Probability and Statistics

Exercise sheet 8 - Solutions

MC 8.1. Let X and Y be two independent and identically distributed random variables taking values in $\{1, 2\}$, such that

$$\mathbb{P}[X = i] = \mathbb{P}[Y = i] = \frac{1}{2}, \quad i \in \{1, 2\}.$$

Define the random variable

$$Z := X + Y.$$

(Exactly one answer is correct for each question.)

1. What is the cumulative distribution function of Z?

   (a) $F_Z(a) = \begin{cases} 0, & a < 2, \\ \frac{1}{4}, & 2 \leq a < 3, \\ \frac{3}{4}, & 3 \leq a < 4, \\ 1, & 4 \leq a. \end{cases}$

   (b) $F_Z(a) = \begin{cases} 0, & a < 2, \\ \frac{1}{4}, & 2 \leq a < 3, \\ \frac{1}{2}, & 3 \leq a < 4, \\ 1, & 4 \leq a. \end{cases}$

   (c) $F_Z(a) = \begin{cases} 0, & a < 2, \\ \frac{1}{3}, & 2 \leq a < 3, \\ \frac{3}{4}, & 3 \leq a < 4, \\ 1, & 4 \leq a. \end{cases}$

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

2. What is the value of $\mathrm{Cov}(X, Z)$?

   (a) $\mathrm{Cov}(X, Z) = \frac{1}{4}$.
   (b) $\mathrm{Cov}(X, Z) = 0$.
   (c) $\mathrm{Cov}(X, Z) = \frac{1}{2}$.
   (d) $\mathrm{Cov}(X, Z) = \frac{19}{4}$.

Solution:

1. (a). We compute:

$$\begin{aligned}
\mathbb{P}[Z=2] &= \mathbb{P}[X+Y=2] = \mathbb{P}[X=1, Y=1] = \mathbb{P}[X=1] \times \mathbb{P}[Y=1] = \frac{1}{4}, \\
\mathbb{P}[Z=3] &= \mathbb{P}[X+Y=3] = \mathbb{P}[X=1, Y=2] + \mathbb{P}[X=2, Y=1] \\
&= \mathbb{P}[X=1] \times \mathbb{P}[Y=2] + \mathbb{P}[X=2] \times \mathbb{P}[Y=1] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\
\mathbb{P}[Z=4] &= \mathbb{P}[X+Y=4] = \mathbb{P}[X=2, Y=2] = \mathbb{P}[X=2] \times \mathbb{P}[Y=2] = \frac{1}{4}.
\end{aligned}$$

Therefore,

$$F_Z(a) = \mathbb{P}[Z \leq a] = \begin{cases} 0, & a < 2, \\ \frac{1}{4}, & 2 \leq a < 3, \\ \frac{3}{4}, & 3 \leq a < 4, \\ 1, & 4 \leq a. \end{cases}$$

2. (a). We compute:

$$\mathbb{E}[XZ] = \mathbb{E}[X(X+Y)] = \mathbb{E}[X^2] + \mathbb{E}[XY] = \mathbb{E}[X^2] + \mathbb{E}[X] \times \mathbb{E}[Y],$$

where we used independence of X and Y.

Now, compute each term:

$$\begin{aligned}
\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{2} = \frac{5}{2}, \\
\mathbb{E}[X] = \mathbb{E}[Y] = 1 \times \frac{1}{2} + 2 \times \frac{1}{2} = \frac{3}{2}, \\
\mathbb{E}[Z] = 2 \times \frac{1}{4} + 3 \times \frac{1}{2} + 4 \times \frac{1}{4} = 3, \\
(\text{Alternatively, } \mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y] = 2 \times \frac{3}{2} = 3).
\end{aligned}$$

Therefore:

$$\mathbb{E}[XZ] = \frac{5}{2} + \left(\frac{3}{2}\right)^2 = \frac{5}{2} + \frac{9}{4} = \frac{19}{4}$$

and

$$\mathrm{Cov}(X, Z) = \mathbb{E}[XZ] - \mathbb{E}[X] \times \mathbb{E}[Z] = \frac{19}{4} - \frac{3}{2} \times 3 = \frac{19}{4} - \frac{9}{2} = \frac{1}{4}.$$

Alternative solution: Note that:

$$\mathrm{Cov}(X, Z) = \mathrm{Cov}(X, X+Y) = \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y)$$

Since X and Y are independent, $\mathrm{Cov}(X, Y) = 0$, so:

$$\mathrm{Cov}(X, Z) = \mathrm{Var}(X)$$

From earlier calculations:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{5}{2} - \left(\frac{3}{2}\right)^2 = \frac{5}{2} - \frac{9}{4} = \frac{1}{4}$$

Therefore,

$$\mathrm{Cov}(X, Z) = \frac{1}{4}.$$

MC 8.2. Let X and Y be random variable with joint density

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{9}, & 1 \leq x \leq 4 \text{ and } 1 \leq y \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

(Exactly one answer is correct for each question.)

1. Are X and Y identically distributed, i.e., do X and Y have the same distribution?
   (a) Yes.
   (b) No.
2. Are X and Y independent?
   (a) Yes.
   (b) No.
3. Are X and Y i.i.d.?
   (a) Yes.
   (b) No.
4. Which of the following functions is the density function f_X of X?
   (a) $x \mapsto 1$ for $x \in \mathbb{R}$.
   (b) $x \mapsto \frac{1}{9}$ for $x \in \mathbb{R}$.
   (c) $x \mapsto \frac{1}{3}$ for $x \in \mathbb{R}$.
   (d) $x \mapsto \begin{cases} \frac{x}{9}, & \text{if } x \in [1,4], \\ 1, & \text{if } x > 4, \\ 0, & \text{otherwise.} \end{cases}$
   (e) $x \mapsto \begin{cases} \frac{x-1}{3}, & \text{if } x \in [1,4], \\ 1, & \text{if } x > 4, \\ 0, & \text{otherwise.} \end{cases}$
   (f) $x \mapsto \begin{cases} \frac{1}{9}, & \text{if } x \in [1,4], \\ 0, & \text{otherwise.} \end{cases}$
   (g) $x \mapsto \begin{cases} \frac{1}{3}, & \text{if } x \in [1,4], \\ 0, & \text{otherwise.} \end{cases}$

(h) $x \mapsto \begin{cases} \frac{x}{9}, & \text{if } x \in [1,4], \\ 0, & \text{otherwise.} \end{cases}$

5. Which of the functions from Question 4 is the distribution function F_X of X?

Solution:

(1) (a). Yes, X and Y are identically distributed because $F_X = F_Y$ (see the solutions to 4 and 5).

(2) (a). Yes, X and Y are independent. This follows from the fact that $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$ (see the solution to 4 for f_X and f_Y).

(3) (a). Yes, X and Y are i.i.d. This follows directly from (1) and (2), since i.i.d. means independent and identically distributed.

(4) (g). The marginal density of X is computed as:

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}y = \int_1^4 \frac{1}{9} \cdot \mathbf{1}_{[1,4]}(x)\mathrm{d}y = \frac{3}{9} \cdot \mathbf{1}_{[1,4]}(x) = \begin{cases} \frac{1}{3}, & \text{if } x \in [1,4], \\ 0, & \text{otherwise.} \end{cases}$$

The marginal density f_Y of Y is computed analogously and, due to symmetry, gives the same result.

(5) (e). The cumulative distribution function F_X of X is computed as:

$$F_X(a) = \int_{-\infty}^a f_X(x)\mathrm{d}x = \int_{-\infty}^a \frac{1}{3}\mathbf{1}_{[1,4]}(x)\mathrm{d}x = \begin{cases} 0, & \text{if } a < 1, \\ \int_1^a \frac{1}{3}\mathrm{d}x = \frac{a-1}{3}, & \text{if } a \in [1,4], \\ \int_1^4 \frac{1}{3}\mathrm{d}x = 1, & \text{if } a > 4. \end{cases}$$

The cumulative distribution function F_Y of Y is identical due to symmetry.

MC 8.3. Let X and Y be two random variables with $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[Y^2] < \infty$. Which of the following statements is generally true? (Exactly one answer is correct.)

(a) $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

(b) If X and Y are independent, then $\mathrm{Var}[X-Y] = \mathrm{Var}[X] - \mathrm{Var}[Y]$.

(c) $\mathrm{Var}[X] = \mathrm{Var}[-X]$.

(d) The equality $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$ is only true if X and Y are independent.

Solution:

(a) is not generally true. It holds in specific cases, e.g., when X and Y are independent (or uncorrelated).

(b) is almost never true. Note that $\mathrm{Var}[X-Y] \geq 0$, but $\mathrm{Var}[X] - \mathrm{Var}[Y]$ might be negative.

(c) is true, since

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

(d) is not true. It is also true for instance when X and Y are uncorrelated, see Exercise 8.7.

MC 8.4. Let X be a random variable with $\mathbb{E}[X^2] < \infty$. Which of the following statements are true? (The number of correct answers is between 0 and 4.)

(a) $\mathbb{E}[X^2] = (\mathbb{E}[X])^2$.

(b) $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$.

(c) If X is centered (i.e., $\mathbb{E}[X] = 0$), then $\mathrm{Var}[X] = \mathbb{E}[X^2]$.

(d) $\mathrm{Var}[X] > 0$.

(e) The random variable $Y \coloneqq X - \mathbb{E}[X]$ has the same variance as X.

Solution:

(a) is almost never true. It is true if and only if there exists a constant $c \in \mathbb{R}$ such that $\mathbb{P}[X = c] = 1$.

(b) is true. The map $x \mapsto x^2$ is convex, and so the result follows from Jensen's inequality. Alternatively, the inequality $(X - \mathbb{E}[X])^2 \geq 0$ implies

$$0 \leq \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2,$$

which gives the result.

(c) is true, since $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - 0^2 = \mathbb{E}[X^2]$.

(d) is not generally true. If there exists a constant $c \in \mathbb{R}$ such that $\mathbb{P}[X = c] = 1$, then $\mathrm{Var}[X] = 0$. In fact, this is the only case where (d) does not hold.

(e) is true. We have that $\mathbb{E}[Y] = \mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0$. It follows

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

More generally, it can be seen that adding or subtracting constants from a random variable never changes its variance.

MC 8.5. Let X be a random variable that takes values in the set $\{0, 1, 3\}$ with $\mathbb{E}[X] = 2$. Which of the following statements are true? (The number of correct answers is between 0 and 4.)

(a) $\mathbb{P}[X = 0] \geq \frac{1}{3}$.

(b) $\mathbb{P}[X = 1] \geq \frac{1}{2}$.

(c) $\mathbb{P}[X = 0] \leq \frac{1}{6}$.

(d) $\mathbb{P}[X = 3] \geq \frac{1}{2}$.

Solution:

(a) is not true, for example if $\mathbb{P}[X = 1] = \mathbb{P}[X = 3] = \frac{1}{2}$.

(b) is not true, for example if $\mathbb{P}[X = 0] = \frac{1}{3}$ and $\mathbb{P}[X = 3] = \frac{2}{3}$.

(c) is not true, for example if $\mathbb{P}[X = 0] = \frac{1}{3}$ and $\mathbb{P}[X = 3] = \frac{2}{3}$.

(d) is true. For any choice of $\mathbb{P}$, let $p_i := \mathbb{P}[X = i]$ for $i = 0, 1, 3$, then

$$2 = \mathbb{E}[X] = p_1 + 3p_3 = (1 - p_0 - p_3) + 3p_3 = 1 - p_0 + 2p_3.$$

Thus, $2p_3 = 1 + p_0 \geq 1$, and therefore $\mathbb{P}[X = 3] = p_3 \geq \frac{1}{2}$

Exercise 8.6. Based on many years of research, it is known that the lead concentration X in a soil sample is approximately normally distributed. It is also known that the expected value is 32 ppb (parts per billion) and that the standard deviation is 6 ppb. (The standard deviation is defined as the square root of the variance, i.e. $\mathrm{sd}(X) := \sqrt{\mathrm{Var}[X]}$.)

(a) Sketch the density of X and indicate in the sketch the probability that a soil sample contains between 26 and 38 ppb of lead.

(b) What is the probability that a soil sample contains at most 40 ppb of lead?

Hint: Standardize the random variable and use the table of the standard normal distribution below.

(c) What is the probability that a soil sample contains at most 27 ppb of lead?

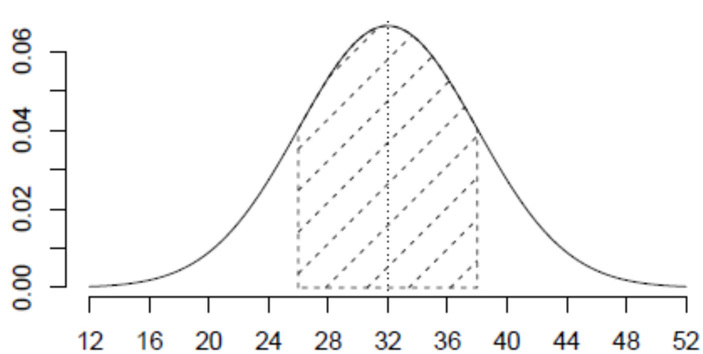
(d) What lead concentration is not exceeded with 97.5% probability? That is, find the value c such that the probability that the lead concentration is less than or equal to c is exactly 97.5%.

(e) What lead concentration is not exceeded with 10% probability?

(f) What is the value of the probability indicated in part (a)?

Solution:

(a) We have



.

(b) From the information above, we have that

$$X \sim \mathcal{N}(\mu, \sigma^2), \qquad \text{with } \mu = 32, \text{ and } \sigma^2 = 6^2.$$

Without a computer, for practical reasons, one usually standardizes to the random variable $Z = (X - \mu)/\sigma$. Then $Z \sim \mathcal{N}(0, 1)$ and:

$$\mathbb{P}[X \leq 40] = \mathbb{P}\left[Z \leq \frac{40 - 32}{6}\right] \approx \mathbb{P}[Z \leq 1.33] = \Phi(1.33) \approx 0.9082.$$

(c) As before, we have $\mathbb{P}[X \leq 27] \approx \mathbb{P}[Z \leq -0.83] = \Phi(-0.83) = 1 - \Phi(0.83) \approx 0.2033$.

(d) Generally, we have

$$\mathbb{P}[X \leq c] = \mathbb{P}\Big[Z \leq \frac{c - 32}{6}\Big] = \Phi\left(\frac{c - 32}{6}\right).$$

Using the table, we find $\Phi(1.96) \approx 0.975$ (That means that 1.96 is the 97.5% quantile of the standard normal distribution). So:

$$\mathbb{P}[X \leq c] \approx 0.975 \iff \frac{c - 32}{6} \approx 1.96 \iff c \approx 32 + 1.96 \times 6 = 43.76.$$

In words: the lead concentration does not exceed 43.76 with a probability of 97.5%.

(e) From the table we have $\Phi(1.28) \approx 0.9$.The standard normal distribution is symmetric around 0, and so $\mathbb{P}[Z \leq c] = 1 - \mathbb{P}[Z \geq -c]$. It follows that $\Phi(-1.28) = 1 - \Phi(1.28) \approx 1 - 0.9 = 0.1$. Thus, arguing as in (d) gives:

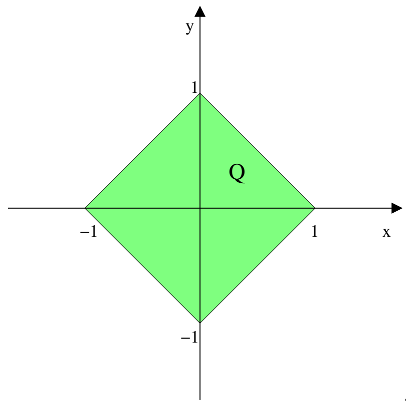
$$c \approx 32 - 1.28 \times 6 = 24.32.$$

In words: the lead concentration does not exceed 24.32 with a probability of 10%.

(f) We have

$$\begin{aligned}\mathbb{P}[26 \leq X \leq 38] &= \mathbb{P}\left[\frac{26 - 32}{6} \leq Z \leq \frac{38 - 32}{6}\right] = \mathbb{P}[-1 \leq Z \leq 1] = \Phi(1) - \Phi(-1) \\ &= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 \approx 2 \times 0.8413 - 1 = 0.6826.\end{aligned}$$

Exercise 8.7. The joint density $f(x, y)$ of two continuous random variables X and Y is constant on the square Q (see sketch) and zero outside of Q.



(a) Determine the joint density of (X, Y).

(b) Determine the marginal densities f_X and f_Y of the random variables X and Y.

(c) Are X and Y independent?

(d) What is the answer to (c) if the square Q is rotated by 45 degrees?

Solution:

(a) The area of Q is $4 \times \frac{1\times 1}{2} = 2$. Therefore, the joint density f is given by

$$f(x, y) = \begin{cases} \frac{1}{2}, & \text{if } (x, y) \in Q, \\ 0, & \text{otherwise.} \end{cases}$$

(b) For the marginal density f_X, we distinguish two cases:

For $-1 \leq x \leq 0$:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y = \int_{-1-x}^{1+x} \frac{1}{2}\mathrm{d}y = \frac{1}{2}(1 + x + 1 + x) = 1 + x.$$

For $0 \leq x \leq 1$:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y = \int_{-1+x}^{1-x} \frac{1}{2}\mathrm{d}y = \frac{1}{2}(1 - x + 1 - x) = 1 - x.$$

Thus,

$$f_X(x) = \begin{cases} 1 + x, & \text{if } -1 \leq x \leq 0, \\ 1 - x, & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

By symmetry, $f_Y = f_X$, so

$$f_Y(y) = \begin{cases} 1+y, & \text{if } -1 \le y \le 0, \\ 1-y, & \text{if } 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) We see that $f_X(x)f_Y(y) \not\equiv \frac{1}{2} = f(x,y)$ for every $(x,y) \in Q$, and so X and Y are not independent.

(d) Due to symmetry, the marginal density of X will be the same as that of Y, but now they are uniform on the interval $\left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$:

$$f_X(x) = f_Y(x) = \begin{cases} \dfrac{1}{\sqrt{2}}, & \text{if } -\dfrac{1}{\sqrt{2}} \le x \le \dfrac{1}{\sqrt{2}}, \\ 0, & \text{otherwise.} \end{cases}$$

Since $f_X(x)f_Y(y) = f(x,y)$ for every $x, y \in \mathbb{R}$, we conclude that X and Y are independent in this case.

Exercise 8.8. For two independent random variables X and Y, it is known from the lecture that

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

i.e., the random variables are uncorrelated. In this problem, we show that the converse is not true in general.

(a) Let $X \sim \mathcal{U}([-\pi,\pi])$. Show that $Y := \cos(X)$ and $Z := \sin(X)$ are uncorrelated, i.e.,

$$\operatorname{Cov}(Y,Z) = 0.$$

(b) Show that Y and Z are not independent.

Hint: If Y and Z were independent, then Y^2 and Z^2 would also be independent. Disprove the latter by considering $\mathbb{P}[Y^2 \le 1/2,\, Z^2 \le 1/2]$.

Solution:

(a) First, we show that Y and Z are uncorrelated. By the definition of covariance,

$$\begin{aligned} \operatorname{Cov}(Y,Z) &= \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z] \\ &= \mathbb{E}[\cos(X)\sin(X)] - \mathbb{E}[\cos(X)]\mathbb{E}[\sin(X)] \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(x)\sin(x)\mathrm{d}x - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(x)\mathrm{d}x\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin(x)\mathrm{d}x\right) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(x)\sin(x)\mathrm{d}x. \end{aligned}$$

Integration by parts gives

$$\int_{-\pi}^{\pi}\cos(x)\sin(x)\mathrm{d}x = \sin^2(x)\Big|_{x=-\pi}^{\pi} - \int_{-\pi}^{\pi}\sin(x)\cos(x)\mathrm{d}x = -\int_{-\pi}^{\pi}\sin(x)\cos(x)\mathrm{d}x.$$

Thus,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(x)\sin(x)\mathrm{d}x = 0,$$

and so

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

Alternative: Using $\sin(x)\cos(x) = \frac{1}{2}\sin(2x)$, we get

$$\int_{-\pi}^{\pi}\cos(x)\sin(x)\mathrm{d}x = \frac{1}{2}\int_{-\pi}^{\pi}\sin(2x)\mathrm{d}x = -\frac{1}{4}\cos(2x)\Big|_{x=-\pi}^{\pi} = 0,$$

since $\cos(2\pi) = \cos(-2\pi) = 1$.

(b) Now we show that Y and Z are not independent. Since

$$Y^2 + Z^2 = \cos^2(X) + \sin^2(X) = 1 \quad \mathbb{P}\text{–a.s.},$$

and we have $\{Y^2 < 1/2,\, Z^2 < 1/2\} \subseteq \{Y^2 + Z^2 < 1\}$, it holds

$$0 \le \mathbb{P}[Y^2 < 1/2,\, Z^2 < 1/2] \le \mathbb{P}[Y^2 + Z^2 < 1] = 0.$$

However,

$$\begin{aligned}
\mathbb{P}[Y^2 < 1/2] &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{1}_{\{\cos^2(x)<1/2\}}\,dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{1}_{\{-1/\sqrt{2}<\cos(x)<1/\sqrt{2}\}}\,dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{1}_{\{\pi/4<x<3\pi/4\}\cup\{-3\pi/4<x<-\pi/4\}}\,dx \\
&= \frac{1}{2\pi}\left(\frac{3\pi}{4} - \frac{\pi}{4} + \frac{3\pi}{4} - \frac{\pi}{4}\right) = \frac{1}{2}.
\end{aligned}$$

By symmetry, we also get

$$\mathbb{P}[Z^2 < 1/2] = \frac{1}{2}.$$

Assume that Y and Z are independent. Then Y^2 and Z^2 are also independent, so

$$\mathbb{P}[Y^2 < 1/2, Z^2 < 1/2] = \mathbb{P}[Y^2 < 1/2] \times \mathbb{P}[Z^2 < 1/2] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

But earlier we saw this probability is 0. This is a contradiction. Hence, Y and Z are not independent.

Exercise 8.9. Let X and Y be two random variables that can only take the values 0 and 1. The joint distribution of (X, Y) satisfies:

$$\mathbb{P}[X = 0] = \frac{1}{2}, \quad \mathbb{P}[Y = 0] = \frac{1}{3}, \quad \text{and} \quad \mathbb{P}[X = 0, Y = 0] = p.$$

(a) What values can p take, and for which values of p are X and Y independent?

(b) Compute the expectations $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[XY]$, as well as the variances $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$, as functions of p. When does $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ hold?

(c) Give an example of random variables U and V such that $\mathbb{E}[UV] = \mathbb{E}[U]\mathbb{E}[V]$, but U and V are not independent.

Solution:

(a) The joint distribution of (X, Y) is:

$$\begin{aligned}
\mathbb{P}[X = 0, Y = 0] &= p, \\
\mathbb{P}[X = 0, Y = 1] &= \mathbb{P}[X = 0] - \mathbb{P}[X = 0, Y = 0] = \frac{1}{2} - p, \\
\mathbb{P}[X = 1, Y = 0] &= \mathbb{P}[Y = 0] - \mathbb{P}[X = 0, Y = 0] = \frac{1}{3} - p, \\
\mathbb{P}[X = 1, Y = 1] &= 1 - \mathbb{P}[X = 0, Y = 0] - \mathbb{P}[X = 0, Y = 1] - \mathbb{P}[X = 1, Y = 1] \\
&= 1 - \left(\frac{1}{2} - p + \frac{1}{3} - p + p\right) = \frac{1}{6} + p.
\end{aligned}$$

Since all of these are probabilities, they must lie in $[0, 1]$. Therefore, p must satisfy:

$$0 \leq p \leq \frac{1}{3}.$$

Conversely, it is easy to check that if $p \in [0, 1/3]$, then all probabilities are valid (non-negative and sum to 1).

X and Y are independent if and only if for all $i, j \in \{0, 1\}$,

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

We have

$$\begin{aligned}
\mathbb{P}[X = 0, Y = 1] &= \frac{1}{2} - p \quad \text{and} \quad \mathbb{P}[X = 0]\mathbb{P}[Y = 1] = \frac{1}{2} \times \frac{2}{3}, \\
\mathbb{P}[X = 1, Y = 0] &= \frac{1}{3} - p \quad \text{and} \quad \mathbb{P}[X = 1]\mathbb{P}[Y = 0] = \frac{1}{2} \times \frac{1}{3}, \\
\mathbb{P}[X = 1, Y = 1] &= \frac{1}{6} + p \quad \text{and} \quad \mathbb{P}[X = 1]\mathbb{P}[Y = 1] = \frac{1}{2} \times \frac{2}{3}.
\end{aligned}$$

We thus have to solve

$$\frac{1}{2} - p = \frac{1}{2} \times \frac{2}{3}, \quad \frac{1}{3} - p = \frac{1}{2} \times \frac{1}{3}, \quad \frac{1}{6} + p = \frac{1}{2} \times \frac{2}{3}.$$

This system is satisfied exactly when $p = \frac{1}{6}$.

(b) We compute

$$
\begin{aligned}
\mathbb{E}[X] &= 0 \times \mathbb{P}[X = 0] + 1 \times \mathbb{P}[X = 1] = 1 - \frac{1}{2} = \frac{1}{2}, \\
\mathbb{E}[Y] &= 0 \times \mathbb{P}[Y = 0] + 1 \times \mathbb{P}[Y = 1] = 1 - \frac{1}{3} = \frac{2}{3}, \\
\mathbb{E}[XY] &= 0 \times (\mathbb{P}[X = 0, Y = 0] + \mathbb{P}[X = 1, Y = 0] + \mathbb{P}[X = 0, Y = 1]) + 1 \times \mathbb{P}[X = 1, Y = 1] \\
&= \frac{1}{6} + p.
\end{aligned}
$$

Since X and Y take values in $\{0, 1\}$, we have $X^2 = X$ and $Y^2 = Y$ $\mathbb{P}$–a.s., and in particular, $\mathbb{E}[X^2] = \mathbb{E}[X]$ and $\mathbb{E}[Y^2] = \mathbb{E}[Y]$. It follows

$$
\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}, \\
\mathrm{Var}[Y] &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}.
\end{aligned}
$$

We have $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if and only if $\frac{1}{6} + p = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$, which is equivalent to $p = \frac{1}{6}$.

(c) Let U be a discrete random variable with values in $\{-1, 0, 1\}$, each with probability $1/3$. Define $V \coloneqq U^2$.

Clearly U and V are not independent. For example:

$$
\mathbb{P}[U = 1, V = 1] = \mathbb{P}[U = 1] = \frac{1}{3} \neq \mathbb{P}[U = 1] \times \mathbb{P}[V = 1].
$$

However, we have:

$$
\mathbb{E}[UV] = \mathbb{E}[U^3] = \mathbb{E}[U] = 0 = \mathbb{E}[U] \times \mathbb{E}[V].
$$

Quantile table for the standard normal distribution

0.5	0.75	0.9	0.95	0.975	0.99	0.995	0.999
0	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

For instance, $\Phi^{-1}(0.9) = 1.2816$, where Φ is the distribution function of $\mathcal{N}(0,1)$.

Table of standard normal distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

For instance, $\mathbb{P}[Z \leq 1.96] = 0.975$.