Probability and Statistics

Exercise sheet 9 - Solutions

MC 9.1. Let $X \sim \mathcal{N}(0,1)$ and let Φ denote the distribution function of X. Which of the following statements are true? (The number of correct answers is between 0 and 4.)

(a) For every $a < b$, we have $\mathbb{P}[X \in (a,b]] = \Phi(b) - \Phi(a)$.

(b) The random variable $Z := 2X - 3$ has distribution $\mathcal{N}(-3,2)$.

(c) For every $a < b$, we have $\mathbb{P}[X \in (a,b]] > 0$.

(d) $\mathbb{P}[X \leq 34] = 1 - \Phi(-34)$.

Solution:

(a) is true.

(b) is **not** true because $Z \sim \mathcal{N}(-3,4)$.

(c) is true, since $\varphi(x) > 0$ for every $x \in \mathbb{R}$, where $\varphi = \Phi'$ is the density function of X.

(d) is true. By symmetry and continuity of the normal distribution,

$$\mathbb{P}[X \leq 34] = \mathbb{P}[X \geq -34] = 1 - \mathbb{P}[X \leq -34] = 1 - \Phi(-34).$$

Exercise 9.2. Let X and Y be random variables with joint density

$$f_{X,Y}(x,y) = cxye^{-y}\mathbf{1}_{[0,y]}(x)\mathbf{1}_{[0,\infty)}(y) = \begin{cases} cxye^{-y}, & \text{if } 0 \leq x \leq y \text{ and } 0 \leq y, \\ 0, & \text{otherwise,} \end{cases}$$

for a constant $c \in \mathbb{R}$.

(a) Find the value of c.

Hint: You may use the identity

$$\int_0^\infty y^n e^{-y} \mathrm{d}y = n! \quad \text{for } n \in \mathbb{N}.$$

(b) Find the marginal density f_Y of Y.

(c) Compute the expectation $\mathbb{E}[X^2/Y]$.

Solution:

(a) We must have $1 = \iint f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$.

$$\begin{aligned}
\iint f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y &= c\int_0^\infty \int_0^y xye^{-y}\mathrm{d}x\mathrm{d}y \\
&= c\int_0^\infty ye^{-y} \times \frac{x^2}{2}\Big|_{x=0}^{y}\mathrm{d}y \\
&= c\int_0^\infty \frac{y^3}{2}e^{-y}\mathrm{d}y \\
&= \frac{c}{2}\int_0^\infty y^3 e^{-y}\mathrm{d}y \\
&= 3c.
\end{aligned}$$

So we get $c = \frac{1}{3}$.

(b) We have

$$\begin{aligned}
f_Y(y) &= \int f_{X,Y}(x,y)\mathrm{d}x \\
&= c\mathbf{1}_{[0,\infty)}(y)\int_0^y xye^{-y}\mathrm{d}x \\
&= cye^{-y}\mathbf{1}_{[0,\infty)}(y) \times \frac{x^2}{2}\Big|_{x=0}^{y} \\
&= c\frac{y^3}{2}e^{-y}\mathbf{1}_{[0,\infty)}(y) \\
&= \frac{y^3}{6}e^{-y}\mathbf{1}_{[0,\infty)}(y), \quad y \in \mathbb{R}.
\end{aligned}$$

(c) We compute

$$\begin{aligned}
\mathbb{E}\left[\frac{X^2}{Y}\right] &= \iint \frac{x^2}{y} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\
&= c\int_0^\infty \int_0^y x^3 e^{-y}\mathrm{d}x\mathrm{d}y \\
&= c\int_0^\infty \frac{y^4}{4}e^{-y}\mathrm{d}y \\
&= \frac{c}{4} \times 4! \\
&= 6c \\
&= 2.
\end{aligned}$$

Exercise 9.3. Let $S \sim \mathcal{N}(-5, 4^2)$ and $T \sim \mathcal{N}(10, 3^2)$ be independent.

(a) Compute $\mathbb{P}[S < T]$.

(b) Would the computation of $\mathbb{P}[S < T]$ also be correct without the assumption of independence?

(c) Compute the variance $\mathrm{Var}[R]$ of $R := S - 2T$.

(d) Would the computation of $\mathrm{Var}[R]$ also be correct without the assumption of independence?

Hint: You may use the following fact: If X and Y are **independent** and normally distributed, then $X+Y$ is also normally distributed.

Let further $U \sim \mathcal{U}([1,3])$ and $V \sim \mathcal{U}([0,4])$ (i.e., $f_U(u) = \frac{1}{2}\mathbf{1}_{[1,3]}(u)$ and $f_V(v) = \frac{1}{4}\mathbf{1}_{[0,4]}(v)$) be independent.

(e) Compute $\mathbb{E}[2U + V^3]$.

(f) Would the computation of $\mathbb{E}[2U + V^3]$ also be correct without the assumption of independence?

Solution:

(a) $\mathbb{P}[S < T] = \mathbb{P}[S - T < 0]$. Since S and $-T$ are independent (just like S and T), and $-T \sim \mathcal{N}(-10, 3^2)$, we have $S - T \sim \mathcal{N}(-5-10, 4^2+3^2) = \mathcal{N}(-15, 25)$.

We consider the standardized variable $Z := \frac{S-T+15}{5} \sim \mathcal{N}(0,1)$ and compute

$$\mathbb{P}[S < T] = \mathbb{P}[S - T < 0] = \mathbb{P}\left[\frac{S-T+15}{5} < \frac{15}{5}\right] = \Phi(3) \approx 0.9987.$$

(b) No. For example, consider $T := \frac{55}{4} + \frac{3}{4}S$, which implies $T \sim \mathcal{N}(10, 3^2)$ and $S - T = \frac{1}{4}S - \frac{55}{4} \sim \mathcal{N}(-15, 1)$, hence $S - T + 15 \sim \mathcal{N}(0,1)$, and thus $\mathbb{P}[S - T < 0] = \Phi(15) > \Phi(3)$.

(c) Since S and $-2T$ are independent, we have $\mathrm{Var}[R] = \mathrm{Var}[S - 2T] = \mathrm{Var}[S] + 2^2\mathrm{Var}[T] = 4^2 + 2^2 \times 3^2 = 16 + 36 = 52$.

(d) No. You can try computing the probability for instance for the case described in (b), which gives a different result.

(e) First, compute

$$\mathbb{E}[U] = \int_{-\infty}^{\infty} u f_U(u)\mathrm{d}u = \int_1^3 \frac{u}{2}\mathrm{d}u = \frac{u^2}{4}\Big|_{u=1}^{3} = 2.$$

Alternatively, since $U \sim \mathcal{U}([a,b])$, we can use the formula $\mathbb{E}[U] = \frac{a+b}{2}$ to compute

$$\mathbb{E}[U] = \frac{1+3}{2} = 2.$$

Further,

$$\mathbb{E}[V^3] = \int_{-\infty}^{\infty} v^3 f_V(v)\mathrm{d}v = \int_0^4 \frac{v^3}{4}\mathrm{d}v = \frac{1}{4 \times 4} \times 4^4 = 16.$$

Using linearity of expectation,

$$\mathbb{E}[2U + V^3] = 2\mathbb{E}[U] + \mathbb{E}[V^3] = 4 + 16 = 20.$$

(f) Yes, because we only used linearity of expectation, not independence. Therefore, the result remains valid even if U and V are not independent.

Exercise 9.4. A rectangle is given with random side lengths X and Y. The joint density of X and Y is given by

$$f_{X,Y}(x,y) := \begin{cases} C(x^2+y^2), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine the constant C.

(b) Compute the marginal densities of X and Y.

(c) Are X and Y independent? Justify your answer.

(d) Compute the probability that side X is more than twice as long as side Y.

(e) Compute the expected area of the rectangle.

Solution:

(a) Since $\iint f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = 1$, we compute:

$$\int_0^1 \int_0^1 C(x^2+y^2)\mathrm{d}x\mathrm{d}y = C\int_0^1 \left(\frac{1}{3}+y^2\right)\mathrm{d}y = C\left(\frac{1}{3}+\frac{1}{3}\right) = \frac{2}{3}C.$$

So we find $C = \frac{3}{2}$.

(b) For the marginal density of X, for $0 \le x \le 1$, we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y = \frac{3}{2}\int_0^1 (x^2+y^2)\mathrm{d}y = \frac{3}{2}x^2+\frac{1}{2}.$$

Thus,

$$f_X(x) = \begin{cases} \frac{3}{2}x^2+\frac{1}{2}, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Analogously, for the marginal density of Y, for $0 \le y \le 1$, we have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x = \frac{3}{2}\int_0^1 (x^2+y^2)\mathrm{d}x = \frac{3}{2}y^2+\frac{1}{2}.$$

Thus,

$$f_Y(y) = \begin{cases} \frac{3}{2}y^2+\frac{1}{2}, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, which is not the case here. Therefore, X and Y are not independent.

(d)

$$\begin{aligned}
\mathbb{P}[X > 2Y] &= \iint \mathbf{1}_{\{x>2y\}} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \frac{3}{2}\int_0^1 \int_0^{x/2} (x^2+y^2)\mathrm{d}y\mathrm{d}x \\
&= \frac{3}{2}\int_0^1 \left(\frac{x^3}{2}+\frac{x^3}{8\times 3}\right)\mathrm{d}x = \frac{3}{2}\left(\frac{1^4}{4\times 2}+\frac{1^4}{4\times 4\times 3}\right) \\
&= \frac{3}{16}\left(1+\frac{1}{12}\right) = \frac{3}{16}\times\frac{13}{12} = \frac{13}{64} \approx 0.2031.
\end{aligned}$$

(e)

$$\begin{aligned}
\mathbb{E}[XY] &= \iint xy f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \frac{3}{2}\int_0^1 \int_0^1 (x^3y+xy^3)\mathrm{d}x\mathrm{d}y \\
&= \frac{3}{2}\int_0^1 \left(\frac{1}{4}y+\frac{1}{2}y^3\right)\mathrm{d}y = \frac{3}{2}\left(\frac{1}{8}+\frac{1}{8}\right) = \frac{3}{8}.
\end{aligned}$$

Exercise 9.5. Let $X_1, X_2, \dots$ be a sequence of independent, identically distributed random variables with distribution function F. The empirical distribution function $F_n : \mathbb{R} \times \Omega \to [0,1]$ is defined as

$$F_n(t,\omega) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i(\omega)\le t\}}.$$

The value $F_n(t,\omega)$ thus describes the relative frequency of those $X_i(\omega)$ that are less that t among the first n. Therefore, $\omega \mapsto F_n(t,\omega)$ is a random variable for each $t \in \mathbb{R}$.

(a) Let $t \in \mathbb{R}$ be arbitrary and define $Y_i := \mathbf{1}_{\{X_i \le t\}}$ for $i \in \mathbb{N}$. Show that $Y_1, Y_2, \dots$ is a sequence of independent, identically distributed random variables. What is $\mathbb{E}[Y_1]$?

(b) Show that for every $t \in \mathbb{R}$, the empirical distribution function $F_n(t)$ converges almost surely to $F(t)$ as $n \to \infty$.

Solution:

(a) For each $i \in \mathbb{N}$, the random variable Y_i can only take the values 1 and 0. We have

$$\mathbb{P}[Y_i = 0] = \mathbb{P}[X_i > t] = \mathbb{P}[X_1 > t]$$

and

$$\mathbb{P}[Y_i = 1] = \mathbb{P}[X_i \le t] = \mathbb{P}[X_1 \le t],$$

since $X_1, X_2, \dots$ are identically distributed. Hence, the sequence $Y_1, Y_2, \dots$ is also identically distributed, and we have

$$\mathbb{E}[Y_1] = 0 \times \mathbb{P}[Y_i = 0] + 1 \times \mathbb{P}[Y_i = 1] = \mathbb{P}[X_1 \le t] = F(t),$$

where we recall that F denotes the distribution function of X_1. Since $X_1, X_2, \dots$ are independent, the random variables $Y_1, Y_2, \dots$ are also independent.

(b) Let $t \in \mathbb{R}$ be arbitrary, and let $Y_1, Y_2, \dots$ be as defined in (a). We know that $Y_1, Y_2, \dots$ is a sequence of independent and identically distributed random variables. By the law of large numbers, it follows that

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le t\}} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{n\to\infty} \mathbb{E}[Y_1] = F(t) \quad \text{almost surely.}$$

Exercise 9.6. In this problem, we compute the limit

$$\lim_{n\to\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}.$$

(a) Let $X \sim \text{Poisson}(\lambda)$ for some $\lambda > 0$. Compute $\mathbb{E}[X]$ and $\text{Var}[X]$.
Hint: It might be easier to compute $\mathbb{E}[X(X-1)]$ and then use linearity to find $\mathbb{E}[X^2]$.

(b) Use the central limit theorem to show that the limit above is equal to $1/2$.
Hint: Recall that for two independent random variables X, Y with $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$, and $Y \sim \text{Poisson}(\mu)$, $\mu > 0$, the sum satisfies $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Solution:

(a) By the formula for expectation for discrete random variables, we have

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

Similarly, we get

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

Hence,

$$\operatorname{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

(b) Let $X_1, X_2, \ldots$ be independent and identically distributed Poisson(1) random variables. From part (a), we know that $\mathbb{E}[X_1] = \operatorname{Var}[X] = 1$.

Moreover, $S_n = X_1 + \cdots + X_n$ is again Poisson distributed with parameter n (see the hint). By the central limit theorem, we have

$$e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \sum_{k=0}^{n} e^{-n} \frac{n^k}{k!} = \sum_{k=0}^{n} \mathbb{P}[S_n = k] = \mathbb{P}[S_n \le n] = \mathbb{P}\left[\frac{S_n - n}{\sqrt{n}} \le 0\right] \xrightarrow{n\to\infty} \Phi(0) = \frac{1}{2}.$$

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