

# PROBABILITY AND STATISTICS

## Exercise sheet 5 - Solutions

**MC 5.1.** Let  $X$  and  $Y$  be random variables taking values in  $\mathbb{N}$ . Which of the following statements are true? (The number of correct answers is between 0 and 6.)

- (a) If we know all values of  $\mathbb{P}[X = i, Y = j]$  for  $i, j \in \mathbb{N}$ , then we also know all values of  $\mathbb{P}[X = i]$  and  $\mathbb{P}[Y = j]$  for  $i, j \in \mathbb{N}$ .
- (b) If we know all values of  $\mathbb{P}[X = i]$  and  $\mathbb{P}[Y = j]$  for  $i, j \in \mathbb{N}$ , then we also know all values of  $\mathbb{P}[X = i, Y = j]$  for  $i, j \in \mathbb{N}$ .
- (c) Statement (b) is true, provided that we know that  $X$  and  $Y$  are independent.
- (d) Statement (b) is true, provided that we know that  $Z_1 := e^{-X}$  and  $Z_2 := \sqrt{Y}$  are independent.
- (e) Statement (b) is true, provided that we know that  $Z_1 := e^{-X}$  and  $Z_2 := Y \bmod 2$  are independent.
- (f) If we assume  $\mathbb{P}[Y = j] > 0$  and we know all values of  $\mathbb{P}[X = i \mid Y = j]$  and  $\mathbb{P}[Y = j]$  for  $i, j \in \mathbb{N}$ , then we also know all values of  $\mathbb{P}[X = i, Y = j]$  for  $i, j \in \mathbb{N}$ .

**Solution:**

(a) is true. We have

$$\mathbb{P}[X = i] = \sum_{j=1}^{\infty} \mathbb{P}[X = i, Y = j], \quad i \in \mathbb{N}, \quad \text{and} \quad \mathbb{P}[Y = j] = \sum_{i=1}^{\infty} \mathbb{P}[X = i, Y = j], \quad j \in \mathbb{N}.$$

(b) is **not** true. We cannot recover the joint probabilities  $\mathbb{P}[X = i, Y = j]$  knowing only the marginal probabilities  $\mathbb{P}[X = i]$  and  $\mathbb{P}[Y = j]$  for  $i, j \in \mathbb{N}$ .

(c) is true. We have

$$\mathbb{P}[X = i, Y = j] = \mathbb{P}[X = i]\mathbb{P}[Y = j], \quad i, j \in \mathbb{N}.$$

(d) is true. We have

$$\mathbb{P}[X = j, Y = j] = \mathbb{P}[e^X = e^j, \sqrt{Y} = \sqrt{j}] = \mathbb{P}[e^X = e^j]\mathbb{P}[\sqrt{Y} = \sqrt{j}] = \mathbb{P}[X = j]\mathbb{P}[Y = j], \quad i, j \in \mathbb{N}.$$

Equivalently, using Proposition 2.6 from the lecture notes, we have that if  $Z_1 = e^X$  and  $Z_2 = \sqrt{Y}$  are independent, then  $X = \log(Z_1)$  and  $Y = Z_2^2$  are also independent, and so the statement holds since (c) does.

(e) is **not** true. For instance, consider the following probabilities:

$$\mathbb{P}[X = i, Y = j] = \mathbb{P}[X = i]\mathbb{P}[Y = j] \quad \text{and} \quad \mathbb{P}[X = i, Y = j] = \begin{cases} \mathbb{P}[X = i] & \text{if } j = 2i, \\ 0 & \text{otherwise,} \end{cases} \quad i, j \in \mathbb{N},$$

both of which define a valid joint distribution of  $X$  and  $Y$ .

Note that in the first case,  $X$  and  $Y$  are independent, and thus so are  $Z_1$  and  $Z_2$ . In the second case, we have  $\mathbb{P}[2X = Y] = 1$ , meaning  $X$  and  $Y$  are generally not independent. However, since  $\mathbb{P}[Z_2 = 0] = \mathbb{P}[(2X \bmod 2) = 0] = 1$ , it is easy to verify that  $Z_2$  is independent of  $Z_1$ . Similar argument as in (d) does not work here since the map  $n \mapsto n \bmod 2$  does not have an inverse.

(f) is true. We have

$$\mathbb{P}[X = i, Y = j] = \mathbb{P}[X = i | Y = j]\mathbb{P}[Y = j], \quad i, j \in \mathbb{N}.$$

**Exercise 5.2.** Let  $N \in \mathbb{N}$ . Consider the set

$$\Omega = \{0, 1\}^N = \{(a_1, \dots, a_N) : a_1, \dots, a_N \in \{0, 1\}\}$$

and the  $\sigma$ -algebra  $\mathcal{A} = 2^\Omega$ , as well as the probability measure  $\mathbb{P}$  defined by

$$\mathbb{P}[\{(a_1, \dots, a_N)\}] = \begin{cases} \frac{1}{2^{N-1}}, & \text{if } a_1 + \dots + a_N \text{ is even,} \\ 0, & \text{otherwise.} \end{cases}$$

Let  $X_1, \dots, X_N$  be random variables defined by  $X_i((a_1, \dots, a_N)) = a_i$ ,  $(a_1, \dots, a_N) \in \Omega$ ,  $i \in \{1, \dots, N\}$ .

- (a) Show that  $X_1, \dots, X_{N-1}$  are independent.
- (b) Show that  $X_1, \dots, X_N$  are not independent.

**Solution:**

- (a) For each  $a_i \in \{0, 1\}$ , there are  $2^{N-2}$  vectors  $(a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_N) \in \{0, 1\}^{N-1}$  such that  $a_1 + \dots + a_N$  is even. Therefore, we have

$$\mathbb{P}[X_i = a_i] = \frac{2^{N-2}}{2^{N-1}} = \frac{1}{2}.$$

Similarly, for each vector  $(a_1, \dots, a_{N-1}) \in \{0, 1\}^{N-1}$ , there is exactly one value of  $a_N \in \{0, 1\}$  such that  $a_1 + \dots + a_N$  is even. Hence,

$$\mathbb{P}[X_1 = a_1, \dots, X_{N-1} = a_{N-1}] = \frac{1}{2^{N-1}} = \prod_{i=1}^{N-1} \mathbb{P}[X_i = a_i].$$

We conclude that  $X_1, \dots, X_{N-1}$  are independent.

- (b) It holds that

$$\mathbb{P}[X_1 = 0, \dots, X_N = 0] = \frac{1}{2^{N-1}} \neq \prod_{i=1}^N \mathbb{P}[X_i = 0] = \frac{1}{2^N},$$

and thus  $X_1, \dots, X_N$  are not independent.

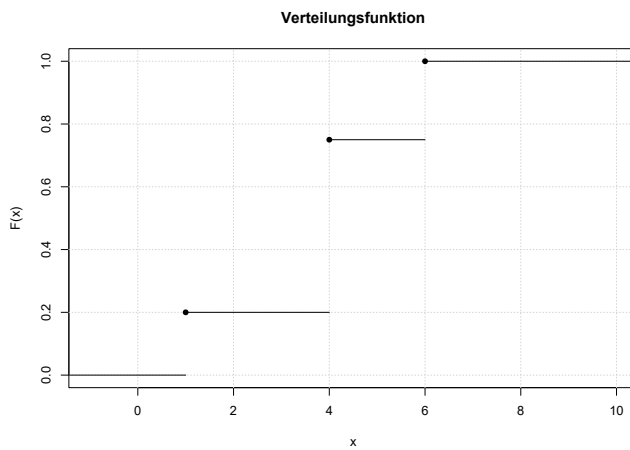
**Exercise 5.3.** Let  $X$  be a discrete random variable with the cumulative distribution function

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

- (a) Sketch the cumulative distribution function of  $X$ .
- (b) Find the distribution of  $X$  (i.e. the values of  $p(x) = \mathbb{P}[X = x]$ ) and sketch it.
- (c) Compute the probabilities  $\mathbb{P}[X = 6]$ ,  $\mathbb{P}[X = 5]$ ,  $\mathbb{P}[2 < X < 5.5]$ ,  $\mathbb{P}[0 \leq X < 4]$ .

**Solution:**

(a) The following sketch illustrates the cumulative distribution function of  $X$ :



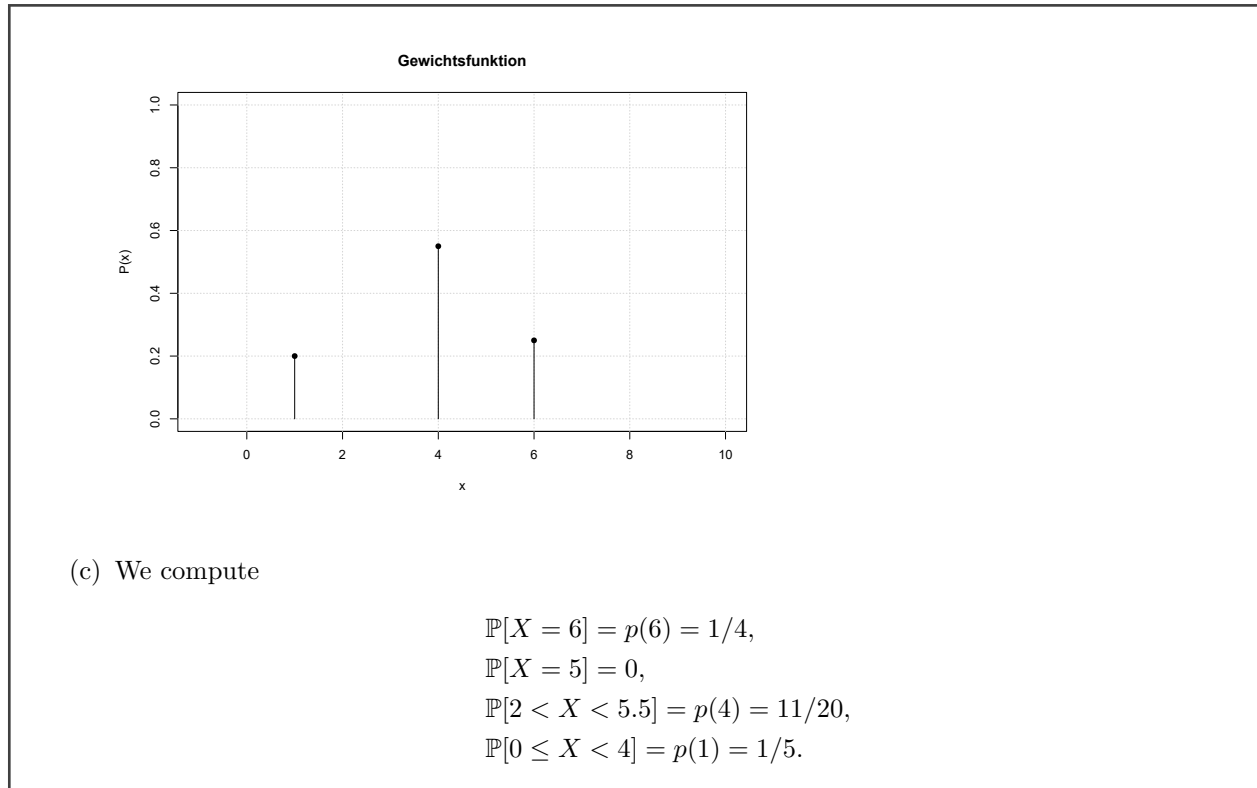
(b) We observe that

$$\mathbb{P}[X = 1] = 1/5, \quad \mathbb{P}[X = 4] = 3/4 - 1/5 = 11/20, \quad \mathbb{P}[X = 6] = 1/4$$

and  $\mathbb{P}[X = x] = 0$  for all  $x \notin \{1, 4, 6\}$ . The probability is given by  $(p(x))_{x \in \{1, 4, 6\}}$  with

$$p(1) = \mathbb{P}[X = 1] = 1/5, \quad p(4) = \mathbb{P}[X = 4] = 3/4 - 1/5 = 11/20, \quad p(6) = \mathbb{P}[X = 6] = 1/4.$$

This is illustrated in the following sketch:



**Exercise 5.4.** Let  $X$  and  $Y$  be two discrete random variables with the following joint distribution:

$$p(j, k) = \mathbb{P}[X = j, Y = k] = \begin{cases} C \left(\frac{1}{2}\right)^k & \text{for } k = 2, 3, \dots \text{ and } j = 1, 2, \dots, k-1 \\ 0 & \text{otherwise,} \end{cases}$$

for some constant  $C \in \mathbb{R}$ .

- (a) Determine the constant  $C$ .
- (b) Compute the (marginal) distributions  $p_X$  and  $p_Y$  of  $X$  and  $Y$ .
- (c) Are  $X$  and  $Y$  independent?

**Solution:**

(a) Since  $p$  is the joint probability mass function of  $X$  and  $Y$ , it must hold that

$$1 = \sum_{k=2}^{\infty} \sum_{j=1}^{k-1} \mathbb{P}[X = j, Y = k].$$

We have

$$\begin{aligned} \sum_{k=2}^{\infty} \sum_{j=1}^{k-1} \mathbb{P}[X = j, Y = k] &= C \sum_{k=2}^{\infty} \sum_{j=1}^{k-1} \left(\frac{1}{2}\right)^k = C \sum_{j=1}^{\infty} \sum_{k=j+1}^{\infty} \left(\frac{1}{2}\right)^k \\ &= C \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^{j+1} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = C \sum_{j=1}^{\infty} \frac{\left(\frac{1}{2}\right)^{j+1}}{1 - \frac{1}{2}} = C \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j = C \frac{\frac{1}{2}}{1 - \frac{1}{2}} = C. \end{aligned}$$

Thus,  $C = 1$ .

(b) We compute for  $j \geq 1$ ,

$$\begin{aligned} p_X(j) &= \mathbb{P}[X = j] = \sum_{k=2}^{\infty} \mathbb{P}[X = j, Y = k] \\ &= \sum_{k=2}^{\infty} p(j, k) = \sum_{k=j+1}^{\infty} \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^{j+1} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^j \end{aligned}$$

and for  $k \geq 2$

$$p_Y(k) = \mathbb{P}[Y = k] = \sum_{j=1}^{k-1} \mathbb{P}[X = j, Y = k] = \sum_{j=1}^{k-1} p(j, k) = (k-1) \left(\frac{1}{2}\right)^k.$$

(c) The random variables  $X$  and  $Y$  are independent if and only if for all  $k \geq 2$  and  $j \geq 1$ ,

$$p(j, k) = p_X(j)p_Y(k).$$

This is not the case, for example, for  $j = 1$  and  $k = 2$ :

$$p(1, 2) = \frac{1}{4} \neq \frac{1}{2} \times \frac{1}{4} = p_X(1)p_Y(2).$$

**Exercise 5.5. [Monty Hall Problem]** You are in a game show and have a choice between three doors. Behind one door is a car, while the other two doors hide goats. You select a door, and the host, who knows what is behind the doors, opens another door revealing a goat. The host then asks you: “Would you like to stick with your originally chosen door or switch to another one?” Assuming you prefer cars over goats, what should you do?

- (a) Construct an appropriate model to answer this question using conditional probabilities.
- (b) Try to find an alternative solution (which must, of course, yield the same answer).

**Solution:**

- (a) We number the doors as 1, 2, 3 in such a way that our initial choice is door 1. We define  $B_i =$  “The car is behind door  $i$ ,” for  $i \in \{1, 2, 3\}$ , and  $A_j =$  “The host opens door  $j$ ,” for  $j \in \{2, 3\}$ . Let  $\Omega = \{1, 2, 3\} \times \{2, 3\}$ . Then  $B_i = \{i\} \times \{2, 3\}$  for  $i \in \{1, 2, 3\}$  and  $A_2 = \{1, 3\} \times \{2\}$  and

$A_3 = \{1, 2\} \times \{3\}$ . We choose the following probabilities:

- $\mathbb{P}[B_1] = \mathbb{P}[B_2] = \mathbb{P}[B_3] = \frac{1}{3}$ , meaning the car is placed randomly.
- $\mathbb{P}[A_2|B_1] = \mathbb{P}[A_3|B_1] = \frac{1}{2}$ , meaning the host randomly picks a goat door if we initially choose the door with the car.
- $\mathbb{P}[A_2|B_2] = 0$  and  $\mathbb{P}[A_3|B_2] = 1$ , since if the car is behind door 2, the host must open door 3.
- $\mathbb{P}[A_2|B_3] = 1$  and  $\mathbb{P}[A_3|B_3] = 0$ .

Using Bayes' theorem, we compute:

$$\begin{aligned}\mathbb{P}[B_1|A_2] &= \frac{\mathbb{P}[A_2|B_1]\mathbb{P}[B_1]}{\mathbb{P}[A_2|B_1]\mathbb{P}[B_1] + \mathbb{P}[A_2|B_2]\mathbb{P}[B_2] + \mathbb{P}[A_2|B_3]\mathbb{P}[B_3]} \\ &= \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} \\ &= \frac{\frac{1}{6}}{\frac{1}{3}} \\ &= \frac{1}{2}\end{aligned}$$

and similarly,

$$\begin{aligned}\mathbb{P}[B_1|A_3] &= \frac{\mathbb{P}[A_3|B_1]\mathbb{P}[B_1]}{\mathbb{P}[A_3|B_1]\mathbb{P}[B_1] + \mathbb{P}[A_3|B_2]\mathbb{P}[B_2] + \mathbb{P}[A_3|B_3]\mathbb{P}[B_3]} \\ &= \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} \\ &= \frac{\frac{1}{6}}{\frac{2}{3}} \\ &= \frac{1}{4}\end{aligned}$$

We see that  $\mathbb{P}[B_1|A_2] = \mathbb{P}[B_1|A_3] = \mathbb{P}[B_1]$ . This also gives  $\mathbb{P}[B_3|A_2] = 1 - \mathbb{P}[B_1|A_2] = \frac{2}{3}$  and  $\mathbb{P}[B_2|A_3] = 1 - \mathbb{P}[B_1|A_3] = \frac{3}{4}$ .

Therefore, you should switch to the other door rather than sticking with your initial choice. You have not gained any additional information about door 1, but you have gained additional information about the final door.

- (b) Consider  $\Omega = \{1, 2, 3\} \times \{1, 2, 3\}$ ,  $\mathcal{F} = 2^\Omega$ , and  $\mathbb{P}$  as the uniform distribution. For  $\omega = (\omega_1, \omega_2)$ ,  $\omega_1$  represents the door number containing the car, and  $\omega_2$  represents the door initially chosen. The decision is whether to switch in the second round or not.

If  $\omega_1 = \omega_2$ , switching results in losing; but if  $\omega_1 \neq \omega_2$ , switching results in winning, as one door is already opened. Thus, the probability of winning the car is  $\frac{6}{9} = \frac{2}{3}$  if we switch, but only  $\frac{3}{9} = \frac{1}{3}$  if we do not switch. Therefore, we should abandon our initial choice and switch.