Probability and Statistics for SIC
Problems

Chapter 1

Problem 1 How many numbers with three distinct digits can be made from the digits 2, 3, 5, 6, 7, 9? How many are (a) less than 500? (b) odd? (c) even? (d) multiples of 5?

Problem 2 How many ways are there to place four rooks (castles) on an ordinary 8×8 chessboard so that no rook threatens another? (One rook threatens another if they share a row or a column.)

Problem 3 In order to better organise the many files associated with a piece of software (data files, code, etc.), we use a new system to name them, in which a filename consists of an identifier and an extension linked by a dot. A file identifier comprises 1 to 5 characters of which the first is a letter; the subsequent characters are either letters or numbers. An extension comprises 1 to 3 characters, the first being a letter, followed by either letters or numbers. In order not to have names that are too unpronounceable, we only use the twenty letters from a to t, all in lowercase.

(a) How many file identifiers are possible?
(b) How many file names are possible?
(c) Files with extensions starting in "d" are data files. How many ways to name data files are there?

Problem 4 A password has six letters, whose case is ignored. How many passwords exist? How many passwords contain only distinct letters?

Problem 5 Four electronic components of type A, three components of type B and five components of type C must be mounted one after another. The only constraint to the system's functioning is that components of the same type must be grouped. How many arrangements can be imagined?

Problem 6 We want to solder 4 electronic components of type A and 4 of type B onto a printed circuit. The printed circuit has only 8 free spaces, which lie in a line. How many ways are there to proceed if:

(a) there are no other restrictions?
(b) the type A components must stay together, and similarly for the type B components?
(c) only the type A components must stay together?
(d) type A and B components must alternate?

Problem 7 Fourteen distinct computers must be distributed in two offices, each of seven people. How many different distributions are possible between the two offices:

(a) supposing that the distribution within each office is important?
(b) supposing that the distribution within each office is not important?
(c) how does the solution in (b) change if we must distribute twenty-one distinct computers in three offices, each of seven people?

Problem 8 A football team of 11 players is to be chosen from 20 candidates. If each player could play in any position, and position matters, how many different teams could be created? If 17 are field players and 3 are goalkeepers, how many distinct teams can be created?

Problem 9 Six servers are connected by a communication line. At any moment, any given server is either ready to transmit, or already occupied.

(a) How many states does the system have?
(b) How many ways are there for exactly 3 servers to be ready to transmit at a given moment?

Problem 10 Raymond has forgotten his email password. He can however remember that it contains r occurrences of the letter "c", of which one is in the final position. The other symbols can only be either "a" or "b". How many passwords of length n could he have?

Problem 11 If $n \geq 3$, how many diagonals has an n-sided polygon? Does there exist a polygon with the same number of sides and diagonals?

Chapter 2

Problem 12 A quality test is performed every hour on computer chips. At each test, the chip tested is classified according to how well it functions. The test stops when two consecutive chips are defective, or when four chips have been tested. Give a sample space for this experiment.

Problem 13 In a chip factory, the three last chips produced are taken every hour and their states are noted, in order. Each chip is classified as either in working order, coded 1, or defective, coded 0.

a) Describe a sample space Ω for this experiment.

b) Describe the following events in terms of the elements of Ω:

- i) the first chip is defective,
- ii) the last chip is in working order,
- iii) exactly two chips are defective,
- iv) at least two chips are in working order.

Problem 14 In the game of lotto six distinct numbers are chosen randomly among the numbers $1, \ldots, 49$. A lotto player must play 6 distinct numbers (forget the notion of "the complementary number"). The jackpot is won if the 6 numbers played correspond to the 6 numbers chosen.

a) What is the sample space linked to the draws?

b) What is the probability of winning the jackpot?

c) The society that organises the lotto is envisaging going from 49 numbers to 60, with a draw of 8 numbers rather than 6. Would winning the jackpot be more likely?

Problem 15 Sophie wants to bet on a horse race but has no idea of the performances of either jockeys or horses. She decides to bet at random.

a) Is Sophie more likely to correctly guess the first three horses (in order) of a race with 17 horses, or the first four horses (in order) of a race of 18 horses?

b) Like a), but the order doesn't matter.

Problem 16 n resistors including A and B are installed in series.

a) Describe the sample space linked to this experiment, and give its number of elements.

b) What is the probability that there are k resistors between A and B $(0 \leq k \leq n-2)$?

c) Check your result for $n = 3$ and explain all possible cases.

Problem 17 The October 13 2010 edition of the Daily Mail reported that the Allali family had just had their third child and that, oddly, all three children were born on the same day (October 7, 2005, 2007 and 2010). The journalist stated that there was just one chance in 48 million that three children from the same family (excluding twins and triplets) should have the same birthday. Do you agree?

Problem 18 Four bits are transmitted through a channel. When they are received, their states are examined, in the order in which they are received. Each bit is either correctly received, or received with errors of transmission. Let d denote a correctly-received bit and e an incorrectly-received one.

(a) What is the sample space Ω for this random experiment?

(b) List the events with at most one bit received with error.

(c) Count the events so that more than half of the bits are incorrectly transmitted.

Problem 19 Use the inclusion-exclusion formula to compute the probability that the decimal expression for a whole number chosen at random between 0 and 9999 will contain the digit 1 at least once.

Problem 20 We choose two numbers between 0 and 9 at random. What is the probability that the first number is 6, given that their sum is 6?

Problem 21 Alan, Boris and Charles have forgotten to write their names on their solutions to a test. The professor knows who these students are, but doesn't know whose copy is whose. She thinks, however, that Alan has an 80% chance of passing the exam, Boris has 70% and Charles has 60%. After correcting the three copies, she discovers that two have passed and one has failed.

Supposing that the students worked independently, what is the probability that Charles failed?

Problem 22 A request is equally likely to be sent on six servers, four of type 1, and two of type 2. Two consecutive requests cannot be sent on the same server. Consider two consecutive requests.

a) Describe the sample space Ω of this random experiment.

b) Let us define the events B_1 to be 'the first request is sent on a server of type 1 and B_2 to be 'the second request is sent on a server of type 1'.

i) Calculate the probabilities $\Pr(B_1)$, $\Pr(B_2 \mid B_1)$, $\Pr(B_2 \mid B_1^c)$ and $\Pr(B_2)$.

ii) Check that $\Pr(B_1 \cap B_2) + \Pr(B_1^c \cap B_2) + \Pr(B_1 \cap B_2^c) + \Pr(B_1^c \cap B_2^c) = 1$.

Problem 23 A packet sent over the internet from server S_1 to server S_2 passes through two intermediate routers R_1 and R_2. The probability that the packet will get lost on each segment of the path $S_1 \to R_1 \to R_2 \to S_2$ is p. If the packet does not arrive at S_2, what is the probability that it was lost at S_1? at R_1? at R_2?

Problem 24 You are invited to play in a game show, with the following rules: three cards marked with the numbers 1, 2, 3 are placed in three identical unmarked envelopes, which are then shuffled. You then choose an envelope and open it. You must then answer the number of questions written on the card inside the envelope. If you answer $r > 0$ of the questions correctly, you win $\$2^r \times 1000$, otherwise you win nothing. You answer each question correctly with probability 0.5, independently of the other questions.

(a) Give the sample space for this random experiment, and give the probabilities of the different outcomes.

(b) Find the probability that you win nothing.

(c) Find the probability that you win \$2000.

(d) Given that you have won at least \$1000, give the probability of winning at least \$4000.

Problem 25 The criminal organization SPECTRE must elect a committee of s people from its assembly of n people, and then choose a president from the committee. They are unable to agree on the membership of the committee, so after several hours of debate they decide to have a random draw, both to choose the s committee members and the president. Ernst Blofeld is a member of the assembly but is not the president. What is the probability that he is, however, a member of the committee?

Problem 26 When 100 computer programs are tested for various errors, 20 have syntax errors, 10 have input/output errors that are not syntax errors, 5 have other types of errors, 6 have both syntax and input/output errors, 3 have syntax and other errors and only 1 has all three types of error. A program is selected randomly among these 100 programs.

a) Express in the form of probabilities the quantities given in the wording.

b) What is the probability that the selected program will have syntax errors or input/output errors or both?

c) We know that the selected program has a syntax error. What is the probability that it will also have an input/output error?

Problem 27 Requests are made on servers via five communication lines. These are not used equally: the percentages of requests sent on lines $1, \dots, 5$ are respectively 20, 30, 10, 15, 25. Furthermore the line on which a request is sent depends on the length of the request: the probability that a request exceeds 100 characters on lines $1, \dots, 5$ is respectively 0.4, 0.5, 0.6, 0.2, 0.8. What is the probability that a given request exceeds 100 characters?

Problem 28 You have just installed a system to detect junk email. The system can identify junk messages in 99% of cases. Nevertheless, the system says that a message is junk when it isn't in 2% of cases. Given that 10% of received emails are junk, what is the probability that a message truly is junk when the system says it is?

Problem 29 We develop an algorithm to detect if a web page written in English was written by a native speaker, i.e., someone whose mother-tongue is English. We use it and decide that 55% of web pages written in English were written by native speakers. The algorithm can correctly detect composition of a page by a native speaker in 95% of cases when it was written by a native speaker, but incorrectly says that a page was written by a native speaker, when in fact it wasn't, with probability 1%. What is the probability that a page was written by a native speaker when the algorithm says so?

Problem 30 A student attempts a multiple choice test; for each question, five answers are suggested, of which only one is correct. If the student knows the right answer, she chooses it, but otherwise she chooses an answer at random from the five. She knows the right answer to 70% of the questions.

a) What probability does she have of correctly answering any given question?

b) If she has answered a question correctly, what is the probability that she did so knowingly (that is, not at random)?

Problem 31 You are private secretary to the Minister for Health. A disease is present in the population, in a proportion of one sick person in 10,000. The head of a large pharmaceutical laboratory comes to praise their new screening test to you: if a person is infected, the test is positive 99% of the time. If a person is not infected, the test is positive 0.1% of the time. Should you advise the minister to authorise marketing of the test?

Problem 32 A router receives most of the packets that it forwards from two other routers R_1 and R_2, in equal proportion. Roughly one packet in 200 is damaged when they come from R_1, and only one packet on 1000 is damaged when they come from R_2. An undamaged packet is received from R_1 or R_2. What is the probability that it came from R_1?

Problem 33 Three terminals are linked to a computer via two communication lines. Terminal 1 has its own line whereas terminals 2 and 3 share a line, so that at each instant only one of these two terminals can be used. During a working day, terminal 1 is used for 30 minutes of each hour, terminal 2 is used for 10 minutes of each hour and terminal 3 is used for 5 minutes of each hour. Supposing that the communication lines work independently, what is the probability that at least one terminal is in use at any given moment?

Problem 34 A hard disk has a 1% probability of failing. It has two backups, each with a 2% probability of failing. Information is lost only when all three components have failed. Supposing that they work independently of each other, what is the probability that the information will be saved?

Problem 35 Bits are transmitted independently through a channel, each equalling 1 with probability p.

a) We record two consecutive bits b_1 and b_2. (i) What is the probability that both bits have different values? (ii) What is the probability that the second bit is equal to 1 if the first is equal to 0?

b) Now we record four consecutive bits. (i) Find the probability of having at least two bits equal to 1 among the four. (ii) Find the probability that at least two bits equal 1 and the fourth bit equals 0.

Problem 36 An electronics shop has noticed that the probability p_n that n clients come into the shop on a given day is $p^n(1-p)$, $n = 0, 1, \ldots$. Furthermore, when a client visits the shop, he leaves with a purchase two times out of three. The client, however, may not be satisfied with this purchase and, one time out of four, returns it to the shop. Suppose that the fact of buying something and the fact of being satisfied with it are independent.

a) What is the probability that a client buys a product and is satisfied?

b) If we know that that k products sold have not been returned to the shop, show that the conditional probability that the shop has been visited by n clients equals

$$C_n^k p^{n-k}(2-p)^{k+1}/2^{n+1}, \quad n = k, k+1, \ldots$$

Chapter 3

Problem 37 The components of the 2×2 matrix X are independent Bernoulli variables with success probability p. Find the probability that $|X| \neq 0$. For what value of p is this probability maximised?

Problem 38 The probability that a component's lifetime exceeds 10 000 hours is 80% and failures of different components are independent.

a) With three components, we undertake "serial" assembly, so if one of the components fails, the whole system fails. What is the probability that the system will function for at least 10,000 hours?

b) With the same components as above, we undertake a "parallel" assembly: the system fails only when all three components fail.

i) What is the probability that the system will function for at least 10,000 hours?

ii) Given that the system is still working after 10,000 hours, what is the probability that at least one component is defective?

Problem 39 For the new year, an internet provider promotes a special subscription offer to attract new customers. Each new customer receives the special offer with a probability of 0.2. Twenty new customers sign up to this provider. What distribution models the number of them who benefit from the special offer? Calculate the probability that at least a quarter of them benefit.

Problem 40 To study a region's soils, tests are performed on samples taken at a distant site. Five samples must be tested, and the probability that a sample is inadequate is 0.2. If we require a probability of at least 0.99 of having at least five adequate samples, how many samples must be taken?

Problem 41 A caretaker goes home in the evening. He has n keys, only one of which opens the door to his home, but he can't remember which one.

a) He tries the keys one after another, eliminating each wrong key as he goes. What is the probability that the kth key he tries opens the door ?

b) After a party he tries to open the door but forgets to eliminate wrong keys after testing them. Give the distribution of the number of attempts needed to find the right key. What is the probability that the door will open on the kth attempt?

Problem 42 A request sent on a network will be answered with probability p, but if no answer is received, the request is resent until it is answered. Give the probability mass function of the number of requests sent until an answer is received.

Problem 43 In the game of craps, two fair six-sided dice are rolled and their scores added. If the sum of the scores equals 2, 3, or 12 the roller loses, if this sum equals 7 or 11 the roller wins, and if this sum equals any other number, the dice are rolled again. What is the distribution of the total number of rolls X until the game ends (i.e., the roller wins or loses)? If we observe $X = 3$, what is the probability that the roller lost?

Problem 44 At a given time t, n servers are trying to send a message each (i.e., a total of n messages for the n servers) via n independent transmission channels. These channels are shared by other servers, and there is a probability p that a channel is already occupied at time t. When a server tries to send an message by an occupied channel, sending fails and the message is put on hold until time $t + 1$, when the server tries sending it again, and so on until the n messages have been sent.

a) Consider a particular message: (i) What is the distribution of the number of time steps necessary to send it? (ii) What is the probability that fewer than k attempts are necessary?

b) Consider the n messages: (i) What is the probability that fewer than k time steps will be necessary to send them? (ii) Deduce the probability that exactly k time steps will be necessary.

Problem 45 The production of a particular component gives 5% of dud components. We need 10 good components for 10 new computers. The components are tested until 10 good ones are found.

a) Give the distribution of the number X of components tested before 10 good ones are found.

b) Give the distribution of the number Y of good components among 12 tested.

c) Calculate in two ways the probability that more than 12 components must be tested in order to find 10 that work.

Problem 46 Code programmed by an EPFL student contains five lines with bugs, uniformly distributed. The code contains 10,000 lines of which 3000 are comments. A particular run of the software goes through one-third of the runnable lines. What is the probability that the program will run correctly?

Problem 47 A company manager wants to build new unloading bays in order to avoid delivery lorries having to wait. Currently, 5 bays are working. Suppose it takes a full day to unload a lorry and that the number of lorries arriving each day is approximately a Poisson variable of parameter 4.

a) What is the probability of having no lorries waiting?

b) How many unloading bays would be needed to raise this probability to over 0.9?

Problem 48 Consider the following game of dice: to participate in the game, you bet 3 CHF. Then you throw the die and recover 15 CHF if the number obtained is 6. Is the game profitable, on average? If not, how much would you have to win from a 6 for it to be profitable?

Problem 49 Here is a simple algorithm to find the largest element, M, in a list of $n \geq 1$ elements:

1. Let M equal the first element in the list.

2. Go through the list, from the second element to the last element, and replace M by this element if the latter is larger.

During treatment of a list of length n:

a) What is the number of comparisons made?

b) What is the number of assignments to M in the most favourable case? What about the least favourable case?

Suppose that the elements of the list are pairwise distinct and that the $n!$ possible permutations of the list are equiprobable. Let $p(n, k)$ be the probability that k assignments will be necessary during a run of the algorithm.

c) Show that

$$p(n,k) = \frac{1}{n}p(n-1,k-1) + \frac{n-1}{n}p(n-1,k), \quad n = 2,3,\ldots,k =,2,\ldots.$$

Hint: Reason according to whether the nth element is maximal or not.

d) Let E_n be the expected number of assignments made during the treatment of a list of size n. Deduce from the above formula that $E_n = E_{n-1} + 1/n$ and give an approximation to E_n when $n \to +\infty$.

Problem 50 A gambler spins a roulette wheel. At each attempt, he has probability $1-p$ of landing on a space that will earn him 1 CHF and probability p of landing on a space that will cause him to lose everything and eliminate him from the game. He starts the game with 0 CHF and desires to obtain G_0 CHF, after which he will stop playing—if he has not already been forced to. The successive results obtained by spinning the wheel are supposed to be independent.

a) What is the probability that he will reach G_0 CHF?

b) Compute the expected final winnings.

Problem 51 The advanced version of a console game is released. 60% of players have passed all the levels of the previous version. 30% of them decide to buy the advanced version, whereas only 10% of players who haven't passed all the levels decide to buy it.

a) What is the probability that a player buys the advanced version of the game?

b) Of 10 players, what is the average number buying the advanced version?

c) What is the probability that at least three of the 10 players buy it?

Problem 52 A linear algebra exam has 20 multiple choice questions, each with four possible choices of which only one is correct. Ten questions must be answered correctly to pass the exam. Albert started revising a bit late, so he has to answer the questions at random. Find the expectation and variance of the total points he will obtain. What is the probability that he will pass? Once the exam is over, he chats with his friends, and they work out that he correctly answered six questions, and wrongly answered four others. What is the expectation of the total points he obtained, conditioned on this new information? What is then the probability that he will pass?

Problem 53 n servers are connected via a communication line. Every minute a server receives a request on this line with probability p. Furthermore, a server can't receive more than one request per minute.

a) What is the distribution of the number of requests received in a minute by the n servers?

b) Find the expectation and the variance of the proportion of requests received by each server in a minute.

Problem 54 Consider a class made up of $N \geq 60$ students, for N even, who are randomly grouped into pairs. We are interested in the $N/2$ distinct "couples" thus formed.

a) Give the distribution of the number of "couples" having their birthday on the same day, that is to say the number of couples of two people born on the same day of the year (but not necessarily the same year). Suppose that a year has 365 days.

b) Calculate (in two ways), as a function of N, the probability that at least one couple has their birthdays on the same day.

NB: use Poisson approximation (and justify it).

c) What must be the minimal number of students in the class for the probability in b) to exceed 60%? Calculate it in two ways.

Problem 55 The use of character recognition software has yielded 1500 randomly distributed errors in 1000 pages of text.

a) Give an exact value of the probability that a page thus transcribed contains fewer than 2 errors.

b) Calculate the probability of this event using a Poisson approximation.

Problem 56 A server serves 1000 independent workstations via 50 high output lines. At peak times, each workstations is busy for 2.5 seconds every minute, on average. What is the probability of network saturation for a minute at peak time?

Problem 57 To win the Euromillions lottery, you must correctly select five distinct numbers in the range $1, \ldots, 50$ and two distinct stars labelled $1, \ldots, 11$. All combinations of numbers and stars are equally likely, the order of drawing does not matter, and the outcomes of different draws are independent.

(a) What is the probability of winning the lottery on the next draw?

(b) If 80 million people play Euromillions independently on each draw, give the expected number of winners at the next draw.

(c) Give approximate probabilities that there will be no winner, one winner, and more than one winner at the next draw.

(d) What is the distribution of the number of draws until the next winner?

Problem 58 An electrician buys components in packets of 10. He examines three components picked at random from one packet, and he only accepts it if the three he picks are perfect. If 30% of the packets contain 4 bad components and the other 70% contain just one bad component, what proportion of packets will he reject?

Problem 59 A pipe-smoking mathematician has a matchbox in each pocket of his jacket, one on the right side and one on the left. Both boxes contain m matches initially. Every time he lights his pipe, he picks a matchbox at random, and throws away the used match. After a while he notices that the matchbox he has picked is empty. What then is the distribution of the number of matches in the other box? (Hint: you may find it helpful to first consider the special cases where the other box has m matches and zero matches.)

Chapter 4

Problem 60 The number of kilometers a car battery lasts before it fails is exponentially distributed with mean 10000 kilometers. A person wants to embark on a 5000-kilometer journey. What is the probability of reaching their destination without a battery failure?

Problem 61 Car components are often copied and sold as original pieces. We want to replace some components. Assume a probability $1/4$ of buying a pirated component, and a probability $3/4$ of buying an original component. The lifespan of a component is an exponential random variable, with parameter 5 for a pirated component, and of parameter 2 for an original component. Let T denote the lifespan of the component that has been bought, and that the component lasts until time t after being installed. What is the probability $\pi(t)$ that it is a pirated component? Find the limit of $\pi(t)$ when $t \to \infty$.

Problem 62 The average waiting time between two successive printer jobs is 15 minutes.

a) Which distribution should one use for the waiting time (hours) between two jobs?

b) What is the probability that the next job will be sent within the next 5 minutes?

Problem 63 In finance the level of a risk of a portfolio is often measured by the value-at-risk, i.e., the loss that is exceeded with probability α, or by the expected shortfall, which is the expected loss, conditional on the loss exceeding the value at risk. If the loss distribution is exponential with mean μ, find the value-at-risk and the expected shortfall.

Problem 64 We have two electronic components, the lifespans of which, given in years, are modelled by independent random variables X and Y. Suppose that the average lifespan of the first component is one year, but that the second component has an average lifespan of 6 months.

a) Give the distributions of X and Y.

b) Calculate their distribution functions F_X and F_Y.

Two assemblies are considered possible for these components: in series or in parallel. The series system breaks down as soon as one of the components fails. The parallel system breaks down when both components fail. Let S denote the lifespan of the series system and T that of the parallel system.

c) Calculate the distribution function F_T of T.

d) Calculate the distribution function F_S of S.

e) Calculate the probability that the component with lifespan X is still working at time x ($x > 0$), given that the series system has already failed.

f) Given that the component with lifespan X failed before time x, what is the probability that the parallel system is still working at time $t > x$?

Problem 65 A chip's lifetime has an average of $\mu = 12$ years and a standard deviation of $\sigma = 4$ years.

a) Specify a suitable distribution for the lifetime, and give the values of its parameters.

b) Calculate the probability that the lifetime will lie between 10 and 15 years.

Problem 66 A software download may be made from two possible mirrors, one in Europe and one in the United States. According to the state of the network, a download from one of the two mirrors takes between a and b seconds, any period between these two values being equiprobable. The times of downloads from the two mirrors are independent.

a) Give the distribution function of the download time for one of the two mirrors.

b) Adrian tries to download simultaneously from both mirrors.

i) What is the distribution of the time for the first download to finish?

ii) Calculate its expectation and its variance.

Problem 67 The two-parameter Pareto distribution function is

$$F(x) = \begin{cases} 1 - (\alpha/x)^{\theta}, & x > \alpha, \\ 0, & x \le \alpha, \end{cases} \qquad \theta, \alpha > 0.$$

a) Check that this is a distribution function.

b) Calculate the first and second moments of the corresponding random variable, X. Under what conditions are they finite?

c) The Pareto distribution is often used to model rare events (big stock market losses, heavy rainfall, ...). Find the distribution function of X, conditional on the event $X > c$, where $c > \alpha$, and find $\mathrm{E}(X \mid X > c)$. Comment.

Problem 68 Two large jobs are submitted simultaneously to a computer with two cores. The run times for the jobs are independent random variables uniformly distributed from zero to three hours. Let X denote the longest run time.

(a) Show that the probability density function of X is

$$f(x) = \begin{cases} 2x/9, & 0 < x < 3, \\ 0, & \text{elsewhere.} \end{cases}$$

(b) Find the expected waiting time until both processors are free.

(c) When I check after two hours, both jobs are done. What is the probability that the longest run time was less than one hour?

(d) The computer is still running after two hours. At that time, what is the expected remaining waiting time until both cores are free?

Problem 69 The stamps produced by the Post Office should be square with sides of length $l = 1.5$ cm. The machine making them is badly calibrated, however, so although the stamps are square, their sides may not be of length l. Suppose that the production error is ± 0.2 cm on each side, and that all values in this range are equiprobable.

a) By what distribution can we model the lengths? Give its distribution function.

b) What is the distribution function of a stamp's area?

Problem 70 If $X \sim \exp(\lambda)$, find the distribution function of $Y = 1/X$, and obtain $\mathrm{E}(Y)$.

Problem 71 Suppose that the average winter temperature in Rome follows a normal distribution with mean 10°C and standard deviation 4°C. A state of extreme cold is decreed if the average winter temperature is lower than 4°C. Suppose that temperatures from one winter to another are independent.

a) What is the probability that a state of extreme cold will be decreed in any given winter?
b) What is the probability that the state of extreme cold is not decreed at all in 10 years?

Problem 72 A survey conducted shows that 40% of customers will make a purchase once they are inside a specific shop. Let X denote the number of customers who buy something, out of 1000 in the shop.

a) What is the mass function of X? Calculate its expectation and its variance.
b) Calculate the probability that fewer than 300 people have bought anything.
c) Calculate the probability that more than 500 people have bought anything.

Problem 73 A server receives 1000 requests per minute, but due to communication problems the probability of a request failing is 4%. Suppose that the failure of different requests is independent.

(a) What is the exact probability that more than 30 requests fail within a minute?
(b) Give two approximations to the probability (a), and justify them.

Problem 74 Each weekday morning I take the bus at 7.45 in order to catch the train at 8.00. The bus journey should take 10 minutes, but on one weekday each week, on average, it takes an additional X minutes, where $X \sim N(3, 1)$. I miss my train if the bus is more than 4 minutes late.

(a) What is the probability that I will catch the train tomorrow?

(b) If the bus is already 3 minutes late, what is the probability that I will miss the train?

(c) I caught the train this morning. What is the probability that the bus was late?

(d) At what time do I expect to arrive at the station?

Chapter 5

Problem 75 If B and C are independent and uniformly distributed on $(0, 1)$, what is the probability that the equation $x^2 + Bx + C = 0$ has two real roots?

Problem 76 If you enjoyed the previous problem, here's another: if A, B and C are independent and uniformly distributed on $(0, 1)$, what is the probability that the equation $Ax^2 + Bx + C = 0$ has two real roots? (First compute the densities of B^2 and AC.)

Problem 77 Suppose that the weather in a city is measured by the sunshine quotient, i.e., the proportion of hours of sunshine to the total hours of daylight on a given day. Research has shown that in many cities, the sunshine quotients on two successive summer days, X and Y, have joint probability density function

$$f_{X,Y}(x, y) = c(1 + 4\theta xy), \quad 0 \le x, y \le 1, \quad \theta \ge 0,$$

where the parameter θ varies from city to city and the positive quantity $c \equiv c(\theta)$ depends on θ.

(a) Find the marginal probability density functions of X and Y.

(b) For what value(s) of θ are the sunshine quotients on two successive days independent?

(c) Give the expected sunshine quotient for London, where $\theta = 1$.

(d) Given that the sunshine quotient today in London equalled $x = 1$, give the conditional density of the sunshine quotient there tomorrow.

Problem 78 The random variables X and Y are discrete with joint probability function:

	Y		
X	0	1	2
0	p	$2p$	p
1	$3p$	$2p$	p

What is p? Calculate the marginal probability function $f_X(x)$ of X. Calculate $\mathrm{E}(Y^2)$ and $\mathrm{E}(XY)$.

Problem 79 In a game of heads or tails, the result is $X = 0$ for tails and $X = 1$ for heads. If the coin falls on heads, a fair die is thrown, and the result $Y \in \{1, \dots, 6\}$ is obtained. Let $Z = X + Y$ denote the total score. Calculate the conditional expectation and variance of Z given X and hence check Theorem 168.

Problem 80 The download time in hours for a software upgrade is random with probability density function $10 \exp(-10x)$, where $x > 0$.

(a) Find the probability that the download time exceeds t hours, and give the conditional probability that it exceeds $t + x$ hours, given that it exceeds t hours. Comment.

Frustrated after waiting 10 minutes, I decide to try downloading from another server, without stopping the first download. The joint probability density function of the remaining time for the first download, X, and the time for the new download, Y, may be written in the form

$$f_{X,Y}(x, y) = \begin{cases} c \exp(-10x - 20y), & x, y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(b) Give the marginal density of Y. Are X and Y independent?

(c) Find the probability that $Y < X$.

Problem 81 A lake is fed by two rivers, and their daily flows X and Y (millions of litres) into the lake have joint probability distribution

$$\Pr(X \le x, Y \le y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}), & x, y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the joint probability density function of X and Y. Are these flows independent?

(b) Find the marginal density of X, and give the probabilities that X exceeds 3 million litres, and that both X and Y exceed 3 million litres.

(c) An alert is sounded if the total inflow in a given day, $X + Y$, exceeds 6 million litres. Find the probability that no alert is sounded today. How does this compare to $\Pr(X > 3, Y > 3)$? Explain.

Problem 82 At lunchtime I shall have to queue at the post office and then at the Epicure to buy a sandwich. My past experience is that the time (in minutes) spent at the post office will have a gamma distribution with shape parameter $\alpha = 5$ and scale parameter $\lambda = 1$, and the time spent buying a sandwich will have a gamma distribution with shape parameter $\beta = 10$ and $\lambda = 1$.

(a) What is the mean time I shall spend queueing, and what is its variance? Give an approximate probability that I shall spend a total of less than 20 minutes queueing.

(b) Find the joint distribution of the total time spent, and the proportion of this time that I shall spend at the post office. Comment.

Problem 83 A file-sharing website has a very large number of files, whose sizes S (GB) and download times T (min) have joint probability density function

$$f_{S,T}(s, t) = \frac{c}{(1 + s)^4} \exp\{-t/(1 + s)\}, \quad s, t > 0.$$

a) Give the probability density function of the size of a file selected at random. Is download time independent of file size?

b) Give the mean file size.

c) Give the conditional density function of the download time for a file of size 1 GB.

d) What is the expected time to download a file of size 2 GB?

e) Give the expected download time for files from the site.

Problem 84 When you arrive at the post office, there are 8 people already waiting to be served. There is just one queue, but there are three servers, and the service time for each customer is exponential with mean 3 minutes. Find the distribution of the time until you are served, and give its mean and variance.

(Bonus) After 5 minutes one server leaves when her customer finishes. If you are still waiting in the queue, by how much does your mean waiting time increase?

Problem 85 Once night has fallen, the number N of mosquitos in my hotel room follows a Poisson distribution of mean λ, and each of them bites me independently with probability p. Find the conditional mean and variance of the number Z of bites I will have in the morning, given that there were $N = n$ mosquitos. Using these results, find the unconditional mean and variance of Z. Use the conditional expectations to calculate the moment generating function of Z. What is the distribution of Z?

Problem 86 A computation requires n functions, whose calculation times are independent $\mathcal{N}(\mu_i, \sigma_i^2)$ variables ($i = 1, \dots, n$). Suppose that the program calling these functions adds a (fixed) time b to the total calculation time. What is the distribution of the total calculation time?

Problem 87 Adele is waiting for the lift, which has a maximum capacity of 750 kg. The lift arrives with ten adults already in it. Suppose that the weight of an adult is distributed as a normal variable of mean 70 kg and standard deviation 12 kg.

a) What is the distribution of the weight of the 10 people in the lift?

b) Is it reasonable for Adele, 55 kg, to get into the lift?

Problem 88 Civil engineers who took part in the construction of the Millau bridge estimate that the weight W (tonnes) that the bridge's span can bear without its structure becoming damaged follows a normal distribution of mean 400 tonnes and of standard deviation 40 tonnes. Suppose that the weight of a car is a normal random variable of mean 3 tonnes and standard deviation 0.3 tonnes. How many cars would have to be on the span for the probability of failure to exceed 0.1?

Problem 89 A very simple model for a quantitative genetic trait, such as height, is as follows. Let X_1 denote the value for a parent, and X_2 the value for their child, and suppose that their joint distribution is bivariable normal with $\mathrm{E}(X_1) = \mathrm{E}(X_2) = \mu$, $\mathrm{var}(X_1) = \mathrm{var}(X_2) = \sigma^2$, and correlation $\rho \in [0, 1)$. Thus the distribution for the trait is the same in both generations.

(a) Show that the conditional distribution of X_2 given that $X_1 = x_1$ is normal with mean $\mu + \rho(x_1 - \mu)$ and variance $\sigma^2(1 - \rho^2)$, and hence show that a parent whose value x_1 is greater than the population average μ will have offspring whose conditional expectation for X_2 satisfies $\mu < \mathrm{E}(X_2 \mid X_1 = x_1) \leq x_1$, whereas a parent for whom $x_1 < \mu$ will have offspring for whom $x_1 < \mathrm{E}(X_2 \mid X_1 = x_1) \leq \mu$; this is 'regression towards the mean'. Discuss the relation between the marginal and conditional variances $\mathrm{var}(X_2)$ and $\mathrm{var}(X_2 \mid X_1 = x_1)$.

(b) Use the conditional mean and variance of X_2 given $X_1 = x_1$ to find the unconditional mean and variance $\mathrm{E}(X_2)$ and $\mathrm{var}(X_2)$. Comment.

Problem 90 Let T denote the average number of non-timetabled hours a student spends studying each week during his first year at EPFL, and let N denote his average mark in the first-year exams. Suppose that T and N have a bivariable normal distribution with respective means 10 and 4, respective variances 25 and 1, and covariance 4.

(a) Alphonse does only 5 hours of non-timetabled work per week on average. Give the distribution of his average mark N. What is the probability that $N < 4$?

(b) Bertil wants to achieve an average of 5. How many extra hours should he work each week, in order that his value of N has median 5?

(c) Camille wants to have a probability 0.9 of passing her first year. How many extra hours should she work each week, on average?

Problem 91 Two successive software downloads take times (minutes) $X_1 \sim \mathcal{N}(8, 3^2)$ and $X_2 \sim \mathcal{N}(16, 4^2)$. The download times are independent.

(a) What is the distribution of the total download time $T = X_1 + X_2$?

(b) What is the probability that the total download time exceeds 30 minutes?

(c) Given that $X_1 = 10$, find the probability that the total download time exceeds 30 minutes.

(d) Given that the total download time was 30 minutes, what is the probability that X_1 was less than 7 minutes?

Chapter 6

Problem 92 On average a particular software application makes 25 errors in every 1000 hours of use, with a standard deviation of 2 errors. Give an upper bound for the probability of having more than 30 errors in 1000 hours of use. What if the standard deviation is 4?

Problem 93 A server receives requests from 8 independent communication lines. On average there are 20 requests per second from each of these lines. We want to know if the total number of requests received by the server does not exceed the critical threshold of 500 requests per second.

a) Use Markov's inequality to obtain an upper bound for this probability.
b) What do you get if you use Chebyshov's inequality?

Problem 94 A hard disk has 350 GB free. Is this likely to be enough to stock 300 films of average size 1.1 GB and of standard deviation 0.4 GB?

Problem 95 A software update requires the installation of 85 files, installed sequentially. The installation time is random, but on average takes 15 seconds per file, with a standard deviation of 4 seconds. What is the probability that the software will be updated in less than 20 minutes?

Problem 96 280 independent messages are sent from a transmission centre. The messages are treated sequentially and the transmission time of each message has an exponential distribution with $\lambda = 5$ min^{-1}. What is the probability that the 280 messages will be transmitted in less than an hour?

Problem 97 N tourists want to go to Zermatt and must therefore get the train at Täsch. The train only has two wagons, each of n seats. Suppose that tourists get into the front or rear wagons independently and with equal probability.

(a) What is the probability that at least one tourist cannot get into their chosen wagon?
(b) How many seats would be needed if $N = 1000$ and if we want p to be less than 0.01?

Problem 98 Each day a company assembles 25 computers, of which 95% work correctly. Give a good approximation of the probability that at least 600 faultless computers have been assembled after 25 days.

Problem 99 The number of clients entering a shop on a given day follows a Poisson distribution of parameter 12. Using appropriate hypotheses of independence, find the probability of having at least 250 clients in 22 working days.

Problem 100 My local sandwich bar serves only soup and sandwiches.

a) The chef makes 40 litres of soup each day. Customers each independently choose one of two sizes of soup bowl, and fill it with a random amount of soup. The amount of soup put into a small bowl has expectation 300 ml and standard deviation of 30 ml, and the amount put into a large bowl has expectation 600 ml and standard deviation 60 ml. If 50 clients choose small bowls and 40 clients choose large bowls, give the distribution of the total amount of soup consumed. What is the probability that there will not be enough soup today?

b) When the bar opens, there are 100 sandwiches for sale, and after t hours each has been sold independently with probability $1 - e^{-t}$. What is the probability that no sandwiches are left, if I arrive 4 hours after opening time?

Chapter 8

Problem 101 Consider a random sample $X_1, \ldots, X_n$ with density function

$$f(x) = \begin{cases} \alpha^{-1} x^{-1/\alpha - 1}, & x \geq 1, \\ 0, & x < 1, \end{cases} \qquad \alpha > 0.$$

(a) Check that f is a probability density.
(b) Determine the maximum likelihood estimate of α.

Problem 102 Each year, the maximum height X of a river is measured. A value over 6 m would be catastrophic. The distribution of X is supposed to be Rayleigh, i.e., its density is

$$f_X(x) = (x/a)e^{-x^2/(2a)}, \quad x > 0, \quad a > 0,$$

where a is unknown. We observed the following heights (m) over 8 years:

$$2.5, \quad 2.9, \quad 1.8, \quad 0.9, \quad 1.7, \quad 2.1 \quad 2.2, \quad 2.8.$$

(a) Give the maximum likelihood estimate for a.

(b) An insurance company estimates that a catastrophe only happens on average once every thousand years. Can this be justified by the data?

(c) Compute a standard error for the estimated probability p of a catastrophe in any given year, and use it to give a confidence interval for p. Discuss.

Problem 103 An electronic component production line produces a proportion p of defective products, where p is unknown but we know that $p \leq 0.08$. We test n components.

(a) What is the maximum likelihood estimator of p?

(b) What size must n be for the mean square error for estimation to be less than or equal to 0.002?

Problem 104 Suppose the lifetime (months) of a component has the $\mathcal{N}(\mu, 9)$ distribution, and that 20 such components sampled independently have average lifetime $\overline{y} = 100.9$ months. Do we reject the hypothesis that $\mu = 100$ at level $\alpha = 1\%$?

Problem 105 Internet connections are sometimes slowed by the network's transmission latency. In order to quantify this, we send 500 packets and get an average latency of 0.8 seconds, with a sample standard deviation of 0.1 seconds. Give a 99.5% confidence interval for the expected latency.

Problem 106 A survey of 123 students showed that during the night of the 12–13 February 2015, 62 of them went to bed after 3 am. Estimate the probability that a student went to bed after 3 am that night. Give an approximate 95% confidence interval for the proportion of the population of all students at this university who went to bed after 3 am. State the assumptions underlying your interval. Do you think they are reasonable?

Problem 107 Suppose that the random variables X_1 and X_2 have means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, with $\mathrm{corr}(X_1, X_2) = \rho$.

(a) If a_1, a_2, b_1 and b_2 are constant, prove that

$$\mathrm{cov}(a_1X_1 + a_2X_2, b_1X_1 + b_2X_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_ib_j\mathrm{cov}(X_i, X_j),$$

precisely quoting the results used.

(b) Prove the statement below, with or without using (a),

$$\mathrm{var}(a_1X_1 + a_2X_2) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2,$$

precisely quoting the results used.

(c) What is the distribution of $\overline{X}_1 - \overline{X}_2$, for two independent averages $\overline{X}_1$ and $\overline{X}_2$ satisfying

$$\overline{X}_1 \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \quad \overline{X}_2 \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)?$$

What is $\mathrm{cov}(\overline{X}_1, \overline{X}_2)$?

(d) The table below gives the commute times of 10 female and 10 male students, rounded to the nearest 5 minutes:

Female student	10	40	20	10	20	45	20	5	5	5
Male student	20	25	30	25	40	20	40	20	15	40

Test if the average commute time for female and male students is identical, at level 5%.

(e) Why is it reasonable to think that the data in (d) can be considered as 'two independent samples'?

Problem 108 After comparing the thickness of paint (mm) on cars from two randomly selected production lines, an engineer has the following data:

Line A	0.23	0.13	0.28	0.19	0.22	0.34	0.28	0.34	0.17	0.22
Line B	0.11	0.23	0.32	0.06	0.16	0.21	0.15	0.20	0.27	0.19

Is there a significant difference in the paint thickness on the two lines? What assumptions have you made? Obtain a 98% confidence interval for the mean paint thickness on line A.

Problem 109 The average of a random sample of 500 server response times is 5 milliseconds and the sample standard deviation is 2 milliseconds.

(a) Give a 95% confidence interval for the underlying mean response time.

(b) If we want the confidence interval to be of length 0.3 ms at most, what size should the sample be?

Problem 110 Here are daily costs (CHF) of electricity consumption from nine computers:

$$1.2 \quad 0.8 \quad 0.6 \quad 1.1 \quad 1.2 \quad 0.9 \quad 1.5 \quad 0.9 \quad 1.0.$$

(a) Calculate the sample average and variance.

(b) Suppose that the daily costs follows a normal distribution. Do we reject the hypothesis of a mean daily cost of 1 CHF at level $\alpha = 5\%$?

Problem 111 We say that Y has a log-normal distribution of parameters (μ, σ^2) if $X = \log Y$ has an $\mathcal{N}(\mu, \sigma^2)$ distribution. We have a sample of nine log-normal observations of unknown (μ, σ^2):

$$11.3 \quad 2.1 \quad 1.1 \quad 8.9 \quad 4.6 \quad 5.7 \quad 13.5 \quad 24.5 \quad 16.4.$$

(a) Construct a 95% confidence interval for μ.

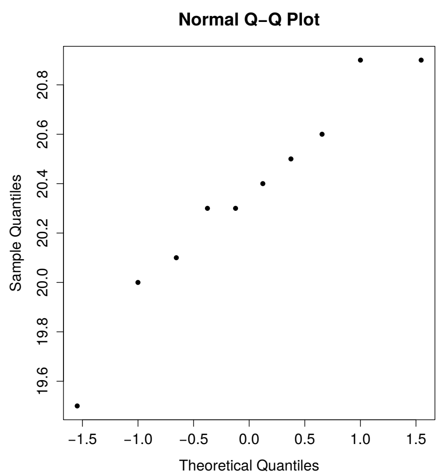
(b) Construct a 95% confidence interval for σ.

Problem 112 Now that their exams are over, Marco and Chris are each packing a suitcase before flying to Thailand for a holiday. The airline fines travellers whose suitcases weigh more than 20 kg.

(a) Marco packs his suitcase, and then weighs it using the bathroom scales. He knows that these scales are very unreliable, so he decides to weigh the suitcase 10 times, resulting in weights (kg)

$$20.0 \quad 21.0 \quad 20.8 \quad 20.7 \quad 19.6 \quad 20.9 \quad 20.0 \quad 20.7 \quad 21.0 \quad 20.8$$

He then produces this plot:



(i) Explain what such a plot is, and why it is useful.

(ii) What do you deduce in this case?

(b) Test at level 95% whether Marco's suitcase is overweight.

(c) Chris is more careful; he uses the kitchen scales to weigh individually each item he will take, including his suitcase. The total weight of these 49 objects equals 19.9 kg. The manufacturer claims that the kitchen scales have a standard deviation of 4 g for each item weighed. Test at level 99% whether Chris is likely to be fined.

Problem 113 According to US federal law, action should be taken if levels of lead in 10% or more samples of drinking water in a town exceed 15 parts per billion (ppb). The data shown in Figure 1 below are samples of tap water in Flint, Michigan, where residents complained about discoloured drinking water. Three hundred sampling kits were sent out to households chosen at random, and 271 households each returned three samples of tap water: sample 1, taken from the first water to come out of the tap; sample 2, taken after letting the water run for 15 seconds; and sample 3, taken after letting the water run for two minutes.

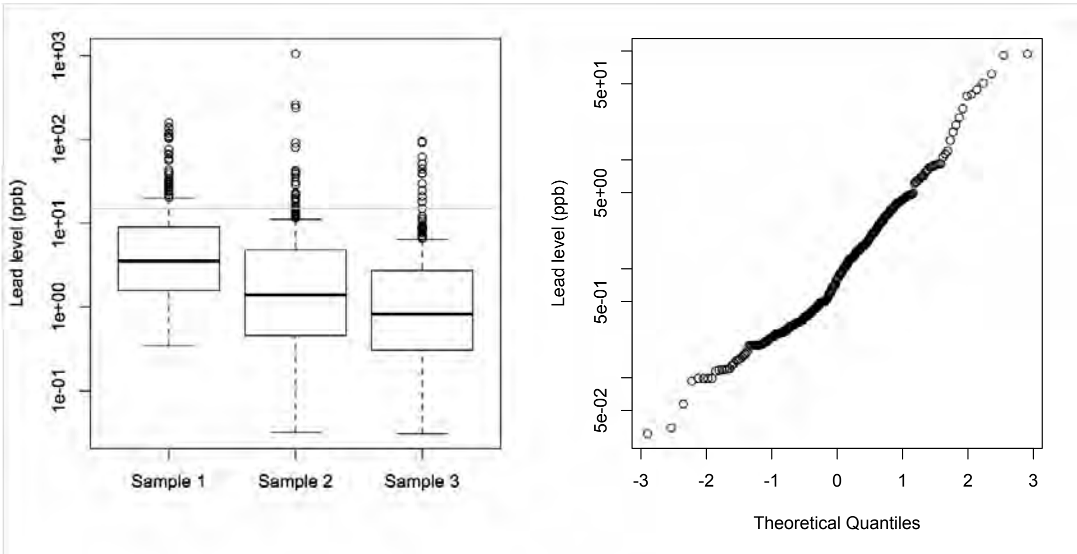


Figure 1: Flint lead data. Left: boxplots of the sampled lead levels (ppb), with horizontal line showing 15 ppb. Right: lead levels (ppb) for sample 1. Note the log scales in both panels.

(a) Consider the left panel of Figure 1.

(i) Explain how a boxplot is constructed.

(ii) What do you learn from the panel?

(b) The level 15 ppb is exceeded by 45 of the elements of sample 1, and

$$\sum_{r=45}^{271} \binom{271}{r} 0.1^r 0.9^{271-r} \approx 0.0003.$$

What does this calculation represent? Do you think that action should be taken?

(c) What do you learn from the right panel of Figure 1?

(d) Let $x_1, \ldots, x_{271}$ represent the log lead levels in sample 1. Given that their average is 1.40 and their sample variance is 1.68, give a 95% confidence interval for the mean log level of lead.

(e) On what assumptions is the interval in (d) based? Do you find them reasonable?

Problem 114 In a trial in which an individual is accused of a crime, which wrong judgement corresponds to a Type I statistical error? (Consider the null hypothesis to be the accused's innocence.)

A law student is interested in sentences passed on young adults convicted of serious knife crime in England and Wales. He randomly selected a sample of 26 cases from judicial files for the end of the year 2009; he finds that the average sentence length is 6.87 months, with standard deviation 2.2 months.

a) Construct a confidence interval at 95% for the mean length of a prison sentence.

b) What precisely are the hypotheses made in a)?

c) Suppose that all such convicts in fact served one-half of their sentences. What are the average and standard deviation of the time spent in prison for his random sample of 26 cases?

d) Test the hypothesis that the average time spent in prison is at least 5 months for such young adults, at level 1%.

Problem 115 The first four lambs have been born, and their weights $Y_1, \ldots, Y_4$ (in mina) are thought to be independent and with the $\mathcal{N}(\mu, 1)$ distribution. It is desired to test the hypothesis that $\mu = 10$ mina.

(a) Abel decides to reject the null hypothesis if $\overline{Y} > 11$ mina. Compute the size of his test.

(b) Cain decides to reject the null hypothesis if $|\overline{Y} - 10| > 0.8$ mina. What is the size of his test?

(c) Seth thinks that in fact $\mu = 10.5$ mina, and decides to use the most powerful of the tests suggested by his older brothers. Which does he use?

Problem 116 Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Pois}(\theta)$. Find the most powerful test of the hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta > \theta_0$. Show that it is not always possible to find a rejection region of size α, and suggest an approximation for the rejection region if an exact region does not exist.

Chapter 9

Problem 117 The times $T_1, \ldots, T_n$ between two successive computer failures are independent random variables with a gamma distribution of parameters $\alpha = 2$ and $\lambda = 1/\theta$, so

$$f(y; \theta) = y\theta^{-2}e^{-y/\theta}, \quad y > 0,$$

where the parameter $\theta > 0$ is unknown.

(a) Find the maximum likelihood estimate of θ based on the first n times $t_1, \ldots, t_n$ between two successive failures.

(b) Give the bias and mean square error of the maximum likelihood estimator of θ.

Problem 118 Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, with μ unknown but σ known.

(a) Determine the maximum likelihood estimator of μ. Is it biased? Calculate its variance.

(b) Calculate the observed information for μ.

Problem 119 100 electric lightbulbs chosen at random have total lifetime of 88,912 hours. The individual lifetimes can be modelled by an exponential distribution of unknown parameter λ.

(a) Give the maximum likelihood estimate of λ.

(b) Give a 90% confidence interval for λ.

(c) Do we reject the hypothesis $\lambda = 0.001$ at level 90%?

Problem 120 The number of fouls x and the number of yellow cards y distributed in group B during the 2006 World Cup were

Match	1	2	3	4	5	6
x	26	19	34	34	31	39
y	3	4	6	8	3	4

(a) Calculate the covariance and correlation of the number of fouls and the number of yellow cards.

(b) Estimate the regression line predicting the number of yellow cards as a function of the number of fouls.

(c) Predict the number of yellow cards in a match with 30 fouls.

Problem 121 Let data $x_1, \ldots, x_n$ be assumed to be a random sample from a population with the shifted exponential density function

$$f_{\theta,\gamma}(x) = \begin{cases} \theta^{-1}e^{-(x-\gamma)/\theta}, & x \geq \gamma, \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$ and $\gamma \in \mathbb{R}$. With $\gamma \geq 0$, this distribution could arise as the time in years before needing to buy a new fridge, if there was period of γ years after its installation during which it could not fail, and the subsequent time to failure was exponential with mean θ years. Thus the total time to buying the next fridge could be written as $\gamma + \theta E$ years, where $E \sim \exp(1)$ is a standard exponential variable.

Determine the maximum likelihood estimates of θ and γ based $x_1, \ldots, x_n$.

Problem 122 The number of failures of an operating system observed over a period of time t has the Poisson distribution with mean λt.

(a) Find the maximum likelihood estimate of λ and the observed information, based on data $(t_1, x_1), \ldots, (t_n, x_n)$ for independent periods of observation.

(b) Use your results from (a) to give a 90% confidence interval for λ based on the data below:

t (days)	2.5	2.0	7.0	0.5	3.0	4.0	3.0	3.0	4.0	7.0
x	0	1	3	0	5	0	0	2	3	10

(c) Suppose that in addition to the data in (a), data $y_1, \ldots, y_m$ are available representing the times to first failure for the operating system. Assuming that these times are independent, and independent of the data in (a), find the maximum likelihood estimate of λ when the sets of data are combined.

Chapter 10

Problem 123 Consider the average number of daily failures of a certain network. In a Bayesian framework we consider that this average number, θ, is treated as a random variable.

(a) Let X denote the number of network failures in a given week, and suppose that the distribution of the failures over the week is random. What distribution do you suggest for X, conditional on θ?

(b) On similar networks we have observed an average of 4 failures per week, with a standard deviation of 2 failures a week. Suggest a prior distribution for θ and give the values of its parameters.

(c) Two failures occurred on the network last week, and none this week. Give the posterior distribution of θ, and its posterior expectation and variance.

Problem 124 A telephone operator wants to update the statistics of the calls it deals with. The latest statistics give an average of 1000 calls an hour, with standard deviation 200 calls. Since the extension of the network, the average number θ of calls dealt with may have changed, and the operator wants to estimate it by Bayesian inference. A new survey reveals that, over 10 randomly chosen hours, 7265 calls were dealt with. Let X denote the number of calls during a given hour, and suppose that conditional on θ, $X \sim \text{Pois}(\theta)$:

$$\Pr(X = x \mid \theta) = \theta^x e^{-\theta}/x!, \quad x = 0, 1, \ldots, \theta > 0.$$

Also suppose that θ has a Gamma(α, λ) distribution.

(a) Based on the numbers of the first survey, what values of α and λ do you predict?

(b) Determine the posterior distribution of θ (name it and give its parameters).

(c) Give the posterior mean and the standard deviation of θ.

(d) Determine a 95% credibility interval for θ.

Problem 125 Packets sent over a communication network are independently corrupted with unknown probability θ. Let $x_1, \ldots, x_n$ be the observed values of indicator variables, with $x_j = 1$ if packet j is corrupted and $x_j = 0$ otherwise, and suppose that for similar networks θ has density function $\pi(\theta) = m(1-\theta)^{m-1}$, for $0 < \theta < 1$ and some $m > 0$.

(a) By writing the joint density of the data as $f(x_1, \ldots, x_n \mid \theta) = \theta^c(1-\theta)^{n-c}$, find the posterior density of θ given $x_1, \ldots, x_n$. What does c represent?

(b) Find the maximum a posteriori estimate of θ based on the data.

(c) Find the probability that the next packet to arrive will be corrupted, conditional on $x_1, \ldots, x_n$.

Note: For $a, b > 0$, $\int_0^1 u^{a-1}(1-u)^{b-1}\,du = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, where the Gamma function $\Gamma(a) = (a-1)!$ for $a = 1, 2, \ldots$, and $\Gamma(a+1) = a\Gamma(a)$ for $a > 0$.