Ve401 Probabilistic Methods in Engineering

Sample Midterm Exam Exercises

The following exercises have been compiled from past first midterm exams of Ve401. A first midterm will usually consist of 25-30 Points worth of such exercises to be completed in 100 minutes. In the actual exam, necessary tables of values of distributions will be provided. A PDF file with all necessary tables has been made available on Canvas.

Multiple Choice

Exercise 1.
In the following exercises, mark the boxes corresponding to true statements with a cross (☒). In each case, exactly one of the provided statements is true.

i) Let X and Y be independent random variables. Then

- ☒ $\operatorname{Var}[X - Y] = \operatorname{Var}[X] + \operatorname{Var}[Y]$
- ☐ $\operatorname{Var}[X - Y] = \operatorname{Var}[X] - \operatorname{Var}[Y]$
- ☐ $\operatorname{Var}[X - Y]^2 = \operatorname{Var}[X]^2 + \operatorname{Var}[Y]^2$
- ☐ $\operatorname{Var}[X - Y]^2 = \operatorname{Var}[X]^2 - \operatorname{Var}[Y]^2$

ii) Let X and Y be two random variables. Which of the following properties guarantees the independence of X and Y?

- ☐ $\varrho_{XY} = 0$.
- ☐ $\operatorname{E}[XY] = \operatorname{E}[X]\operatorname{E}[Y]$.
- ☒ $f_{XY}(x, y) = f_X(x)f_Y(y)$.
- ☐ $P[Y = y \mid X = x] = P[Y = y]$ and $P[X = x \mid Y = y] = P[X = x]$.

iii) An unfair coin has probability $p \neq 1/2$ for returning "heads" when tossed. It is tossed twice. Compared with a fair coin ($p = 1/2$), the probability of obtaining "two heads or two tails"

- ☐ is the same.
- ☒ is higher.
- ☐ is lower.
- ☐ can be higher or lower, depending on p.

iv) Which of the following is a property of the exponential distribution?

- ☐ The sample variance and sample mean are independent.
- ☐ The sum of two exponential random variables follows an exponential distribution.
- ☒ The mode is zero.
- ☐ The mean and the median have the same value.

v) Suppose that X follows a normal distribution. Which of the following random variables does **not** follow a normal distribution?

- ☐ $Y = X + 1$.
- ☐ $Y = 2X$.
- ☒ $Y = X^2$.
- ☐ $Y = -X$.

(5 Marks)

Elementary Probability

Exercise 2.

A company produces widgets in three factories, A, B and C. Factory A produces 20% of the widgets, factory B produces 45% of the widgets and factory C produces the remaining 35%. Of all widgets produced, 5% fail tolerance. Of those that fail tolerance, 25% were produced in factory A, 35% were produced in factory B and 40% were produced in factory C. In each factory, what percentage of the widgets produced fails tolerance?
(3 Marks)

Solution 2.

We have

$$\begin{aligned} P[A] &= 0.2, & P[B] &= 0.45, & P[C] &= 0.35, & P[\text{fail}] &= 0.05 \\ P[A \mid \text{fail}] &= 0.25, & P[B \mid \text{fail}] &= 0.35, & P[C \mid \text{fail}] &= 0.40. \end{aligned}$$

(1 Mark) Hence,

$$\begin{aligned} P[\text{fail} \mid A] &= \frac{P[A \mid \text{fail}] \cdot P[\text{fail}]}{P[A]} = 0.0625 \\ P[\text{fail} \mid B] &= \frac{P[B \mid \text{fail}] \cdot P[\text{fail}]}{P[B]} = 0.0039 \\ P[\text{fail} \mid C] &= \frac{P[C \mid \text{fail}] \cdot P[\text{fail}]}{P[C]} = 0.0057 \end{aligned}$$

(2 Marks)

Exercise 3.

A fair coin is tossed repeatedly.

i) What is the probability that the 5^{th} head occurs on the 10^{th} toss?

ii) Suppose that we know that the 10^{th} head occurs on the 25^{th} toss. Find the probability density function of the toss number of the 5^{th} head.

(4 Marks)

Solution 3.

i) The probability is given by the Pascal distribution with parameters $p = 1/2$ and $r = 5$. The probability is

$$P[X = 10] = \binom{10-1}{5-1} \cdot \frac{1}{2^{10}} = 0.123.$$

(1 Mark)

ii) The tenth head occurs on the 25^{th} toss, so nine heads occur in the first 24 tosses. There are $\binom{24}{9}$ ways of selecting the toss numbers for these nine heads. If the 5^{th} head occurs on the x^{th} toss, then 4 heads must have occurred in the first $x - 1$ tosses and 4 heads must occur in the following $24 - x$ tosses. Here $5 \le x \le 20$. It follows that $f_X : \Omega \to \mathbb{R}$ given by

$$f_X(x) = P[X = x] = P[5^{\text{th}} \text{ head on } x^{\text{th}} \text{ toss} \mid 10^{\text{th}} \text{ head on } 25^{\text{th}} \text{ toss}] = \frac{\binom{x-1}{4}\binom{24-x}{4}}{\binom{24}{9}}$$

is the density of the random variable $X : S \to \Omega = \{5, 6, \dots, 20\}$, where X is the toss number of the 5^{th} head and S is the sample space for the experiment.
(3 Marks)

Exercise 4.

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?
(5 Marks)

Solution 4.
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 5.
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?
(5 Marks)

Solution 5.
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,

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

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.*

Approximating Distributions

Exercise 6.
A coin is tossed 28800 times. Assuming that the coin is fair (i.e., heads and tails each have a 50% chance of coming up), use the normal approximation to the binomial distribution to give the probability that between 14380 and 14399 heads come up.
(4 Marks)

Solution 6.
We want to approximate

$$\mathcal{P} = \sum_{k=14380}^{14399} \binom{28800}{k} 0.5^k 0.5^{28800-k}.$$

We use the normal approximation, so that

$$\mu = np = 28800 \cdot 0.5 = 14400$$

and

$$\sigma^2 = np(1-p) = 0.5 \cdot 28800 \cdot (1-0.5) = 7200.$$

(**1 Mark**) Note that $\sigma = 60\sqrt{2}$. Taking into account the half-unit correction, (**1 Mark**) we have ($Z = (X-14400)/(60\sqrt{2})$)

$$\begin{aligned}\mathcal{P} &\approx P[14379.5 \le X \le 14399.5] \\ &= P\left[\frac{14379.5-14400}{60\sqrt{2}} \le Z \le \frac{14399.5-14400}{60\sqrt{2}}\right] \\ &= P[-0.24 \le Z \le -0.0059] \\ &= P[0 \le Z \le 0.24] - P[0 \le Z \le 0.0059] \\ &= 0.0948 - 0.0000 = 0.0948 = 9.48\%.\end{aligned}$$

(**2 Marks**)

Exercise 7.
Suppose that a hotel has 200 equally-sized rooms. From experience, it is known that 10% of travellers who have reserved a room do not turn up. Use the normal approximation to the binomial distribution to answer the following questions.

i) Suppose that the hotel has received reservations for 215 rooms. What is the probability that there will enough rooms for all travelers who turn up?

ii) Suppose that the hotel has received reservations for 215 rooms. What is the probability that more than 190 rooms (but of course not more than 200) will be occupied?

iii) At what point should the hotel stop accepting reservations if management wants to be at least 99% certain to have enough rooms for all travelers who show up?

(2+2+2 Marks)

Solution 7.
We use the normal approximation to the binomial distribution with parameter $p = 0.1$ and n depending on the situation.

i) We want to find $P[T \le 200]$, where T denotes the number of travelers who actually turn up. There are $n = 215$ travelers who each have a probability of $p = 0.9$ of turning up. Thus, T follows a binomial distribution with these parameters. Since $nq = 21.5 > 5$ we can approximate this distribution using a normal random variable X with mean $\mu = np = 193.5$ and variance $\sigma^2 = npq = 19.35$. We denote by Z the standard normal random variable. Taking into account the half-unit correction,

$$P[T \le 200] \approx P[X \le 200.5] = P\left[Z \le \frac{200.5-193.5}{4.399}\right] = P[Z \le 1.59] = 0.94$$

so there is a 94% chance that there will be enough rooms.

ii) As above, we calculate $P[T \geq 191]$. Taking into account the half-unit correction,

$$P[T \geq 191] \approx P[X \geq 190.5] = P\left[Z \geq \frac{190.5 - 193.5}{4.399}\right] = P\,[Z \geq -0.68] = 0.752$$

so there is a 75% chance that more than 190 rooms will be occupied.

iii) We need to solve

$$0.99 \leq P[T \leq 200] \approx P[X \leq 200 + 0.5]$$

where X is a normal random variable with mean $0.9R$ and standard deviation $0.3\sqrt{R}$, R being the number of reservations accepted. Hence we need to calculate

$$0.99 \leq P\left[Z \leq \frac{200.5 - 0.9R}{0.3\sqrt{R}}\right]$$

From the table we see that this implies

$$2.33 \leq \frac{200.5 - 0.9R}{0.3\sqrt{R}}.$$

There is only one solution to this equation, sind the right hand side is monotonically decreasing in R. To find the point of equality, we solve $0.2097R \leq (200.5 - 0.9R)^2 = 0.81R^2 + 360.9R + (200.5)^2$. We find the roots of

$$0 = R^2 - 446.16R + 49630,$$

yielding $R = 223.08 \pm 11.6$. The smaller of the roots is 211.48, so the hotel should not accept more than 211 reservations.

Exercise 8.
A factory produces drill bits. It is known that 2% of the drill bits are defective. The drill bits are shipped in packages of 100. Use the Poisson approximation to the binomial distribution to answer the following questions.

i) What is the probability that a package contains no defective drill bits?

ii) What is the probability that there are no more than three defective drill bits in a package?

iii) How many drill bits must a package contain so that with probability greater than 0.9 there are at least 100 non-defective drill bits in the package?

(6 Marks)

Solution 8.
We use the Poisson approximation with $\lambda = n \cdot p = 100 \cdot 0.02 = 2$. **(2 Marks)**

i) $P[X = 0] = 0.135$ **(1 Mark)**

ii) $P[X \leq 3] = 0.857$ **(1 Mark)**

iii) Since $P[X \leq 4] = 0.947$ and from the above $P[X \leq 3] = 0.857$ we susect that 104 drill bits might be sufficient. In order to verify this, we actually need to check $P[X \leq 4 \mid \lambda = 104 \cdot 0.02 = 2.08]$. **(1 Mark)** We interpolate between $P[X \leq 4 \mid \lambda = 2.0] = 0.947$ and $P[X \leq 4 \mid \lambda = 2.5] = 0.891$ to obtain

$$P[X \leq 4 \mid \lambda = 2.08] \approx 0.891 + (0.947 - 0.891)\frac{2.5 - 2.08}{2.5 - 2} = 0.938 > 0.90.$$

(Using the value $P[X \leq 4 \mid \lambda = 2.0] = 0.947$ as an approximation is also OK.) **(1 Mark)**

Individual Random Variables

Exercise 9.
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

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

(4 Marks)

Solution 9.
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

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

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

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

(4 Marks)

Solution 10.
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

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

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 11.
The target value for the thickness of a machined cylinder is 8 cm. The upper specification limit is 8.2 cm and the lower specification limit is 7.9 cm. A machine produces cylinders that have a mean thickness of 8.1 cm and a standard deviation of 0.1 cm. Assume that the thickness of the cylinders produced by this machine follows a normal distribution.

i) What is the probability that the thickness of a randomly selected cylinder is within specification?

ii) What is the probability that the thickness of a randomly selected cylinder exceeds the target value?

iii) Find a bound $L > 0$ such that the thickness of 90% of all cylinders lies within $(8.1 \pm L)$ cm.

(2+2+2 Marks)

Solution 11.

i) Denote the random variable "thickness" by T. Then the variable $Z = \frac{T-8.1}{0.1}$ is standard normal and

$$\begin{aligned} P[7.9 \le T \le 8.2] &= P\left[\frac{7.9-8.1}{0.1} \le \frac{T-8.1}{0.1} \le \frac{8.2-8.1}{0.1}\right] \\ &= P\left[\frac{7.9-8.1}{0.1} \le Z \le \frac{8.2-8.1}{0.1}\right] \\ &= P\left[-2 \le Z \le 1\right] \\ &= P[0 \le Z \le 1] + P[0 \le Z \le 2] \\ &= 0.3413 + 0.4772 = 0.8185 \approx 82\%. \end{aligned}$$

(2 Marks)

ii) We have

$$\begin{aligned} P[T \ge 8] &= P\left[\frac{T-8.1}{0.1} \ge \frac{8-8.1}{0.1}\right] \\ &= P\left[Z \ge -1\right] \\ &= P[0 \le Z \le 1] + P[0 \le Z] \\ &= 0.3413 + 0.5 = 0.8413 \approx 84\%. \end{aligned}$$

(2 Marks)

iii) We want to find L such that

$$0.05 = P[T > 8.1 + L] = P\left[\frac{T-8.1}{0.1} > 10L\right] = 1 - P[Z \le 10L] = 0.5 - P[0 \le Z \le 10L]$$

(1 Mark) From the table, $10L = 1.645$ (interpolating between 1.64 and 1.65), so $L = 0.1645$. Thus the thickness of 90% of all cylinders lies within (8.1 ± 0.16) cm. **(1 Mark)**

Exercise 12.

Consider the continuous random variable X with density

$$f_X(x) = \frac{c}{e^{-x} + e^{x}} \qquad \text{for } x \in \mathbb{R}.$$

i) Determine the constant $c \in \mathbb{R}$.

ii) Find $P[X \le 1]$.

iii) Find the density of the random variable X^2.

(1+1+2 Marks)

Solution 12.

i) The density must satisfy $\int_{\mathbb{R}} f_X(x)\,dx = 1$. Since

$$\begin{aligned} \int_{\mathbb{R}} f_X(x)\,dx &= c\int_{-\infty}^{\infty} \frac{dx}{e^{-x}+e^{x}} \\ &= c\int_{-\infty}^{\infty} \frac{e^{x}\,dx}{1+e^{2x}} \\ &= c\int_{0}^{\infty} \frac{dy}{1+y^2} \\ &= c \cdot \arctan(y)\big|_0^{\infty} = c\frac{\pi}{2} \end{aligned}$$

we obtain $c = 2/\pi$. **(1 Mark)**

ii) The probability is given by

$$\begin{aligned} P[X \leq 1] &= \frac{2}{\pi} \int_{-\infty}^{1} \frac{dx}{e^{-x} + e^{x}} \\ &= \frac{2}{\pi} \int_{-\infty}^{1} \frac{e^{x}\, dx}{1 + e^{2x}} \\ &= \frac{2}{\pi} \int_{0}^{e} \frac{dy}{1+y^2} \\ &= \frac{2}{\pi} \arctan(y)\Big|_0^e = \frac{2}{\pi} \arctan(e). \end{aligned}$$

(1 Mark)

iii) Note that the function $\varphi\colon \mathbb{R} \to \mathbb{R}_+ \cup \{0\}$, $\varphi(x) = x^2$ is not bijective, so we can't simply apply the theorem for transforming random variables from the lecture. Let $y > 0$. Then, using the fact that f_X is even,

$$F_Y(y) = P[Y \leq y] = P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\, dx$$
$$= 2 \int_0^{\sqrt{y}} f_X(x)\, dx.$$

(1/2 Mark) It follows that

$$f_Y(y) = F_Y'(y) = 2 f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{2}{\pi\sqrt{y}} \frac{1}{e^{-\sqrt{y}} + e^{\sqrt{y}}}$$

for $y > 0$. **(1 Mark)** For $y \leq 0$ we have

$$F_Y(y) = P[Y \leq y] = P[X^2 \leq y] = 0,$$

so $f_Y(y) = 0$ for $y \leq 0$. **(1/2 Mark)**

Exercise 13.
The distribution function of the speed (modulus of the velocity) V of a gas molecule is described by the Maxwell-Boltzmann law

$$f_V(v) = \begin{cases} \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{3/2} v^2 e^{-\frac{m}{kT} v^2/2} & v > 0 \\ 0 & v \leq 0 \end{cases}$$

where $m > 0$ is the mass of the molecule, $T > 0$ is its temperature and $k > 0$ is the Boltzmann constant.

i) Find the mean and variance of V.

ii) Find the mean of the kinetic energy $E = mV^2/2$.

iii) Find the probability density f_E of E.

(3 + 2 + 2 Marks)

Solution 13.

i) The mean of V is given by

$$\begin{aligned} \mathrm{E}[V] &= \int_{\mathbb{R}} v f_V(v)\, dv \\ &= \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{1/2} \int_0^\infty \frac{m}{kT} v^3 e^{-\frac{m}{kT} v^2/2}\, dv \\ &= 2 \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{1/2} \int_0^\infty v e^{-\frac{m}{kT} v^2/2}\, dv \\ &= 2 \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{-1/2} [-e^{-\frac{m}{kT} v^2/2}]_0^\infty \\ &= 2 \left(\frac{2kT}{m\pi}\right)^{1/2}. \end{aligned}$$

(1 Mark) Furthermore,

$$\begin{aligned}
\mathrm{E}[V^2] &= \int_{\mathbb{R}} v^2 f_V(v)\,dv \\
&= \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{1/2} \int_0^\infty \frac{m}{kT} v^4 e^{-\frac{m}{kT}v^2/2}\,dv \\
&= 3\left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{1/2} \int_0^\infty v^2 e^{-\frac{m}{kT}v^2/2}\,dv \\
&= 3\left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{-1/2} \int_0^\infty e^{-\frac{m}{kT}v^2/2}\,dv.
\end{aligned}$$

Setting $w = \sqrt{m/(kT)}v$, we have

$$\mathrm{E}[V^2] = 6\left(\frac{m}{kT}\right)^{-1} \underbrace{\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-w^2/2}\,dw}_{=1/2} = \frac{3kT}{m}.$$

(1 Mark) It follows that

$$\mathrm{Var}[V] = \mathrm{E}[V^2] - \mathrm{E}[V]^2 = \frac{kT}{m}\left(3 - \frac{8}{\pi}\right)$$

(1 Mark)

ii) The expectation value of the kinetic energy is given by

$$\mathrm{E}[E] = \frac{m}{2}\,\mathrm{E}[V^2] = \frac{m}{2}\frac{3kT}{m} = \frac{3}{2}kT.$$

(2 Marks)

iii) Note that the function $\varphi\colon \mathbb{R} \to \mathbb{R}_+ \cup \{0\}$, $\varphi(v) = \frac{m}{2}v^2$ is not bijective, so we can't simply apply the theorem for transforming random variables from the lecture. Let $\varepsilon > 0$. Then

$$F_E(\varepsilon) = P[E \le \varepsilon] = P\left[\frac{m}{2}V^2 \le \varepsilon\right] = P\left[-\sqrt{2\varepsilon/m} \le V \le \sqrt{2\varepsilon/m}\right] = \int_{-\sqrt{2\varepsilon/m}}^{\sqrt{2\varepsilon/m}} f_V(v)\,dv$$
$$= \int_0^{\sqrt{2\varepsilon/m}} f_V(v)\,dx.$$

(1/2 Mark) It follows that

$$\begin{aligned}
f_E(\varepsilon) = F_E'(\varepsilon) &= f_V(\sqrt{2\varepsilon/m}) \cdot \frac{1}{\sqrt{2m\varepsilon}} \\
&= \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{3/2} \frac{2\varepsilon}{m} e^{-\frac{\varepsilon}{kT}} \cdot \frac{1}{\sqrt{2m\varepsilon}} \\
&= \frac{2}{\sqrt{\pi}}(kT)^{-3/2}\sqrt{\varepsilon}e^{-\frac{\varepsilon}{kT}}
\end{aligned}$$

for $\varepsilon > 0$. **(1 Mark)** For $\varepsilon \le 0$ we have

$$F_E(\varepsilon) = P[E \le \varepsilon] = P\left[\frac{m}{2}V^2 \le \varepsilon\right] \le P\left[\frac{m}{2}V^2 \le 0\right] = 0,$$

so $f_E(\varepsilon) = 0$ for $\varepsilon \le 0$. **(1/2 Mark)**

Multivariate Random Variables

Exercise 14.
A fair four-sided die is rolled and the result (a number from 1,. . . ,4) recorded. If the result is odd, that number is retained. But if the result is even, then a fair, four-sided die is rolled and that number is added to the first result. So the final result is a number between 1 and 8.

Define the two random variables

$$X = \begin{cases} 1 & \text{first die roll was even,} \\ 0 & \text{first die roll was odd,} \end{cases}$$
$$Y = \text{final result after the die roll(s).}$$

Calculate the correlation coefficient of X and Y.
(5 Marks)

Solution 14.
The random variable X follows a Bernoulli distribution with $P[X = 0] = 1/2$, $P[X = 1] = 1/2$. Thus,

$$\mathrm{E}[X] = \frac{1}{2}, \qquad \mathrm{Var}[X] = \frac{1}{4}.$$

(1 Mark) The random variable Y has a discrete distribution with

$$P[Y = 1] = \frac{1}{4}, \quad P[Y = 2] = 0, \quad P[Y = 3] = \frac{1}{4} + \frac{1}{16} = \frac{5}{16}, \quad P[Y = 4] = \frac{1}{16},$$
$$P[Y = 5] = \frac{2}{16}, \quad P[Y = 6] = \frac{2}{16}, \quad P[Y = 7] = \frac{1}{16}, \quad P[Y = 8] = \frac{1}{16}$$

Hence,

$$\mathrm{E}[Y] = \frac{4 + 15 + 4 + 10 + 12 + 7 + 8}{16} = \frac{15}{4},$$
$$\mathrm{E}[Y^2] = \frac{4 + 5 \cdot 9 + 1 \cdot 16 + 2 \cdot 25 + 2 \cdot 36 + 1 \cdot 49 + 1 \cdot 64}{16} = \frac{300}{16},$$
$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 = \frac{75}{16}.$$

(1 Mark) each for the expectation and variance of *Y*. For the covariance we need the random variable XY, where we see

$$P[XY = 2] = \frac{1}{2}, \quad P[XY = 1] = 0, \quad P[XY = 2] = 0, \quad P[XY = 3] = \frac{1}{16}, \quad P[XY = 4] = \frac{1}{16},$$
$$P[XY = 5] = \frac{2}{16}, \quad P[XY = 6] = \frac{2}{16}, \quad P[XY = 7] = \frac{1}{16}, \quad P[XY = 8] = \frac{1}{16}$$

so

$$\mathrm{E}[XY] = \frac{0 + 3 + 4 + 10 + 12 + 7 + 8}{16} = \frac{44}{16}$$

(1 Mark) Hence,

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \frac{44}{16} - \frac{1}{2} \cdot \frac{15}{4} = \frac{14}{16}$$

(1 Mark) The correlation coefficient is found to be

$$\varrho_{XY} = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{44}{\sqrt{4 \cdot 75}} = \frac{7}{5\sqrt{3}}.$$

(1 Mark)

Exercise 15.
Two weighted coins are available: a black coin has probability p of turning up heads, while a red coin has probability $1 - p$ of turning up heads.

On the first toss, the black coin is used. If it turns up heads, the red coin is used in the second toss. But if the black coin turns up tails, the black coin is used again in the second toss and the red coin is not used.

Let X be the number of heads on the first toss (0 or 1) and Y be the number of heads on the second toss (0 or 1).

Calculate the correlation coefficient of X and Y.
(5 Marks)

Solution 15.
The random variable X follows a Bernoulli distribution with $P[X = 0] = 1 - p$, $P[X = 1] = p]$. Thus,

$$\mathrm{E}[X] = p, \qquad \mathrm{Var}[X] = p(1-p).$$

(1 Mark) The random variable Y has a Bernoulli distribution with

$$\begin{aligned} P[Y = 1] &= P[Y = 1 \mid X = 0] \cdot P[X = 0] + P[Y = 1 \mid X = 1] \cdot P[X = 1] \\ &= p \cdot (1-p) + (1-p) \cdot p = 2p(1-p) \end{aligned}$$

so that

$$\mathrm{E}[Y] = 2p(1-p), \qquad \mathrm{Var}[Y] = 2p(1-p)(1 - 2p(1-p)).$$

(1 Mark) each for the expectation and variance of Y. For the covariance we need the random variable XY. This is also a Bernoulli random variable with

$$\mathrm{E}[XY] = P[XY = 1] = p(1-p).$$

(1 Mark) Hence,

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = p(1-p) - p \cdot 2p(1-p) = p(1-p)(1-2p)$$

(1 Mark) The correlation coefficient is found to be

$$\varrho_{XY} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{1-2p}{\sqrt{2\big((1-p)^2 + p^2\big)}}.$$

(1 Mark)

Exercise 16.
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: $\int_0^{2\pi} \cos^2(\theta)\,d\theta = \pi$.
(12 Marks)

Solution 16.

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

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

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 17.
Let X be a continuous random variable following a uniform distribution in the interval $[1,2]$. Let Z be an exponential random variable with random parameter $\beta = X$, i.e., Z is a random variable whose conditional density is that of an exponential distribution with parameter x.

i) Find the density function f_Z of Z.

ii) Find the expectation $\mathrm{E}[Z]$.

You may use the following integral identity, which holds for any $b > a > 0$:

$$\int_0^\infty \frac{e^{-ax} - e^{-bx}}{x}\, dx = \ln\left(\frac{b}{a}\right).$$

(8 Marks)

Solution 17.
The density of X is given by

$$f_X(x) = \begin{cases} 1 & 1 < x < 2 \\ ,0 & \text{otherwise} \end{cases}$$

and the conditional density of Z is

$$f_{Z|x}(z) = \begin{cases} xe^{-zx} & z > 0 \\ ,0 & \text{otherwise.} \end{cases}$$

(1 Mark) Then

$$f_{XZ}(x,z) = f_{Z|x}(z) \cdot f_X(x) = \begin{cases} xe^{-zx} & z > 0,\ 1 < x < 2, \\ ,0 & \text{otherwise.} \end{cases}$$

(2 Marks) including (1 Mark) for correctly indicating the region of x and z. The (marginal) density of Z is given by

$$\begin{aligned} f_Z(z) &= \int_{\mathbb{R}} f_{XZ}(x,z)\, dx && \textbf{(1 Mark) for the approach} \\ &= \begin{cases} \int_1^2 xe^{-zx}\, dx & z > 0 \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} \frac{1}{z}(e^{-z} - 2e^{-2z}) + \frac{1}{z^2}(e^{-z} - e^{-2z}) & z > 0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

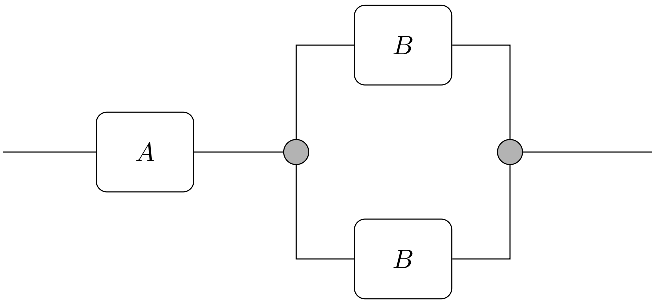
(2 Marks) for the result, including (1 Mark) for correctly indicating the region for z. Furthermore,

$$\begin{aligned} \mathrm{E}[Z] &= \int_{\mathbb{R}} z f_Z(z)\, dz \\ &= \underbrace{\int_0^\infty e^{-z}\, dz}_{=1} - \underbrace{\int_0^\infty 2e^{-2z}\, dz}_{=1} + \int_0^\infty \frac{e^{-z} - e^{-2z}}{z}\, dz \\ &= \ln(2) \end{aligned}$$

(2 Marks)

Reliability

Exercise 18.
Consider the following system of components:



The system will fail if either component A or both components marked B fail. The components A and B have failure densities

$$f_A(t) = \frac{1}{100}e^{-t/100}, \qquad f_B(t) = \frac{1}{50}e^{-t/50}, \qquad t \geq 0,$$

respectively.

i) Find the reliability functions $R_A(t)$ and $R_B(t)$ of component A and each component B.

ii) Find the reliability function $R(t)$ of the system.

iii) What is the expected time of failure of each individual component?

iv) What is the expected time of failure of the system?

(2+2+2+2 Marks)

Solution 18.

i) We have

$$\begin{aligned} P[t \leq T] &= \int_0^T f_A(t)\,dt = \frac{1}{100}\int_0^T e^{-t/100}\,dt = 1 - e^{-T/100} \\ R_A(T) &= 1 - P[t \leq T] = e^{-T/100} \\ R_B(T) &= e^{-T/50} \end{aligned}$$

(2 Marks)

ii) The reliability R_2 of the subsystem containing the two B components is given by

$$1 - R_2(t) = (1 - R_B(t))^2 = 1 - 2e^{-t/50} + e^{-t/25}$$

so $R_2(t) = 2e^{-t/50} - e^{-t/25}$. **(1 Mark)** The reliability of the entire system is given by

$$R(t) = R_A(t)R_2(T) = 2e^{-3t/100} - e^{-t/20}.$$

(1 Mark)

iii) The expected time of failure for component A is $\mathrm{E}[A] = 100$, for component B $\mathrm{E}[B] = 50$, since their time-to-failures follow exponential distributions with parameters $\beta = 10$ and 5, respectively. **(2 Marks)**

iv) The failure density for the system is

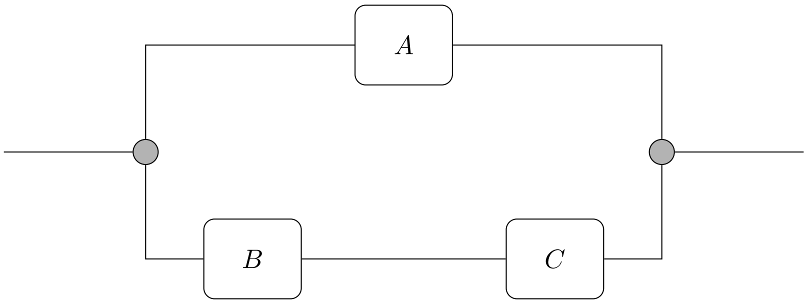
$$f(t) = -R'(t) = \frac{3}{50}e^{-3t/100} - \frac{1}{20}e^{-t/20}$$

(1 Mark) The expected time of failure is

$$2\frac{3}{100}\int_0^\infty te^{-3t/100}\,dt - \frac{1}{20}\int_0^\infty te^{-t/20}\,dt = 2\frac{100}{3} - 20 = \frac{140}{3} = 46.67.$$

(1 Mark)

Exercise 19.
Consider the following system of components:



The components A, B and C have failure densities

$$f_A(t) = \frac{t}{50}e^{-t^2/100}, \qquad f_B(t) = \frac{1}{40}e^{-t/40}, \qquad f_C(t) = \frac{1}{25}e^{-t/25}, \qquad t \geq 0,$$

respectively.

i) Find the reliability functions $R_A(t)$, $R_B(t)$ and $R_C(t)$.

ii) Find the reliability function $R(t)$ of the system.

iii) What is the expected time of failure of each component? Evaluate all integrals explicitly; there should be no gamma functions or other integrals in your answer.

iv) What is the expected time of failure of the system? Evaluate all integrals explicitly; there should be no gamma functions or other integrals in your answer.

(2+2+2+3 Marks)

Solution 19.

i) We have $P[t \leq T] = \int_0^T f(t)\,dt$ and $R(T) = 1 - P[t \leq T]$, so we obtain

$$\begin{aligned} R_A(T) &= 1 - \int_0^T f_A(t)\,dt = 1 - \int_0^T \frac{t}{50}e^{-t^2/100}\,dt \\ &= 1 - [-e^{-t^2/100}]_0^T = 1 - (-e^{-T^2/100} + 1) = e^{-T^2/100}, && \textbf{(1 Mark)} \\ R_B(T) &= 1 - \int_0^T f_B(t)\,dt = 1 - \int_0^T \frac{1}{40}e^{-t/40}\,dt = e^{-T/40} && \textbf{(1/2 Mark)} \\ R_C(T) &= e^{-T/25} && \textbf{(1/2 Mark)} \end{aligned}$$

ii) The reliability R_2 of the subsystem containing the components B and C is given by

$$R_2(t) = R_B(t)R_C(T) = e^{-t/40}e^{-t/25} = e^{-13t/200}$$

and the reliability $R(t)$ of the entire system is given by

$$\begin{aligned} 1 - R(t) &= (1 - R_A(t))(1 - R_2(t)) = (1 - e^{-t^2/100})(1 - e^{-13t/200}) \\ &= 1 - e^{-13t/200} - e^{-t^2/100} + e^{-(2t^2+13t)/200} \end{aligned}$$

so

$$R(t) = e^{-13t/200} + e^{-t^2/100} - e^{-(2t^2+13t)/200}$$

(2 Marks)

iii) Component A has the failure density of a gamma random variable with $\beta = 2$ and $\alpha = 1/100$. Thus the expected time of failure is

$$\begin{aligned}
\mathrm{E}[X_A] &= \alpha^{-1/\beta}\Gamma(1+1/\beta) = 10\Gamma(3/2) \\
&= 10\int_0^\infty z^{1/2}e^{-z}\,dz \\
&= 20\int_0^\infty w^2 e^{-w^2}\,dw \\
&= 10\int_0^\infty w(-2w)e^{-w^2}\,dw \\
&= 10\int_0^\infty e^{-w^2}\,dw \\
&= \frac{10}{\sqrt{2}}\int_0^\infty e^{-z^2/2}\,dz \\
&= 5\sqrt{\pi}
\end{aligned}$$

(1 Mark) Components B and C have failure densities of an exponential random variable with $\beta = 40$ and 25, respectively, so that

$$\mathrm{E}[X_B] = \beta = 40, \qquad \mathrm{E}[X_C] = 25.$$

(1 Mark)

iv) The expected time of failure of the system is

$$\begin{aligned}
E[X] &= \int_0^\infty tf(t)\,dt = -\int_0^\infty tR'(t)\,dt = -tR(t)|_0^\infty + \int_0^\infty R(t)\,dt \\
&= -0+0+\int_0^\infty e^{-13t/200} + e^{-t^2/100} - e^{-(2t^2+13t)/200}\,dt \\
&= \left[\frac{-200}{13}e^{-13t/200}\right]_0^\infty + \sqrt{50}\int_0^\infty e^{-s^2/2}\,ds - e^{169/1600}\int_0^\infty e^{-(t+13/4)^2/100}\,dt \\
&= \frac{200}{13} + \sqrt{25\pi} - e^{169/1600}\int_{13/4}^\infty e^{-s^2/100}\,ds \\
&= \underbrace{\frac{200}{13}}_{\textbf{(1 Mark)}} + \underbrace{5\sqrt{\pi}}_{\textbf{(1 Mark)}} - \sqrt{50}e^{169/1600}\int_{13/(4\sqrt{50})}^\infty e^{-t^2/2}\,dt \\
&= 15.3846 + 8.8623 - e^{169/1600}\cdot 5\sqrt{2}\cdot\frac{\sqrt{2\pi}}{\sqrt{2\pi}}\int_{0.460}^\infty e^{-t^2/2}\,dt \\
&= 24.2469 - e^{169/1600}\cdot 10\sqrt{\pi}\cdot\left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}}\int_0^{0.460} e^{-t^2/2}\,dt\right) \\
&= 24.2469 - \underbrace{19.6991\cdot(0.5-0.1772)}_{\textbf{(1 Mark)}} \\
&= 17.89.
\end{aligned}$$

Estimators

Exercise 20.
Let X be a continuous random variable with density

$$f_\theta(x) = \begin{cases} \dfrac{\theta+1}{x^{\theta+2}} & \text{for } x > 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter. Find the method-of-moments estimator for the parameter θ.
(3 Marks)

Solution 20.
We calculate

$$E[X] = \int_1^\infty \frac{\theta+1}{x^{\theta+1}}\,dx = (\theta+1)\int_1^\infty x^{-\theta-1}\,dx = \frac{\theta+1}{-\theta}[x^{-\theta}]_1^\infty = 1 + \frac{1}{\theta}$$

so that

$$\theta = \frac{1}{\mathrm{E}[X]-1}.$$

Using $\widehat{\mathrm{E}[x]} = \overline{X}$ we see that the method-of-moments estmator for θ is

$$\widehat{\theta} = \frac{1}{\overline{X}-1}$$

Exercise 21.
Let (X, f_X) be a continuous random variable following a uniform distribution on the interval $[0, \theta]$, $\theta > 0$, i.e.,

$$f_X(x) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{otherwise.} \end{cases}$$

i) Find the method-of-moments estimator for θ.

ii) Find the maximum-likelihood estimator for θ.

(2+3 Marks)

Solution 21.

i) The mean of X is given by

$$\mathrm{E}[X] = \int_0^\theta \frac{x}{\theta}\,dx = \frac{\theta}{2}.$$

Thus,

$$\theta = 2\,\mathrm{E}[X].$$

The method-of-moments estimator for θ is then

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

ii) The likelihood function for a random sample $X_1, \ldots, X_n$ of X is given by

$$L(\theta) = \prod_{i=1}^n f_X(x_i) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 < x_1, \ldots, x_n < \theta \\ 0 & \text{otherwise} \end{cases}$$

In order for $L(\theta) > 0$ we need $\theta \ge x_i$ for $i = 1, \ldots, n$, or $\theta \ge \max_{1\le i\le n} x_i$. The likelihood function is then maximized if $\theta = \max_{1\le i\le n} x_i$. The maximum-likelihood estimator hence is

$$\hat{\theta} = \max_{1\le i\le n} x_i$$

Exercise 22.
A certain type of harddrive is known to have a lifetime given by an exponential distribution with (unknown) parameter $\beta > 0$. To estimate β, n identical and independent hard drives are tested and their times of failure recorded. After time $T > 0$ the test is stopped and $n - m$ hard drives are found to be still working.

Find the maximum likelihood estimator for β.
(5 Marks)

Solution 22.
The density of the exponential distribution is given by

$$f(t) = \begin{cases} \frac{1}{\beta} e^{-t/\beta} & t > 0, \\ 0 & t \leq 0. \end{cases}$$

(1/2 Mark) This gives

$$P[t \leq T] = \int_0^T f(t)\,dt = 1 - e^{-T_k/\beta}.$$

The probability of failing after time T is

$$P[t > T] = 1 - P[t \leq T_k] = e^{-T/\beta}$$

(1/2 Mark) Suppose that the first m hard drives fail at times $T_1, \ldots, T_m < T$. The likelihood function is then

$$\begin{aligned} L(\beta) &= f_1(T_1) \cdot f_2(T_2) \ldots f_m(T_m) \cdot \underbrace{e^{-T/\beta} \cdots e^{-T/\beta}}_{n-m \text{ terms}} \\ &= \frac{1}{\beta^m} e^{-(T_1 + \cdots + T_m + (n-m)T)/\beta} \end{aligned}$$

(1 Mark) To find the maximum of L, we take the logarithm,

$$\ln(L(\beta)) = -(T_1 + \cdots + T_m + (n-m)T)/\beta - m\ln(\beta)$$

(1/2 Mark) and set the derivative equal to zero,

$$\frac{T_1 + \cdots + T_m + (n-m)T}{\beta^2} - \frac{m}{\beta} = 0.$$

(1/2 Mark) Solving for β gives

$$\widehat{\beta} = \frac{T_1 + \cdots + T_m + (n-m)T}{m}.$$

(1 Mark)

Exercise 23.
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.
(5 Marks)

Solution 23.
Let $X_1, \ldots, X_n$ be a random sample of size n. The Likelihood function is then

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

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

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

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.*

Data Visualization, Interpretation and Confidence Intervals

Exercise 24.
Consider the following set of 36 data:

9.2	6.7	7.6	6.2	7.8	6.0
4.6	4.4	7.9	10.1	5.6	7.8
6.1	9.2	5.4	6.0	7.6	3.5
6.6	8.2	3.9	4.9	8.8	5.4
4.4	7.3	7.2	12.2	4.6	4.2
9.8	3.0	6.0	8.6	4.8	8.2

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!
(6 Marks)

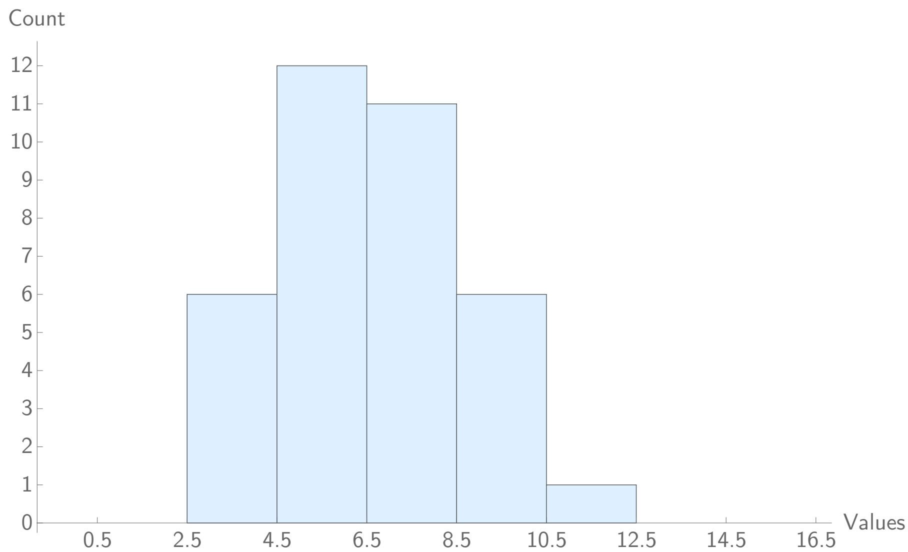
Solution 24.
We find

$$q_1 = 4.85, \qquad q_2 = 6.4, \qquad q_3 = 8.05, \qquad \text{iqr} = q_3 - q_1 = 3.2$$

(1 Mark) The bin widths should be given by

$$h = \frac{2\,\text{iqr}}{\sqrt[3]{36}} = 1.94$$

and it is reasonable to round to 1.9 or 2. **(1 Mark)**



(2 Marks) for drawing the histogram correctly: (1/2 Mark) for labelling the axes, (1/2 Mark) for starting the histogram at some point that prevents data from falling on the boundary or using a bin width with the same result. (1 Mark) is for the general shape and correctness of the histogram.
In case of an excessively messy/unreasonably small/unintelligible plot, from (0 Marks) to (2 Marks) may be deducted.

This histogram has a *unimodal shape* **(1/2 Mark)** which is consistent with a normal distribution. It is *not significantly skewed*, **(1/2 Mark)** again consistent with a normal distribution. Therefore, there is no evidence that the data does not come from a normal distribution. **(1 Mark)**

Exercise 25.
Consider the following set of 36 data:

9.2	6.7	7.6	6.2	7.8	6.0
4.6	4.4	7.9	10.1	5.6	7.8
6.1	9.2	5.4	6.0	7.6	3.5
6.6	8.2	3.9	4.9	8.8	5.4
4.4	7.3	7.2	12.2	4.6	4.2
9.8	3.0	6.0	8.6	4.8	8.2

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!
(6 Marks)

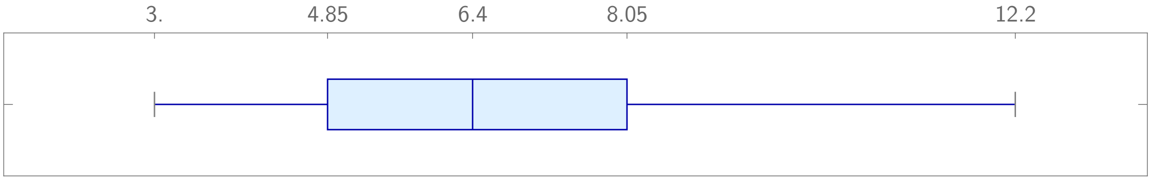
Solution 25.
We find

$$q_1 = 4.85, \qquad q_2 = 6.4, \qquad q_3 = 8.05, \qquad \text{iqr} = q_3 - q_1 = 3.2$$

(1 Mark) and

$$f_1 = q_1 - \frac{3}{2}\,\text{iqr} = 0.05, \qquad f_3 = q_3 + \frac{3}{2}\,\text{iqr} = 12.85,$$

(1 Mark)



(2 Marks) for drawing the boxplot correctly: (1/2 Mark) for the general shape of the boxplot, (1/2 Mark) for labelling the ordinate, (1 Mark) for the correct drawing indicating the whisker values, q_1, q_2, q_3 and for correctly identified outlier(s), if any.
In case of an excessively messy/unreasonably small/unintelligible plot, from (0 Marks) to (2 Marks) may be deducted.

The *whiskers are moderately asymmetric* **(1/2 Mark)** but the *median line is not too far from the center of the box*, **(1/2 Mark)**. There is *no outlier*, **(1/2 Mark)** and in summary no strong evidence that the data does not come from a normal distribution. **(1/2 Mark)**

Exercise 26.
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 two-sided, 90% confidence interval for the variance.
(3 Marks)

Solution 26.
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.*

Exercise 27.
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 two-sided, 95% confidence interval for the mean.
(3 Marks)

Solution 27.
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.*