Exercises

1. Lifelengths, systems and the Poisson process

Lifelengths and systems

1. If human lifelength is assumed to have the intensity function

$$\lambda(t) = 3 \cdot 10^{-3} + 6 \cdot 10^{-5} \cdot e^{t/10}, \quad t \geq 0,$$

   what is the chance that a human attains 50 years of age?

2. Assume that a lifelength variable T has the intensity

$$\lambda(t) = \frac{kt^{k-1}}{a^k}, \quad t \geq 0,$$

   where k and a are positive parameters. Find the distribution function and the density function of T.

3. Suppose that the hazard rate function, $\lambda(t)$, for a smoker of age t to contract lung cancer is modeled by the function

$$\lambda(t) = 0.027 + 0.00024(t - 40)^2, \quad t \geq 0.$$

   Under the assumption that a smoker of age 40 survives all other risks, what is the probability that the smoker will survive to a) age 50 and b) age 60, without contracting lung cancer?

4. Suppose bacteria that reproduce by cell-division are such that each individual has a lifelength T given by the intensity function $\lambda(t) = t$, per hour. When a cell dies (the failure event) it is replaced by two new cells. Starting with a single bacterium, what is the expected time until the first cell-division?

5. As part of a project to monitor the movement of polar bear a radio transmitter sourced by two independent batteries is attached to a bear. The batteries have exponential lifelengths with a mean value of 8 months. As soon as both batteries are out the transmitter silences. What is the intensity that the transmitter goes silent?

6. (Continuation of the previous example) Suppose instead that only one of the batteries is required to drive the radio transmitter and that the second battery is connected only when the first battery empties. In this case, what is the transmitter silence intensity?

7. Consider a system of two components connected in parallel. Assume that the components have independent lifelengths with constant failure rates $\lambda_1 = 1$ and $\lambda_2 = 2$.

   (a) Find the survival function of the system life-length.

   (b) What is the expected system life time?

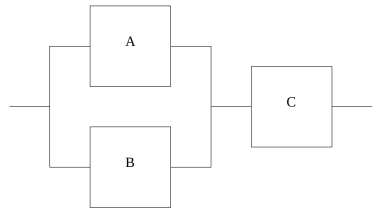
   (c) Compute the failure rate for the parallel system. (This example can be used to show that the property IFR is in general *not* preserved for parallel systems.)

8. Prove that the function $F_{\text{eq}}(t)$ defined by

$$F_{\text{eq}}(t) = \frac{1}{\nu}\int_0^t (1 - F(s))\,ds, \qquad \nu = \int_0^\infty (1 - F(s))\,ds < \infty,$$

is a distribution function, for any given distribution function $F(t)$, $t > 0$.

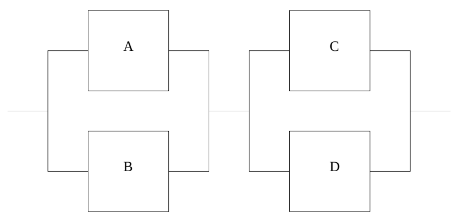
9. The system below works if C and at least one of A and B works. The components have independent life lengths with constant failure intensities; $\lambda_A = \lambda_B = 2$ and $\lambda_C = 1$.



(a) What is the probability that the system works at time t?

(b) Determine the expected life length of the system.

10. The system below works if at least one of A and B, and at least one of C and D works.

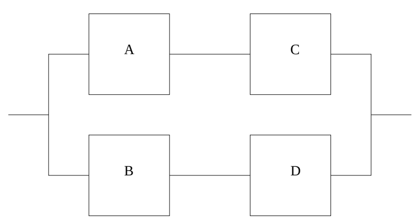


(a) What is the probability that the system works if the components function independently with probability p?

(b) What is the probability that the system works at time t if the components function independently and fail with common constant intensity $\lambda = 2$?

(c) Determine the expected life length of the system under the assumptions in (b).

11. The system below works if A and C, or B and D works.

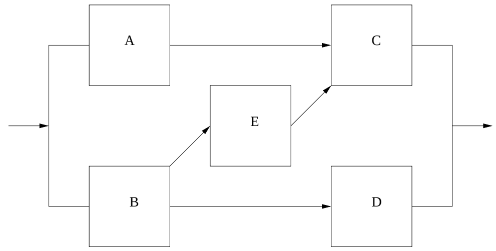


(a) What is the probability that the system works if the components function independently with probability p?

(b) What is the probability that the system works at time t if the components function independently and fail with common constant intensity $\lambda = 2$?

(c) Determine the expected life length of the system under the assumptions in (b). *Compare with the answer in the previous problem!*

12. The system below works if there is a path from left to right through the system.

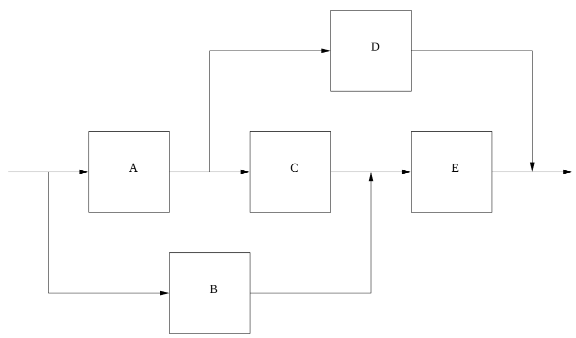


(a) What is the probability that the system works if the components function independently with probability p?

(b) What is the probability that the system works at time t if the components function independently and fail with common constant intensity $\lambda = 1$?

(c) Determine the expected life length of the system under the assumptions in (b).

13. The system below works if there is a path from left to right through the system.



What is the probability that the system works if the components function independently with probability p?

14. Verify the expression $\mathrm{E}(T_{\text{par}}) = \mu \sum_{k=1}^{n} 1/k$ for the mean system time to failure of a n independent components in parallel, all with constant failure rate $1/\mu$.

The Poisson process

15. Assume that vehicle traffic along a road is described by the Poisson process with intensity 20 cars per hour. A measuring device placed at a fixed observation site registers the time of arrival of each passing car.

(a) Find the probability that at least 2 cars are registered during a period of 6 minutes.

(b) Find the probability that the time periods from the beginning of the observation session to the first car and from the first to the second car, are both at most 3 minutes in length.

16. Events occur for $t \geq 0$ according to a Poisson process. During the interval $(0, 4]$ it is observed that four events occur. Find the probability that two of these events occur in the interval $(0, 1]$ and the other two in the interval $(1, 2]$.

17. During shotgun sequencing in genome research, enzymes are used to cut DNA sequences in shorter fragments. According to a commonly used model, the distances between consecutive cut points are independent and given by exponentially distributed random variables with expected value 2 (in some unit). What is the probability of at least 4 cut points in a DNA string of length 8 (measured in the same unit).

18. Suppose that $N(t)$ is a Poisson process with intensity λ, and put $p_n(t) = P(N(t) = n)$, $n \geq 0$. Verify that these probabilities solve the ordinary differential equations

$$\frac{d}{dt}p_n(t) = \lambda p_{n-1}(t) - \lambda p_n(t), \quad n \geq 1.$$

19. Local and long-distance calls arrive at a telephone exchange according to two independent Poisson processes with intensities two per minute for local calls and one every second minute for long-distance calls. What is the probability that

    (a) it takes more than one minute before the first call arrives?
    (b) the first call is a long-distance call?
    (c) the five first calls are local?

20. Tomas likes to go fishing. While waiting for the fishes to bite, he formulates the following model for the process: Fishes bite according to a Poisson process with intensity 4 bites per hour. Biting fishes are caught independently, and on average only every second time. Compute the probability that

    (a) six fishes bite during the first hour,
    (b) he fails to catch any fishes during the first hour,
    (c) six fishes bite and two of these are caught during the first hour.

21. A power plant is subject to incidents according to a Poisson process with intensity 6 (per year). Compute the probability that

    (a) there are no incidents during the first quarter of the year,
    (b) there is exactly one incident during the first half of the year,
    (c) there are no incidents during the first quarter given that there is exactly one incident during the first half of the year.

22. A DNA sequence of length 1 (in some unit) is subject to shotgun sequencing, that is it is cut at points given by a Poisson process with intensity $\lambda = 2$. Let Y_1 be the length of the segment up to the first cut point. ($Y_1 = 1$ if there are no cut points at all in the sequence.)

    (a) Compute $P(Y_1 > 0.5)$.
    (b) Compute $P(Y_1 = 1)$.
    (c) Compute the probability of exactly two cut points.
    (d) Compute the probability of exactly two cut points and $Y_1 > 0.5$.
    (e) Compute the probability that $Y_1 > 0.5$ given that there are exactly two cut points.
    (f) If there are exactly two cut points, the sequence is divided into three segments. One can show that the lengths of these have the same distribution. Use this to compute the probability that there is a segment longer than 0.5 given that there are exactly two cut points.

23. On a road with sparse traffic, passing cars can be described by a Poisson process with intensity 0.4 cars per minute.
    Compute the probability that

    (a) no cars pass during a five minute period,
    (b) the time gap between two succesive cars is at least five minutes,
    (c) at least three cars pass during a five minute period,
    (d) given that three cars pass during a five minute period, none passed during the first two minutes.

2. Markov chain models

Discrete time

1. Per, Pål and Petter are playing with a ball. Per throws with probability 0.3 to Pål and with probability 0.7 to Petter. Pål throws the ball to Per with probability 0.6 and to Petter with probability 0.4. Petter throws the ball with equal probabilities to his two friends. All throws are independent of each other. This can be thought of as a Markov chain. Introduce appropriate states and set up the transition probability matrix.

2. On a library table is a stack of three volumes of an encyclopedia. Visitors to the library use the encyclopedia independently as follows. Each user takes with equal probability one of the volumes, looks in it and puts it back on top of the pile. Two users are never using the books simultaneously. Considering this as a Markov chain, what are the possible states of the chain? Find the transition probabilities.

3. A Markov chain has transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

   Find the transition probabilities of the second order.

4. Suppose that the Markov chain in the previous example has the initial distribution $(1/2, 1/2, 0)$. Find the absolute probabilities $p_i^{(2)}$, $i = 1, 2, 3$.

5. Consider the weather during a number of days as a stochastic process with the only possible states 0 : *sun* and 1 : *rain*. We assume for simplicity that the process is a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}$$

   (a) Find the probability that a rainy day is followed by a sunny.
   (b) Formulate in words the event $\{X_{62} = 1\}$.
   (c) Find the conditional probability $P(X_{62} = 1 | X_{61} = 0)$.
   (d) Find the probability that a rainy day is followed by two sunny days.
   (e) Find $P(X_{62} = 1 | X_{60} = 0)$.
   (f) If Friday is sunny, what is the probability that the next following Sunday is also a sunny day?

6. (Continuation of the previous problem) Suppose we start the chain a sunny Friday, so that $X_0 = 0$.

   (a) What is the vector $p^{(0)}$?
   (b) Find $p^{(1)}$ and the probability that Saturday is sunny.
   (c) What is the probability for rain on Sunday?
   (d) What is $p^{(1)}$ if $p^{(0)} = (1/3, 2/3)$?

7. A Markov chain has transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

   Determine all stationary distributions of the chain.

8. A sequence of electrical impulses passes through a measuring device, which registers the largest value observed so far in the sequence. Assume that the impulse values at times $1, 2\dots$ are described by independent random variables that are uniformly distributed on $\{1, 2, 3, 4, 5\}$. This means that if $X_1, X_2, \dots$ are the registered values then

$$X_n = \max(Y_1, \dots, Y_n), \quad n \geq 1.$$

   (a) Find the probability function for the random variable X_n.
   (b) Motivate that $\{X_n\}$ is a Markov chain.
   (c) Determine the transition probability matrix.
   (d) Check the result in a) by computing the distribution of X_3 with the help of the initial distribution and the transition probability matrix.

9. The weather changes at a tourist resort from one day to the next can somewhat simplified be described as a Markov chain with the three states: E_1 : sun, E_2 : clouds, E_3 : rain. Using weather statistics of the area the following transition probability matrix has been estimated:

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.3 & 0.5 & 0.2 \\ 0.7 & 0 & 0.3 \end{pmatrix}$$

   A vacationer intends to visit the resort during 24-26 December. Under the assumption that there is still a lot of time before Christmas, derive the probability

   (a) that there will be three sunny days in a row;
   (b) of no rain at least during the first two days.

10. Assuming that the game in Exercise 1 has gone on for a long time, what are the probabilities of possessing the ball for Per, Pål and Petter respectively?

11. A particle is placed uniformly at one of the nine points in a 3×3 square grid. The particle then performs a random walk such that at each step one of the adjacent points (to the right or left, upwards or downwards) is chosen with equal probabilities. This means that the particle never remains in a point or moves diagonally. Find the probability that the particle after three steps is at the central point.

12. A Markov chain has the transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

   Determine if the chain has an asymptotic distribution. If it does, derive this distribution.

13. A Markov chain has the transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 1/3 & 2/3 & 0 \end{pmatrix}$$

   Determine if the chain has an asymptotic distribution and if so, find it.

14. A sequence $X_1, X_2, \ldots$ of binary symbols is modeled according to a Markov chain with possible values zero or one and transition matrix

$$\mathbf{P} = \begin{pmatrix} 4/5 & 1/5 \\ 2/5 & 3/5 \end{pmatrix}$$

A reading device that moves along the sequence of symbols is subject to reading errors as follows: The symbol Y_k observed as output of the device at site k is correctly read with probability $1 - \epsilon$ and miss-read with probability ϵ, independent of any other symbols. When the system is in equilibrium, find the conditional probability $P(X_k = 1 | Y_k = 1)$ as a function of ε.

15. A Markov chain in discrete time with states $\{1, 2, 3, 4\}$ is defined by the transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0.3 & 0.5 & 0.2 & 0 \\ 0.3 & 0.6 & 0 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Clearly, states 3 and 4 are absorbing. We wish to find the probability that the chain is absorbed in state 3 if it starts in state 1. Therefore, put

$$p_{3|j} = \text{probability of absorption in state 3, given } X_0 = j, \quad j = 1, 2,$$

and use the principle of conditioning on the first jump. In other words, derive two equations for the probabilities $p_{3|1}$ and $p_{3|2}$, then find these probabilities by solving the system of equations.

16. A very simple model for DNA sequences is that the letters, $\{X_n\}$, are independently chosen from $\{A, C, G, T\}$ with probabilities 0.25, 0.30, 0.15, 0.30 respectively. Suppose that there is an enzyme that breaks up the sequence as soon as the "word" AC appears in the sequence. To study the effect of the enzyme, let $\{Y_n\}$ be a process with the three states $\{A, AC, B\}$, so that $Y_n = A$ if $X_n = A$, $Y_n = AC$ if $X_{n-1} = A$ and $X_n = C$, and $Y_n = B$ otherwise.

    (a) Motivate that $\{Y_n\}$ is a Markov chain, and determine its transition matrix.
    (b) Compute the asymptotic distribution for $\{Y_n\}$.
    (c) What is the average length of an AC-fragment (in equilibrium)?

17. In a long DNA sequence, it has been recorded how often a certain base is followed by another. This is summarized in the matrix

$$\mathbf{P} = \begin{pmatrix} 0.300 & 0.205 & 0.285 & 0.210 \\ 0.322 & 0.298 & 0.078 & 0.302 \\ 0.248 & 0.246 & 0.298 & 0.208 \\ 0.177 & 0.239 & 0.292 & 0.292 \end{pmatrix},$$

where the rows and columns represent A, C, G, T in that order, so that the number 0.078 in the second row gives the relative frequency with which a C is followed by a G. Note that the rows sum to 1.
A simplified model for the DNA sequence is to consider the matrix $\mathbf{P}$ as a transition model for a Markov chain with state space $E = \{A, C, G, T\}$.
For this simplified model,

    (a) compute the asymptotic distribution and interpret it,
    (b) compute the probability of observing an A followed by another A at a randomly chosen position in the sequence.

18. In certain parts of a DNA sequence, so called CpG islands appear. In these, the letters C and G are much more common than usual. One such example is given by the matrix

$$\mathbf{P} = \begin{pmatrix} 0.180 & 0.274 & 0.426 & 0.120 \\ 0.171 & 0.368 & 0.274 & 0.187 \\ 0.161 & 0.339 & 0.375 & 0.125 \\ 0.079 & 0.355 & 0.384 & 0.182 \end{pmatrix},$$

where the entries should be interpreted as in the previous exercise. Compute the asymptotic distribution.

19. A certain letter sequence only contains the two letters A and B. They occur with the same probability and the letters in different positions are independent.
Study the process $\{X_n\}$, where X_n denotes the letter in position n, $n = 1, 2, 3, \ldots$.
Let $T_{AB} = \min\{n : X_{n-1} = A \text{ and } X_n = B\}$ and $T_{BB} = \min\{n : X_{n-1} = B \text{ and } X_n = B\}$, that is the positions when the sequences AB and BB occur for the first time in the sequence. *Note that* $T_{AB} \geq 2$ *and* $T_{BB} \geq 2$.

   (a) Determine $P(T_{AB} = 2)$ and $P(T_{BB} = 2)$.
   (b) Determine $E(T_{AB})$ and $E(T_{BB})$.
   Hint: Introduce a Markov chain with four states, $\{AA, AB, BA, BB\}$, describing the last two letters in the sequence.
   (c) Compute the probability that $T_{AB} < T_{BB}$, that is that the sequence AB occurs before the sequence BB.

Continuous time

20. An automatic switch in a control system can be in one of the three states *on*, *off* or *stand-by*. When the switch is *on* it changes to *off* after an exponentially distributed time with expected value 2 seconds. If the switch is *off* it either turns *on*, with intensity 0.2 per second, or changes into *stand-by*, with intensity 0.4 per second. Finally, in state *stand-by* the switch turns to *on* after an exponential time with expected value 1 second. Draw a model graph of a Markov process that describes the system and indicate all transition intensities. In equilibrium, what fraction of time is the switch in the three different states?

21. A tiger is always in one of the three states *sleeping*, *hunting* and *eating*. This tiger always goes from sleeping to hunting to eating and back to sleeping. On average, the tiger sleeps for 3 hours, hunts for 2 hours, and eats for 30 minutes. Assuming the tiger remains in each state for an exponential time, model the tiger's life as a continuous time Markov chain. What are the stationary probabilities?

22. A birth-and-death process with states $0, 1, 2 \ldots$ has birth intensities $\lambda_n = \lambda/\sqrt{n+1}$ and death intensities $\mu_n = \mu\sqrt{n}$, where λ and μ are positive parameters. Determine the asymptotic distribution.

23. The choice of birth and death intensities $\lambda_i = \lambda$, $i \geq 0$, and $\mu_i = \mu i$, $i \geq 1$, gives a continuous time birth-death process $\{X(t)\}$, which is known as the M/M/∞ service model. To interpret this Markov chain as a service model imagine Poisson events of intensity λ forming a stream of customer arrivals, and a service unit that provides service for each customer immediately upon arrival. Each customer leaves after spending an exponential service time of mean $1/\mu$ in the system. Then the varying number of currently served customers in the system is given by the Markov process $\{X(t)\}$. Show that the process has a limit distribution for any values of the parameters λ and μ, and find the equilibrium distribution.

24. Let $\lambda_k = a^k$, $k \geq 0$, and $\mu_k = a^k$, $k \geq 1$, where a is a parameter, be the jump intensities for a birth-and-death process. For which parameters a does the process have a unique asymptotic distribution? For such a values, what is the equilibrium distribution?

25. A birth-death process has the jump intensities $\lambda_k = \lambda(N-k)$ and $\mu_k = \mu N k$. Here λ and μ are positive parameters, N is a positive integer, and k varies between 0 and N. Show that there is an asymptotic distribution, which is given by a particular standard distribution. What happens to the asymptotic distribution as $N \to \infty$?

26. At a factory there are two machines (A and B) that fail independently with intensity 2 (failures per day) each. There is one service man who only can repair one machine at the time. It takes an exponentially distributed time with mean 4 hours to repair a machine. These times are independent of each other and of the state of the other machine. Machine A is considered more important than B, so that when A fails, the service man always starts repairing it even if he already is working on B; in that case the work on B is resumed when the work on A is finished.

    (a) Describe the system with an appropriate Markov process.

    (b) Compute the asymptotic availability for the two machines.

    (c) Determine the proportion of time that the service man is busy with repairs.

27. A fisherman puts out his net in the evening and empties it the following morning. During the night, fishes are caught in the net according to a Poisson process with intensity λ (per hour). A caught fish escapes with intensity $\lambda/3$. Compute the probability that there are at least 4 fishes in the net when it is emptied in the morning. *(You can, just like the poor fishes, consider the night as very long.)*

28. At a small company there is a telephone exchange with 3 lines. Calls arrive according to a Poisson process with intensity 0.5 calls per minute and last an exponentially distributed time with mean 4 minutes. The lengths of different calls are independent of each other and of the arrival process. When all lines are busy the exchange is blocked and new calls are rejected. Consider the exchange in equilibrium.

    (a) Determine the probability that the exchange is blocked.

    (b) The staff at the company feel that blocking occurs too often, and consider increasing the number of lines so that the blocking probability gets lower than 5%. How many lines do they need?

29. Customers arrive to a service system according to a Poisson process with intensity 2 (per hour). The system has two servers, and service times are exponentially distributed with mean 30 minutes. The service times are independent of each other and of the arrival process. There is one waiting space for customers that arrive when both servers are busy. There may thus be a maximum of three customers in the system; two being served and one waiting. Customers that arrive when the waiting space is occupied are rejected and leave the system without being served.
    Let $X(t)$ be the number of customers in the system at time t. Then, $\{X(t)\}$ is a discrete Markov process with continuous time.

    (a) Illustrate the process, indicating all possible transitions and their intensities.

    (b) Motivate that there is an asymptotic distribution and compute it.

    (c) What is the average number of rejected customers per hour?

3. Non-Markov models

1. Let $\{X_n, -\infty < n < \infty\}$ be independent random variables, all such that $P(X_n = 0) = P(X_n = 1) = 1/2$. Form $Y_n = X_n - X_{n-1}$, $n \geq 1$. What is the mean value function and the covariance function of $\{Y_n\}$.

2. Two processes $\{X(t)\}$ and $\{Y(t)\}$ are independent and weakly stationary with expected value 0 and covariance functions r_X and r_Y. Form the product $Z(t) = X(t)Y(t)$, $t \geq 0$. Find the mean value function and the covariance function of $\{Z(t)\}$. Is the product process weakly stationary?

3. Let $\{N(t), t \geq 0\}$, where $N(0) = 0$, denote a Poisson process with intensity λ, and let Y be a random variable, which is independent of $\{N(t)\}$ and with equal probabilities take the values 0 och 1. Put $Z(t) = (-1)^{Y+N(t)}$. Find the mean and covariance functions of $\{Z(t), t \geq 0\}$. Is this process weakly stationary?

4. A 150 W outdoor light bulb costs 40 kr and has an average lifetime of 720 hours. Assume that the normal usage of a bulb is 4 hours per day. If electricity costs 1 kr per kW-hr, what is the average cost per year of outdoor lighting?

5. Customers arrive at a service station with one server according to a Poisson process with intensity 2 (per hour). The service times are exponentially distributed with mean 20 minutes and independent of each other and of the arrival process. There is unlimited queuing space and customers who arrive when the server is busy always wait in line. *(This is a so called $M/M/1$ system.)* Seen from the server's perspective, time can be divided into *busy periods* when he is active and *vacant periods* when there are no customers in the system.
   Consider the system in equilibrium.

   (a) What proportion of the time is the server active?

   (b) Determine the expected length of a vacant period.

   (c) Determine the expected length of a busy period.
   Hint: It may be useful to note that the time points when customers arrive to an empty system are renewals!

6. A scientist uses a sensor driven video camera (with large capacity) to study butterflies at an inaccessible location. As soon as a butterfly touches the sensor, the camera starts and records during one minute. If the sensor is activated during a recording it continues one minute from the last activation. Assume that the butterflies appears according to a Poisson process with intensity one per minute. What proportion of time will the camera be recording?

7. An AR(2) process is defined by the relation $Y_n = X_n - 0.25Y_{n-2}$, where the X_n's are uncorrelated and have $\mathrm{E}(X_n) = 0$ and $\mathrm{V}(X_n) = 15$. Derive the covariance function of $\{Y_n\}$.

8. Show that the AR(2) filter is stable for the parameter values in Section 3.4, and sketch the stability area in the (a_1, a_2)-plane.

9. Let $(X_n)_{n=-\infty}^{\infty}$ denote Gaussian white noise with expected value 0 and variance σ^2. A stationary time series in discrete time of the type MA(2) is defined by

$$Y_n = X_n + c_1 X_{n-1} + c_2 X_{n-2}.$$

   It is known that the covariance function of $\{Y_n\}$ is given by $r_Y(0) = 27$, $r_Y(1) = 18$ and $r_Y(2) = 9$. Find the variance σ^2 of the input signal and the constants c_1 and c_2.

10. The input data sequence $\{X_n\}$ of an autoregressive filter of order 2,

$$Y_n + 0.5Y_{n-1} + 0.8Y_{n-2} = X_n,$$

is Gaussian white noise. If we start the recursion with Gaussian variables Y_0 and Y_1, it is clear that the resulting output sequence $\{Y_n\}$ is also Gaussian. It has been noted during a long series of observations that the absolute value $|Y_n|$ in such a series is larger than 2 approximately 10% of the time. Determine the variance $\mathrm{V}(X_n)$.

11. (Continuation of the previous problem)
Find the best linear predictor of Y_3 given that

   (a) $Y_2 = 1.0$ is observed,
   (b) $Y_1 = -0.5$ is observed,
   (c) $Y_1 = -0.5$ and $Y_2 = 1.0$ is observed,
   (d) $Y_0 = 0.8$, $Y_1 = -0.5$ and $Y_2 = 1.0$ is observed.

12. In a simple econometric model it is assumed that the price of a commodity one month depends on the supply during that month, whereas the supply in its turn depends on the price one month ago. If the price during month n is X_n and the supply of month n is Y_n, then it is assumed that

$$X_n = -c_1 Y_n + U_n, \quad Y_n = c_2 X_{n-1} + V_n,$$

where $\{U_n\}$ and $\{V_n\}$ denote independent noise variables with zero mean and unit variance. Show that both $\{X_n\}$ and $\{Y_n\}$ are AR(1) processes and find their covariance functions r_X and r_Y.

13. The weakly stationary process $\{Y_n\}$ has covariance function $r(\tau)$ given by $r(0) = 2$, $r(\pm 1) = 1$ and $r(k) = 0$ for all other k. Find the best linear predictor of Y_n in the case when the previous value Y_{n-1} has been observed, and in the case when both of Y_{n-1} and Y_{n-2} have been observed. Compare the variances of the prediction errors in the two cases.

14. The stationary process $\{Y_n\}$ is defined as

$$Y_n + 0.5Y_{n-1} = X_n,$$

where $\{X_n\}$ is Gaussian white noise with variance 4.

   (a) Determine the covariance function of $\{Y_n\}$.
   (b) What is the distribution of Y_n?
   (c) What proportion of time will the process $\{Y_n\}$ spend above the level 5?

4. Applications in computer communications

1. Generalize the $m \times m$ crossbar model to an $m \times n$ model, with m input lines and n output lines. The arrival structure is the same, hence one cell per slot arrives with probability p at each input, independently of each other. Arriving cells are immediately directed to a randomly chosen output line and if more than one cell ends up at the same output, excess cells are lost. Find the throughput per line and the loss probability of the switch, as functions of n and m.

2. Consider the Go-Back-N protocol described in the text.

   (a) What is the probability that a complete round of N packets is delivered without any losses?

   (b) Consider the random variable K = the number of rounds until the first loss occurs for the GBN protocol. Which distribution does it have?

   (c) Just as for the Go-Back-1 protocol, let $\{N_t\}$ be the renewal process counting the number of cycles to time t. Find the expected inter-renewal time ν_{GBN}.

   (d) How many packets are successfully delivered during each round? Hence, what is the appropriate reward variable R_i associated to cycle i?

   (e) Find the expected reward $\text{E}(R_i)$.

   (f) Compute the asymptotic throughput for the GBN model.

   (g) Take $T_0 = 3$. Produce function curves of the throughput as a function of p, for three different values of N, e.g. $N = 1$, $N = 4$, $N = 12$, displayed in the same graph.

5. Applications in biology

1. A certain type of bacteria reproduces by cell division so that each individual splits into two new ones with a constant intensity λ. The reproduction is independent for different individuals. Let $X(t)$ denote the number of individuals at time t.

   (a) The time until an individual splits is random. What is its distribution?

   (b) Motivate that $\{X(t)\}$ is a Markov chain in continuous time with state space $\{1, 2, \dots\}$. Determine the transition intensities.

   (c) Suppose in addition that the bacteria can die with intensity μ. Sketch in a model graph the possible transitions in this case and find the transition intensities. How is the state space modified?

2. Verify the formula for the expected value in the steady state distribution for the Wright-Fisher model with mutation.

3. Consider the Wright-Fisher model with population size $N = 3$. Find the expected time to fixation, m_i, as a function of the initial distribution $X_0 = i$.

4. For the case when the Wright-Fisher model starts with the two alleles in equal proportions, find an approximation to the mean time to fixation in terms of N, that is, find a constant c such that when $i \approx N/2$ then $m_i \approx c \cdot N$. Also, find an approximation if the allele A_1 is introduced at frequency $1/N$.

5. Consider the Wright-Fisher model for a population of size N. The *heterozygosity* in the population is described by the function

$$h(n) = P(\text{two randomly chosen individuals in generation } n \text{ are of different types}).$$

(a) Explain the relation $h(n) = (1 - 1/N)h(n-1)$.

(b) Suppose the population consists of $N/2$ individuals of each type at time $n = 0$ (N can be assumed to be an even number) and find the heterozygosity $h(n)$ and the asymptotic heterozygosity, i.e. the limit of $h([Nt])$, $t \geq 0$, as $N \to \infty$.

(c) An alternative view is to define an empirical heterozygosity $H(n)$ as the conditional probability

$$H(n) = P(\text{two randomly chosen individuals in generation } n \text{ are of different types} | X_n).$$

Show that $H(n) = 2X_n(N - X_n)/N(N-1)$.

(d) Show that $E(H(n)) = h(n)$. *Hint:* Use that the conditional distribution of X_n given X_{n-1} is $\text{Bin}(N, X_{n-1}/N)$ to show that $E(H(n))$ satisfies relation a).

6. We assume that a particular organism is subject to death with constant intensity $\mu > 0$, and consider a population that at time $t = 0$ consists of n independent organisms. Put

$$X(t) = \#\text{ organisms alive at time } t$$

and

$$T_0 = \text{extinction time of the population.}$$

(a) Find $p(t) = P(\text{a particular organism is alive at } t)$.

(b) Find $P(X(t) = j)$. *Hint:* Use the binomial distribution.

(c) What are the associated mean and variance functions $m(t) = E(X(t))$ and $V(t) = V(X(t))$?

(d) Now think of $X(t)$ as a pure death process. What are the death intensities? How much time does $X(t)$ spend in state j, $1 \leq j \leq n$? What is the expected time spent in j? What is the expected time to extinction?

(e) In a 1978 study of ringed lapwing birds (*Vanellus vanellus*) found dead in Europe, the population dropped from 100 at time 4.1 years to 1 at time 14.2 years of observation. Using $m(t)$, give an estimate $\hat{\mu}$ of the death intensity μ. (Example taken from Renshaw (1991)).

(f) If the lapwing population starts with $n = 1000$, what is the expected time to extinction? If n is 10^6?

7. The Moran model is a birth-and-death process in continuous time used in population genetics. A simple version of the model has two parameters: an integer N for population size and a real number $u > 0$ for mutation intensity. The state space is $\{0, 1, \ldots, N\}$ and the birth and death rates are

$$\lambda_i = \left(1 - \frac{i}{N}\right) p_i, \qquad \mu_i = \frac{i}{N}(1 - p_i),$$

where

$$p_i = \frac{i}{N}(1 - u) + \left(1 - \frac{i}{N}\right) u.$$

Find the asymptotic distribution for the case $N = 2$. Sketch graphically the form of this distribution, indicating how the character of the distribution differs when you select smaller or larger values of the mutation parameter.

8. Verify the asymptotic distributions for the Jukes-Cantor and the Kimura models. Suggest a measure of heterozygosity using the asymptotic distributions in these models.

9. With reference to the model for shotgun sequencing, what is the minimum required coverage to have a risk of less than 1% that an arbitrary site on the sequence is not covered by at least one fragment? What minimum coverage is needed to reduce this risk to 0.1%?

10. Consider covering a genome of length 100,000 by fragments of length 500. Make a table of the mean number of contigs as the level of coverage a is varied from 0.5 to 7. Explain the shape of this quantity as a function of the parameter a. Find the coverage for which the mean number of contigs is maximized.

11. Consider the random model, described in Section 5.4 of the lecture notes, for joining N fragments of length L to cover a genom of length G. Let $M =$ "number of contigs". One can argue that a better model than the one discussed in the lecture notes is obtained by assuming that
$$M = 1 + Z, \quad \text{where } Z \in \text{Bin}(N-1, e^{-a}),$$
and a is the coverage.

    (a) Determine the mean number of contigs as a function of a.
    (b) Determine the variance as a function of a.
    (c) Compute the mean and variance when $N = 100$ and a is chosen as the coverage by which there is a 75% probability that the whole genom is covered by one single contig.

12. Suppose that we have used the so called Viterbi algorithm on the three sequence

CAETPDH CAEFDH CDAEFPDH

with a Hidden Markov Model and obtained the corresponding sequences of states

$$m_0m_1m_2m_3d_4m_5m_6m_7m_8m_9$$
$$m_0m_1m_2m_3m_4d_5d_6m_7m_8m_9$$
$$m_0m_1i_1m_2m_3m_4d_5m_6m_7m_8m_9$$

Give the induced alignment of the three sequences of amino acids.

Answers to the exercises

1. Lifelengths, systems and the Poisson process

Lifelengths and systems

1.1 0.788

1.2 $F_T(t) = 1 - e^{-(t/a)^k}, \qquad f_T(t) = kt^{k-1}a^{-k}e^{-(t/a)^k}, \qquad t \geq 0.$

1.3 (a) 0.705
(b) 0.307

1.4 $\sqrt{\pi/2}$

1.5 $\lambda(t) = \dfrac{(1-e^{-t/8})}{4(2-e^{-t/8})}$ (time unit: months)

1.6 $\lambda(t) = \dfrac{t}{64+8t}$

1.7 (a) $R(t) = e^{-t} + e^{-2t} - e^{-3t}$
(b) 7/6
(c) $\lambda_{par}(t) = \dfrac{e^{2t}+2e^t-3}{e^{2t}+e^t-1}$ (Show that this function is *not* increasing)

1.8 A distribution function must be nonnegative and increasing from 0 to 1 (and in addition right-continuous). Check also a specific example. For example, what is F_{eq} if F is the uniform distribution on $[0,1]$?

1.9 (a) $R(t) = 2e^{-3t} - e^{-5t}$
(b) 7/15

1.10 (a) $4p^2 - 4p^3 + p^4$
(b) $4e^{-4t} - 4e^{-6t} + e^{-8t}$
(c) $11/24 \approx 0.458$

1.11 (a) $2p^2 - p^4$
(b) $2e^{-4t} - e^{-8t}$
(c) $3/8 = 0.375$

1.12 (a) $2p^2 + p^3 - 3p^4 + p^5$
(b) $2e^{-2t} + e^{-3t} - 3e^{-4t} + e^{-5t}$
(c) $47/60 \approx 0.783$

1.13 $2p^2 + p^3 - 3p^4 + p^5$

1.14 One can use

$$E(T_{par}) = \int_0^\infty (1 - (1-e^{-t/\mu})^n)\, dt$$

and make the change of variables $u = 1 - e^{-t/\mu}$ in the integral. Then use the formula for a geometric sum.

The Poisson process

1.15 (a) 0.5940
(b) 0.3996

1.16 3/128

1.17 0.56653

1.18 Take the time derivative and check that the expressions on the left and the right are equal.

1.19 (a) $e^{-5/2} \approx 0.082$
(b) 1/5
(c) $(4/5)^5 \approx 0.328$

1.20 (a) 0.1042
(b) 0.1353
(c) 0.0244

1.21 (a) $e^{-3/2} \approx 0.223$
(b) $3e^{-3} \approx 0.149$
(c) 1/2

1.22 (a) $e^{-1} \approx 0.368$
(b) $e^{-2} \approx 0.135$
(c) $2e^{-2} \approx 0.271$
(d) $e^{-2}/2 \approx 0.0677$
(e) 1/4
(f) 3/4

1.23 (a) 0.135
(b) 0.135
(c) 0.323
(d) 0.216

2. Markov chain models

Discrete time

2.1 $\mathbf{P} = \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0.6 & 0 & 0.4 \\ 0.5 & 0.5 & 0 \end{pmatrix}, \qquad E = \{\text{Per, Pål, Petter}\}.$

2.2 $\mathbf{P} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$

2.3 $\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{pmatrix} 4/9 & 5/18 & 5/18 \\ 5/12 & 5/12 & 1/6 \\ 5/12 & 1/6 & 5/12 \end{pmatrix}$

2.4 $\frac{1}{72}(31,\ 25,\ 16)$

2.5 (a) 0.2
(b) Day 62 rainy.
(c) 0.3
(d) 0.14
(e) 0.45
(f) 0.55

2.6 (a) (1, 0)
(b) 0.7
(c) 0.45
(d) $\frac{1}{30}(11,\ 19)$

2.7 $\pi_1 = \pi_2 = \pi_3 = 1/3$

2.8 (a) $P(X_n = k) = \dfrac{k^n - (k-1)^n}{5^n}, \qquad k = 1, \ldots, 5.$

(b) $X_n = \max(X_{n-1},\ Y_n)$

(c) $\mathbf{P} = \dfrac{1}{5}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 1 \\ 0 & 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$

(d) $p(3) = \frac{1}{125}(1,\ 7,\ 19,\ 37,\ 61)$

2.9 (a) 1/5
(b) $28/45 \approx 0.622$

2.10 (0.352, 0.286, 0.361)

2.11 4/27

2.12 $\pi = (1/3,\ 2/3,\ 0)$

2.13 $\pi = (0.3,\ 0.4,\ 0.3)$

2.14 $(1-\epsilon)/(1+\epsilon)$

2.15 $p_{3|1} = 8/13 \approx 0.6154, \qquad p_{3|2} = 6/13 \approx 0.4615.$

2.16 (a) $\mathbf{P} = \begin{pmatrix} 0.25 & 0.30 & 0.45 \\ 0.25 & 0 & 0.75 \\ 0.25 & 0 & 0.75 \end{pmatrix}$

(b) (1/4, 3/40, 27/40)
(c) $40/3 \approx 13.33$

2.17 (a) (0.262, 0.246, 0.239, 0.253)
(b) 0.079

2.18 (0.159, 0.340, 0.399, 0.101)

2.19 (a) $P(T_{AB} = 2) = P(T_{BB} = 2) = 1/4$
(b) $E(T_{AB}) = 4, \qquad E(T_{BB}) = 6.$
(c) 3/4

Continuous time

2.20 $\frac{1}{13}(6,\ 5,\ 2)$

2.21 $\frac{1}{11}(6,\ 4,\ 1)$

2.22 $\mathrm{Po}(\lambda/\mu)$

2.23 $\mathrm{Po}(\lambda/\mu)$

2.24 $a > 1, \qquad \mathrm{Ge}(1 - 1/a)$.

2.25 $\mathrm{Bin}(N, p), \qquad p = \lambda/(\lambda + \mu N), \qquad \mathrm{Po}(\lambda/\mu)$.

2.26 (a) Introduce 4 states: $\{0, A, B, 2\}$ that show how many (and which) machines are non-working. Intensity matrix:

$$\mathbf{A} = \begin{pmatrix} -4 & 2 & 2 & 0 \\ 6 & -8 & 0 & 2 \\ 6 & 0 & -8 & 2 \\ 0 & 0 & 6 & -6 \end{pmatrix}.$$

(b) A: 3/4, $\qquad B$: $45/68 \approx 0.662$.
(c) $8/17 \approx 0.471$

2.27 0.353

2.28 (a) 0.211
(b) 5 lines are required; gives blocking probability 0.037.

2.29 (a) Birth and death process with $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = 2$ and $\mu_2 = \mu_3 = 4$.
(b) $\frac{1}{11}(4,\ 4,\ 2,\ 1)$
(c) 2/11

3. Non-Markov models

3.1 $m(n) = E(Y_n) = 0$.
$r(0) = C(Y_n,\ Y_n) = 1, \qquad r(\pm 1) = C(Y_n,\ Y_{n+1}) = -1/2,$
$r(\pm k) = C(Y_n,\ Y_{n+k}) = 0, \qquad k = 2, 3, ...$

3.2 Weakly stationary with $m(t) = 0$ and $r_Z(s,\ s+t) = r_X(s,\ s+t) \cdot r_Y(s,\ s+t)$.
(Thus $r_Z(s,\ s+t)$ only depends on t and not on s.)

3.3 Weakly stationary with $m(t) = 0$ and $r(s,\ s+t) = e^{-2\lambda t}$.

3.4 $E(R) = 148$ and $E(T) = 180/365$. Cost approximately 300 kr per year.

3.5 (a) 2/3
(b) 1/2
(c) 1

3.6 $\dfrac{e}{e+1} \approx 0.731$.

3.7 $r_Y(2k+1) = 0, \qquad r_Y(2k) = 16(-0.25)^{|k|}, \qquad k = 0, \pm 1, \ldots$

3.8 -

3.9 $c_1 = c_2 = 1, \qquad \sigma = 3$.

3.10 $\sigma^2 = 0.4911$.
(The relation $P(|Y_n| > 2) = \cdots = 2(1 - \Phi(2/\sqrt{r(0)})) = 0.10$ yields $r(0) = 1.47818$. The Yule-Walker equations yields $r(0) = 3.01003/\sigma^2$. Thus $\sigma^2 = 0.4911$.)

3.11 (a) $\hat{Y}_3 = -\frac{1}{3.6} Y_2 \approx -0.278$
(b) $\hat{Y}_3 = -\frac{2.38}{3.6} Y_1 \approx 0.331$
(c) $\hat{Y}_3 = -0.5Y_2 - 0.8Y_1 \approx -0.100$
(d) Same answer as in (c).

3.12 $r_X(\tau) = \dfrac{1 + c_1^2}{1 - c_1^2 c_2^2}(-c_1c_2)^{|\tau|}, \qquad r_Y(\tau) = \dfrac{1 + c_2^2}{1 - c_1^2 c_2^2}(-c_1c_2)^{|\tau|}.$

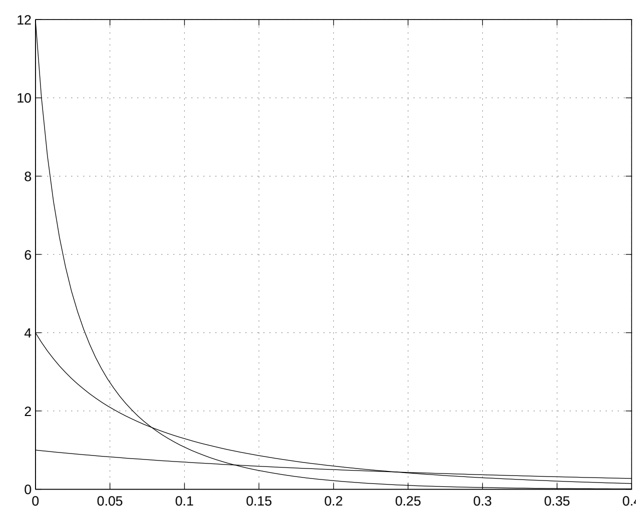
3.13 One previous observation: $\hat{Y}_n = \frac{1}{2} \cdot Y_{n-1}, \qquad V(Y_n - \hat{Y}_n) = 3/2.$
Two previous observations: $\hat{Y}_n = \frac{2}{3} \cdot Y_{n-1} - \frac{1}{3} \cdot Y_{n-2}, \qquad V(Y_n - \hat{Y}_n) = 4/3.$

3.14 (a) $r(0) = 16/3, \qquad r(k) = (-0.5)^k \cdot \frac{16}{3}, \qquad k = 1, 2, 3, \ldots, \qquad r(-k) = r(k).$
(b) $Y_n \in N(0, \sqrt{16/3})$
(c) 0.015

4. Applications in computer communications

4.1 Utilization: $1 - (1 - p/n)^m$.
Traffic intensity: mp/n.
Loss probability: $1 - \dfrac{1 - (1 - p/n)^m}{mp/n}$.

4.2 (a) $p_N = (1 - p)^N$
(b) $P(K = k) = p_N^{k-1}(1 - p_N)$
(c) $\nu_{\text{GBN}} = E(K_i) + T_0 = \frac{1}{1-(1-p)^N} + T_0$
(d) $R_i = N(K_i - 1)$
(e) $E(R_I) = N(E(K_i) - 1) = N\frac{(1-p)^N}{1-(1-p)^N}$
(f) $\text{Throughput}_{\text{GBN}} = \lim_{t\to\infty} \dfrac{1}{t} \sum_{i=1}^{N_t} R_i = \dfrac{1}{\nu} E(R_1) = \dfrac{N\frac{(1-p)^N}{1-(1-p)^N}}{\frac{1}{1-(1-p)^N} + T_0} =$
$= \frac{N(1-p)^N}{1+(1-(1-p)^N)T_0}$
(g) See figure.



5. Applications in biology

5.1 (a) $\mathrm{Exp}(1/\lambda)$
(b) Pure birth-process: $\lambda_i = \lambda \cdot i, \qquad i = 1, 2, 3, \ldots$
(c) Birth and death-process: $\lambda_i = \lambda \cdot i, \qquad \mu_i = \mu \cdot i, \qquad i = 0, 1, 2, \ldots$

5.2 -

5.3 $m_1 = m_2 = 3$

5.4 $m(N/2) \approx 1.386N$ generations, $\qquad m(1) \approx 2 \log N$ generations.

5.5 (a) -
(b) $e^{-t}/2$

5.6 (a) $e^{-\mu t}$
(b) Given by the distribution $\mathrm{Bin}(n, e^{-\mu t})$
(c) Follows from (b)
(d) $\mu_j = \mu j, \qquad \mathrm{Exp}(1/\mu_j), \qquad 1/\mu j, \qquad \mu^{-1} \sum_{j=1}^{n} 1/j.$
(e) $\hat{\mu} = 0.456$
(f) 16.4 respectively 31.6 years.

5.7 $\pi = \left(\frac{1}{2(1+2u)}, \frac{2u}{1+2u}, \frac{1}{2(1+2u)} \right)$

5.8 $1 - (\pi_A^2 + \pi_G^2 + \pi_C^2 + \pi_T^2) = \ldots$

5.9 4.6 respectively 6.9.

5.10 -

5.11 (a) $E(M) = 1 + (N-1) \cdot e^{-a}$
(b) $V(M) = (N-1) \cdot e^{-a} \cdot (1 - e^{-a})$
(c) $e^{-a} = 1 - 0.75^{\frac{1}{N-1}} = 1 - 0.75^{\frac{1}{99}} \approx 0.9971 \qquad \Rightarrow \qquad a \approx 5.84$

$$E(M) = 1 + 99 \cdot (1 - 0.75^{\frac{1}{99}}) \approx 1.5095$$

$$V(M) = 99 \cdot 0.75^{\frac{1}{99}} \cdot (1 - 0.75^{\frac{1}{99}}) \approx 0.5069.$$

5.12

m_0	m_1	$\mathbf{i_1}$	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
	C	–	A	E	–	T	P	D	H	
	C	–	A	E	F	–	–	D	H	
	C	D	A	E	F	–	P	D	H	