Ve401 Probabilistic Methods in Engineering

Spring 2020 Midterm Exam

Questions and Solutions

Exercise 1.1
Let (X, Y) be a continuous bivariate random variable with density $f_{XY}\colon S \to \mathbb{R}^2$ given by

$$f_{XY}(x,y) = \begin{cases} c \cdot (x^2+y^2) & \text{for } x^2+y^2 \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $c \in \mathbb{R}$ is a suitable constant.

i) Determine the constant c.

ii) Find $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.

iii) Find $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$.

iv) Find the correlation coefficient ϱ_{XY}.

v) Find the density of $U = X/Y$.

Helpful note: $\int_0^{2\pi} \cos^2(\theta)\, d\theta = \pi$.

Solution.

i) We must have $\int_{\mathbb{R}^2} f(x,y)\, dx\, dy = 1$, so

$$\begin{aligned}\int_{\mathbb{R}^2} f(x,y)\, dx\, dy &= \iint_{x^2+y^2\le 1} c\cdot(x^2+y^2)\, dx\, dy \\ &= \int_0^{2\pi}\int_0^1 c\cdot r^2 \cdot r\, dr\, d\theta \\ &= c \cdot 2\pi \cdot \frac{1}{4}\end{aligned}$$

implies that $c = 2/\pi$.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

ii) We calculate

$$\mathrm{E}[X] = \int_{\mathbb{R}^2} x f(x,y)\, dx\, dy = \frac{2}{\pi}\int_0^{2\pi}\int_0^1 r\cos(\theta)\cdot r^3\, dr\, d\theta = 0$$

by symmetry of the cosine function. The expectation of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. It is acceptable to say that the expectation is zero "by symmetry" without doing a calculation. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

iii) We calculate

$$\begin{aligned}\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 &= \mathrm{E}[X^2] \\ &= \int_{\mathbb{R}^2} x^2 f(x,y)\, dx\, dy \\ &= \frac{2}{\pi}\int_0^{2\pi}\int_0^1 r^2\cos^2(\theta)\cdot r^3\, dr\, d\theta \\ &= \frac{2}{\pi}\underbrace{\int_0^{2\pi}\cos^2(\theta)\, d\theta}_{=\pi}\int_0^1 r^5\, dr \\ &= \frac{1}{3}.\end{aligned}$$

The variance of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

iv) We calculate

$$\begin{aligned}
\mathrm{Cov}[X,Y] &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \mathrm{E}[XY] \\
&= \int_{\mathbb{R}^2} xyf(x,y)\,dx\,dy \\
&= \frac{2}{\pi}\int_0^{2\pi}\int_0^1 r^2\cos(\theta)\sin(\theta)\cdot r^3\,dr\,d\theta \\
&= 0
\end{aligned}$$

by orthogonality of the sine and cosine functions. It follows that the correlation coefficient is zero.

Give 0.5 Marks for finding the correlation coefficient from the covariance. Give 1.5 Marks for the covariance if the calculation and the answer are correct. It is acceptable to say that the expectation is zero "by symmetry" or "orthogonality" without doing a calculation. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

v) Formally, we have, for suitable $u \in \mathbb{R}$

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v)\cdot|v|\,dv = \frac{2}{\pi}\int_I (u^2v^2 + v^2)\cdot|v|\,dv$$

Here the integral is taken over all v such that $u^2v^2 + v^2 \le 1$. We note that this is possible for all $u \in \mathbb{R}$ and that

$$I = \left[-\frac{1}{\sqrt{1+u^2}}, \frac{1}{\sqrt{1+u^2}}\right].$$

It follows that, using the fact that the integrand is even

$$\begin{aligned}
f_U(u) &= \frac{2}{\pi}(1+u^2)\int_{-\frac{1}{\sqrt{1+u^2}}}^{\frac{1}{\sqrt{1+u^2}}} v^2\cdot|v|\,dv \\
&= \frac{4}{\pi}(1+u^2)\int_0^{\frac{1}{\sqrt{1+u^2}}} v^3\,dv \\
&= \frac{4}{\pi}(1+u^2)\frac{1}{4(\sqrt{1+u^2})^4} \\
&= \frac{1}{\pi}\frac{1}{1+u^2}
\end{aligned}$$

for $u \in \mathbb{R}$.

- *1 Mark for correctly identifying that this is the density for all $u \in \mathbb{R}$.*
- *1 Marks for correctly finding the interval I.*
- *2 Marks for the remainder of the calculation involving finding the density f_U.*
- *As always, no marks for those parts where reasoning or calculation is missing.*

Exercise 1.2

Let (X, Y) be a continuous bivariate random variable with density $f_{XY} : S \to \mathbb{R}^2$ given by

$$f_{XY}(x,y) = \begin{cases} c\cdot\sqrt{x^2+y^2} & \text{for } x^2+y^2 \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $c \in \mathbb{R}$ is a suitable constant.

i) Determine the constant c.

ii) Find $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.

iii) Find $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$.

iv) Find the correlation coefficient ϱ_{XY}.

v) Find the density of $U = X/Y$.

Helpful note: $\int_0^{2\pi} \cos^2(\theta)\, d\theta = \pi$.

Solution.

i) We must have $\int_{\mathbb{R}^2} f(x,y)\, dx\, dy = 1$, so

$$\begin{aligned}\int_{\mathbb{R}^2} f(x,y)\, dx\, dy &= \iint_{x^2+y^2\leq 1} c \cdot \sqrt{x^2+y^2}\, dx\, dy \\ &= \int_0^{2\pi}\int_0^1 c \cdot r \cdot r\, dr\, d\theta \\ &= c \cdot \frac{2\pi}{3}\end{aligned}$$

implies that $c = 3/(2\pi)$.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

ii) We calculate

$$\mathrm{E}[X] = \int_{\mathbb{R}^2} x f(x,y)\, dx\, dy = \frac{3}{2\pi}\int_0^{2\pi}\int_0^1 r\cos(\theta) \cdot r^2\, dr\, d\theta = 0$$

by symmetry of the cosine function. The expectation of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. It is acceptable to say that the expectation is zero "by symmetry" without doing a calculation. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

iii) We calculate

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \mathrm{E}[X^2] \\ &= \int_{\mathbb{R}^2} x^2 f(x,y)\, dx\, dy \\ &= \frac{3}{2\pi}\int_0^{2\pi}\int_0^1 r^2\cos^2(\theta) \cdot r^2\, dr\, d\theta \\ &= \frac{3}{2\pi}\underbrace{\int_0^{2\pi}\cos^2(\theta)\, d\theta}_{=\pi}\int_0^1 r^4\, dr \\ &= \frac{3}{10}.\end{aligned}$$

The variance of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

iv) We calculate

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \mathrm{E}[XY] \\ &= \int_{\mathbb{R}^2} xy f(x,y)\, dx\, dy \\ &= \frac{3}{2\pi}\int_0^{2\pi}\int_0^1 r^2\cos(\theta)\sin(\theta) \cdot r^2\, dr\, d\theta \\ &= 0\end{aligned}$$

by orthogonality of the sine and cosine functions. It follows that the correlation coefficient is zero.

Give 0.5 Marks for finding the correlation coefficient from the covariance. Give 1.5 Marks for the covariance if the calculation and the answer are correct. It is acceptable to say that the expectation is zero "by symmetry" or "orthogonality" without doing a calculation. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

v) Formally, we have, for suitable $u \in \mathbb{R}$

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v) \cdot |v| \, dv = \frac{2}{\pi} \int_I \sqrt{u^2v^2 + v^2} \cdot |v| \, dv$$

Here the integral is taken over all v such that $u^2v^2 + v^2 \leq 1$. We note that this is possible for all $u \in \mathbb{R}$ and that

$$I = \left[-\frac{1}{\sqrt{1+u^2}}, \frac{1}{\sqrt{1+u^2}} \right].$$

It follows that, using the fact that the integrand is even

$$\begin{aligned} f_U(u) &= \frac{3}{2\pi}\sqrt{1+u^2} \int_{-\frac{1}{\sqrt{1+u^2}}}^{\frac{1}{\sqrt{1+u^2}}} |v|^2 \, dv \\ &= \frac{6}{2\pi}\sqrt{1+u^2} \int_0^{\frac{1}{\sqrt{1+u^2}}} v^2 \, dv \\ &= \frac{3}{\pi}\sqrt{1+u^2} \frac{1}{3(\sqrt{1+u^2})^3} \\ &= \frac{1}{\pi} \frac{1}{1+u^2} \end{aligned}$$

for $u \in \mathbb{R}$.

- *1 Mark for correctly identifying that this is the density for all $u \in \mathbb{R}$.*
- *1 Marks for correctly finding the interval I.*
- *2 Marks for the remainder of the calculation involving finding the density f_U.*
- *As always, no marks for those parts where reasoning or calculation is missing.*

Exercise 1.3

Let (X, Y) be a continuous bivariate random variable with density $f_{XY} : S \to \mathbb{R}^2$ given by

$$f_{XY}(x, y) = \begin{cases} c \cdot (x + y) & \text{for } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $c \in \mathbb{R}$ is a suitable constant.

i) Determine the constant c.

ii) Find $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.

iii) Find $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$.

iv) Find the correlation coefficient ϱ_{XY}.

v) Find the density of $U = X/Y$.

Solution.

i) We must have $\int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1$, so

$$\begin{aligned} \int_{\mathbb{R}^2} f(x, y) \, dx \, dy &= \int_0^1 \int_0^1 c \cdot (x + y) \, dx \, dy \\ &= c \int_0^1 \int_0^1 x \, dx \, dy + c \int_0^1 \int_0^1 y \, dx \, dy \\ &= \frac{c}{2} + \frac{c}{2} = c. \end{aligned}$$

implies that $c = 1$.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

ii) We calculate

$$\mathrm{E}[X] = \int_0^1 \int_0^1 x \cdot (x+y)\,dx\,dy = \int_0^1 \Big(\frac{1}{3} + \frac{1}{2}y\Big)\,dy = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

The expectation of Y is the same, since X and Y enter into the density in the same way.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

iii) We calculate

$$\mathrm{E}[X^2] = \int_0^1 \int_0^1 x^2(x+y)\,dx\,dy = \int_0^1 \Big(\frac{1}{4} + \frac{1}{3}y\Big)\,dy = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

It follows that

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{11}{144}.$$

The variance of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

iv) We calculate

$$\mathrm{E}[XY] = \int_0^1 \int_0^1 xy(x+y)\,dx\,dy = \int_0^1 y\Big(\frac{1}{3} + \frac{1}{2}y\Big)\,dy = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

It follows that

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = -\frac{1}{144}.$$

It follows that the correlation coefficient is

$$\varrho_{XY} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = -\frac{1}{11}.$$

Give 0.5 Marks for finding the correlation coefficient from the covariance. Give 1.5 Marks for the covariance if the calculation and the answer are correct. Give no marks if there is no calculation or other reasoning.

v) Formally, we have, for suitable $u \in \mathbb{R}$

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v) \cdot |v|\,dv = \frac{2}{\pi} \int_{I_u} (uv + v) \cdot |v|\,dv$$

Here the integral is taken over all v such that $(uv, v) \in [0,1] \times [0,1]$. We first note that if $u < 0$, then $v < 0$ to ensure $uv \in [0,1]$. But then $v \notin [0,1]$ (the second component). Therefore, this density is only defined for $u \geq 0$.

Now if $0 \leq u \leq 1$, then $0 \leq v \leq 1$. If $u > 1$, then $0 \leq v \leq 1/u$ is required.

For $0 \leq u \leq 1$ we have

$$f_U(u) = \int_0^1 (uv + v) \cdot |v|\,dv = \frac{1}{3}(u+1)$$

while for $1 \leq u < \infty$ we have

$$f_U(u) = \int_0^{1/u} (uv + v) \cdot |v|\,dv = \frac{1}{3u^3}(u+1).$$

In summary,

$$f_u(u) = \begin{cases} 0 & u < 0, \\ (u+1)/3 & 0 \leq u < 1, \\ (u+1)/(3u^3) & 1 \leq u < \infty. \end{cases}$$

- *1 Mark for correctly identifying that the density vanishes for $u < 0$.*
- *1 Marks for correctly finding the two intervals for v based on u.*
- *2 Marks for the remainder of the calculation involving finding the density f_U.*
- *As always, no marks for those parts where reasoning or calculation is missing.*

Exercise 1.4
Let (X, Y) be a continuous bivariate random variable
with density $f_{XY} : S \to \mathbb{R}^2$ given by

$$f_{XY}(x, y) = \begin{cases} c \cdot (x - y) & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le x, \\ 0 & \text{otherwise,} \end{cases}$$

where $c \in \mathbb{R}$ is a suitable constant.

i) Determine the constant c.

ii) Find $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.

iii) Find $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$.

iv) Find the correlation coefficient ϱ_{XY}.

v) Find the density of $U = X/Y$.

Solution.

i) We must have $\int_{\mathbb{R}^2} f(x, y)\, dx\, dy = 1$, so

$$\int_{\mathbb{R}^2} f(x, y)\, dx\, dy = \int_0^1 \int_0^x c \cdot (x - y)\, dy\, dx = c \int_0^1 (x^2 - x^2/2)\, dx = \frac{c}{6}.$$

implies that $c = 6$.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

ii) We calculate

$$\mathrm{E}[X] = 6 \int_0^1 \int_0^x x(x - y)\, dy\, dx = 6 \int_0^1 x \cdot \frac{1}{2} x^2\, dx = \frac{3}{4},$$
$$\mathrm{E}[Y] = 6 \int_0^1 \int_0^x y(x - y)\, dy\, dx = 6 \int_0^1 \left(\frac{1}{2} x^3 - \frac{1}{3} x^3\right) dx = \frac{1}{4}.$$

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

iii) We calculate

$$\mathrm{E}[X^2] = 6 \int_0^1 \int_0^x x^2(x - y)\, dy\, dx = \frac{6}{2} \int_0^1 x^4\, dx = \frac{3}{5},$$
$$\mathrm{E}[Y^2] = 6 \int_0^1 \int_0^x y^2(x - y)\, dy\, dx = 6 \int_0^1 \frac{1}{12} x^4\, dx = \frac{1}{10},$$

It follows that

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80},$$
$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 = \frac{1}{10} - \frac{1}{16} = \frac{3}{80}.$$

The variance of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

iv) We calculate

$$\mathrm{E}[XY] = 6\int_0^1 \int_0^x xy(x-y)\,dy\,dx = 6\int_0^1 \frac{1}{6}x^4\,dx = \frac{1}{5}.$$

It follows that

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \frac{16}{80} - \frac{15}{80} = \frac{1}{80}.$$

It follows that the correlation coefficient is

$$\varrho_{XY} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{1}{3}.$$

Give 0.5 Marks for finding the correlation coefficient from the covariance. Give 1.5 Marks for the covariance if the calculation and the answer are correct. Give no marks if there is no calculation or other reasoning.

v) Formally, we have, for suitable $u \in \mathbb{R}$

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv,v)\cdot|v|\,dv = 6\int_{I_u}(uv-v)\cdot|v|\,dv$$

Here the integral is taken over all v such that $0 \le uv \le 1$ and $0 \le v \le uv$. It follows that $u \ge 1$ and $v \le 1/u$ (the density is zero otherwise). The density is then given by

$$f_U(u) = 6(u-1)\int_0^{1/u} v^2\,dv = 2\frac{u-1}{u^3}.$$

so that

$$f_u(u) = \begin{cases} 0 & u < 1, \\ 2\cdot\frac{u-1}{u^3} & u \ge 1. \end{cases}$$

- *1 Mark for correctly identifying that the density vanishes for $u < 0$ and $u > 1$.*
- *1 Marks for correctly finding the interval for v based on u.*
- *2 Marks for the remainder of the calculation involving finding the density f_U.*
- *As always, no marks for those parts where reasoning or calculation is missing.*

Exercise 1.5
Let (X,Y) be a continuous bivariate random variable with density $f_{XY}\colon S \to \mathbb{R}^2$ given by

$$f_{XY}(x,y) = \begin{cases} c\cdot(1+x^2+y^2)^{-1} & \text{for } x^2+y^2 \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $c \in \mathbb{R}$ is a suitable constant.

i) Determine the constant c.

ii) Find $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.

iii) Find $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$.

iv) Find the correlation coefficient ϱ_{XY}.

v) Find the density of $U = X/Y$.

Helpful note: $\int_0^1 r^3/(1+r^2)\,dr = \frac{1}{2}(1-\ln(2))$.

Solution.

i) We must have $\int_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$, so

$$\begin{aligned}\int_{\mathbb{R}^2} f(x,y)\,dx\,dy &= \iint_{x^2+y^2\le 1} c \cdot frac11 + x^2 + y^2\,dx\,dy \\ &= c\int_0^{2\pi}\int_0^1 \frac{1}{1+r^2}\cdot r\,dr\,d\theta \\ &= c\int_0^{2\pi} \frac{1}{2}\ln(1+r^2)|_0^1\,dr\,d\theta \\ &= c\cdot\pi\ln(2)\end{aligned}$$

implies that $c = 1/(\pi\ln(2))$.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

ii) We calculate

$$\mathrm{E}[X] = \int_{\mathbb{R}^2} xf(x,y)\,dx\,dy = \frac{1}{\pi\ln(2)}\int_0^{2\pi}\int_0^1 \frac{r\cos(\theta)}{1+r^2}\cdot\cdot r\,dr\,d\theta = 0$$

by symmetry of the cosine function. The expectation of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. It is acceptable to say that the expectation is zero "by symmetry" without doing a calculation. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

iii) We calculate

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \mathrm{E}[X^2] \\ &= \int_{\mathbb{R}^2} x^2 f(x,y)\,dx\,dy \\ &= \frac{1}{\pi\ln(2)}\int_0^{2\pi}\int_0^1 \frac{r^2\cos^2(\theta)}{1+r^2}\cdot r\,dr\,d\theta \\ &= \frac{1}{\pi\ln(2)}\underbrace{\int_0^{2\pi}\cos^2(\theta)\,d\theta}_{=\pi}\int_0^1 \frac{r^3}{1+r^2}\,dr \\ &= \frac{1}{2}\Big(\frac{1}{\ln(2)} - 1\Big).\end{aligned}$$

The variance of Y is the same, since X and Y enter into the density the same way.

Give 2 Marks if the calculation and the answer are correct. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation.

iv) We calculate

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \mathrm{E}[XY] \\ &= \int_{\mathbb{R}^2} xyf(x,y)\,dx\,dy \\ &= \frac{1}{\pi\ln(2)}\int_0^{2\pi}\int_0^1 \frac{r\cos(\theta)\sin(\theta)}{1+r^2}\cdot r\,dr\,d\theta \\ &= 0\end{aligned}$$

by orthogonality of the sine and cosine functions. It follows that the correlation coefficient is zero.

Give 0.5 Marks for finding the correlation coefficient from the covariance. Give 1.5 Marks for the covariance if the calculation and the answer are correct. It is acceptable to say that the expectation is zero "by symmetry" or "orthogonality" without doing a calculation. Give 1 Mark if the calculation is present but the answer is wrong. Give no marks if there is no calculation or other reasoning.

v) Formally, we have, for suitable $u \in \mathbb{R}$

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v) \cdot |v| \, dv = \frac{1}{\pi \ln(2)} \int_I \frac{1}{1 + u^2v^2 + v^2}) \cdot |v| \, dv$$

Here the integral is taken over all v such that $u^2v^2 + v^2 \leq 1$. We note that this is possible for all $u \in \mathbb{R}$ and that

$$I = \left[-\frac{1}{\sqrt{1+u^2}}, \frac{1}{\sqrt{1+u^2}} \right].$$

It follows that, using the fact that the integrand is even

$$\begin{aligned} f_U(u) &= \frac{1}{\pi \ln(2)} \int_{-\frac{1}{\sqrt{1+u^2}}}^{\frac{1}{\sqrt{1+u^2}}} \frac{1}{1 + (1+u^2)v^2} \cdot |v| \, dv \\ &= \frac{2}{\pi \ln(2)} \int_0^{\frac{1}{\sqrt{1+u^2}}} \frac{v}{1 + (1+u^2)v^2} \, dv \\ &= \frac{2}{\pi \ln(2)} \frac{1}{1+u^2} \frac{1}{2} \ln\big(1 + (1+u^2)v^2)\big)\Big|_0^1 \\ &= \frac{1}{\pi} \frac{1}{1+u^2} \end{aligned}$$

for $u \in \mathbb{R}$.

- *1 Mark for correctly identifying that this is the density for all $u \in \mathbb{R}$.*
- *1 Marks for correctly finding the interval I.*
- *2 Marks for the remainder of the calculation involving finding the density f_U.*
- *As always, no marks for those parts where reasoning or calculation is missing.*

Exercise 2.1

Let X be a discrete random variable following a Bernoulli distribution with parameter $p = 1/2$ and let $X_1, \ldots, X_{10}$ be a random sample of size 10. Calculate the probability that the sample mean is greater than 3/4, i.e., find

$$P[\overline{X} > 3/4].$$

Solution. We note that $X_1 + \cdots + X_{10}$ follows a binomial distribution with $n = 10$ and $p = 1/2$.

This observation is worth 2 Marks.

Then

$$\begin{aligned} P[\overline{X} > 3/4] &= P[X_1 + \cdots + X_{10} > 3/4 \cdot 10] \\ &= P[X_1 + \cdots + X_{10} > 7.5] \\ &= 1 - P[X_1 + \cdots + X_{10} < 7.5] \\ &= 1 - P[X_1 + \cdots + X_{10} \leq 7] \\ &= 0.0547 \end{aligned}$$

Give a total of 2 Marks for the calculation, subtract 1 Mark if the final result is incorrect. Give 0 marks if there is no calculation.

Exercise 2.2

Let X be a discrete random variable following a geometric distribution with parameter $p = 1/2$ and let $X_1, \ldots, X_{10}$ be a random sample of size 10. Calculate the probability that the sample mean is no more than than 1.5, i.e., find

$$P[\overline{X} \leq 1.5].$$

Solution. We note that $X_1 + \cdots + X_{10}$ follows a Pascal distribution with $r = 10$ and $p = 1/2$.

This observation is worth 2 Marks.

Then

$$\begin{aligned}
P[\overline{X} < 1.5] &= P[X_1 + \cdots + X_{10} \leq 10 \cdot 1.5] \\
&= P[X_1 + \cdots + X_{10} \leq 15] \\
&= \sum_{x=10}^{15} \binom{x-1}{9} \frac{1}{2^x} \\
&= \frac{1}{1024} \left(\binom{9}{9} + \frac{1}{2}\binom{10}{9} + \frac{1}{4}\binom{11}{9} + \frac{1}{8}\binom{12}{9} + \frac{1}{16}\binom{13}{9} + \frac{1}{32}\binom{14}{9} + \frac{1}{64}\binom{15}{9} \right) \\
&= \frac{1}{1024} \left(1 + \frac{10}{2} + \frac{55}{4} + \frac{220}{8} + \frac{715}{16} + \frac{2002}{32} + \frac{5005}{64} \right) \\
&= \frac{309}{2048} = 0.15
\end{aligned}$$

Give a total of 2 Marks for the calculation, subtract 1 Mark if the final result is incorrect. Give 0 marks if there is no calculation.

Exercise 2.3

Let X be a discrete random variable following a Poisson distribution with parameter $k = 2$ and let $X_1, \ldots, X_{10}$ be a random sample of size 10. Calculate the probability that the sample mean is no more than than 1.5, i.e., find

$$P[\overline{X} \leq 1.5].$$

Solution. We note that $X_1 + \cdots + X_{10}$ follows a Poisson distribution with $k = 20$.

This observation is worth 2 Marks.

Then

$$\begin{aligned}
P[\overline{X} < 1.5] &= P[X_1 + \cdots + X_{10} \leq 10 \cdot 1.5] \\
&= P[X_1 + \cdots + X_{10} \leq 15] \\
&= 0.156
\end{aligned}$$

Give a total of 2 Marks for the calculation, subtract 1 Mark if the final result is incorrect. Give 0 marks if there is no calculation.

Exercise 3.1

Let X be a continuous random variable with density

$$f_\theta(x) = \begin{cases} (\theta+1)x^\theta & \text{for } 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter. Find the maximum likelihood estimator for θ.

Solution. Let $X_1, \ldots, X_n$ be a random sample of size n. The Likelihood function is then

$$L(\theta) = \begin{cases} (\theta+1)^n (x_1 x_2 \ldots x_n)^\theta & 0 < x_1, x_2, \ldots, x_n < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The logarithm of L is

$$\ln(L(\theta)) = n \ln(\theta + 1) + \theta \ln(x_1 x_2 \ldots x_n).$$

Differentiating with respect to θ, we have

$$\frac{dL}{d\theta} = \frac{n}{\theta+1} + \ln(x_1 x_2 \ldots x_n) = 0,$$

which gives the estimator

$$\widehat{\theta =} - \frac{n}{\ln(x_1 x_2 \ldots x_n)} - 1.$$

i) *2 Marks for writing down the likelihood function correctly.*

ii) *2 Marks for the calculation of the maximum.*

iii) *1 Mark for the correct identification of the estimator.*

Exercise 3.2
Let X be a continuous random variable with density

$$f_\theta(x) = \begin{cases} \dfrac{x}{\theta^2} e^{-x/\theta} & \text{for } x > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter. Find the maximum likelihood estimator for θ.

Solution. Let $X_1, \dots, X_n$ be a random sample of size n. The likelihood function is then

$$L(\theta) = \begin{cases} \frac{x_1 x_2 \dots x_n}{\theta^{2n}} e^{-n\overline{x}/\theta} & 0 < x_1, x_2, \dots, x_n, \\ 0 & \text{otherwise.} \end{cases}$$

The logarithm of L (in the domain $x_1, x_2, \dots, x_n > 0$) is

$$\ln(L(\theta)) = \ln(x_1 x_2 \dots x_n) - 2n \ln(\theta) - n\overline{x}/\theta$$

Differentiating with respect to θ, we have

$$\frac{dL}{d\theta} = \frac{-2n}{\theta} + n\overline{x}/\theta^2 = 0,$$

which gives

$$\hat{\theta} = \frac{1}{2}\overline{X}.$$

i) *2 Marks for writing down the likelihood function correctly.*

ii) *2 Marks for the calculation of the maximum.*

iii) *1 Mark for the correct identification of the estimator.*

Exercise 3.3
Let X be a continuous random variable with density

$$f_\theta(x) = \begin{cases} \dfrac{x}{\theta} e^{-x^2/(2\theta)} & \text{for } x > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter. Find the maximum likelihood estimator for θ.

Solution. Let $X_1, \dots, X_n$ be a random sample of size n. The Likelihood function is then

$$L(\theta) = \begin{cases} \frac{x_1 x_2 \dots x_n}{\theta^n} e^{-\sum_{i=1}^n x_i^2/(2\theta)} & 0 < x_1, x_2, \dots, x_n, \\ 0 & \text{otherwise.} \end{cases}$$

The logarithm of L (in the domain $x_1, x_2, \dots, x_n > 0$) is

$$\ln(L(\theta)) = \ln(x_1 x_2 \dots x_n) - n \ln(\theta) - \frac{1}{2\theta}\sum_{i=1}^n x_i^2$$

Differentiating with respect to θ, we have

$$\frac{dL}{d\theta} = \frac{-n}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 = 0,$$

which gives

$$\widehat{\theta} = \frac{1}{2n}\sum_{i=1}^n x_i^2.$$

i) *2 Marks for writing down the likelihood function correctly.*

ii) *2 Marks for the calculation of the maximum.*

iii) *1 Mark for the correct identification of the estimator.*

Exercise 3.4
Let X be a continuous random variable with density

$$f_\theta(x) = \begin{cases} \theta x & \text{for } 0 < x < 2/\theta, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter. Find the maximum likelihood estimator for θ.

Solution. This exercise suffers from an error. The density should be as follows:

$$f_\theta(x) = \begin{cases} \dfrac{\theta^2}{2} x & \text{for } 0 < x < 2/\theta, \\ 0 & \text{otherwise,} \end{cases}$$

Give full marks to students who noticed this and did something else, e.g., calculate the value of theta that allows f_θ as given to actually be a density function.

Given the incorrect function f_θ as stated in the problem, the likelihood function would be

$$L(\theta) = \begin{cases} x_1 x_2 \dots x_n \cdot \theta^n & 0 < x_1, x_2, \dots, x_n < 2/\theta, \\ 0 & \text{otherwise.} \end{cases}$$

Then this function is maximized if θ is chosen as large as possible. However, the likelihood function will vanish unless $\max x_i \leq 2/\theta$, i.e., $\theta \leq 2/\max x_i$. Hence, one should choose

$$\widehat{\theta} = \frac{2}{\max\{x_1, \dots, x_n\}}.$$

(For solving this exercise, the scaling factor θ vs. $\theta^2/2$ is actually irrelevant.)

i) *2 Marks for writing down the likelihood function correctly.*

ii) *2 Marks for the calculation of the maximum.*

iii) *1 Mark for the correct identification of the estimator.*

Exercise 4.1
A certain widget uses a random number generator that generates random digits 0-9. However, it is known that in 20% of the widgets, the random number generator is defective and does not generate any 0s. However, the remaining digits 1-9 are generated with equal probability in the defective widgets.

A test run of the RNG in a given widget comprising twenty random numbers failed to yield the digit 0. What is the probability that this widget is one of the defective RNGs?

Solution. We know that $P[\text{defective}] = 0.2$. Furthermore,

$$P[\text{result} \mid \text{not defective}] = 0.9^{20} = 0.1216$$

Then

$$\begin{aligned} P[\text{defective} \mid \text{result}] &= \frac{P[\text{result} \mid \text{defective}] \cdot P[\text{defective}]]}{P[\text{result} \mid \text{defective}] \cdot P[\text{defective}]] + P[\text{result} \mid \text{not defective}] \cdot P[\text{not defective}]]} \\ &= \frac{1 \cdot 0.2}{1 \cdot 0.2 + 0.1216 \cdot 0.8} \\ &= 0.6728 \end{aligned}$$

i) *2 Marks for writing down the correct probabilities based on the exercise description.*

ii) *2 Marks for the calculation using conditional probabiliy/Bayes's theorem.*

iii) *1 Mark for the correct correct result, if it is supported by calculation above.*

Exercise 4.2

A company produces toy plastic coins for use in board games. They will be tossed and should return either "heads" or "tails" withe qual probability $p_0 = 0.5$. Most of the coins are fine, but due to a molding process fault, 5% of the coins are defective and have a $p = 0.7$ chance of returning "heads".

A coin is tested by tossing it 100 times and recording the number of heads. It will be deemed defective and discarded if "heads" occurs at least 70 times.

Given that a coin is discarded, what is the probability that it was defective?

Solution. We know that $P[\text{defective}] = 0.05$. Furthermore,

$$P[\text{discarded} \mid \text{not defective}] = \frac{1}{2^n} \sum_{x=70}^{100} \binom{100}{x} = 0.000039$$

and

$$P[\text{discarded} \mid \text{defective}] = \sum_{x=70}^{100} \binom{100}{i} 0.7^x 0.3^{100-x} = 0.549$$

Then

$$\begin{aligned} P[\text{defective} \mid \text{discarded}] &= \frac{P[\text{discarded} \mid \text{defective}] \cdot P[\text{defective}]]}{P[\text{discarded} \mid \text{defective}] \cdot P[\text{defective}]] + P[\text{discarded} \mid \text{not defective}] \cdot P[\text{not defective}]]} \\ &= \frac{0.549 \cdot 0.05}{0.549 \cdot 0.05 + 0.000039 \cdot 0.95} \\ &= 0.999 \end{aligned}$$

i) *2 Marks for writing down the correct probabilities based on the exercise description.*

ii) *2 Marks for the calculation using conditional probabiliy/Bayes's theorem.*

iii) *1 Mark for the correct correct result, if it is supported by calculation above.*

Exercise 4.3

A certain widget has a mean time between failures of 24 hours, i.e., failures occur at a constant rate of one failure every 24 hours.

One evening, the widget was observed to be working at 10 pm and then left unobserved for the night. The next morning at 6 am, it was observed to have failed earlier. What is the probability that it was still working at 5 am that morning?

Solution. The failures of the widget follow a Poisson distribution with a rate $\lambda = 1/24$. The time to failure is therefore exponentially distributed with parameter $\beta = \lambda = 1/24$.
Generally, the failure density is $f_\beta(t) = \beta e^{-\beta t}$ so that

$$P[\text{widget fails between time } T_1 \text{ and } T_2] = \int_{T_1}^{T_2} \beta e^{-\beta t}\, dt$$

Therefore,

$$\begin{aligned} &P[\text{widget fails between time } T_1 \text{ and } T_2 \mid \text{widget has failed not after time } T_2] \\ &= \frac{P[\text{widget fails between time } T_1 \text{ and } T_2]}{P[\text{widget has failed not after time } T_2]} \\ &= \frac{\int_{T_1}^{T_2} \beta e^{-\beta t}\, dt}{\int_0^{T_2} \beta e^{-\beta t}\, dt} \\ &= \frac{-e^{-\beta t}\big|_{T_1}^{T_2}}{-e^{-\beta t}\big|_0^{T_2}} \\ &= \frac{e^{-\beta T_1} - e^{-\beta T_2}}{1 - e^{-\beta T_2}} \end{aligned}$$

Inserting our values, the probability is

$$\frac{e^{-7/24} - e^{-1/3}}{1 - e^{-1/3}} = 0.108.$$

i) *2 Marks for writing down the correct probability that the widget fails between time T_1 and T_2.*

ii) *2 Marks for the calculation using conditional probabiliy/Bayes's theorem.*

iii) *1 Mark for the correct correct result, if it is supported by calculation above.*

Exercise 5.1

Consider the following set of 34 data:

2	2	3	7	9	14
1	4	6	4	15	3
10	4	9	4	7	3
3	12	7	5	4	7
5	3	8	6	12	7
8	5	4	6		

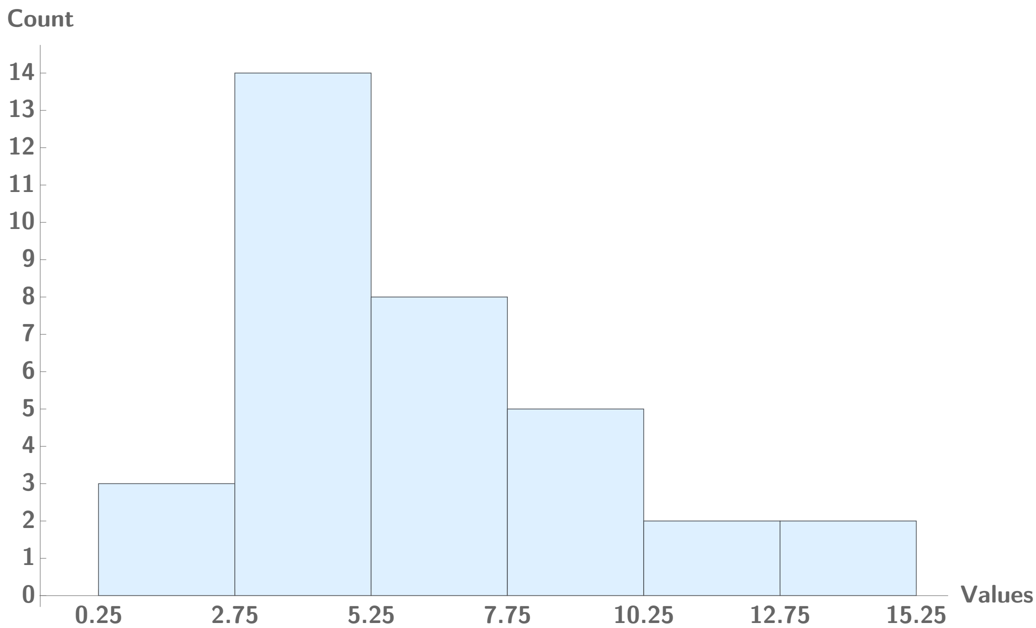
Find the quartiles and the interquartile range for the data. Create a histogram using the Freedman-Diaconis bin widths. Does the data appear to come from a normal distribution? Give your reasoning!

Important: you must use pencil-and-paper to create the plot. Computer graphics are not acceptable!

Solution. The quartiles are 4, 5.5, and 8. The inter-quartile range is $8 - 4 = 4$. Giuven the sample size $n = 34$, the Freedman-Diaconis rule gives a bin width of

$$\frac{2\,\mathrm{IQR}}{\sqrt[3]{n}} = 2.47 \approx 2.5$$

Given the data, it is reasonable to start at the bins at zero. However, then some data would fall onto the boundaries of the bins, so we start at 1/4 instead. Hence, we obtain the following histogram:



The histogram has a single mode but is strongly skewed to one side. Even given the small sample size, the strong skew makes it unlikely that the data comes from a normal distribution.

i) *1 Mark for finding the correct quartiles (a calculation is not necessary in this case).*

ii) *1.5 Marks for drawing the histogram correctly, including 1/2 Mark for finding the correct bin width (either 2.5 or 2.47) and 1/2 Mark for starting the histogram at some point that prevents data from falling on the boundary. (If 2.47 is used for the bin width, then the histogram can start at 0, for instance.) 1/2 Mark is for the general shape and correctness of the histogram.*

iii) *1.5 Marks for discussing that the data does not come from a normal distribution, including 1/2 Mark for mentioning the relatively small sample size, 1/2 Mark for mentioning the skew and 1/2 Mark for saying it is not likely that the data comes from a normal distribution (or something along these lines).*

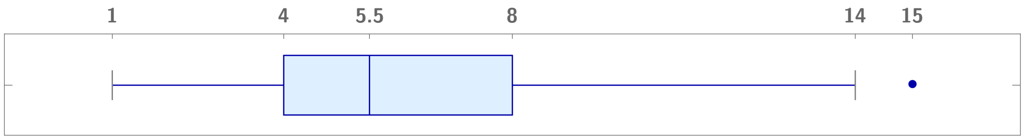
Exercise 5.2
Consider the following set of 34 data:

2	2	3	7	9	14
1	4	6	4	15	3
10	4	9	4	7	3
3	12	7	5	4	7
5	3	8	6	12	7
8	5	4	6		

Find the quartiles and the interquartile range for the data. Create a box-and-whisker diagram. Does the data appear to come from a normal distribution? Give your reasoning!

Important: you must use pencil-and-paper to create the plot. Computer graphics are not acceptable!

Solution. The quartiles are 4, 5.5, and 8. The inter-quartile range is $8 - 4 = 4$. The boxplot is therefore as follows:



The data has an outlier, despite the small sample size. Furthermore, the boxplot is skewed to one side. This makes it unlikely that the data comes from a normal distribution.

i) *1 Mark for finding the correct quartiles (a calculation is not necessary in this case).*

ii) *1.5 Marks for drawing the boxplot correctly, including 1/2 Mark each for the outlier and the whisker boundaries (both need to be correct to get the 1/2 Mark). The remaining 1/2 Mark is allocated for the general shape of the plot.*

iii) *1.5 Marks for discussing that the data does not come from a normal distribution, including 1/2 Mark for mentioning the presence of the outlier, 1/2 Mark for mentioning the skew and 1/2 Mark for saying it is not likely that the data comes from a normal distribution (or something along these lines).*

Exercise 5.3
Consider the following set of 34 data, taken from a normal distribution:

5.8	6.0	7.8	7.4	6.2	6.1
8.6	9.7	4.4	4.4	7.6	9.0
9.2	6.4	7.0	8.7	6.0	6.1
6.8	10.	5.3	8.1	8.9	5.3
6.7	5.6	3.7	8.7	6.6	10.1
6.9	6.6	6.4	5.1		

Find a 95% confidence interval for the mean.

Solution. Although not explicitly mentioned, a two-sided interval is sought. However, if a student decides to calculate a one-sided interval, that is also OK.

We first calculate

$$\overline{x} = 6.98, \qquad s^2 = 2.778, \qquad s = 1.67.$$

Given the sample $n = 34$, we need the value of $t_{0.025,33} = 2.0345$ and obtain the confidence interval

$$\mu = \overline{x} \pm t_{0.025,33} \frac{s}{\sqrt{n}} = 6.98 \pm 0.58$$

i) *1 Mark each for correctly calculating the sample mean and the sample variance.*

ii) *1 Mark finding the correct value for $t_{0.025,33}$.*

iii) *1 Marks for writing down the correct confidence interval.*

Exercise 5.4
Consider the following set of 34 data, taken from a normal distribution:

5.8	6.0	7.8	7.4	6.2	6.1
8.6	9.7	4.4	4.4	7.6	9.0
9.2	6.4	7.0	8.7	6.0	6.1
6.8	10.	5.3	8.1	8.9	5.3
6.7	5.6	3.7	8.7	6.6	10.1
6.9	6.6	6.4	5.1		

Find a 90% confidence interval for the variance.

Solution. Although not explicitly mentioned, a two-sided interval is sought. However, if a student decides to calculate a one-sided interval, that is also OK.

We first calculate $s^2 = 2.778$. Given the sample $n = 34$, we need the values of $\chi^2_{1-0.05,33} = 20.9$ and $\chi^2_{0.05,33} = 47.4$. We then obtain the confidence interval

$$\left[\frac{(n-1)S^2}{\chi^2_{0.05,33}}, \frac{(n-1)S^2}{\chi^2_{0.95,33}}\right] = [1.93, 4.39]$$

i) *1 Mark for correctly calculating the sample variance.*

ii) *1 Mark each for finding the correct values for $\chi^2_{1-0.025,33}$ and $\chi^2_{0.025,33}$.*

iii) *1 Marks for writing down the correct confidence interval.*