Probability and Statistics

Exercise sheet 8

Exercise 8.1

(a) Let

$$f(x) := \frac{1}{x^k} I_{[1,+\infty)}(x).$$

For which values of k, if any, is f a density function? Whatwould change if we consider instead $g(x) := cf(x)$ with $c > 0$?

(b) Give an example of a density function f such that $c\sqrt{f}$ cannot be a density function for any $c > 0$.

(c) Let

$$f(x) = c\,|x|\,(1 - x^2) I_{[-1,1]}(x).$$

(i) Find $c > 0$ such that f is a density function.

(ii) Find the cdf corresponding to this density.

(iii) Compute $\mathbb{P}\left[X < -\frac{1}{2}\right]$ and $\mathbb{P}\left[|X| \leq \frac{1}{2}\right]$.

Solution 8.1

(a) From the lecture, we know that for a measurable $f \geq 0$ to be a density on $\mathbb{R}$, it has to satisfy

$$\int_{\mathbb{R}} f(t)dt = 1.$$

f is measurable since it is piecewise continuous. We see that for f to be at all integrable, k has to be strictly larger than 1.

Let $k > 1$. Then

$$\int_{\mathbb{R}} \frac{dx}{x^k} I_{[1,\infty)}(x) = \int_1^\infty \frac{dx}{x^k} = \frac{1}{k-1}.$$

Then $k - 1 = 1$ implies $k = 2$. Hence, $k = 2$ is the only possibility for f to be a density.

For g, the condition becomes instead $\frac{c}{k-1} = 1$ or $c = k - 1$; so for any $k > 1$, $g(x) = \frac{k-1}{x^k} I_{[1,+\infty)}(x)$ is a density function.

(b) If we take

$$f(x) = \frac{1}{x^2} I_{[1,+\infty)}(x)$$

then

$$\sqrt{f(x)} = \frac{1}{x} I_{[1,+\infty)}(x)$$

is not integrable, which implies that there is no $c \in \mathbb{R}$ such that $c\sqrt{f}$ is a density.

(c)
$$f(x) = c|x|(1-x^2)I_{[-1,1]}(x).$$

(i) To find c, we use the condition

$$1 = \int_{\mathbb{R}} f(x)dx = 2c\int_0^1 x(1-x^2)dx = 2c\left[-\frac{(1-x^2)^2}{4}\right]_0^1 = \frac{c}{2}(0+1) = \frac{c}{2}.$$

Thus, $c = 2$.

(ii) By definition, the cdf of the random variable whose density is f is given by

$$\begin{aligned} F(x) = \mathbb{P}[X \le x] &= \int_{-\infty}^{x} f(t)dt \\ &= \begin{cases} 0 & \text{if } x < -1, \\ 2\int_{-1}^{x} |t|(1-t^2)dt & \text{if } -1 \le x < 1, \\ 1 & \text{if } x \ge 1. \end{cases} \end{aligned}$$

In the interval $[-1, 1]$, there are two further cases:

$$F(x) = \begin{cases} 0 & \text{if } x < -1, \\ 2\int_{-1}^{x}(-t)(1-t^2)dt & \text{if } -1 \le x \le 0, \\ 2\int_{-1}^{0}(-t)(1-t^2)dt + 2\int_0^x t(1-t^2)dt & \text{if } 0 \le x \le 1, \\ 1 & \text{if } x \ge 1, \end{cases}$$

where

$$2\int_{-1}^{x}(-t)(1-t^2)dt = \left[\frac{(1-t^2)^2}{2}\right]_{-1}^{x} = \frac{(1-x^2)^2}{2} \quad \text{if } -1 \le x \le 0$$

and

$$2\int_0^x t(1-t^2)dt = \left[-\frac{(1-t^2)^2}{2}\right]_0^x = \frac{1}{2}(1-(1-x^2)^2) \quad \text{if } 0 \le x \le 1.$$

Thus

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

To conclude,

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

As a quick check of monotonicity, one can observe that $(1-x^2)^2$ decreases as $|x|$ increases, and therefore each of the branches is monotonically increasing. Moreover, one sees that $F(0) = \frac{1}{2}$ is the same on the two middle branches.

(iii) We compute

$$\mathbb{P}\left[X < -\frac{1}{2}\right] = \mathbb{P}\left[X \leq -\frac{1}{2}\right] = F\left(-\frac{1}{2}\right) = \frac{1}{2}\left(1 - \frac{1}{4}\right)^2 = \frac{9}{32}$$

(in the first step we used the fact that X is absolutely continuous, implying that $\mathbb{P}[X = a] = 0$ for any $a \in \mathbb{R}$). Moreover, using symmetry to get $F(x) = 1 - F(-x)$

$$\begin{aligned}\mathbb{P}\left[|X| \leq \frac{1}{2}\right] &= F\left(\frac{1}{2}\right) - F\left(-\frac{1}{2}\right) \\ &= 1 - F\left(-\frac{1}{2}\right) - F\left(-\frac{1}{2}\right) \\ &= 1 - 2 \times \frac{9}{32} \\ &= 1 - \frac{9}{16} = \frac{7}{16}.\end{aligned}$$

Exercise 8.2

(a) Consider a random variable $X \sim \mathcal{U}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Find $\mathbb{E}[\sin X]$ and $\mathrm{Var}[\sin X]$.
Hint: $\sin(x)\sin(y) = 1/2(\cos(x-y) - \cos(x+y))$

(b) The lengths of the sides of a triangle are X, $2X$ and $2.5X$ with $X \sim \mathcal{U}(0, \alpha)$ for some $\alpha > 0$.

- Find the mean and variance of its area.
  Hint: Recall that if
  $$s = \frac{a+b+c}{2}$$
  with a, b, c the lengths of the sides, then the area of the triangle is
  $$|\Delta| = \sqrt{s(s-a)(s-b)(s-c)}$$
  (Heron's formula).
- How should we choose α so that the mean area is ≥ 1?

Solution 8.2

(a) We have

$$\mathbb{E}[\sin X] = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin x}{\pi} dx = 0$$

and by using the hint we obtain

$$\mathrm{Var}[\sin X] = \mathbb{E}\left[\sin^2 X\right] = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^2 X}{\pi} dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1 - \cos(2x)}{2\pi} dx = \frac{1}{2}.$$

(b) The area of the triangle is given by Heron's formula as

$$|\Delta| = X^2\sqrt{2.75 \times 1.75 \times 0.75 \times 0.25} = AX^2$$

for a constant $A \approx 0.95$.

- Note that if $X \sim \mathcal{U}(0, \alpha)$, then $Y = \frac{X}{\alpha} \sim \mathcal{U}(0,1)$. Therefore,

$$\mathbb{E}[|\Delta|] = \mathbb{E}\left[AX^2\right] = A\alpha^2\mathbb{E}\left[Y^2\right] = \frac{A\alpha^2}{3}$$

as we found in Exercise 7.3(a).
Moreover,

$$\mathbb{E}\left[|\Delta|^2\right] = A^2\alpha^4\mathbb{E}\left[Y^4\right] = \frac{A^2\alpha^4}{5}$$

also by Exercise 7.3(a). Finally

$$\mathrm{Var}[|\Delta|] = \mathbb{E}\left[|\Delta|^2\right] - \mathbb{E}[|\Delta|]^2 = \frac{4A^2\alpha^4}{45}.$$

- We need to ensure that

$$\frac{A\alpha^2}{3} \geq 1 \Leftrightarrow \alpha \geq \sqrt{\frac{3}{A}} \approx 1.78,$$

so this is the suitable value of α.

Exercise 8.3 It costs 1 dollar to play a certain slot machine in Las Vegas. The machine is set by the house to pay 2 dollars with probability 0.45 and nothing with probability 0.55.

Let X_i be the house's net winnings on the i^{th} play of the machine.

Let $S_n := \sum_{i=1}^n X_i$ be the house's winnings after n plays of the machine. Assuming that successive plays are independent, find:

(a) $\mathbb{E}[S_n]$;

(b) $\mathrm{Var}[S_n]$;

(c) the approximate probability that after $10,000$ rounds of the machine, the house's winnings are between 800 and $1,100$ dollars.

Solution 8.3 The house's winnings resulting from each play are independent, and take values

$$X_i = \begin{cases} 1, & \text{with probability } 0.55, \\ -1, & \text{with probability } 0.45. \end{cases}$$

Therefore, one can observe that

$$B_i := \frac{X_i + 1}{2}$$

are independent Ber(0.55) random variables, and that

$$\tilde{S}_n = \frac{S_n + n}{2} \text{ or } S_n = 2\tilde{S}_n - n,$$

where $\tilde{S}_n := \sum_{i=1}^n B_i$ has a $\mathrm{Bin}(n, 0.55)$ distribution.

From the considerations above,

(a)

$$\mathbb{E}[S_n] = 2\mathbb{E}\left[\tilde{S}_n\right] - n = 1.1n - n = 0.1n$$

and

(b)
$$\mathrm{Var}\left[S_n\right] = 4\,\mathrm{Var}\left[\tilde{S}_n\right] = 4 \times 0.55 \times 0.45n = 0.99n,$$
by the linearity properties of expectations and variance, and using the known values for a binomial distribution.

Alternatively one could easily calculate the expectation $\mathbb{E}\left[X_i\right] = 0.55 - 0.45 = 0.1$ and the variance $\mathrm{Var}\left[X_i\right] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[X_i\right]^2 = 1 - 0.1^2 = 0.99$ of X_i to directly use the linearity properties of the expectation and (due to independence) the variance to get the results of (a) and (b).

(c) Since the X_n are i.i.d., we can find a good Gaussian approximation to this probability by the central limit theorem. Since we calculated the expectation and variance before, for large n,
$$S_n^* := \frac{S_n - 0.1n}{\sqrt{0.99n}} \overset{\text{approx}}{\sim} \mathcal{N}(0,1),$$
and in particular
$$S_{10{,}000}^* = \frac{S_{10{,}000} - 1{,}000}{\sqrt{9{,}900}} \overset{\text{approx}}{\sim} \mathcal{N}(0,1).$$
Therefore we obtain
$$\begin{aligned}\mathbb{P}\left[800 \le S_{10{,}000} \le 1{,}100\right] &= \mathbb{P}\left[-\frac{200}{\sqrt{9{,}900}} \le \frac{S_n - 1{,}000}{\sqrt{9{,}900}} \le \frac{100}{\sqrt{9{,}900}}\right] \\ &\approx \Phi\left(\frac{100}{\sqrt{9{,}900}}\right) - \Phi\left(-\frac{200}{\sqrt{9{,}900}}\right) \\ &\approx 0.82.\end{aligned}$$

Exercise 8.4 Consider the joint density
$$f_{X,Y}(x,y) = \begin{cases} cxy, & 1 \le x \le 3 \text{ and } 1 \le y \le 3, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the normalising constant c.

(b) Are X and Y independent? Why?

(c) Find $\mathbb{E}\left[X\right]$, $\mathbb{E}\left[Y\right]$ and $\mathbb{E}\left[XY\right]$.

Solution 8.4
$$f_{X,Y}(x,y) = \begin{cases} cxy, & 1 \le x \le 3 \text{ and } 1 \le y \le 3, \\ 0, & \text{otherwise.} \end{cases}$$
(a) We find the normalising constant by integrating via
$$1 = \int_{\mathbb{R}^2} f_{X,Y}(x,y)dxdy = \int_1^3 \int_1^3 cxy\,dxdy = c\int_1^3 xdx \int_1^3 ydy = c \times 4 \times 4 = 16c,$$
so $c = \frac{1}{16}$.

(b) Yes, X and Y are independent. We can see this immediately since the joint density can be factorized as
$$f_{X,Y}(x,y) = \frac{xy}{16} I_{[1,3]\times[1,3]}(x,y) = \left(\frac{x}{4} I_{[1,3]}(x)\right)\left(\frac{y}{4} I_{[1,3]}(y)\right).$$

To give some more detail, note that the marginal density of X is

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy = \int_1^3 \frac{xy}{16} I_{[1,3]}(x)dy = \frac{x}{4} I_{[1,3]}(x),$$

and by a similar calculation

$$f_Y(y) = \frac{y}{4} I_{[1,3]}(y),$$

and indeed $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, showing that they are independent.

(c) We compute

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)dx = \int_1^3 x\frac{x}{4}dx = \frac{13}{6}$$

and analogously

$$\mathbb{E}[Y] = \frac{13}{6}.$$

By independence,

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] = \left(\frac{13}{6}\right)^2.$$

If you have feedback regarding the exercise sheets, please send a mail to Jakob Heiss.