = 1.
\end{aligned}$$

where we changed variables: $y' = (y - 100(x - 1))/5$, and $x' = (x - 1.7)/0.2$ and used the indication.

2. X_1 is the first marginal of X. So, its density is given by

$$f_{X_1}(x) = \int_{-\infty}^{\infty} f(x,y)dy$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-(x-1.7)^2/0.08} e^{-(y-100(x-1))^2/50} dy$$
$$= \frac{1}{2\pi} e^{-(x-1.7)^2/0.08} \int_{-\infty}^{\infty} e^{-(y')^2/2} 5 \cdot dy' = \frac{5}{\sqrt{2\pi}} e^{-(x-1.7)^2/0.08}$$

where we changed variable to $y' = (y - 100(x-1))/5$.

3. The average height in that population is given by the expected value of X_1 which is (using the previous point):

$$\int_{-\infty}^{\infty} x f_{X_1}(x)dx = \frac{5}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-1.7)^2/0.08} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\tfrac{x'}{5} + 1.7) e^{-(x')^2/2} dx'$$
$$= 1.7 + \frac{1}{5\sqrt{2\pi}} \int_{-\infty}^{\infty} x' e^{-(x')^2/2} dx' = 1.7$$

where we changed variable to $x' = (x - 1.7)/0.2$, used the indication, and used that the integral of an odd function on a symmetric interval is 0.

4. The average weight in that population is given by the expected value of X_2 which is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)dydx = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y e^{-(x-1.7)^2/0.08} e^{-(y-100(x-1))^2/50} dydx$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-(x')^2/2} \int_{-\infty}^{\infty} (5y' + 100(0.2 \cdot x' + 1.7 - 1)) e^{-(y')^2/2} dy'dx'$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-x^2/2} \int_{-\infty}^{\infty} (5y + 20x + 70) e^{-y^2/2} dydx$$
$$= 70 + \frac{20}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} + \frac{5}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy = 70,$$

where we changed variable to $y' = (y - 100(x-1))/5$ and $x' = (x - 1.7)/0.2$.

5. This is the probability that $X_1 \geq 1.8$ and that $X_2 \leq 70$. This gives the expression

$$P(X_1 \geq 1.8,\ X_2 \leq 70) = P\big(X \in [1.8, +\infty)\times(-\infty, 70]\big) = \int_{1.8}^{+\infty}\int_{-\infty}^{70} f(x,y)dydx.$$

Exercise 2.18. Let $X = (X_1, X_2)$ be a random vector with density

$$f_X(x,y) = \frac{15}{31}\mathbb{1}_{[0,1]}(x)\mathbb{1}_{[0,1]}(y)(x+y)^4.$$

1. What is the density of X_1?
2. Give a density for the random vector

$$Y = (Y_1, Y_2) := (X_1 + X_2, X_1 - X_2).$$

3. Give a density for the random vector

$$Z = (Z_1, Z_2) := \big((X_1 + X_2)^2, (X_1 - X_2)^3\big).$$

Solution 2.18. 1. X_1 is the first marginal of X. Its density is thus given by

$$f_{X_1}(x) = \int_{-\infty}^{+\infty} f_X(x,y)dy = \frac{15}{31}\mathbb{1}_{[0,1]}(x)\int_{-\infty}^{+\infty} \mathbb{1}_{[0,1]}(y)(x+y)^4 dy$$
$$= \frac{15}{31}\mathbb{1}_{[0,1]}(x)\int_0^1 (x+y)^4 dy = \frac{15}{31}\mathbb{1}_{[0,1]}(x)\big[\tfrac{1}{5}(x+y)^5\big]_0^1 = \frac{3}{31}\mathbb{1}_{[0,1]}(x)((x+1)^5 - x^5)$$

2. We have that

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{=:M} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \qquad \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}}_{=M^{-1}} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$

In particular, Corollary **??** gives that the density for Y is

$$f_Y(y) = f_X(M^{-1}y)|\det(M^{-1})| = \frac{1}{2} f_X\big(\tfrac{y_1+y_2}{2}, \tfrac{y_1-y_2}{2}\big) = \frac{15}{62}\mathbb{1}_{[0,1]}\big(\tfrac{y_1+y_2}{2}\big)\mathbb{1}_{[0,1]}\big(\tfrac{y_1-y_2}{2}\big)y_1^4$$

3. We have that

$$Z = \big(Y_1^2, Y_2^3\big).$$

So, considering $\varphi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+ \times \mathbb{R}$, $\varphi(x,y) = (x^2, y^3)$, we have $Z = \varphi(Y)$ and

$$\varphi^{-1}(x,y) = (\sqrt{x}, y^{1/3}), \quad D_\varphi(x,y) = \begin{pmatrix} 2x & 0 \\ 0 & 3y^2 \end{pmatrix}.$$

So, $\det D_\varphi(x,y) = 6xy^2$, and thus Corollary **??** gives that the density for Z is

$$f_Z(x,y) = f_Y(\sqrt{x}, y^{1/3})\frac{1}{|\det D_\varphi(\sqrt{x}, y^{1/3})|}$$
$$= \frac{1}{6|\sqrt{x}y^{2/3}|}\frac{15}{62}\mathbb{1}_{[0,1]}\big(\tfrac{\sqrt{x}+y^{1/3}}{2}\big)\mathbb{1}_{[0,1]}\big(\tfrac{\sqrt{x}-y^{1/3}}{2}\big)x^2$$
$$= \frac{5x^{3/2}}{124|y|^{2/3}}\mathbb{1}_{[0,2]}\big(\sqrt{x} + y^{1/3}\big)\mathbb{1}_{[0,2]}\big(\sqrt{x} - y^{1/3}\big).$$

Exercise 2.19. Let X be a continuous random variable with density f_X. In each of the following cases, compute the density f_Y of $Y = g(X)$.

1. $f_X(x) = \frac{1}{\pi(1+x^2)}$ and $g(x) = x^3$.
2. $f_X(x) = \mathbb{1}_{[0,+\infty)}(x)e^{-x}$ and $g(x) = 25x^2$.
3. $f_X(x) = \frac{2}{\pi}\mathbb{1}_{[0,\pi/2]}(x)$ and $g(x) = \sin(x)$.

Solution 2.19. In each case, we apply Corollary **??** with $\varphi = g$.

1. We have $U = \mathbb{R}$, $V = \mathbb{R}$, $\varphi(x) = x^3$, $\varphi^{-1}(x) = x^{1/3}$, and thus
$$f_Y(x) = \frac{1}{3|x|^{2/3}} \frac{1}{\pi(1+x^{2/3})}.$$
2. We have $U = (0, +\infty)$, $V = (0, +\infty)$, $\varphi(x) = 25x^2$, $\varphi^{-1}(x) = \frac{1}{5}x^{1/2}$, and thus
$$f_Y(x) = \frac{1}{10|x|^{1/2}}\mathbb{1}_{[0,+\infty)}(x)e^{-\sqrt{x}/5}.$$
3. We have $U = (0, \pi/2)$, $V = (0, 1)$, $\varphi(x) = \sin(x)$, $\varphi^{-1}(x) = \arcsin(x)$, and thus
$$f_Y(x) = |\arcsin'(x)|\frac{2}{\pi}\mathbb{1}_{[0,\pi/2]}(\arcsin(x)) = \frac{1}{\sqrt{1-x^2}}\frac{2}{\pi}\mathbb{1}_{[0,1]}(x).$$

Exercise 2.20. Compute the first three moments of the random variable X in each of the following cases.

1. X is a continuous random variable with density $f_X(x) = \frac{9}{10}\mathbb{1}_{[1,10]}(x)\frac{1}{x^2}$.
2. X is a discrete random variable with $P(X = -1) = P(X = +2) = P(X = 0) = \frac{1}{3}$.
3. X is a discrete random variable with $P(X = -1) = P(X = +1) = \frac{1}{2}$.
4. X is a continuous random variable with $f_X(x) = \frac{1}{2}e^{-|x|}$.

Solution 2.20. The transfer theorem (Theorem **??**) gives that the pth moment of a continuous random variable X is given by
$$E(X^p) = \int_{-\infty}^{\infty} f_X(x)x^p dx,$$
and similarly for discrete random variables.

1. We have that the first moment of X is
$$E(X) = \int_{-\infty}^{\infty} x\frac{9}{10}\mathbb{1}_{[1,10]}(x)\frac{1}{x^2}dx = \frac{9}{10}\int_1^{10}\frac{1}{x}dx = \frac{9}{10}\ln(10).$$
The second moment is
$$E(X^2) = \int_{-\infty}^{\infty} x^2\frac{9}{10}\mathbb{1}_{[1,10]}(x)\frac{1}{x^2}dx = \frac{9}{10}\int_1^{10} dx = \frac{81}{10}.$$

And the third moment is

$$E(X^3) = \int_{-\infty}^{\infty} x^3 \frac{9}{10} \mathbb{1}_{[1,10]}(x) \frac{1}{x^2} dx = \frac{9}{10} \int_1^{10} x dx = \frac{9 \cdot 99}{20}.$$

a continuous random variable with density $f_X(x) = \frac{9}{10} \mathbb{1}_{[1,10]}(x) \frac{1}{x^2}$.

2. We have that $P(X \in \{-1, 0, 2\}) = 1$ so,

$$\begin{aligned} E(X) &= (-1) \cdot P(X = -1) + 0 \cdot P(X = 0) + 2P(X = 2) = \tfrac{1}{3}(-1 + 0 + 2) = \tfrac{1}{3}, \\ E(X^2) &= (-1)^2 \cdot P(X = -1) + 0^2 \cdot P(X = 0) + 2^2 P(X = 2) = \tfrac{1}{3}(1 + 0 + 4) = \tfrac{5}{3}, \\ E(X^3) &= (-1)^3 \cdot P(X = -1) + 0^3 \cdot P(X = 0) + 2^3 P(X = 2) = \tfrac{1}{3}(-1 + 0 + 8) = \tfrac{7}{3}. \end{aligned}$$

3. We have that $P(X \in \{-1, 1\}) = 1$ so,

$$\begin{aligned} E(X) &= 1 \cdot P(X = 1) - 1 \cdot P(X = -1) = \tfrac{1}{2} - \tfrac{1}{2} = 0, \\ E(X^2) &= 1^2 \cdot P(X = 1) + (-1)^2 \cdot P(X = -1) = \tfrac{1}{2} + \tfrac{1}{2} = 1, \\ E(X^3) &= 1^3 \cdot P(X = 1) + (-1)^3 \cdot P(X = -1) = \tfrac{1}{2} - \tfrac{1}{2} = 0. \end{aligned}$$

4. First note that as $x\frac{1}{2}e^{-|x|}$ and $x^3\frac{1}{2}e^{-|x|}$ are odd functions, we have $E(X) = E(X^3) = 0$. Let us finally compute the second moment of X.

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \frac{1}{2} \int_{-\infty}^{+\infty} x^2 e^{-|x|} dx = \int_0^{+\infty} x^2 e^{-x} dx \\ &= [-x^2 e^{-x}]_0^\infty + 2 \int_0^{+\infty} x e^{-x} dx = 2[-x e^{-x}]_0^\infty + 2 \int_0^{+\infty} e^{-x} dx = 2[-e^{-x}]_0^\infty = 2. \end{aligned}$$

Exercise 2.21. Let $X : \Omega \to \mathbb{R}$ be a random variable. Show that if there is $\delta > 0$ such that $E(e^{\delta|X|}) < \infty$, then, for any $p \in \mathbb{N}$,

$$E(|X|^p) < \infty.$$

Can you find explicit bounds on $E(|X|^p)$ depending on p and $C = E(e^{\delta|X|})$?

Solution 2.21. Let $C = E(e^{\delta|X|}) < \infty$. We have that for any $x \geq 0$, and any $p \in \mathbb{N}$,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \geq \frac{x^p}{p!}.$$

So, for $p \in \mathbb{N}$, $e^{\delta|X|} \geq \frac{\delta^p |X|^p}{p!}$. In particular,

$$C = E(e^{\delta|X|}) \geq E\Big(\frac{\delta^p |X|^p}{p!}\Big) = \frac{\delta^p}{p!} E(|X|^p).$$

So,

$$E(|X|^p) \leq C p! \delta^{-p} < \infty.$$

Exercise 2.22. 1. Let Ω be a finite set. Let P be the uniform probability measure on Ω, and let $A \subset \Omega$ be non-empty. Show that $P(\ \mid A)$ can be seen as the uniform probability measure on A (i.e.: for every $B \subset A, P(B \mid A) = |B|/|A|$).

2. Let P be the law of a fair (6 faces) dice roll. Let even be the event that we obtain a face with an even number of points. What is the probability mass function associated to $P(\mid \text{even})$?

Solution 2.22. 1. Let $B \subset A \subset \Omega$. We have that the probability of B under $P(\ \mid A)$ is (using the definition):

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{|B \cap A| \cdot |\Omega|}{|\Omega| \cdot |A|} = \frac{|B|}{|A|}$$

as we have $B \subset A$, so $B \cap A = B$.

2. Our realisation set is $\Omega = \{F_1, F_2, \dots, F_6\}$ (F_i stands for "face with i points"). We then compute the probability of each realisation in Ω under $P(\mid \text{even})$: recalling that $\text{even} = \{F_2, F_4, F_6\}$,

$$P\big(\{F_i\} \mid \text{even}\big) = \frac{P\big(\{F_i\} \cap \text{even}\big)}{P(\text{even})} = \begin{cases} \frac{1/6}{1/2} & \text{if } i \text{ is even,} \\ 0 & \text{if } i \text{ is odd.} \end{cases}$$

So, the probability mass function

$$p(F_i) = \begin{cases} \frac{1}{3} & \text{if } i \text{ is even,} \\ 0 & \text{if } i \text{ is odd,} \end{cases}$$

is the probability mass function associated to the probability measure $P(\ \mid \text{even})$.

Exercise 2.23. Nicole would like to go to the hairdresser, but she can't decide between hairdresser A and hairdresser B. So she roll a fair 6-faces dice: if she obtains a 5 or a 6, she goes to hairdresser A, and if she obtains a 1,2,3, or 4 she goes to hairdresser B. Suppose that the waiting time (in minutes) is random and is a uniform random variable on $[0, 30]$ for hairdresser A, and that it is a uniform random variable on $[0, 20]$ for hairdresser B. *The following questions are independent of each others.*

1. Nicole obtains a 5. What is the probability that she waits at least 25 minutes?
2. What is the probability that she waits at least 15 minutes?
3. Nicole waited for at least 15 minutes. What is the probability that she rolled a 4?

Solution 2.23. Denote T the time that Nicole waits, H_A the event that she goes to hairdresser A, H_B the event that she goes to hairdresser B, and D the result of the dice. From the pieces of information we have, we get

$$P(H_A) = P(D \in \{5, 6\}) = \tfrac{2}{6}, \quad P(H_B) = P(D \in \{1, 2, 3, 4\}) = \tfrac{4}{6},$$
$$E(T \mid H_A) = \int_0^{30} \tfrac{1}{30} x dx = 15, \quad E(T \mid H_B) = \int_0^{20} \tfrac{1}{20} x dx = 10.$$

1. We want to compute $P(T \geq 25 \,|\, D = 5)$. Conditionally on H_A, T is a uniform random variable on $[0, 30]$, so using that $\{D = 5\} \subset H_A$,

$$P(T \geq 25 \,|\, D = 5) = P(T \geq 25 \,|\, D = 5, H_A) = \int_0^{30} \tfrac{1}{30} \mathbb{1}_{[25,+\infty)}(x) dx = \tfrac{1}{30} \int_{25}^{30} dx = \tfrac{1}{6}.$$

2. As $H_A \cap H_B = \varnothing$ (Nicole goes to at most one of the two hairdresser), we have $\mathbb{1}_{H_A} + \mathbb{1}_{H_B} \leq 1$. On the other hand, she goes to at least one the two so $H_A \cup H_B = \Omega$, and $\mathbb{1}_{H_A} + \mathbb{1}_{H_B} \geq 1$. So, $1 = \mathbb{1}_{H_A} + \mathbb{1}_{H_B}$, and thus

$$\begin{aligned} P(T \geq 15) &= E\big(\mathbb{1}_{\{T \geq 15\}} (\mathbb{1}_{H_A} + \mathbb{1}_{H_B})\big) \\ &= P(T \geq 15, H_A) + P(T \geq 15, H_B) \\ &= P(T \geq 15 \,|\, H_A) P(H_A) + P(T \geq 15 \,|\, H_B) P(H_B) \\ &= \tfrac{2}{6} \int_0^{30} \tfrac{1}{30} \mathbb{1}_{[15,+\infty)}(x) dx + \tfrac{4}{6} \int_0^{20} \tfrac{1}{20} \mathbb{1}_{[15,+\infty)}(x) dx \\ &= \tfrac{1}{3} \tfrac{15}{30} + \tfrac{2}{3} \tfrac{5}{20} = \tfrac{1}{3}. \end{aligned}$$

3. We use the definition of conditional probability twice:

$$\begin{aligned} P(D = 4 \,|\, T \geq 15) &= \frac{P(T \geq 15, D = 4)}{P(T \geq 15)} \\ &= \frac{P(T \geq 15 \,|\, D = 4) P(D = 4)}{P(T \geq 15)} \\ &= \frac{1/6}{1/3} P(T \geq 15 \,|\, D = 4, H_B) = \frac{1}{2} \int_0^{20} \tfrac{1}{20} \mathbb{1}_{[15,+\infty)}(x) dx = \frac{1}{8}. \end{aligned}$$

***Exercise 2.24** (The two envelopes problem). The problem is a famous seemingly paradoxical situation. It goes as follows.

> magine you are given two identical envelopes, each containing money. One contains twice as much as the other. You may pick one envelope and keep the money it contains. Having chosen an envelope at will, but before inspecting it, you are given the chance to switch envelopes. Should you switch?
>
> (I)

The two contradictory conclusions are

1. The situation is obviously symmetric, so it does not matter whether you switch or not. Formally, let $x, 2x$ be the money in the two envelopes, and S be the money in the first envelope you pick. Then, one the one hand, if you keep your envelope, you have an expected gain of
$$2xP(S = 2x) + xP(S = x) = \frac{3}{2}x,$$
as, by symmetry, $P(S = 2x) = P(S = x) = 1/2$. On the other hand, if you switch, you have an expected gain of
$$xP(S = 2x) + 2xP(S = x) = \frac{3}{2}x.$$

2. You should always switch as, given the money in the envelope you have, you can double the sum with probability 1/2 or divide it by two with probability 1/2, which has higher average gain than the number you see. More formally, let A be the money in the first envelope you pick. Then, the other envelope contains either $2A$ or $A/2$ each with probability 1/2, so you gain A if you keep your envelope, but on average you gain
$$\frac{1}{2}\frac{A}{2} + \frac{1}{2}2A = \frac{5}{4}A > A$$
if you switch.

Which one of the two reasoning is incorrect? can you spot why?

Solution 2.24. The second reasoning is incorrect. The problem lies in the modelisation step: in the first case, we define everything from the start by fixing the total amount of money in the two envelopes. The second reasoning assumes that the amount is random (which can sound reasonable as we do not know how much money there is in total). Letting A be the amount in the envelope you first pick, the second reasoning gives the money distributions $(A, 2A)$ and $(A, A/2)$ to be equally likely. But we do not know which envelop we picked, so if we assume that these two situations are equally likely, we picked the envelop containing $2A$ with probability $\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$ (proba $\frac{1}{2}$ of picking situation $(A, 2A)$ times proba $\frac{1}{2}$ of picking the envelop containing A)... which is a contradiction with the fact that A was the amount of money in the envelop you picked from the start, so that you should have a probability 1 of picking an envelop containing A.

The only value of A not leading to a contradiction is $A = 0$, which is the only value of the total money for which the two reasoning give the same solution!

Exercise 2.25. A communication system is composed of n components. Each components works with probability p, independently of the others.

1. What is the law of the random variable X which gives the number of working components? (i.e.: give $P(X = x)$ for all x).

2. For the system to work properly, we need that at least half of the components are working. For which values of p does a system with 5 components has bigger chances of working that a system with 3 components?

Solution 2.25.

1. Let $Y_1, \dots, Y_n$ be the random variables corresponding to the status of each components: $Y_i = 1$ is component i is working, and $Y_i = 0$ else. As the components are all independent from the others, for any $\epsilon_1, \dots \epsilon_n \in \{0, 1\}$,
$$P(Y_1 = \epsilon_1, \dots Y_n = \epsilon_n) = \prod_{i=1}^{n} P(Y_i = \epsilon_i) = \prod_{i=1}^{n} p^{\epsilon_i}(1-p)^{1-\epsilon_i}$$
In particular, for any sequence $\epsilon_1, \dots, \epsilon_n \in \{0, 1\}$ containing exactly k 1's,
$$P(Y_1 = \epsilon_1, \dots Y_n = \epsilon_n) = p^k(1-p)^{n-k}.$$
Then, we have that the number of working component is given by $X = \sum_{i=1}^{n} Y_i \in \{0, 1 \dots, n\}$. The probability of $X = k$ is then given by the probability that exactly k of the Y_i's are 1. There are $\binom{n}{k}$ possibilities for choosing which of the Y_i's give 1, and by the previous computation, the probability of each choice is $p^k(1-p)^{n-k}$, so
$$P(X = k) = \mathbb{1}_{\{0,\dots,n\}}(k) \binom{n}{k} p^k (1-p)^{n-k}.$$
This is a *Binomial distribution.*

2. The probability of working for the 3 components system is
$$\binom{3}{2} p^2(1-p) + \binom{3}{3} p^3 = 3p^2(1-p) + p^3.$$
The probability of working for the 5 components system is
$$\binom{5}{3} p^3(1-p)^2 + \binom{5}{4} p^4(1-p) + \binom{5}{5} p^5 = 10p^3(1-p)^2 + 5p^4(1-p) + p^5.$$
We thus want to find for which values of p we have
$$\begin{aligned} & 10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3 \\ \iff & (1-p)(10p(1-p) + 5p^2 - 3) + p^3 - p > 0 \\ \iff & (1-p)(10p(1-p) + 5p^2 - 3) + p(p^2 - 1) > 0 \\ \iff & (1-p)(10p(1-p) + 5p^2 - 3) - p(1-p)(p+1) > 0 \\ \iff & 10p(1-p) + 5p^2 - 3 - p(p+1) > 0 \\ \iff & 9p - 6p^2 - 3 > 0 \\ \iff & 3p - 2p^2 - 1 > 0. \end{aligned}$$
As $3p - 2p^2 - 1 = (2p-1)(1-p)$, the last condition is equivalent to $p > 1/2$.

Exercise 2.26. Bob has an appointment with the dentist at 10:00. The waiting time between the appointment time and the actual moment Bob will be able to see the dentist is random and assumed to be a continuous random variable with density $f(x) = c\mathbb{1}_{[0,+\infty)}(x)e^{-2x}$ for some $c > 0$. The time the meeting takes is independent of the waiting time, and is uniformly distributed on $[5, 15]$.

1. What is the value of c?

2. What is the density of the meeting duration?

3. On average, at what time will Bob get out of the dentist office?

At 10:20, a gas leak in the building of the dentist office kills everyone in the office.

4. What is the probability that Bob survives?

Solution 2.26. Denote W the time Bob waits and T the time duration of the meeting.

1. f has to be a density function, so $f(x)$ should be non-negative, so $c \geq 0$. Then, f must integrate to 1, which implies

$$c^{-1} = \int_{-\infty}^{+\infty} \mathbb{1}_{[0,+\infty)}(x)e^{-2x}dx = \int_{0}^{+\infty} e^{-2x}dx = \left[-\tfrac{1}{2}e^{-2x}\right]_0^{\infty} = \tfrac{1}{2}$$

   so $c = 2$.

2. The meeting duration is uniform over $[5, 15]$, so a density for the meeting duration T is

$$f_T(x) = \frac{1}{10}\mathbb{1}_{[5,15]}(x)$$

   as the length of $[5, 15]$ is 10.

3. It is given by the expectation of $W + T$,

$$\begin{aligned} E(W+T) = E(W) + E(T) &= \int_{-\infty}^{+\infty} xf(x)dx + \int_{-\infty}^{+\infty} xf_T(x)dx \\ &= 2\int_{0}^{+\infty} xe^{-2x}dx + \frac{1}{10}\int_{5}^{15} xdx = \frac{1}{2} + 10 = 10.5. \end{aligned}$$

4. To survive, Bob must leave the dentist office before 10:20, which is equivalent to say that the waiting time + meeting time should be less than 20. We thus want

to compute $P(W+T<20)$.

$$\begin{aligned}
P(W+T<20) &= \int_{-\infty}^{+\infty} dw \int_{-\infty}^{+\infty} dt f(w) f_T(t) \mathbb{1}_{[0,20)}(w+t) \\
&= \frac{2}{10} \int_0^{+\infty} dw \int_5^{15} dt e^{-2w} \mathbb{1}_{[0,20)}(w+t) \\
&= \frac{2}{10} \int_5^{15} dt \int_0^{+\infty} dw e^{-2w} \mathbb{1}_{[-t,20-t)}(w) \\
&= \frac{2}{10} \int_5^{15} dt \int_0^{20-t} dw e^{-2w} \\
&= \frac{1}{10} \int_5^{15} dt (1 - e^{-2(20-t)}) \\
&= 1 - \frac{e^{-40}}{10} \int_5^{15} dt e^{2t} \\
&= 1 - \frac{e^{-40}}{10} \frac{1}{2} (e^{30} - e^{10}) = 1 - \tfrac{1}{20}(e^{-10} - e^{-30}) \approx 0.99995.
\end{aligned}$$

Exercise 2.27. Let P be the uniform probability measure on $\{1,2,3,4\}$. Define the events

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

1. Show that A and B are independent, that A and C are independent, and that B and C are independent.
2. Show that $P(A \cap B \cap C) \neq P(A)P(B)P(C)$. Are A, B, C forming an independent family?
3. Based on the two previous points, construct three random variables X_1, X_2, X_3 that are two-by-two independent but that do not form an independent family.

Solution 2.27.

1. We have $P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}$. Then,

$$\begin{aligned}
P(A \cap B) = P(\{1\}) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2} = P(A)P(B), \\
P(A \cap C) = P(\{2\}) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2} = P(A)P(C), \\
P(B \cap C) = P(\{3\}) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2} = P(B)P(C).
\end{aligned}$$

2. We have

$$P(A \cap B \cap C) = P(\varnothing) = 0 \neq \tfrac{1}{8} = \tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = P(A)P(B)P(C),$$

so A, B, C are not forming an independent family.

3. We can take the random variables

$$X_1 = \mathbb{1}_A,\ X_2 = \mathbb{1}_B,\ X_3 = \mathbb{1}_C.$$

Exercise 2.28. 1. Let $X, Y : \Omega \to \mathbb{R}$ be two independent random variables. Show that
$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

2. Let $n \geq 2$, let $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ be an independent family of random variables. Show that
$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t).$$

Solution 2.28. 1. As X, Y are independent, $E(f(X)g(Y)) = E(f(X))E(g(Y))$ for any functions f, g. Applying this with $f(x) = g(x) = e^{tx}$, we get
$$M_{X+Y}(t) = E(e^{(X+Y)t}) = E(e^{Xt}e^{Yt}) = E(e^{Xt})E(e^{Yt}) = M_X(t)M_Y(t).$$

2. Same reasoning as in the previous point using the "expectation of product of functions" definition of being an independent family.

Exercise 2.29. 1. Let $X, Y : \Omega \to \mathbb{R}$ be two independent random variables. Show that
$$P\big(\max(X, Y) \leq t\big) = F_X(t)F_Y(t).$$

2. Let $n \geq 2$, let $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ be an independent family of random variables. Show that
$$P\big(\max_{i=1,\ldots,n} X_i \leq t\big) = \prod_{i=1}^n F_{X_i}(t).$$

Solution 2.29. 1. As X, Y are independent $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any $A, B \subset \mathbb{R}$. In particular,
$$P\big(\max(X, Y) \leq t\big) = P(X \leq t, Y \leq t) = P(X \leq t)P(Y \leq t) = F_X(t)F_Y(t).$$

2. We use the same reasoning as in the previous point: as $X_1, \ldots, X_n$ form an independent family, $P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \ldots P(X_n \in A_n)$ for any $A_1, \ldots, A_n \subset \mathbb{R}$. In particular,
$$\begin{aligned}P\big(\max_{i=1,\ldots,n} X_i \leq t\big) &= P(X_1 \leq t, X_2 \leq t, \ldots, X_n \leq t) \\ &= P(X_1 \leq t)P(X_2 \leq t) \ldots P(X_n \leq t) = \prod_{i=1}^n F_{X_i}(t).\end{aligned}$$

Exercise 2.30. Three teams, numbered $1, 2, 3$ are handling customer service. They all have different efficiency. We suppose that the treatment time (in hours) of team i to a given customer request is a continuous random variable T_i with density

$$f_{T_1}(x) = \mathbb{1}_{[0,+\infty)}(x)e^{-x}, \ f_{T_2}(x) = 2\mathbb{1}_{[0,+\infty)}(x)e^{-2x}, \ f_{T_3}(x) = 3\mathbb{1}_{[0,+\infty)}(x)e^{-3x}.$$

When a customer submits a request, it is assigned uniformly at random to one of the three teams. You submit a request to customer service. Denote X the random variable corresponding to the time that your request takes to be treated.

1. What is the expected value of X?
2. What is the probability that X is less than 1?

You now have waited for an hour and your request has not yet been treated.

3. What is the probability that the team assigned to your request is team 1? team 2? teams 3?

Solution 2.30. Let us first compute the expected treatment time of team $i \in \{1, 2, 3\}$. It is given by

$$E(T_i) = i \int_0^\infty x e^{-ix} dx = \frac{1}{i}.$$

Let U be a uniform random variable on $\{1, 2, 3\}$ independent of T_1, T_2, T_3, which represent the team assigned to the request. We then have that $X = T_U$.

1. Using the formula of total probability (total expectation in this case), we get

$$E(X) = E(T_U) = \sum_{i=1}^{3} E(T_i \,|\, U = i)P(U = i).$$

Then, we know that $P(U = i) = 1/3$ for $i = 1, 2, 3$, and that T_i is independent of U, so

$$E(T_i \,|\, U = i) = \frac{E(T_i \mathbb{1}_{\{i\}}(U))}{P(U = i)} = \frac{E(T_i)E(\mathbb{1}_{\{i\}}(U))}{P(U = i)} = \frac{E(T_i)P(U = i)}{P(U = i)} = E(T_i) = \frac{1}{i}.$$

Putting things together,

$$E(X) = \sum_{i=1}^{3} \frac{1}{i} \cdot \frac{1}{3} = \frac{11}{18}.$$

2. Using the formula of total probability, we get

$$P(X \le 1) = P(T_U \le 1) = \sum_{i=1}^{3} P(T_i \le 1 \,|\, U = i)P(U = i).$$

As in the previous point, T_i being independent from U gives that

$$P(T_i \le 1 \,|\, U = i) = P(T_i \le 1) = i \int_0^1 e^{-ix} dx = [-e^{-ix}]_0^1 = 1 - e^{-i}.$$

So, putting things together,

$$P(X \leq 1) = \sum_{i=1}^{3}(1 - e^{-i}) \cdot \frac{1}{3} = 1 - \tfrac{1}{3}(e^{-1} + e^{-2} + e^{-3}) \approx 0.8157.$$

3. We know that $X > 1$ and we want to know the probability that $U = i$ for $i \in \{1, 2, 3\}$. So, we want to compute $P(U = i \,|\, X > 1)$. Using Bayes formula,

$$\begin{aligned} P(U = i \,|\, X > 1) &= \frac{P(X > 1 \,|\, U = i)P(U = i)}{P(X > 1)} \\ &= \frac{P(T_i > 1 \,|\, U = i)\frac{1}{3}}{1 - P(X \leq 1)} = \frac{P(T_i > 1)}{e^{-1} + e^{-2} + e^{-3}} \end{aligned}$$

where we used the previous point, that under the event $U = i$, $X = T_U = T_i$, and that T_i and U are independent. Now, using the computation of the previous point,

$$P(T_i > 1) = 1 - P(T_i \leq 1) = 1 - 1 + e^{-i} = e^{-i}.$$

So,

$$P(U = i \,|\, X > 1) = \frac{e^{-i}}{e^{-1} + e^{-2} + e^{-3}}.$$

Exercise 2.31. Let X be a discrete random variable with $P(X = -1) = P(X = 3) = \frac{1}{4}$, and $P(X = 0) = \frac{1}{2}$. Let Y be a continuous random variable with density $f_Y(x) = \mathbb{1}_{[0,+\infty)}(x)\frac{3}{(x+1)^4}$. Finally, let Z be a discrete random variable with $P(Z = 0) = P(Z = 1) = \frac{1}{2}$. Suppose that X, Y, Z form an independent family. Define the random variable

$$W = ZX + (1 - Z)Y.$$

Compute $E(W)$.

Solution 2.31. Note that W is neither discrete nor continuous. Nevertheless, using the formula of total probability (total expectation in this case),

$$\begin{aligned} E(W) &= E(ZX + (1 - Z)Y) = P(Z = 1)E(X \,|\, Z = 1) + P(Z = 0)E(Y \,|\, Z = 0) \\ &= \tfrac{1}{2}E(X) + \tfrac{1}{2}E(Y) = \frac{1}{2}\Big(\tfrac{1}{4} \cdot (-1) + \tfrac{1}{4} \cdot 3 + \tfrac{1}{2} \cdot 0 + \int_0^{+\infty} \frac{3x}{(x+1)^4}dx\Big) \\ &= \frac{1}{2}\Big(\tfrac{1}{2} + \int_0^{+\infty} \frac{1}{(x+1)^3}dx\Big) = \frac{1}{2}, \end{aligned}$$

as Z is independent from X, Y.

Exercise 2.32. Suppose that we have two identical bags contain each 30 marbles. In one of the bag (bag 1), there are 10 red, 10 green, and 10 blue marbles. In the other one (bag 2), there are 3 red, 8 green, and 19 blue. You chose one of the bags uniformly at random and start drawing marbles out of the bag, one by one without putting them back in the bag. The first marble you pick is blue.

1. What is the probability that you chose bag 1?
2. What is the probability that the next marble you pick is red?

You draw a second marble. It is again blue.

3. What is now the probability that you chose bag 1?
4. What is now the probability that the next marble you pick is red?

Based on the first two draws, you evaluate that the bag you picked is bag 2.

5. What is the probability that your evaluation of picking bag 2 is correct?

You draw a third marble. It is red.

6. What is now the probability that your evaluation of picking bag 2 is correct?

Solution 2.32. Denote Bag the bag that you pick. Denote $M_1, M_2, \ldots$ the colours of the marble that you pick first, second,...

1. We know that $M_1 = \text{blue}$, we thus want to compute $P(\text{Bag} = 1 \,|\, M_1 = \text{blue})$. We first compute

$$P(M_1 = \text{blue} \,|\, \text{Bag} = 1) = \tfrac{10}{30}, \quad P(M_1 = \text{blue} \,|\, \text{Bag} = 2) = \tfrac{19}{30}.$$

So, using Bayes' law, and the formula of total probability,

$$\begin{aligned} P(\text{Bag} = 1 \,|\, M_1 = \text{blue}) &= \frac{P(M_1 = \text{blue} \,|\, \text{Bag} = 1) P(\text{Bag} = 1)}{P(M_1 = \text{blue})} \\ &= \frac{P(M_1 = \text{blue} \,|\, \text{Bag} = 1) P(\text{Bag} = 1)}{P(M_1 = \text{blue} \,|\, \text{Bag} = 1) P(\text{Bag} = 1) + P(M_1 = \text{blue} \,|\, \text{Bag} = 2) P(\text{Bag} = 2)} \\ &= \frac{\frac{10}{30}}{\frac{10}{30} + \frac{19}{30}} = \frac{10}{29} \end{aligned}$$

as $P(\text{Bag} = 1) = P(\text{Bag} = 2) = 1/2$.

2. We use the formula of total probability:

$$\begin{aligned} P(M_2 = \text{red} \,|\, M_1 = \text{blue}) = {} & P(M_2 = \text{red} \,|\, M_1 = \text{blue}, \text{Bag} = 1) P(\text{Bag} = 1 \,|\, M_1 = \text{blue}) \\ & + P(M_2 = \text{red} \,|\, M_1 = \text{blue}, \text{Bag} = 2) P(\text{Bag} = 2 \,|\, M_1 = \text{blue}). \end{aligned}$$

Now, using the first point,

$$\begin{gathered} P(\text{Bag} = 1 \,|\, M_1 = \text{blue}) = \tfrac{10}{29}, \\ P(\text{Bag} = 2 \,|\, M_1 = \text{blue}) = 1 - P(\text{Bag} = 1 \,|\, M_1 = \text{blue}) = \tfrac{19}{29}. \end{gathered}$$

Moreover, counting the balls of each colour in each bags, we have

$$P(M_2 = \text{red} \mid M_1 = \text{blue},\ \text{Bag} = 1) = \tfrac{10}{29},$$
$$P(M_2 = \text{red} \mid M_1 = \text{blue}, \text{Bag} = 2) = \tfrac{3}{29}.$$

So,

$$P(M_2 = \text{red} \mid M_1 = \text{blue}) = \tfrac{10}{29} \cdot \frac{10}{29} + \tfrac{3}{29} \cdot \tfrac{19}{29} = \tfrac{157}{841}.$$

3. We now know $M_1 = M_2 = \text{blue}$. We proceed as in the first point: we first compute

$$P(M_1 = M_2 = \text{blue} \mid \text{Bag} = 1) = \tfrac{10 \cdot 9}{30 \cdot 29}, \quad P(M_1 = M_2 = \text{blue} \mid \text{Bag} = 2) = \tfrac{19 \cdot 18}{30 \cdot 29}.$$

Then, using Bayes' law, the formula of total probability, and $P(\text{Bag} = 1) = P(\text{Bag} = 2) = 1/2$,

$$\begin{aligned} &P(\text{Bag} = 1 \mid M_1 = M_2 = \text{blue}) \\ &= \frac{P(M_1 = M_2 = \text{blue} \mid \text{Bag} = 1)}{P(M_1 = M_2 = \text{blue} \mid \text{Bag} = 1) + P(M_1 = M_2 = \text{blue} \mid \text{Bag} = 2)} \\ &= \frac{\frac{10 \cdot 9}{30 \cdot 29}}{\frac{10 \cdot 9}{30 \cdot 29} + \frac{19 \cdot 18}{30 \cdot 29}} = \frac{45}{216} \end{aligned}$$

4. We proceed as in the second point: using the third point,

$$P(\text{Bag} = 1 \mid M_1 = M_2 = \text{blue}) = \tfrac{45}{216},$$
$$P(\text{Bag} = 2 \mid M_1 = M_2 = \text{blue}) = 1 - \tfrac{45}{216} = \tfrac{171}{216}.$$

Moreover, counting the balls of each colour in each bags, we have

$$P(M_3 = \text{red} \mid M_1 = M_2 = \text{blue},\ \text{Bag} = 1) = \tfrac{10}{28},$$
$$P(M_3 = \text{red} \mid M_1 = M_2 = \text{blue}, \text{Bag} = 2) = \tfrac{3}{28}.$$

So, by the formula of total probability (as in the second point)

$$P(M_3 = \text{red} \mid M_1 = M_2 = \text{blue}) = \tfrac{10}{28} \cdot \tfrac{45}{216} + \tfrac{3}{28} \cdot \tfrac{171}{216} = \tfrac{963}{28 \cdot 216} = \tfrac{107}{672} \approx 0.16.$$

5. This probability is the probability that you picked bag 2 given what you already saw: that is $M_1 = M_2 = \text{blue}$. From the computations of the previous point, this probability is equal to

$$P(\text{Bag} = 2 \mid M_1 = M_2 = \text{blue}) = \tfrac{171}{216} \approx 0.79.$$

6. This probability is the probability that you picked bag 2 given that $M_1 = M_2 = \text{blue}$ and that $M_3 = \text{red}$. This is given by (again using Bayes' law)

$$\begin{aligned} &P(\text{Bag} = 2 \mid M_1 = M_2 = \text{blue}, M_3 = \text{red}) \\ &= \frac{P(M_3 = \text{red} \mid M_1 = M_2 = \text{blue}, \text{Bag} = 2) P(\text{Bag} = 2 \mid M_1 = M_2 = \text{blue})}{P(M_3 = \text{red} \mid M_1 = M_2 = \text{blue})} \\ &= \frac{\frac{3}{28} \frac{171}{216}}{\frac{107}{672}} = \tfrac{57}{107} \approx 0.53, \end{aligned}$$

where we used the quantities computed at the fourth point.

Exercise 2.33. Let $X, Y : \Omega \to \mathbb{R}$ be two random variables such that $P(X \in \{0,1\}) = P(Y \in \{0,1\}) = 1$. Suppose that

$$\mathrm{Cov}(X, Y) = 0.$$

1. Show that for any $a, b \in \{0,1\}$, $P(X = a, Y = b) = P(X = a)P(Y = b)$.

2. Using the first point, show that for any events $A, B \subset \mathbb{R}$, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.

Hint: what is $P\big(X \in (A \cap \{0,1\})\big)$*?*

Solution 2.33. 1. We have that for any $x, y \in \{0,1\}$,

$$2xy - x - y + 1 = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{else.} \end{cases}$$

So, for any $a, b \in \{0,1\}$,

$$\begin{aligned}
&P(X = a, Y = b) - P(X = a)P(Y = b) \\
&= E\big((2aX - X - a + 1)(2bY - Y - b + 1)\big) \\
&\quad - E(2aX - X - a + 1)E(2bY - Y - b + 1) \\
&= E(XY)(2a-1)(2b-1) + E(X)(2a-1)(b-1) + (1-a)E(Y)(2b-1) \\
&\quad + (1-a)(1-b) - E(X)E(Y)(2a-1)(2b-1) - E(X)(2a-1)(b-1) \\
&\quad - (1-a)E(Y)(2b-1) - (1-a)(1-b) \\
&= (2a-1)(2b-1)\mathrm{Cov}(X, Y) = 0
\end{aligned}$$

where we used that $P(X \in \{0,1\}) = P(Y \in \{0,1\}) = 1$.

2. As $P(X \in \{0,1\}) = 1 = P(Y \in \{0,1\})$, we have that (X, Y) is a discrete random vector with $P\big((X, Y) \in \{0,1\}^2\big) = 1$. So, for any $A, B \subset \mathbb{R}$

$$\begin{aligned}
&P(X \in A, Y \in B) \\
&= P(X = 0, Y = 0)\mathbb{1}_{A\times B}(0,0) + P(X = 1, Y = 0)\mathbb{1}_{A\times B}(1,0) \\
&\quad + P(X = 0, Y = 1)\mathbb{1}_{A\times B}(0,1) + P(X = 1, Y = 1)\mathbb{1}_{A\times B}(1,1) \\
&= P(X = 0)P(Y = 0)\mathbb{1}_A(0)\mathbb{1}_B(0) + P(X = 1)P(Y = 0)\mathbb{1}_A(1)\mathbb{1}_B(0) \\
&\quad + P(X = 0)P(Y = 1)\mathbb{1}_A(0)\mathbb{1}_B(1) + P(X = 1)P(Y = 1)\mathbb{1}_A(1)\mathbb{1}_B(1) \\
&= P(X \in A)P(Y \in B).
\end{aligned}$$

by the first point.

Exercise 2.34. Let (X,Y) and (X',Y') be two independent vector such that $(X,Y) \overset{\text{Law}}{=} (X',Y')$. Show the following points.

1. $\text{Var}(X) = E(X^2) - E(X)^2 = \frac{1}{2}E\big((X-X')^2\big) = \text{Cov}(X,X)$;
2. $\text{Cov}(X,Y) = E\big((X-E(X))(Y-E(Y))\big)$;
3. $2\text{Cov}(X,Y) = E\big((X-X')(Y-Y')\big)$;
4. $\text{Cov}(X,Y) = \text{Cov}(Y,X)$;
5. $\text{Cov}(aX,bY) = ab\text{Cov}(X,Y)$;
6. $\text{Cov}(X,Y+Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z)$.

Solution 2.34. We detail the calculations of the first point, the others follow the exact same scheme: expand, use linearity, re-combine. First, expanding the square and using linearity of the expectation,

$$\begin{aligned}\text{Var}(X) = E\big((X-E(X))^2\big) &= E\big(X^2 - 2XE(X) + E(X)^2\big)\\ &= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2.\end{aligned}$$

Then, using the definition,

$$\text{Cov}(X,X) = E\big((X_E(X))(X-E(X))\big) = E\big((X_E(X))^2\big) = \text{Var}(X).$$

Finally, as X, X' are independent,

$$\begin{aligned}E\big((X-X')^2\big) = E\big(X^2 - 2XX' + (X')^2\big) &= E(X^2) - 2E(XX') + E((X')^2)\\ &= 2E(X^2) - 2E(X)E(X') = 2\big(E(X^2) - E(X)^2\big) = 2\text{Var}(X),\end{aligned}$$

as $X \overset{\text{Law}}{=} X'$ implies that $E(X) = E(X')$, and $E(X^2) = E((X')^2)$.

Exercise 2.35. In each of the following cases, compute the covariance between X and Y.

1. The random vector (X,Y) is a discrete random vector with
$$P\big((X,Y)=(2,1)\big) = \tfrac{1}{4}, \quad P\big((X,Y)=(1,0)\big) = \tfrac{1}{2}, \quad P\big((X,Y)=(0,0)\big) = \tfrac{1}{4}.$$
2. The random vector (X,Y) is a discrete random vector with
$$\begin{gathered}P\big((X,Y)=(0,1)\big) = \tfrac{1}{15}, \quad P\big((X,Y)=(2,3)\big) = \tfrac{7}{15}, \quad P\big((X,Y)=(5,5)\big) = \tfrac{4}{15},\\ P\big((X,Y)=(1,0)\big) = \tfrac{2}{15}, \quad P\big((X,Y)=(0,0)\big) = \tfrac{1}{15}.\end{gathered}$$
3. The random vector (X,Y) is a continuous random vector with density
$$f_{(X,Y)}(x,y) = 96\mathbb{1}_{[1,+\infty)}(x)\mathbb{1}_{[1,+\infty)}(y)\frac{1}{(x+y)^5}.$$

Solution 2.35. 1. We compute

$$E(X) = 2 \cdot P\big((X,Y) = (2,1)\big) + 1 \cdot P\big((X,Y) = (1,0)\big) + 0 \cdot P\big((X,Y) = (0,0)\big) = 1$$
$$E(Y) = 1 \cdot P\big((X,Y) = (2,1)\big) + 0 \cdot P\big((X,Y) = (1,0)\big) + 0 \cdot P\big((X,Y) = (0,0)\big) = \tfrac{1}{4},$$
$$E(XY) = 2 \cdot 1 \cdot P\big((X,Y) = (2,1)\big) + 1 \cdot 0 \cdot P\big((X,Y) = (1,0)\big) + 0 \cdot 0 \cdot P\big((X,Y) = (0,0)\big) = \tfrac{1}{2}.$$

Thus,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \tfrac{1}{2} - 1 \cdot \tfrac{1}{4} = \tfrac{1}{4}.$$

2. Proceed as in the previous point:

$$E(X) = 0 \cdot \tfrac{1}{15} + 2 \cdot \tfrac{7}{15} + 5 \cdot \tfrac{4}{15} + 1 \cdot \tfrac{2}{15} + 0 \cdot \tfrac{1}{15} = \tfrac{12}{5},$$
$$E(Y) = 1 \cdot \tfrac{1}{15} + 3 \cdot \tfrac{7}{15} + 5 \cdot \tfrac{4}{15} + 0 \cdot \tfrac{2}{15} + 0 \cdot \tfrac{1}{15} = \tfrac{14}{5},$$
$$E(XY) = 0 \cdot \tfrac{1}{15} + 6 \cdot \tfrac{7}{15} + 25 \cdot \tfrac{4}{15} + 0 \cdot \tfrac{2}{15} + 0 \cdot \tfrac{1}{15} = \tfrac{142}{15}.$$

Thus,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \tfrac{142}{15} - \tfrac{12}{5} \cdot \tfrac{14}{5} = \tfrac{206}{75}.$$

3. We note that $f_{(X,Y)}$ is invariant under the exchange of x, y, for $(X,Y) \overset{\text{Law}}{=} (Y,X)$ and thus $X \overset{\text{Law}}{=} Y$, so $E(X) = E(Y)$. We then compute

$$\begin{aligned} E(X) = E(Y) &= 96 \int_1^\infty dx \int_1^\infty dy \frac{x}{(x+y)^5} \\ &= \frac{96}{4} \int_1^\infty dx \frac{x}{(x+1)^4} \\ &= \frac{96}{4} \Big(\big[\tfrac{-x}{3(x+1)^3}\big]_1^\infty + \tfrac{1}{3} \int_1^\infty dx \frac{1}{(x+1)^3} \Big) \\ &= \frac{96}{4} \big(\tfrac{1}{24} + \tfrac{1}{24} \big) = 2. \end{aligned}$$

Also,

$$E(XY) = 96 \int_1^\infty dx \int_1^\infty dy \frac{xy}{(x+y)^5} = 5$$

by successive integration by parts (or asking WolframAlpha). Thus,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 5 - 2 \cdot 2 = 1.$$

Exercise 2.36. We do a study on the processors produced by a small firm. We study the price and durability of each models. The firm proposes three models, and the observed data are given in the following table.

Model	Price	Average lifetime	Number produced
A	537	5.6	135k
B	425	6	208k
C	368	3.7	67k

We pick a processor produced by this firm uniformly at random, what is the covariance between its price and average lifetime?

Solution 2.36. Let us formalize the situation. Given the numbers, the picked processor model is A with probability $\frac{135}{410}$, B with probability $\frac{208}{410}$, and C with probability $\frac{67}{410}$. In particular, writing X for the price of the picked processor and Y for its expected lifetime, we have that (X, Y) is a discrete random vector with

$$P\big((X,Y) = (537, 5.6)\big) = \frac{135}{410}, \quad P\big((X,Y) = (425, 6)\big) = \frac{208}{410},$$
$$P\big((X,Y) = (368, 3.7)\big) = \frac{67}{410}.$$

In particular, we have:

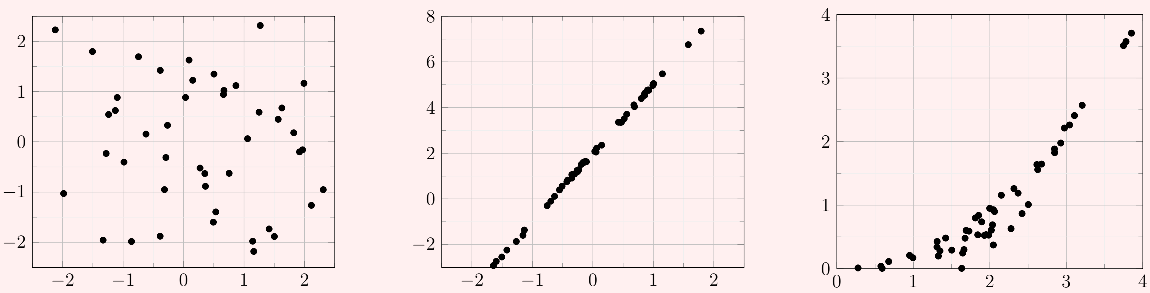
$$E(X) = 537 \cdot \frac{135}{410} + 425 \cdot \frac{208}{410} + 368 \cdot \frac{67}{410} = \frac{185551}{410},$$
$$E(Y) = 5.6 \cdot \frac{135}{410} + 6 \cdot \frac{208}{410} + 3.7 \cdot \frac{67}{410} = \frac{22519}{4100}$$
$$E(XY) = 537 \cdot 5.6 \cdot \frac{135}{410} + 425 \cdot 6 \cdot \frac{208}{410} + 368 \cdot 3.7 \cdot \frac{67}{410} = \frac{10275992}{4100}.$$

So,

$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{10275992}{4100} - \frac{185551}{410} \cdot \frac{22519}{4100} \approx 20.66.$$

Exercise 2.37. Associate each of the values of ρ_{XY} to a graph of samples from (X, Y).

$$\text{(A) } \rho_{XY} = 0.66, \quad \text{(B) } \rho_{XY} = 0, \quad \text{(C) } \rho_{XY} = 1.$$



(a) (b) (c)

Solution 2.37. (A) with (c), (B) with (a), and (C) with (b).

Exercise 2.38. Compute ρ_{XY} in each of the following situations.

1. $Y = aX + b$ for some $a, b \in \mathbb{R}$ with $a \neq 0$.

2. X is a continuous random variable with density $f_X(x) = \mathbb{1}_{[0,1]}(x)$, and $Y = \mathbb{1}_{[0,1/2]}(X)$.

3. U, V are two independent continuous random variables with density $f(x) = \mathbb{1}_{[0,1]}(x)$. X, Y are given by $X = U + V$, $Y = UV$.

4. X is a continuous random variable with density $f_X(x) = \frac{1}{3}\mathbb{1}_{[0,3]}(x)$, and $Y = X^2$.

5. (X, Y) is a continuous random vector with density $f_{(X,Y)}(x,y) = \frac{1}{2\pi} e^{-\sqrt{x^2+y^2}}$.

Solution 2.38. In each cases, we need to compute the covariance between X and Y, as well as the variance of X and of Y.

1. We have

$$\operatorname{Cov}(X,Y) = \operatorname{Cov}(X, aX+b) = a\operatorname{Var}(X),$$
$$\operatorname{Var}(Y) = \operatorname{Var}(aX+b) = a^2\operatorname{Var}(X).$$

Thus,

$$\rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{a\operatorname{Var}(X)}{\sqrt{a^2\operatorname{Var}(X)^2}} = 1.$$

2. We have

$$E(X) = \tfrac{1}{2}, \quad E(Y) = P(X \leq 1/2) = \tfrac{1}{2}, \quad E(XY) = \int_0^{1/2} x dx = \tfrac{1}{8},$$
$$\operatorname{Cov}(X,Y) = \tfrac{1}{8} - \tfrac{1}{2}\cdot\tfrac{1}{2} = -\tfrac{1}{8},$$
$$\operatorname{Var}(X) = E(X^2) - E(X)^2 = \int_0^1 x^2 dx - \tfrac{1}{4} = \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12},$$
$$\operatorname{Var}(Y) = E(Y^2) - E(Y)^2 = E(Y) - E(Y)^2 = \tfrac{1}{4}.$$

as $Y^2 = \mathbb{1}_{[0,1/2]}(X)\cdot\mathbb{1}_{[0,1/2]}(X) = Y$. Thus,

$$\rho_{XY} = \frac{-\frac{1}{8}}{\sqrt{\frac{1}{12}\cdot\frac{1}{4}}} = -\sqrt{\tfrac{48}{64}} = -\tfrac{\sqrt{3}}{2}.$$

3. We have

$$E(U) = E(V) = \tfrac{1}{2}, \quad E(U^2) = E(V^2) = \int_0^1 x^2 dx = \tfrac{1}{3},$$
$$\operatorname{Cov}(X,Y) = \operatorname{Cov}(U+V, UV) = \operatorname{Cov}(U, UV) + \operatorname{Cov}(V, UV)$$
$$= 2E(U^2V) - 2E(U)E(UV) = 2E(U^2)E(V) - 2E(U)^2E(V) = \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12},$$
$$\operatorname{Var}(X) = \operatorname{Var}(U+V) = \operatorname{Var}(U) + \operatorname{Var}(V) + \operatorname{Cov}(U,V) = 2(E(U^2) - E(U)^2) = \tfrac{1}{6},$$
$$\operatorname{Var}(Y) = \operatorname{Var}(UV) = E(U^2V^2) - E(UV)^2 = E(U^2)E(V^2) - E(U)^2E(V)^2 = \tfrac{1}{9} - \tfrac{1}{16} = \tfrac{7}{144}.$$

as U, V are independent and identically distributed. Thus,

$$\rho_{XY} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{6}\cdot\frac{7}{144}}} = \sqrt{\tfrac{6\cdot 144}{7\cdot 144}} = \sqrt{\tfrac{6}{7}}.$$

4. We have

$$E(X) = \tfrac{3}{2}, \quad E(Y) = E(X^2) = 3, \quad E(XY) = E(X^3) = \tfrac{27}{4}, \quad E(Y^2) = E(X^4) = \tfrac{81}{5},$$
$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y) = \tfrac{27}{4} - \tfrac{3}{2}\cdot 3 = \tfrac{9}{4},$$
$$\operatorname{Var}(X) = E(X^2) - E(X)^2 = \tfrac{3}{4},$$
$$\operatorname{Var}(Y) = E(Y^2) - E(Y)^2 = \tfrac{81}{5} - 9 = \tfrac{36}{5}.$$

Thus,

$$\rho_{XY} = \frac{\frac{9}{4}}{\sqrt{\frac{3}{4}\cdot\frac{36}{5}}} = \tfrac{\sqrt{15}}{4}.$$

5. As $f_{(X,Y)}(x,y) = f_{(X,Y)}(y,x)$, we have $(X,Y) \overset{\text{Law}}{=} (Y,X)$ and in particular, $X \overset{\text{Law}}{=} Y$. Moreover, $f_{(X,Y)}(-x,y) = f_{(X,Y)}(x,y)$ for every y, thus $x \mapsto x f_{(X,Y)}(x,y)$ is an odd function so

$$E(X) = E(Y) = \frac{1}{2\pi} \int_{\mathbb{R}} dx \int_{\mathbb{R}} dy e^{-\sqrt{x^2+y^2}} x = 0,$$

and

$$E(XY) = \frac{1}{2\pi} \int_{\mathbb{R}} dx \int_{\mathbb{R}} dy e^{-\sqrt{x^2+y^2}} xy = 0.$$

Thus, $\text{Cov}(X,Y) = 0$, and $\rho_{XY} = 0$.

Exercise 2.39. 1. Prove Lemma ??.

2. Let $X \sim \text{Bin}(n,p)$. Show that

$$E(X) = np, \quad \text{Var}(X) = np(1-p), \quad E(e^{tX}) = \big(1 + p(e^t - 1)\big)^n$$

for any $t \in \mathbb{R}$.

Solution 2.39. 1. As the X_i's only take values in $\{0,1\}$, the variable $Y = \sum_{k=1}^n$ only takes values in $\{0,1,\dots,n\}$. We just need to check that for any $k \in \{0,1,\dots,n\}$, $P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Now, for any set $I \subset \{1,\dots,n\}$ with $|I| = k$, we have

$$\begin{aligned} P\big(\cap_{i\in I} \{X_i = 1\}, \cap_{j\in I^c}\{X_j = 0\}\big) &= \prod_{i\in I} P(X_i = 1) \cdot \prod_{j\in I^c} P(X_j = 0) \\ &= p^{|I|}(1-p)^{n-|I|} = p^k(1-p)^{n-k}, \end{aligned}$$

where we used independence between the X_i's. Now, the event $Y = k$ is the event that there is exactly one set $I \subset \{1,\dots,n\}$ with $|I| = k$ such that $(\cap_{i\in I}\{X_i = 1\}) \cap (\cap_{j\in I^c}\{X_j = 0\})$ occurs. The probability this event for a given I is $p^k(1-p)^{n-k}$ as computed previously, and there are $\binom{n}{k}$ subsets of $\{1,\dots,n\}$ with size k, so

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

as wanted.

2. We can use the previous point. Let $X_1,\dots,X_n$ be i.i.d. random variables with $X_i \sim \text{Bern}(p)$. By the previous point, $X \overset{\text{Law}}{=} \sum_{i=1}^n X_i$, so

$$E(X) = E\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np,$$

where we used the properties of Bernoulli random variables. Also,

$$\text{Var}(X) = \text{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p),$$

where we used that the X_i's are independent in the third equality, and where we again used the properties of Bernoulli random variables in the last equality. Finally,

$$E(e^{tX}) = E\Big(\exp\Big(t\sum_{i=1}^{n} X_i\Big)\Big) = E\Big(\prod_{i=1}^{n} e^{tX_i}\Big) = \prod_{i=1}^{n} E(e^{tX_i}) = \big(1 + p(e^t - 1)\big)^n$$

where we used that the X_i's are independent in the third equality, and where we again used the properties of Bernoulli random variables in the last equality.

Exercise 2.40. Let $X \sim \mathrm{Geo}(p)$.

1. Show that
$$E(X) = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$
2. Show that
$$E(e^{tX}) = \begin{cases} \frac{pe^t}{1-(1-p)e^t} & \text{for } t < |\ln(1-p)|, \\ +\infty & \text{for } t \geq |\ln(1-p)|. \end{cases}$$
3. Show Lemma **??**.

Solution 2.40. 1. We first compute $E(X)$:

$$\begin{aligned} E(X) &= \sum_{n=1}^{\infty} p(1-p)^{n-1} n = p\sum_{n=1}^{\infty} x^{n-1} n\Big|_{x=1-p} = p\frac{d}{dx}\sum_{n=0}^{\infty} x^n\Big|_{x=1-p} \\ &= p\frac{d}{dx}\frac{1}{1-x}\Big|_{x=1-p} = p\frac{1}{(1-x)^2}\Big|_{x=1-p} = p\frac{1}{p^2} = \frac{1}{p}. \end{aligned}$$

Then, we compute the second moment $E(X^2)$: using the computation of the expectation, we have

$$\begin{aligned} E(X^2) &= \sum_{n=1}^{\infty} p(1-p)^{n-1} n^2 = \sum_{n=1}^{\infty} p(1-p)^{n-1} n(n+1) - \sum_{n=1}^{\infty} p(1-p)^{n-1} n \\ &= \sum_{n=1}^{\infty} p(1-p)^{n-1} n(n+1) - \frac{1}{p} = p\sum_{n=1}^{\infty} x^{n-1} n(n+1)\Big|_{x=1-p} - \frac{1}{p} \\ &= p\frac{d^2}{dx^2}\sum_{n=0}^{\infty} x^{n+1}\Big|_{x=1-p} - \frac{1}{p} = p\frac{d^2}{dx^2}\frac{x}{1-x}\Big|_{x=1-p} - \frac{1}{p} \\ &= p\frac{2}{(1-x)^3}\Big|_{x=1-p} - \frac{1}{p} = p\frac{2}{p^3} - \frac{1}{p} = \frac{2-p}{p^2}. \end{aligned}$$

as

$$\frac{d^2}{dx^2}\frac{x}{1-x} = \frac{d}{dx}\Big(\frac{1}{1-x} + \frac{x}{(1-x)^2}\Big) = \frac{1}{(1-x)^2} + \frac{1}{(1-x)^2} + \frac{2x}{(1-x)^3} = \frac{2}{(1-x)^3}.$$

The variance is then given by

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

2. For $t \geq |\ln(1-p)| = -\ln(1-p)$, we have

$$E(e^{tX}) = \sum_{n=1}^{\infty} e^{tn} p(1-p)^{n-1} \geq \frac{p}{1-p} \sum_{n=1}^{\infty} e^{-\ln(1-p)n}(1-p)^n$$
$$= \frac{p}{1-p} \sum_{n=1}^{\infty} (1-p)^{-n}(1-p)^n = \frac{p}{1-p} \sum_{n=1}^{\infty} 1 = +\infty.$$

Now, for $t < -\ln(1-p)$,

$$E(e^{tX}) = \sum_{n=1}^{\infty} e^{tn} p(1-p)^{n-1} = pe^t \sum_{n=1}^{\infty} (e^t(1-p))^{n-1}$$
$$= pe^t \sum_{n=0}^{\infty} (e^t(1-p))^n = pe^t \frac{1}{1-e^t(1-p)},$$

where we were able to use the geometric series as $e^t(1-p) < 1$ by the choice of t.

3. Let $X_1, X_2, \ldots$ be an i.i.d. sequence of Bernoulli random variables with parameter p. Define
$$Y = 1 + \sum_{n \geq 1} \prod_{i=1}^{n} (1 - X_i).$$
Then, as $\prod_{i=1}^{n}(1-X_i)$ takes values in $\{0,1\}$ (it is 1 if X_i, $i = 1, \ldots, n$ all take value 0, and is 0 otherwise), one has that $P(Y \in \mathbb{N}^*) = 1$, as does a geometric random variable. It is thus sufficient to check that $P(Y = n) = p(1-p)^n$ for all $n \in \mathbb{N}^*$. For $n \geq 1$, we have that $Y_=n$ if and only if $\prod_{i=1}^{k}(1-X_i) = 1$ for all $k \leq n-1$ and $\prod_{i=1}^{n}(1-X_i) = 0$. This happens if and only if $X_1 = X_2 = \cdots = X_{n-1} = 0$, and $X_n = 1$. Thus, using independence,

$$P(Y = n) = P(X_1 = \cdots = X_{n-1} = 0,\ X_n = 1)$$
$$= P(X_1 = 0) \cdots P(X_{n-1} = 0) P(X_n = 1) = (1-p)^{n-1} p,$$

which is indeed the probability that a geometric of parameter p takes the value n.

Exercise 2.41. Let $\lambda \in [0, +\infty)$, and $X \sim \text{Poi}(\lambda)$.

1. Show that
$$E(X) = \lambda, \quad \text{Var}(X) = \lambda, \quad E(e^{tX}) = \exp\big(\lambda(e^t - 1)\big)\ \forall t \in \mathbb{R}.$$

2. Show Lemma **??**.

Solution 2.41. 1. We use a slightly different trick than in the previous exercises: we first compute the moment generating function.

$$M_X(t) = E(e^{tX}) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} e^{tn} = \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda e^t)^n}{n!} = e^{\lambda(e^t - 1)}$$

where we used the exponential series: $\sum_{n=0}^{\infty} \frac{z^n}{n!} = e^z$ for all $z \in \mathbb{R}$. We then use that

$$E(X) = M_X'(0), \quad E(X^2) = M_X''(0),$$

by Theorem **??**. This gives

$$E(X) = e^{\lambda(e^t-1)}\lambda e^t\Big|_{t=0} = \lambda,$$
$$E(X^2) = \left(e^{\lambda(e^t-1)}\lambda e^t + e^{\lambda(e^t-1)}\lambda^2 e^{2t}\right)\Big|_{t=0} = \lambda + \lambda^2.$$

Thus, the variance is then given by

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

2. We first show that 1. implies 2.: suppose $X \sim \mathrm{Poi}(\lambda)$, then, $P(X=0) = e^{-\lambda}$, and for $k \in \mathbb{N}$,

$$\frac{P(X=k+1)}{P(X=k)} = \frac{e^{-\lambda}\frac{\lambda^{k+1}}{(k+1)!}}{e^{-\lambda}\frac{\lambda^k}{k!}} = \frac{\lambda}{k+1}.$$

We then show that 2. implies 1.: suppose that X is a random variable such that $P(X=0) = e^{-\lambda}$ and for all $k \in \mathbb{N}$, $\frac{P(X=k+1)}{P(X=k)} = \frac{\lambda}{k+1}$. Then, for any $k \geq 1$,

$$\begin{aligned} P(X=k) &= \frac{P(X=k)}{P(X=k-1)}P(X=k-1) = \dots \\ &= \frac{P(X=k)}{P(X=k-1)}\frac{P(X=k-1)}{P(X=k-2)}\cdots\frac{P(X=1)}{P(X=0)}P(X=0) \\ &= \frac{\lambda}{k}\frac{\lambda}{k-1}\cdots\frac{\lambda}{1}e^{-\lambda} = \frac{\lambda^k}{k!}e^{-\lambda}, \end{aligned}$$

which is indeed the probability that a Poisson random variable take value k.

Exercise 2.42. Let $n \geq 1$.

1. Let $X \sim \mathrm{Uni}(\{0, 1, \dots, n\})$. Show that[a]

$$E(X) = \frac{n}{2}, \quad \mathrm{Var}(X) = \frac{n^2+2n}{12}, \quad E(e^{tX}) = \frac{e^{(n+1)t}-1}{(n+1)(e^t-1)} \quad \forall t \in \mathbb{R}.$$

2. Let $Y \sim \mathrm{Uni}(\{1, \dots, n\})$. Show that

$$E(Y) = \frac{n+1}{2}, \quad \mathrm{Var}(Y) = \frac{n^2-1}{12}, \quad E(e^{tY}) = \frac{e^{nt}-1}{n(1-e^{-t})} \quad \forall t \in \mathbb{R}.$$

[a]Hint: first show that $\sum_{k=1}^n k = \frac{n(n+1)}{2}$, that $\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$, and recall the truncated geometric series: $\sum_{k=m}^n z^k = \frac{z^m - z^{n+1}}{1-z}$.

Solution 2.42. 1. Start with the expectation. We have

$$2\sum_{k=1}^n k = \sum_{k=1}^n k + \sum_{k=1}^n (n+1-k) = \sum_{k=1}^n (k + (n+1-k)) = n(n+1),$$

so $\sum_{k=1}^n k = \frac{n(n+1)}{2}$. In particular,

$$E(X) = \sum_{k=0}^{n} \frac{1}{n+1} k = \frac{1}{n+1} \sum_{k=1}^{n} k = \frac{n}{2}.$$

Then, we show $\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$ by induction over n. The case $n = 1$ is

$$1 = \frac{(1+1)(2+1)}{6}.$$

Suppose the equality true for n and let us prove it for $n+1$.

$$\begin{aligned}\sum_{k=1}^{n+1} k^2 &= (n+1)^2 + \sum_{k=1}^{n} k^2 \\ &= (n+1)^2 + \frac{n(n+1)(2n+1)}{6} \\ &= \frac{(n+1)(n(2n+1) + 6(n+1))}{6} \\ &= \frac{(n+1)(2n^2+7n+6)}{6} = \frac{(n+1)(n+2)(2n+3)}{6},\end{aligned}$$

where we used the induction hypotheses in the second line. Using this, we obtain

$$E(X^2) = \sum_{k=0}^{n} \frac{1}{n+1} k^2 = \frac{1}{n+1} \sum_{k=1}^{n} k^2 = \frac{n(2n+1)}{6}.$$

So, the variance is given by

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{n(2n+1)}{6} - \frac{n^2}{4} = \frac{4n(2n+1) - 6n^2}{24} = \frac{n^2+2n}{12}.$$

Finally, let us compute the moment generating function. For $t \in \mathbb{R}$,

$$E(e^{tX}) = \sum_{k=0}^{n} \frac{1}{n+1} e^{tk} == \frac{1}{n+1} \frac{1 - e^{(n+1)t}}{1 - e^t} = \frac{e^{(n+1)t} - 1}{(n+1)(e^t - 1)}$$

where we used the truncated geometric series identity.

2. Proceed exactly as in the first point:

$$\begin{aligned} E(Y) &= \sum_{k=1}^{n} \frac{1}{n} k = \frac{n+1}{2}, \quad E(Y^2) = \sum_{k=1}^{n} \frac{1}{n} k^2 = \frac{(n+1)(2n+1)}{6}, \\ \mathrm{Var}(Y) &= E(Y^2) - E(Y)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= (n+1)\frac{2(2n+1) - 3(n+1)}{12} = (n+1)\frac{n-1}{12} = \frac{n^2-1}{12}, \\ E(e^{tY}) &= \sum_{k=1}^{n} \frac{1}{n} e^{tk} = \frac{1}{n} \frac{e^t - e^{t(n+1)}}{1 - e^t} = \frac{e^{nt} - 1}{n(1 - e^{-t})}.\end{aligned}$$

Exercise 2.43. Let $a < b$ and $X \sim \text{Uni}([a,b])$.

1. Show that

$$E(X) = \frac{b+a}{2}, \quad \text{Var}(X) = \frac{(b-a)^2}{12}, \quad E(e^{tX}) = \begin{cases} 1 & \text{if } t = 0, \\ \frac{e^{bt}-e^{at}}{t(b-a)} & \text{if } t \neq 0. \end{cases}$$

2. Prove Lemma **??**.

Solution 2.43. 1. We have

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\tfrac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}.$$

In the same fashion,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[\tfrac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)}.$$

So,

$$\begin{aligned} \text{Var}(X) = E(X^2) - E(X)^2 &= \frac{b^3-a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{4b^3-4a^3-3(b+a)^2(b-a)}{12(b-a)} \\ &= \frac{4b^3-4a^3-3(b+a)(b^2-a^2)}{12(b-a)} = \frac{4b^3-4a^3-3b^3+3ba^2-3ab^2+3a^3}{12(b-a)} \\ &= \frac{b^3-a^3+3ba^2-3ab^2}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12} \end{aligned}$$

Finally, for $t \neq 0$,

$$E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{1}{b-a} \left[\tfrac{e^{tx}}{t}\right]_a^b = \frac{e^{tb}-e^{ta}}{t(b-a)}.$$

2. As X is a uniform random variable on $[a, d$, for any $a \leq b < c \leq d$,

$$P(X \in [b,c]) = \frac{b-c}{d-a}.$$

So, for any $a < b < t_1 < t_2 < c < d$,

$$\begin{aligned} P(X \in [t_1,t_2] \,|\, X \in [b,c]) = \frac{P(X \in [t_1,t_2] \cap [b,c])}{P(X \in [b,c])} &= \frac{P(X \in [t_1,t_2])}{P(X \in [b,c])} \\ &= \frac{t_2-t_1}{d-a}\frac{d-a}{b-c} = \frac{t_2-t_1}{b-c}. \end{aligned}$$

Exercise 2.44. Let $\mu \in \mathbb{R}$, $\sigma \geq 0$.

1. Show that $\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi\sigma^2}$, so that f_X defined in (??) is indeed a probability density.
   Hint: change variables to reduce to $\mu = 0$, $\sigma = 1$. Then, try to compute the square of this integral as an integral over $\mathbb{R}^2$, and use polar coordinates.

2. Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$E(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2, \quad E(e^{tX}) = \exp\Big(\mu t + \tfrac{\sigma^2 t^2}{2}\Big) \ \forall t \in \mathbb{R}.$$

   Hint: for the last point, try completing the square.

3. Show that if $X \sim \mathcal{N}(0, 1)$, and $m \in \mathbb{N}^*$,

$$E(X^m) = \begin{cases} 0 & \text{if } m \text{ is odd,} \\ \frac{m!}{2^{m/2}(m/2)!} & \text{if } m \text{ is even.} \end{cases}$$

   Hint: remember what is an induction.

Solution 2.44. 1. First using the change of variable $y = (x - \mu)/\sigma$

$$\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy.$$

Now, using the hint, we compute

$$\begin{aligned} \Big(\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy\Big)^2 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{y^2+x^2}{2}} dy dx \\ &= \int_0^{\infty} \int_{-\pi}^{\pi} r e^{-\frac{r^2}{2}} d\theta dr = 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr \end{aligned}$$

where we used polar coordinates $(x, y) = (r\cos(\theta), r\sin(\theta))$, see Remark **??**. Finally, we compute using direct integration (note that $\frac{d}{dr} e^{-r^2/2} = -re^{-r^2/2}$):

$$\int_0^{\infty} r e^{-\frac{r^2}{2}} dr = \big[-e^{-\frac{r^2}{2}}\big]_0^{+\infty} = 1.$$

Gathering everything, we obtain the wanted result.

2. First, using $y = x - \mu$ and the previous point,

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

as the last integral is the integral of an odd function over a symmetric domain. Then, using again $y = \frac{x-\mu}{\sigma}$

$$\begin{aligned}
\mathrm{Var}(X) &= E\big((X - E(X))^2\big) = E\big((X-\mu)^2\big) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \underbrace{\big[-ye^{-\frac{y^2}{2}}\big]_{-\infty}^{+\infty}}_{=0} + \frac{\sigma^2}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy}_{=\sqrt{2\pi}}
\end{aligned}$$

where we used integration by part (primitive $ye^{-\frac{y^2}{2}}$ and differentiate y) in the fifth equality, and the first point. Let us now compute the moment generating function. Using again $y = \frac{x-\mu}{\sigma}$,

$$\begin{aligned}
E(e^{tX}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\sigma t y} e^{-\frac{y^2}{2}} dy \\
&= \frac{e^{t\mu + \sigma^2 \frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(y-\sigma t)^2}{2}} dy = \frac{e^{t\mu + \sigma^2 \frac{t^2}{2}}}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy}_{=\sqrt{2\pi}}
\end{aligned}$$

where we used another shift of the variable y, and a completion of the square:

$$\sigma t y - \frac{y^2}{2} = -\tfrac{1}{2}(\sigma^2 t^2 - 2\sigma t y + y^2) + \tfrac{\sigma^2 t^2}{2} = -\tfrac{1}{2}(y - \sigma t)^2 + \tfrac{\sigma^2 t^2}{2}$$

in the third equality.

3. First, let m be odd. Then, the function $x \mapsto x^m e^{-x^2/2}$ is odd. In particular,

$$E(X^m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^m e^{-\frac{x^2}{2}} dx = 0$$

as it is the integral of an odd function over a symmetric domain. Now, for $m = 2n$ even, define the sequence

$$a_n = E(X^{2n}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n} e^{-\frac{x^2}{2}} dx.$$

We know from the previous point that $a_1 = E(X^2) = 1$ (as we have $\mu = 0$ and $\sigma^2 = 1$). Now, using an integration by part, for $n \geq 2$,

$$\begin{aligned}
a_n &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \underbrace{x^{2n-1}}_{=g(x)} \underbrace{xe^{-\frac{x^2}{2}}}_{=f'(x)} dx \\
&= \frac{1}{\sqrt{2\pi}} \big[-x^{2n-1} e^{-\frac{x^2}{2}}\big]_{-\infty}^{+\infty} + \frac{(2n-1)}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n-2} e^{-\frac{x^2}{2}} dx \\
&= \frac{(2n-1)}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n-2} e^{-\frac{x^2}{2}} dx = (2n-1) a_{n-1}.
\end{aligned}$$

We then prove then wanted result, which is $a_n = \frac{(2n)!}{2^n n!}$, by induction. Then case $n=1$ follows from

$$\frac{2!}{2^1 1!} = \frac{2}{2} = 1 = a_n.$$

Then, suppose that the equality holds for $n \geq 1$, and let us show the equality for $n+1$. By the relation we derived on the sequence (and the fact that $n+1 \geq 2$),

$$a_{n+1} = (2(n+1)-1)a_n = (2n+1)\frac{(2n)!}{2^n n!}$$

by the induction hypotheses. Now, $(2n+1) = \frac{(2n+2)(2n+1)}{2(n+1)}$, so

$$a_{n+1} = \frac{(2n+2)(2n+1)}{2(n+1)} \frac{(2n)!}{2^n n!} = \frac{(2(n+1))!}{2^{n+1}(n+1)!},$$

which is the wanted induction step.

Exercise 2.45. Let $\lambda > 0$ and $X \sim \text{Exp}(\lambda)$.

1. Show that
$$\int_0^{+\infty} e^{-\lambda x} dx = \tfrac{1}{\lambda},$$
so that f_X defined in (**??**) is indeed a probability density function.

2. Show that
$$E(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

3. Show that
$$E(e^{tX}) = \begin{cases} \frac{\lambda}{\lambda - t} & \text{if } t < \lambda, \\ +\infty & \text{if } t \geq \lambda. \end{cases}$$

4. Prove Lemma **??**.

Solution 2.45. 1. We have that $\frac{d}{dx}e^{-\lambda x} = -\lambda e^{-\lambda x}$, so
$$\int_0^{+\infty} e^{-\lambda x} dx = \big[-\tfrac{1}{\lambda}e^{-\lambda x}\big]_0^{+\infty} = \tfrac{1}{\lambda}.$$

2. Using integration by parts, we have
$$E(X) = \lambda \int_0^{+\infty} x e^{-\lambda x} dx = \lambda \underbrace{\big[-\tfrac{x}{\lambda}e^{-\lambda x}\big]_0^{+\infty}}_{=0} + \lambda \underbrace{\int_0^{+\infty} \tfrac{1}{\lambda} e^{-\lambda x} dx}_{=\frac{1}{\lambda}},$$
by the first point. Then, again using integration by parts,
$$E(X^2) = \lambda \int_0^{+\infty} x^2 e^{-\lambda x} dx = \lambda\big[-\tfrac{x^2}{\lambda}e^{-\lambda x}\big]_0^{+\infty} + 2\int_0^{+\infty} x e^{-\lambda x} dx = \tfrac{2}{\lambda}E(X) = \tfrac{2}{\lambda^2}.$$
Thus,
$$\text{Var}(X) = E(X^2) - E(X)^2 = \tfrac{2}{\lambda^2} - \tfrac{1}{\lambda^2} = \tfrac{1}{\lambda^2}.$$

3. We have that for $t \geq \lambda$,

$$E(e^{tX}) = \lambda \int_0^{+\infty} e^{tx} e^{-\lambda x} dx \geq \lambda \int_0^{+\infty} 1 dx = +\infty,$$

as $t - \lambda \geq 0$, so $e^{(t-\lambda)x} \geq 1$ for $x \geq 0$. Now, for $t < \lambda$,

$$E(e^{tX}) = \lambda \int_0^{+\infty} e^{tx} e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{-(\lambda - t)x} dx = \frac{\lambda}{\lambda - t}$$

as in the first point.

4. First, for $t \geq 0$, we compute,

$$P(X \geq t) = \lambda \int_0^{+\infty} \mathbb{1}_{[t,\infty)}(x) e^{-\lambda x} dx = \lambda \int_t^{+\infty} e^{-\lambda x} dx = \lambda \big[-\tfrac{1}{\lambda} e^{-\lambda x} \big]_t^{+\infty} = e^{-\lambda t}.$$

Then, for $0 < a < b$,

$$\begin{aligned} P(X \geq b \,|\, X \geq a) = \frac{P(X \geq b, X \geq a)}{P(X \geq a)} &= \frac{P(X \geq b)}{P(X \geq a)} \\ &= \frac{e^{-\lambda b}}{e^{-\lambda a}} = e^{-\lambda(b-a)} = P(X \geq b - a). \end{aligned}$$

Exercise 2.46. Let $X_1, X_2, \dots$ be an independent family of exponential random variables with rate 1 ($X_i \sim \text{Exp}(1)$ for all i's). Let $S_0 = 0$, and for $n \geq 1$, define

$$S_n = \sum_{i=1}^n X_i.$$

Now, let $A > 0$ and define

$$M = \max\{n \geq 0 : \; S_n \leq A\}.$$

1. What is the value of $P(M = k)$ for $k \in \mathbb{N}$?

2. Deduce the law of M.

Solution 2.46. 1. To have $M = k$, we need that the sum of the first k variables is less than A, and that adding the $k+1$th makes the sum bigger than A. Formally,

$$\{M = k\} = \{S_k \leq A\} \cap \{S_{k+1} > A\} = \{S_k \leq A, \; X_{k+1} > A - S_k\}.$$

We can now compute the probability of this event.

$$\begin{aligned} P(M = k) &= P(S_k \leq A, \; X_{k+1} > A - S_k) \\ &= \int_0^\infty dx_1 e^{-x_1} \cdots \int_0^\infty dx_{k+1} e^{-x_{k+1}} \mathbb{1}_{x_1 + \cdots + x_k \leq A} \mathbb{1}_{x_{k+1} > A - (x_1 + \cdots + x_k)} \\ &= \int_0^\infty dx_1 e^{-x_1} \cdots \int_0^\infty dx_k e^{-x_{k+1}} \mathbb{1}_{x_1 + \cdots + x_k \leq A} P\big(X_{k+1} > A - (x_1 + \cdots + x_k)\big) \\ &= \int_0^\infty dx_1 e^{-x_1} \cdots \int_0^\infty dx_k e^{-x_{k+1}} \mathbb{1}_{x_1 + \cdots + x_k \leq A} e^{-A + x_1 + \cdots + x_k} \\ &= e^{-A} \int_0^\infty dx_1 \cdots \int_0^\infty dx_k \mathbb{1}_{x_1 + \cdots + x_k \leq A} \end{aligned}$$

where in the fourth line we used the fact that X_{k+1} is an exponential random variable with rate 1 so for $t > 0$,

$$P(X_{k+1} > t) = \int_t^\infty e^{-x} dx = e^{-t}.$$

To conclude, we need to compute the value of

$$a_k = \int_0^\infty dx_1 \cdots \int_0^\infty dx_k \mathbb{1}_{x_1+\cdots+x_k \le A}.$$

First, we have

$$a_0 = 1, \quad a_1 = \int_0^\infty dx_1 \mathbb{1}_{x_1 \le A} = A.$$

Then, for $k \geq 2$, we make the change of variable

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = T \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} x_1 \\ x_1 + x_2 \\ \vdots \\ x_1 + x_2 + \cdots + x_k \end{pmatrix}$$

$$T = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

We have $|\det T| = 1$ as it is a lower triangular matrix with 1's on the diagonal, so the change of variable formula gives

$$a_k = \int_0^\infty dx_1 \cdots \int_0^\infty dx_k \mathbb{1}_{x_1+\cdots+x_k \le A} = \int_0^A dy_k \int_0^{y_k} dy_{k-1} \cdots \int_0^{y_2} dy_1$$

as the image of $[0, +\infty)^k$ under T is $\{y \in [0, +\infty)^k : \ y_1 \le y_2 \le \cdots \le y_k\}$, and $y_k = x_1 + \cdots + x_k \le A$. To conclude, we have

$$\begin{aligned} a_k = \int_0^A dy_k \int_0^{y_k} dy_{k-1} \cdots \int_0^{y_2} dy_1 &= \int_0^A dy_k \int_0^{y_k} dy_{k-1} \cdots \int_0^{y_3} dy_2 y_2 \\ &= \int_0^A dy_k \int_0^{y_k} dy_{k-1} \cdots \int_0^{y_4} dy_3 \frac{y_3^2}{2} \\ &= \cdots = \int_0^A dy_k \frac{y_k^{k-1}}{(k-1)!} = \frac{A^k}{k!}. \end{aligned}$$

We obtained

$$P(M = k) = e^{-A} a_k = \begin{cases} e^{-A} & \text{if } k = 0, \\ e^{-A} \frac{A^k}{k!} & \text{if } k \in \mathbb{N}^*. \end{cases}$$

2. We recognize the Poisson law of parameter A, so $M \sim \mathrm{Poi}(A)$. This observation justifies the use of Poisson random variables when we want to model "the number of occurrences of an event in a certain time duration" like the number of cars passing on a given road between 13 : 00 and 14 : 00. It also tells us that the parameter of the Poisson variable should be proportional (linear in) the length of the time interval.

Exercise 2.47. Let $Z \sim \mathcal{N}(0,1)$ and $U \sim \mathrm{Uni}([0,1])$.

1. Show that F_Z is continuous and strictly increasing.
2. Show that $F_Z : \mathbb{R} \to (0,1)$ is invertible. Denote its inverse $F_Z^{-1} : (0,1) \to \mathbb{R}$.
3. Show that the random variable $F_Z^{-1}(U)$ has the same law as Z.

Solution 2.47.

1. As already observed, F_Z is a primitive of the density of Z, which is continuous. In particular, F_Z is differentiable and therefore continuous.
2. The function $F_Z : \mathbb{R} \to (0,1)$ is continuous. Moreover, for $t < t'$,

$$F_Z(t') - F_Z(t) = \int_{-\infty}^{t'} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx - \int_{-\infty}^{t} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_{t}^{t'} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx > 0,$$

   so F_Z is strictly increasing. Moreover,

$$\lim_{t \to -\infty} F_Z(t) = 0, \quad \lim_{t \to +\infty} F_Z(t) = 1.$$

   As seen in analysis I (consequence of the intermediate value theorem), F_Z is thus a bijection between $\mathbb{R}$ and $(0,1)$. Denote its inverse $F_Z^{-1} : (0,1) \to \mathbb{R}$.
3. It is sufficient to see that $F_Z^{-1}(U)$ has the same repartition function as Z. We thus write the definition: for $t \in \mathbb{R}$,

$$P\big(F_Z^{-1}(U) \leq t\big) = P\big(U \leq F_Z(t)\big)$$

   as F_Z is an increasing bijection. Finally, as $U \sim \mathrm{Uni}([0,1])$,

$$P\big(U \leq F_Z(t)\big) = \int_0^{F_Z(t)} du = F_Z(t).$$

   So, we obtained that $F_{F_Z^{-1}(U)}(t) = F_Z(t)$ for all $t \in \mathbb{R}$, thus $F_Z^{-1}(U) \overset{\text{Law}}{=} Z$.

What did we use about Z? Can we do the same procedure for other laws than the Gaussian? This kind of operation is extremely useful when sampling random numbers on a computer: it is sufficient to know how to sample a $\mathrm{Uni}([0,1])$ to be able to sample more complicated random variables by applying a suitable function!

Exercise 2.48. A professor knows from experience that students grades at a final exam are random variables with expectation 4.3.

1. Give an upper bound on the probability that a given student grade is higher than 5.5.

Suppose now that the professor also knows that the variance of the grade is 0.4.

2. What can we say about the probability that a given student obtains a grade between 4 and 5.5?
3. How large should a class of students be in order to have a probability at least 0.9 to have that the average grade of the class is in $[4, 4.6]$?

Solution 2.48. Let G denote the grade of the student.

1. From Markov's inequality, we get

$$P(G \geq 5.5) \leq \frac{E(G)}{5.5} = \frac{4.3}{5.5} \approx 0.78.$$

2. From Chebychev's inequality applied to the random variable $G - E(G) = G - 4.3$, we obtain

$$P(G \in [3.3, 5.3]) = P(|G - 4.3| \leq 1) = 1 - P(|G - E(G)| > 1) \geq 1 - \mathrm{Var}(G) = 0.6.$$

3. Denote n the number of students in the class. Denote $G_1, \ldots, G_n$ the grades of the students (we suppose that they are all independent and identically distributed). Set $\bar{G} = \frac{1}{n}\sum_{i=1}^n G_i$ the mean of the class. We have that the expected value of $\bar{G}$ is

$$E(\bar{G}) = \frac{1}{n}\sum_{i=1} E(G_i) = \frac{1}{n} n 4.3 = 4.3.$$

Its variance is then

$$\mathrm{Var}(\bar{G}) = \mathrm{Cov}\Big(\frac{1}{n}\sum_{i=1}^n G_i, \frac{1}{n}\sum_{i=1}^n G_i\Big) = \frac{1}{n^2}\sum_{i,j=1}^n \mathrm{Cov}(G_i, G_j)$$
$$= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Cov}(G_i, G_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(G_i) = \frac{1}{n}0.4,$$

as $\mathrm{Cov}(G_i, G_j) = 0$ for $i \neq j$ (by independence). We can then used Chebychev's inequality:

$$P(|\bar{G} - 4.3| \geq 0.3) \leq \frac{\frac{1}{n}0.4}{0.09} = \frac{40}{9n}.$$

So, in order for the mean grade to have probability at least 0.9 to be in $[4, 4.6]$, it is sufficient to take n such that $\frac{40}{9n} < 0.1$, which amounts to take $n \geq 45$.

Exercise 2.49. X is a Pareto random variable of parameter $(\alpha, \theta) \in (0, +\infty)^2$ if its CDF is

$$F_X(t) = \begin{cases} 1 - \left(\frac{\alpha}{t}\right)^\theta & \text{if } t > \alpha, \\ 0 & \text{else.} \end{cases}$$

1. Check that F_X is indeed a repartition function (right-continuous, non-decreasing, and tends to 0 at $-\infty$ and to 1 at $+\infty$).

2. Compute the first and second moments of X. Under which conditions are they finite?

3. Pareto random variables are often used to to model rare events (big stock market drop, heavy snowfall,...). For $c > \alpha$, and $t \in \mathbb{R}$, compute

$$P(X \leq t \,|\, X > c), \quad E(X \,|\, X > c).$$

Solution 2.49. 1. from the expression, F_X is continuous on $(-\infty, \alpha)$ and on $(\alpha, +\infty)$. Moreover,

$$\lim_{t \nearrow \alpha} F_X(t) = \lim_{t \nearrow \alpha} 0 = 0, \quad \lim_{t \searrow \alpha} F_X(t) = \lim_{t \searrow \alpha} \left(1 - \left(\tfrac{\alpha}{t}\right)^\theta\right) = 0,$$

so F_X is continuous (in particular, it is right-continuous). F_X is constant 0 on $(-\infty, \alpha]$, and $t \mapsto (\alpha/t)^\theta$ is decreasing to 0 on $(\alpha, +\infty)$, so we have the wanted monotonicity and limits.

2. We saw that the density function for X is given by the derivative of its CDF F_X. So, for $p > 0$,

$$E(X^p) = \int_{-\infty}^{+\infty} x^p F_X'(x) dx = \int_\alpha^{+\infty} x^p \theta \tfrac{\alpha^\theta}{x^{\theta+1}} dx$$

as F_X is constant on $(-\infty, \alpha]$. We then have

$$\int_\alpha^{+\infty} x^p \theta \tfrac{\alpha^\theta}{x^{\theta+1}} dx = \theta \alpha^\theta \int_\alpha^{+\infty} x^{p-\theta-1} dx,$$

which converges if and only if $\theta > p$. In particular, we have

$$E(X) = \frac{\theta}{\theta - 1}\alpha \text{ if } \theta > 1, \quad E(X^2) = \frac{\theta}{\theta - 2}\alpha^2 \text{ if } \theta > 2.$$

Thus,

$$\begin{aligned} \mathrm{Var}(X) = E(X^2) - E(X)^2 &= \frac{\theta}{\theta - 2}\alpha^2 - \frac{\theta^2}{(\theta - 1)^2}\alpha^2 \\ &= \alpha^2 \theta \frac{(\theta - 1)^2 - \theta(\theta - 2)}{(\theta - 2)(\theta - 1)^2} \\ &= \alpha^2 \theta \frac{\alpha^2 \theta}{(\theta - 2)(\theta - 1)^2} \end{aligned}$$

and is finite only if $\theta > 2$.

3. For $c > \alpha$, and $t \in \mathbb{R}$, we have

$$P(X > c) = 1 - P(X \leq c) = 1 - F_X(c) = \frac{\alpha^\theta}{c^\theta},$$

and so, for $t > c$,

$$\begin{aligned} G_X(t) := P(X \leq t \,|\, X > c) = 1 - P(X > t \,|\, X > c) &= 1 - \frac{P(X > t)}{P(X > c)} \\ &= 1 - \frac{\alpha^\theta}{t^\theta} \frac{c^\theta}{\alpha^\theta} = 1 - \frac{c^\theta}{t^\theta}. \end{aligned}$$

Then, $G_X(t) = 0$ for $t \leq c$. Now, realize that $G_X(t)$ is the CDF of X under $P(\ \,|\, X > c)$. And that it is the CDF of a Pareto random variable of parameter (c, θ). Thus, if $\theta > 1$,

$$E(X \,|\, X > c) = \frac{\theta}{\theta - 1} c.$$

Exercise 2.50. Let $X_1, X_2, \dots$ be and independent family of random variables with $X_i \sim \text{Bern}(p)$ for all i's, $p \in (0,1)$. Let $S_0 = 0$, and for $n \geq 1$,

$$S_n = \sum_{i=1}^{n} X_i, \quad \bar{S}_n = \tfrac{1}{n} S_n.$$

1. What are the values of $E(S_n)$ and $E(\bar{S}_n)$?
2. What is the value of $\text{Cov}(S_5, S_{10})$?
3. Give a value of n (as small as you can) such that

$$P\big(|\bar{S}_n - p| \geq 0.01\big) \leq 0.01.$$

Solution 2.50. 1. We compute using linearity of the expectation:

$$E(S_n) = E\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} E(X_i) = np, \quad E\big(\bar{S}_n\big) = \tfrac{1}{n} E(S_n) = p.$$

2. We use bi-linearity of the covariance and the fact that $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$ by independence:

$$\text{Cov}(S_5, S_{10}) = \text{Cov}\Big(\sum_{i=1}^{5} X_i, \sum_{j=1}^{10} X_j\Big) = \sum_{i=1}^{5}\sum_{j=1}^{10} \text{Cov}(X_i, X_j)$$
$$= \sum_{i=1}^{5} \text{Var}(X_i) = 5p(1-p).$$

3. By the weak Law of Large Numbers (Theorem **??**), for any $\epsilon > 0$, and any $n \geq 1$,

$$P\big(|\bar{S}_n - E(X_1)| \geq \epsilon\big) \leq \tfrac{\text{Var}(X_1)}{\epsilon^2 n}.$$

Using $E(X_1) = p$, $\text{Var}(X_1) = p(1-p)$, and $\epsilon = 0.01$, we find that for any $n \geq 1$,

$$P\big(|\bar{S}_n - p| \geq 0.01\big) \leq \tfrac{p(1-p)}{10^{-4} n}.$$

We thus want n such that $10^4 \frac{p(1-p)}{n} \leq 0.01$, which amounts to take

$$n \geq 10^6 p(1-p).$$

Exercise 2.51. Let $X : \Omega \to \mathbb{R}$ be a random variable.

1. Show that if $E(X^2) < \infty$, then $E(|X|) < \infty$.
2. Show that if $E(X) = 0$, $M_X(t) \geq 1$ for any $t \in \mathbb{R}$.

Solution 2.51. 1. Suppose $E(X^2) < \infty$. By Jensen's inequality (using that $x \mapsto \sqrt{x}$ is concave over $\mathbb{R}_+$), we have

$$E(|X|) = E(\sqrt{X^2}) \leq \sqrt{E(X^2)} < \infty.$$

2. The function $x \mapsto e^{tx}$ is convex for any $t \in \mathbb{R}$. Thus, by Jensen's inequality

$$M_X(t) = E(e^{tX}) \geq e^{tE(X)} = e^{t \cdot 0} = 1.$$