Probability and Statistics for SIC
Exercises

Chapter 1

Exercise 1 How many passwords of 8 symbols can be created from 66 characters?

Exercise 2 In a given country, cars have number plates composed of two letters (of a 26 letter alphabet) followed by three numbers. How many possible number plates are there?

Exercise 3 A professor has 23 maths books and 9 physics books on one shelf of her bookcase. She would like to order her books so that all books on the same subject are together. How many different ways of ordering the books are there?

Exercise 4 Four Englishmen, three Irishmen and five Scotsmen sit on the same bench. In order to minimise conflict, people of the same nationality should be grouped. How many placements are possible?

Exercise 5 A group of n people, including A and B, start to queue randomly. How many ways are there to have k people between A and B, in the queue of length n?

Exercise 6 A committee must be formed of 4 people chosen from a group of 23. How many possible committees are there?

Exercise 7 How many distinct hands of poker are there? The game comprises 52 cards, and each hand contains 5 cards.

Exercise 8 In a group of 10 women and 8 men, a committee must be formed of 3 men and 3 women. How many different committees can be formed:
a) in total?
b) if 2 of the men refuse to be on the same committee?
c) if 2 of the women refuse to be on the same committee?
d) if 1 man and 1 woman refuse to be on the same committee?

Exercise 9 A research laboratory in the psychology of dreams has 4 rooms with two beds in each. Three pairs of identical twins are studied. Each pair is to be placed in a room and each twin assigned a bed. In how many different ways can the experiment be set up?

Exercise 10 Prove that the construction of Pascal's triangle is correct, i.e., that

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k.$$

Chapter 2

Exercise 11 A die is cast repeatedly until a 6 is thrown, which ends the experiment.
a) What is the sample space for this experiment?
b) Let E_n denote the event "the experiment ends at the n^{th} cast". Which points of the sample space are contained in E_n? Describe $(\cup_1^\infty E_n)^c$.

Exercise 12 n people including A and B form a line at random.
a) Describe the sample space associated with this experiment, and give the number of events.
b) What is the probability that there are k people between A and B, where $0 \le k \le n-2$?
c) Check your result for $n = 3$, and explain all the possible cases.

Exercise 13 A receptor blocks if the time between the arrivals of two signals is less than a down-time of length θ. Both signals reach the receptor in the time interval $(0, t)$, and arrive independently of each other and at random.

a) Describe the sample space Ω linked with this random experiment.

b) What is the probability that the receptor blocks?

c) Simplify your result if $\theta \ll t$.

Exercise 14 A fair die is thrown three times.

a) Describe the sample space of this experiment.

b) What is the probability that the sum of the dice is greater than or equal to 15?

Exercise 15 An urn contains 25 balls; 15 are red and 10 are green. Five balls are picked at random without replacement.

a) Find the probability that the first, the third and the fifth balls are red, and that the second and fourth are green.

b) Find the probability that the first, second and third balls are red, and the fourth and fifth are green.

c) Find the probability that the second and third balls are red and the first, fourth and fifth are green.

Exercise 16 The colour of a person's eyes is determined by a unique pair of genes. If both are the genes for blue eyes, the person will have blue eyes; if they are both the genes for brown eyes, the person will have brown eyes; if one is a blue-eyed gene and the other brown-eyed, the person will have brown eyes (i.e., the brown-eyed gene is dominant). A newborn baby receives one eye gene independently from each parent, and the gene he receives from each parent has an equal probability of being either one of that parent's two eye genes. Suppose that Xavier and both his parents have brown eyes, but Xavier's sister has blue eyes.

a) What is the probability that Xavier has one blue-eyed gene?

b) If Xavier's wife has blue eyes, what is the probability that their first child will have blue eyes?

Exercise 17 Two distinguishable dice are cast. Let E denote the event "the sum of the dice is odd", F denote the event "one at least of the dice shows a 1", and G denote "the sum of the dice is 5". Describe

a) $E \cap F$,

b) $E \cup F$,

c) $F \cap G$,

d) $E \cap F^c$,

e) $E \cap F \cap G$.

Exercise 18 Three players, A, B and C, take it in turns to toss a coin: A plays first, then B, then C, then A etc. The first to get tails wins. The sample space Ω of this experiment can be described using the following notation:

$$S = \begin{cases} 1, 01, 001, 0001, \ldots, \\ 0000\cdots \end{cases}$$

a) Give an interpretation of the points of S.

b) Describe the following events in terms of these points, using a sequence: (i) "A wins"; (ii) "B wins"; (iii) $(A \cup B)^c$.

Exercise 19 Let S be the sum of the values of three dice thrown simultaneously. There are six different configurations allowing 9 or 10 to be obtained: for $S = 9$, they are (6,2,1), (5,3,1), (5,2,2), (4,4,1), (4,3,2) and (3,3,3); and for $S = 10$ they are (6,3,1), (6,2,2), (5,4,1), (5,3,2), (4,4,2) and (4,3,3). Can we conclude that $\Pr(S = 9) = \Pr(S = 10)$?

Exercise 20 What is the probability of casting at least one 6 when a die is thrown four times?

Exercise 21 Two dice are thrown n times. Calculate the probability that the 6 appears at least once. How large should n so that this probability reaches 1/2?

Exercise 22 How many people are needed for the probability that at least two of them have their birthdays during the same month to be at least 1/2? All months are equally likely.

Exercise 23 Two fair coins are thrown. What is the probability of obtaining at least one tail?

Exercise 24 In 1654 de Mer set Pascal the following problem: is it more probable to obtain at least one 6 with 4 dice than to obtain a double 6 in 24 casts of two dice? What do you think?

Exercise 25 Three balls are chosen without replacement from an urn containing 20 balls numbered from 1 to 20. Someone bets that at least one of the balls chosen will bear a number equal to or greater than 17. What is the probability of winning this bet?

Exercise 26 A couple has two children. What is the probability that both are girls, given that the eldest is a girl?

Exercise 27 Two fair dice are cast. What is the probability of seeing at least one 6, given that they show different results?

Exercise 28 Three cards are chosen at random without replacement from an ordinary pack of 52 cards. Calculate the probability that the first card is a spade, given that the second and third are spades.

Exercise 29 Days can be sunny or cloudy. The probability that tomorrow will have the same weather as today is p, and the probability that it will be different is q, where $p+q=1$. If it is sunny today, show that the probability s_n that it will be sunny on the n^{th} following day satisfies

$$s_n = (p-q)s_{n-1} + q, \quad n = 1, 2, \ldots,$$

where $s_0 = 1$. Deduce that

$$s_n = \tfrac{1}{2}(1 + (p-q)^n); \quad n = 0, 1, \ldots$$

Hint: define the events 'sunny on day n' and 'not sunny on day n'.

Exercise 30 a) A coin is tossed n times, $n \geq 2$. Let P_n denote the probability that successive tails never appear in the n tosses. Show that

$$P_n = \tfrac{1}{2}P_{n-1} + \tfrac{1}{4}P_{n-2} \qquad n \geq 4.$$

Hint: Use two events to help you.

b) Show that $\lim_{n\to\infty} P_n = 0$.

c) Let Q_n be the probability that the series $THTTTHHT$ never appears during n tosses. Show that $\lim_{n\to\infty} Q_n = 0$.

Hint: The probability Q_n that the series $THTTTHHT$ never appears in n tosses is smaller than the probability R_n that the same series never appears between toss number $8i$ and toss number $8(i+1)-1$ where i is a positive integer smaller than or equal to $n/8-8$.

Exercise 31 Consider three playing cards of the same shape. Both sides of the first card are black, both sides of the second card are red, and the third card has one red side and one black side. The three cards are mixed in a hat, then one is drawn at random and placed on the ground. If the visible side is red, what is the probability that the other side is black?

Exercise 32 Suppose that 5% of men and 0.25% of women are colour-blind. If a colour-blind person is selected at random from a group with equal numbers of men and women, what is the probability that the selected person is a man? If there were twice as many women as men in the group, what would the probability be?

Exercise 33 A test is performed to detect a disease. If the patient is indeed infected, the test gives a positive result in 99% of cases. But it is also possible that the result of the test is positive and the patient is in good health, and this occurs in 2% of cases. Given that on average one patient in 1000 is infected with the disease, calculate the probability that a patient is infected given that his test was positive. How do your results reflect on the effectiveness of the test?

Exercise 34 The pieces manufactured in a factory are checked, but the checking procedure is not wholly reliable. If a piece is good, it is accepted with probability 0.9, but if it is defective, it is refused with probability 0.8.

a) If a batch contains three good pieces and one defective piece, what is the probability that all four pieces are accepted by the checking mechanism?

b) What is the probability of an error during the checking process, given that on average 20% of the pieces produced are defective?

c) What is the probability that a piece accepted during the checking process is defective (again given that 20% of production is defective)?

Chapter 3

Exercise 35 A production unit contains two machines functioning independently of each other. Each machine has a reliability p over the course of a day, so the probability of it breaking down during that time is equal to $1-p$. In that case, it will be repaired during the night and will be in working order the next day. Only one machine can be repaired at once.

Let X_n denote the number of machines broken down at the start of the n^{th} day, for $n = 1, 2, \ldots$. Calculate the conditional probabilities $\Pr(X_{n+1} = 0 \mid X_n = 0)$, $\Pr(X_{n+1} = 1 \mid X_n = 0)$, $\Pr(X_{n+1} = 0 \mid X_n = 1)$ and $\Pr(X_{n+1} = 1 \mid X_n = 1)$.

Exercise 36 Throw two fair dice. Let X_1 and X_2 respectively denote the largest and smallest of the two results. Give the mass functions of X_1 and of X_2 and sketch them.

Exercise 37 42800 weddings were celebrated in Switzerland in 2010. Giving explanations for your hypotheses and for the corresponding laws of probability, calculate the probability that for exactly 2 of these couples:

a) both partners were born on a January 1st,

b) both partners have the same birthday.

Exercise 38 a) A park contains L lions and T tigers. A park ranger is tasked with roaming the park. Each time he sees a lion or a tiger, he makes a note of the type of animal (lion or tiger) he has seen. He stops when he has seen n animals and reports to his boss. What is the distribution of the number of lions he has seen? Give the probability that k lions have been noted.

b) In a rival park, another strategy has been adopted: each time that the (brave!) ranger sees a lion or a tiger, he captures it. After having captured n animals, the ranger reports to his boss. What is the distribution of the number of lions he has captured? Give the probability that k lions have been captured.

Exercise 39 Consider the following game, which can be played as many times as desired. Each turn is independent and each time, the probability of wining is p. Each turn, the player bets a certain amount x. If he wins, he recovers double his stake. If he loses, he forfeits his stake.

Arnaud has a simple strategy: he carries on playing until he wins. At the first turn, he bets 100CHF. Every time he loses, he bets twice what he bet at the last turn. As soon as he wins, he stops playing.

a) What is the amount bet by Arnaud at turn n?

b) What is the distribution of the number of turns Arnaud plays until his first win?

c) Show that mean value of Arnaud's stake on the turn at which he stops is

$$\begin{cases} \frac{100p}{1-2(1-p)}, & p > 1/2, \\ \infty, & p \leq 1/2. \end{cases}$$

Interpret this.

Exercise 40 Two six-sided dice are thrown simultaneously until a sum of either 5 or 7 is reached.

a) The game stopped at the jth throw. What is the probability that a sum of 5 was reached at the jth throw?

b) What is the probability of finishing the game with a sum of 5?

c) What is the average number of throws before the end of the game?

Exercise 41 a) n balls are chosen randomly without replacement from an urn containing N white balls and M black balls. Give the probability mass function of the random variable X which counts the number of white balls chosen. Show that the expectation of X is $Nn/(M+N)$.

Hint: Do not use the hypergeometric distribution; instead write $X = X_1 + \cdots + X_N$ where X_i equals 1 if the ith white ball was chosen from among the n balls, and otherwise equals 0.

b) An urn contains N white balls and M black balls. The balls are removed from the urn one by one until a white ball appears. If X denotes the number of balls removed, what is the expectation of X?

Exercise 42 A coin is tossed n times independently. Suppose that each time it is tossed, the probability of getting tails is p, where $p \in [0, 1]$.

a) Give the probability distribution of the variable X which counts the number of tails attained.

b) Give the expectation and variance of X.

Exercise 43 A maths assistant is desperately trying to pass his driving test. Suppose that he has a probability $p \in (0, 1]$ of passing at each attempt, and let T denote the number of attempts necessary in order to pass. Give the mass function of T, and its mean and its variance.

Exercise 44 The mass function of a random variable X is

$$\Pr(X = i) = c\,\lambda^i / i!, \quad i = 0, 1, 2, \ldots,$$

where $\lambda > 0$.

a) Calculate c and give the distribution of X

b) Calculate $\Pr(X = 0)$.

c) Calculate $\Pr(X > 2)$.

d) Give the expectation and the variance of X.

Exercise 45 A book of 350 pages contains 450 randomly distributed printing errors. Calculate, in two different ways, the probability of there being at least three errors on a given page.

Exercise 46 A computer program has two independent modules. The number of errors X_1 in the first module has mass function P_1 and the number of errors X_2 in the second module has mass function P_2, where

x	0	1	2	3
$P_1(x)$	0.5	0.2	0.2	0.1
$P_2(x)$	0.7	0.2	0.1	0

a) Check that P_1 and P_2 are indeed mass functions.

b) Sketch the distribution functions of X_1 and X_2.

c) Calculate the mass function of $Y = X_1 + X_2$, the total number of errors in the program.

d) Sketch the distribution function of Y.

Chapter 4

Exercise 47 If a system has a random life-span X (months) and its density is

$$f(x) = \begin{cases} cxe^{-x/2}, & x > 0, \\ 0, & x \le 0, \end{cases}$$

what is the probability that the system functions for at least 5 months?

Exercise 48 Let X denote a gamma random variable of density

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1}, & x \ge 0, \\ 0, & x < 0, \end{cases}$$

for $\alpha, \lambda > 0$ and where $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$ is the gamma function.

a) By employing integration in parts, show that $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$. Can you find an expression for $\Gamma(n)$ when n is a whole number?

b) Show that the expectation of X is α/λ.

Exercise 49 If X has probability density $f_X(x) = 1/\{\pi(1 + x^2)\}$, for $x \in \mathbb{R}$, we say that X is a Cauchy random variable. Show that $\mathrm{E}(X)$ is undefined.

Exercise 50 Suppose that the lifetime (in years) of an individual has density

$$f(t) = \begin{cases} ct^2(t - 100)^2, & t \in [0, 100], \\ 0, & \text{otherwise.} \end{cases}$$

a) Calculate the constant c.

b) What is the expected average lifetime?

c) What is the probability that an individual has a lifetime between 50 and 80 years?

Exercise 51 The lifetime (years) of a component is a continuous random variable with density function

$$f(x) = \begin{cases} k/x^4, & x \geq 1, \\ 0, & x < 1. \end{cases}$$

a) Find k.
b) Calculate the distribution function of the lifetime.
c) Calculate the probability that the component's lifetime will exceed 3 years.
d) What is the average lifetime of the component?
e) What is the standard deviation of the lifetime?

Exercise 52 If $U \sim U(0,2)$, what distribution has U conditional on (a) $U \leq 1$? (b) $|U-1| \geq 1/2$?

Exercise 53 If $X \sim \exp(1)$, find the density of $Y = \log X$.

Exercise 54 If a continuous random variable X has probability density function x^{-2}, for $x \geq 1$, find the cumulative distribution function of $Y = 1/X$.

Exercise 55 If $X \sim N(0,1)$, use the Jacobian formula to obtain the probability density function of $Y = e^X$; Y is said to be a *lognormal* variable.

Exercise 56 Suppose that the height (cm) of a 25-year-old man is a normal random variable with $\mu = 175$ and $\sigma^2 = 36$. What percentage of 25-year-old men are taller than 185 cm? Among men taller than 180 cm, what percentage are taller than 192 cm?

Exercise 57 The speed of a molecule in a homogeneous gas in equilibrium is a random variable with density function

$$f(x) = \begin{cases} ax^2 e^{-bx^2}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where $b = m/2kT$ and k, T, m respectively are Boltzmann's constant, the absolute temperature, and the molecular mass. What value of a makes this a density function?

Chapter 5

Exercise 58 Give expressions for the density functions of the maximum $Z = \max(X,Y)$ and the minimum $\tilde{Z} = \min(X,Y)$ of two independent continuous random variables X and Y.

Exercise 59 Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \exp(\lambda)$. Calculate the distribution function of $\min(X_1, \ldots, X_n)$ and say what this distribution is.

Exercise 60 Let X and Z denote two random variables taking values in $0, 1, \ldots$. We suppose that

a) Z is a Poisson variable of parameter λ;

b) $X \leq Z$ and

$$\Pr(X = k \mid Z = n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, \ldots, n\}, n = 0, 1, \ldots,$$

i.e., the conditional distribution of X given that $Z = n$ is binomial of parameters n and p, for a fixed $0 < p < 1$.

Prove that X and $Y = Z - X$ are two independent Poisson variables, and give their parameters.

Exercise 61 Let (X,Y) be a random vector with joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} c(x^2 + xy/2), & 0 \leq x \leq 1, 0 \leq y \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

a) Determine the constant c.
b) Determine the marginal densities of X and Y. Are X and Y independent?
c) Calculate $\Pr(X > Y)$.

Exercise 62 Let X and Y have joint density $f(x, y)$. Calculate the joint density function of X and of Y in the following cases, in each case saying whether X and Y are independent: a) $f(x, y) = xe^{-x(1+y)}$ for $x, y \geq 0$, and $f(x, y) = 0$ otherwise; b) $f(x, y) = 60xy^2$ for $x, y \geq 0$ and $x + y \leq 1$, and $f(x, y) = 0$ otherwise.

Exercise 63 If X_1 and X_2 are random variables, show that $X_1 + X_2 = \max(X_1, X_2) + \min(X_1, X_2)$. If $X_1, X_2 \overset{\text{ind}}{\sim} \exp(\lambda_1), \exp(\lambda_2)$, find the distribution of $\min(X_1, X_2)$, and hence show that

$$\mathrm{E}\{\max(X_1, X_2)\} = 1/\lambda_2 + 1/\lambda_2 - 1/(\lambda_1 + \lambda_2).$$

Exercise 64 Let Y be a gamma random variable with $\alpha = 1/2$ and $\lambda = 1$ and density function

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{\pi y}} e^{-y}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

Let X be another random variable, such that $X \sim \mathcal{N}(0, \frac{1}{2y})$, conditional on $Y = y$.

a) Calculate the joint density of (X, Y).
b) Calculate the density of X.
c) Calculate and identify the conditional density of Y given X.

Exercise 65 Let (X, Y) denote a random vector of joint probability density:

$$h(x, y) = \begin{cases} 1/\pi, & x^2 + y^2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

a) Calculate the marginal densities f and g of X and of Y. Are X and Y independent?
b) Find $\mathrm{cov}(X, Y)$ **without calculation**.

Exercise 66 Let X_1 and X_2 denote two independent binomial variables with parameters n and π. What is the variance of $X_1 - X_2$? If $X_1, \ldots, X_n$ are independent binomial variables with parameters n and π, what is the variance of $S = X_1 + \cdots + X_n$?

Exercise 67 The random variables X and Y are such that X has mean 1 and variance 4, Y has mean 2 and variance 9, and $\mathrm{corr}(X, Y) = 1/3$. What is the variance of $3X - 2Y + 1$? What is the covariance of $X + 2Y$ with $X - Y$? If Z is another random variable that satisfies $\mathrm{E}(3X - 2Y + Z) = 0$, what is the mean of Z?

Exercise 68 Let X and Y denote two independent random variables, the first with mean 2 and variance 3, the second with mean 4 and variance 5.

a) What is the mean of $5X - 3Y + 9$? What is the variance of $3Y - 2X$?

b) Supposing that X and Y have the same variance σ^2, if $X + Y$ has variance 96 and $X - Y$ has variance 64, what is $\mathrm{corr}(X, Y)$?

Exercise 69 A fair die is thrown n times. Let X_1 denote the total number of 1s obtained during the n throws, and X_2 the total number of 2s.

a) What are the distributions of X_1 and X_2? Why?

b) What are the variances of X_1 and X_2?

c) Let $U = X_1 + X_2$. What does it represent? By considering the distribution of U and hence obtaining its variance, deduce that $\mathrm{corr}(X_1, X_2) = -1/5$.

d) What is the variance of $V = X_1 - X_2$, and what is its correlation with U?

Exercise 70 We toss a fair coin, giving X, with 0 for tails and 1 for heads, and independently throw a fair die, with Y indicating the number of points obtained. Give the density function of $Z = X + Y$ and the conditional expectation $\mathrm{E}(Z \mid X = x)$. Calculate the conditional variance $\mathrm{var}(Z \mid X = x)$ and check that the results are in accordance with Theorem 168 of the notes.

Exercise 71 Let X and Y denote independent Bernoulli and geometric random variables. Find the density function of $Z = X + Y$ and calculate the conditional expectations of Z given that $X = x$ and of Z given that $Y = y$.

Exercise 72 Let X denote a standard normal variable and $Y = e^X$ (so Y is a lognormal variable). Calculate the mean and variance of Y.
Note: Recall that $M_X(t) = \exp(t^2/2)$, $t \in \mathbb{R}$.

Exercise 73 Let $Z = X + Y$ where X, Y are independent exponential variables of parameters λ_1 , λ_2.
a) Find the distribution function of Z.
b) Calculate its moment generating function.
c) The moment generating function of a gamma random variable G with parameters α, λ, is

$$M_G(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}, \quad t < \lambda.$$

Give conditions on λ_1, λ_2 under which Z has a gamma distribution. What are then its parameters?

Exercise 74 Compute the moment-generating function of a geometric random variable X, and hence obtain its mean and variance.

Exercise 75 Show that if X and Y are two independent Poisson variables with parameters λ_1 and λ_2 then $Z = X + Y$ is Poisson with parameter $\lambda_1 + \lambda_2$: a) by a direct calculation of the mass function; b) using the moment generating function.

Exercise 76 Let X be a standard normal variable, let ϵ be independent of X and such that $\Pr(\epsilon = 1) = \Pr(\epsilon = -1) = 1/2$, and let $Y = \epsilon X$.
a) Calculate $\mathrm{cov}(X, Y)$.
b) Is the vector (X, Y) jointly Gaussian?
c) Are the variables X and Y independent?

Exercise 77 Suppose that $X, Y \overset{\text{iid}}{\sim} U(0, 1)$. Show that $Z = X + Y$ has density

$$f_Z(z) = \begin{cases} z, & 0 < z \le 1, \\ 2 - z, & 1 < z \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

Chapter 6

Exercise 78 Let $\theta > 0$ and $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f$, where

$$f(x) = \begin{cases} 0, & x < 0, \\ \alpha x, & 0 \le x \le \theta, \\ 0, & x > \theta. \end{cases}$$

a) Determine the value of α for which f is a density function.
b) Give the distribution function F_{M_n} of $M_n = \max(X_1, \ldots, X_n)$ and calculate its first two moments.
c) Show that $M_n \xrightarrow{P} \theta$ as $n \to \infty$.

Exercise 79 A professor knows by experience that student marks at a final exam are random variables of expectation 75.
a) Give the upper bound for the probability that a mark will exceed 85.
Now suppose that the professor also knows that the mark has variance 25.
b) What can we say about the probability that a student will obtain a mark between 65 and 85?
c) How many students would have to sit this exam to be sure, with a probability of at least 0,9, that the class average will be of 75, give or take 5? Don't use the central limit theorem.
d) Use the central limit theorem to solve part c).

Exercise 80 What is the probability that fewer than 16 sixes are seen when a fair die is thrown 120 times?

Exercise 81 A coin is tossed 500 times. What is the probability that the number of heads differs from 250 by at most 10?

Exercise 82 Round 50 numbers to the nearest whole number and add them up. If the individual rounding errors are uniformly distributed on $(-0.5,\ 0.5)$, what is the probability that the sum obtained will differ from the exact sum by more than 3?

Exercise 83 We have 100 light-bulbs, whose lifespans are independent exponential random variables with average 5 hours. If we light one light-bulb at a time and a failed bulb is immediately replaced by a new one, what is the probability that after 525 hours there's still at least one intact light-bulb left?

Exercise 84 Premises must be permanently illuminated; when the light-bulb blows, it is immediately replaced by a new one. Two qualities of light-bulb exist: quality A of exponential lifespan (in hours) of parameter $\lambda_A = 0.01$ and quality B of exponential lifespan (hours) of parameter $\lambda_B = 0.02$. What is the probability that 40 light-bulbs of quality A and 60 of quality B are enough to light the premises for at least 6500 hours?

Exercise 85 Let $X \sim U(0,1)$.

a) Calculate the expectation and variance of $Y = \log X$.

b)If $X_1, \ldots, X_{100} \overset{\text{iid}}{\sim} U(0,1)$ and $Z = X_1 \times \cdots \times X_{100}$, give an approximation to $\Pr(Z < 10^{-40})$.

Exercise 86 A computer program has three sections: a start-up section A, a calculation section B comprised of 100 identical and independent calculation cycles, and a final section C. The calculation times of the program for each of the three sections are summarised in the following table.

Section	Mean time (ms)	Standard deviation (ms)
Start-up A	5.5	2.5
Calculation B (100 cycles)	3.4	2.6
Finishing C	4.5	1.3

All the times are independent except those in sections A and C, whose correlation is 0.2.

a) Calculate $\text{cov}(A, C)$.

b) Calculate the mean and variance of total execution time T of the program.

c) The times of sections A and C are normally distributed. Although the calculation time of one cycle of section B is not normally distributed, explain why the total time distribution of section B (which comprises 100 cycles) can be approximated by a normal distribution and give the parameters of this distribution.

d) Calculate the proportion of cases for which the program takes

i) less than 10 ms,

ii) more than 20 ms.

Exercise 87 If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Pois}(\lambda)$, find the approximate distribution of $Y = 2\overline{X}^{1/2}$ for large n.

Exercise 88 If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, find the approximate distribution of $Y = 1/\overline{X}$ for large n.

Chapter 7

Exercise 89 Briefly explain the terms 'robust' and 'location measure' in the phrase 'the median is a robust location measure'.

Exercise 90 Let $x_1, \ldots, x_n$ denote observations with average $\overline{x}$ and sample variance $s^2 = 0$. Which of the following statements are true? Justify your answer(s).

a) The sample size is too small.

b) All the observations are equal.

c) $\overline{x} = 0$.

d) The observations are normally distributed.

Exercise 91 Which of the following statements can be applied to the empirical covariance s_{XY}? Justify your answer(s).

a) It measures the association of the variables X and Y

b) It lies between -1 and 1.

c) It can never be *exactly* zero.

d) It depends on the units in which X and Y are expressed.

Exercise 92 Which of the following statements can be applied to the empirical correlation r_{XY}? Justify your answer(s).

a) It depends on the units in which X and Y are expressed.

b) It measures the association of the variables X and Y.

c) It lies between -1 and 1.

d) It can never be *exactly* zero.

Exercise 93 Define the empirical correlation r for a sample of n observation pairs (x_i, y_i), $i = 1, \dots, n$. Sketch the types of clouds of points x and y we can expect when $r = -1, 0, 1$.

Exercise 94 Data $(x_1, y_1), \dots, (x_n, y_n)$ have empirical correlation r. For each of the following values of r, sketch one possible data configuration: (i) $r = 1$, (ii) $r = -0.5$, (iii) $r = 0$.

Chapter 8

Exercise 95 If $X_1, \dots, X_n$ are independent variables sampled from the gamma distribution with parameters α and λ, find method of moment estimators of the parameters.

Exercise 96 If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = e^X$ is said to have a log-normal distribution.
(a) Use the moment-generating function of X to show that

$$\mathrm{E}(Y) = \exp(\mu + \sigma^2/2), \quad \mathrm{var}(Y) = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1).$$

(b) If $Y_1, \dots, Y_n$ is a random sample from the log-normal distribution, find method of moments estimators of $\psi = \mu + \sigma^2/2$ and σ^2.

Exercise 97 We observe a sample $y_1, \dots, y_n$ from a Gamma$(3, \theta)$ density

$$f(y) = \tfrac{1}{2}\theta(\theta y)^2 e^{-\theta y}, \qquad y > 0.$$

Determine the maximum likelihood estimator of θ.

Exercise 98 Give the maximum likelihood estimator for θ based on a random sample $z_1, \dots, z_n$ from a Poisson distribution with mean 2θ.

Exercise 99 Let $X_1, \dots, X_n$ denote a random sample from the uniform distribution $\mathcal{U}(0, b)$. Find the maximum likelihood estimator of b.

Exercise 100 Consider a random sample $X_1, \dots, X_n$ from the distribution given by $p\mathcal{U}[0, a] + (1 - p)\mathcal{U}[0, b]$, i.e., X_i follows the uniform distribution on $[0, a]$ with probability p and the uniform distribution on $[0, b]$ with probability $1 - p$. Suppose that $0 < a < b$ with a and b fixed and known.
(a) Determine the distribution function and density of X_1.
(b) Let N_a denote the random variable equal to the number of individuals X_i between 0 and a. What is its distribution? Find its expectation and variance.
(c) Determine the maximum likelihood estimator of p.

Exercise 101 Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} U(a, b)$, with mean $\mathrm{E}(X_j) = (a + b)/2 = \theta$, say. Find the mean square error of the average $\overline{X}$ as an estimator of θ.

Exercise 102 We observe the waiting times of n people at traffic lights, denoted $t_1, \dots, t_n$. We hypothesise that the underlying waiting times $T_1, \dots, T_n$ are independent $U(0, \theta)$ variables. The maximum waiting time at these particular lights is therefore $\theta > 0$, which is unknown.

(a) Show that $\widehat{\theta}_1 = 2\overline{T}$, where $\overline{T} = n^{-1} \sum_{i=1}^n T_i$, is an unbiased estimator of θ and that its variance tends to 0 when $n \to \infty$.

(b) Calculate the distribution and density functions of the random variable $M_n = \max_{i=1}^n T_i$, and obtain its expectation and its variance.

(c) Find a second unbiased estimator $\widehat{\theta}_2$ of θ, and show that its variance tends to 0 when $n \to \infty$.

(d) Which of the two estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ would you choose?

(e) Show that M_n and $\widehat{\theta}_2$ converge in probability towards θ, as $n \to \infty$.

Exercise 103 During a test, we asked 120 people who Jean-Jacques Rousseau was, and 12 of them answered that he was a diver (false!). Estimate the proportion of the population who would give a false answer. Obtain a 95% confidence interval for this proportion. (Extra question: who was Jean-Jacques Rousseau?)

Exercise 104 A physical constant μ is measured 25 times in a laboratory. The measurements are assumed to follow a normal distribution $\mathcal{N}(\mu, \sigma^2)$ of known variance σ^2. Based on the measurements, the 90% confidence interval for μ is $[6.04, 6.18]$. How many more measurements must be taken if we want: (a) to halve the length of the 90% confidence interval for μ? (b) to obtain a 95% confidence interval having the same length?

Exercise 105 Suppose $x_1, \dots, x_6$ are a random sample from the $\mathcal{N}(\mu_1, \sigma_1^2)$ distribution and $y_1, \dots, y_{12}$ are a random sample from the $\mathcal{N}(\mu_2, \sigma_2^2)$ distribution, independent of the xs. All the parameters are unknown. We find that $\overline{x} = \frac{1}{6}\sum_{k=1}^6 x_k = 49.2$, $s_1^2 = \frac{1}{5}\sum_{k=1}^6 (x_k - \overline{x})^2 = 2.8$, $\overline{y} = 48.4$ and $s_2^2 = 3.04$.

a) Construct 95% confidence intervals for μ_1 and μ_2.

b) Now information is revealed: the variances σ_1^2 and σ_2^2 equal 2.5 and 3. Give new 95% confidence intervals for μ_1 and μ_2.

c) Still supposing that $\sigma_1^2 = 2.5$ and $\sigma_2^2 = 3$, give the distribution of the difference $\overline{X} - \overline{Y}$ of the averages, and construct a 90% confidence interval for $\mu_1 - \mu_2$.

Exercise 106 A study of 300 employees of the same firm revealed that the average number, X, of coffees drunk annually at the firm's cafeteria by any given employee follows a normal distribution 500 and sample standard deviation 100. Give 95% confidence intervals for the mean and variance of X.
Hint: use the `R` command `qchisq(`α`, df=`ν`)` to compute quantiles that are not in the tables.

Exercise 107 Ten apples randomly picked from a tree have weights in grams

121.9 164.4 167.3 170.1 178.6 99.2 96.4 187.1 144.6 172.9

a) Obtain a 90% confidence interval for the average weight μ of an apple.

b) Explain the meaning of the words '90% confidence interval' in the phrase above.

Exercise 108 The weight in grams of 1000 pots of jam from a packing machine were the following (the results are given by classes of length 2, the first class starting at 2000 and the last class finishing at 2024):

Class	1	2	3	4	5	6	7	8	9	10	11	12
Number of pots	9	21	58	131	204	213	185	110	50	16	3	0

Suppose the weights X of the pots are independent $\mathcal{N}(\mu, \sigma^2)$ variables.

a) Compute the sample average and variance, and hence give estimates of μ and σ^2.

b) Give 95% and 99% confidence intervals for μ.

Exercise 109 A firm wants to hire an IT engineer but hasn't yet decided what annual salary to offer. Not wanting to offer a salary too far removed from the average salaries offered by other firms, the recruiter decides to ask 1000 engineers their annual salaries. The survey reveals that the average salary of the 1000 engineers is 48000 CHF with standard deviation of 12000 CHF.

a) Give a 90% confidence interval for the average salary.

b) Is it reasonable to say that the average salary is roughly 50000 CHF?

Exercise 110 A company analyses the length of phone conversations of its employees. Nine consecutive conversation had lengths (minutes):

$$10.3, \quad 9.4, \quad 9.9, \quad 7.5, \quad 11.7, \quad 3.4, \quad 7.8, \quad 11.0, \quad 10.0.$$

a) Calculate the average $\overline{x}$ and the sample variance s^2. What are the units of $\overline{x}$ and s^2?

b) If we fit a normal distribution to these observations, what is the probability that a phone conversation lasts longer than 10 minutes?

c) Test at level 5% whether the average length of phone conversations equals eight minutes against the alternative that it is less than eight minutes.

Exercise 111 Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \exp(\lambda)$. Find the most powerful test of the null hypothesis $H_0 : \lambda = 1$ against the alternative $H_1 : \lambda > 1$.

Chapter 9

Exercise 112 Consider a disk $\mathcal{D}$ of known centre O and unknown radius $\tau > 0$.

a) We choose a point x uniformly in $\mathcal{D}$, i.e., if A is a subset of $\mathcal{D}$ of area $s(A)$, then $\Pr\{x \in A\} = s(A)/(\pi\tau^2)$. Let R be the distance from x to O. For $r \geq 0$, find $\Pr\{R \leq r\}$ and hence obtain the density of the random variable R.

b) Now we choose n points independently and uniformly in $\mathcal{D}$, and let $R_1, \ldots, R_n$ denote their distances from O. Give the maximum likelihood estimator of τ as a function of $R_1, \ldots, R_n$. Calculate the bias of this estimator and find an unbiased estimator of τ.

Exercise 113 The monthly household savings X (thousands of maravedis) of a faraway country have density function

$$f(x) = k^2 x e^{-kx}, \quad x > 0$$

where the parameter k is unknown. A sample of 400 households has average savings $\overline{x} = 2$.

(a) Estimate k by maximum likelihood.

(b) Give a 95% confidence interval for k.

(c) What percentage of families save less than 1000 maravedis a month (neglect the estimation error for k)?

Chapter 10

Exercise 114 A company produces electronic components. The employee responsible for the production line states that only 5% of the components produced are defective. A quality inspector, on the other hand, believes that the proportion of defective components is 10%. In order to determine which is nearer to the truth, a sample of twenty components is taken randomly and tested, revealing 3 defective components.

In a Bayesian framework, let p denote the unknown probability that a component is defective and let X denote the number of defective components among the 20 chosen. In the absence of any further information, we assume that the defectiveness levels of 5% and 10% are equiprobable.

(a) Translate the problem into probability terms concerning p and X.

(b) Calculate the marginal probability that $X = 3$.

(c) Calculate the posterior mass function of p and interpret the result.

(d) Calculate the posterior mean and variance of p.

(e) Discuss.

Exercise 115 Bill is trying to debug an operating system. He believes that the number of bugs in the code has a geometric distribution with parameter θ, and that he can find each bug independently with probability θ. Past experience suggests that θ has a $U(0, 1)$ distribution. If he has already found x bugs, what is the probability that m bugs remain? Find the total expected number of bugs if $x = 0$.

Exercise 116 The time T between two successive requests to a server follows an exponential distribution of parameter $\theta > 0$ and density

$$f(t \mid \theta) = \theta e^{-\theta t}, \quad t > 0.$$

Suppose that the parameter θ is treated as stochastic with an exponential distribution of known parameter $\lambda > 0$, of density

$$g(\theta) = \lambda e^{-\lambda\theta}, \quad \theta > 0.$$

Thus the density $g(\theta)$ represents the prior density on the parameter θ.

(a) Show that the posterior density of θ is proportional to $\theta e^{-\theta(t+\lambda)}$.

(b) We observe a sample of n times $t_1, \ldots, t_n$. Find the maximum a posteriori (MAP) estimate of θ.