Stochastics
Midterm test 2, solutions
Fall 2021

1. A face paint artist makes two types of face paint: tiger or butterfly. A tiger takes 10 minutes to paint and a butterfly takes 20 minutes to paint. Each child will ask for tiger with probability 2/3 and for butterfly with probability 1/3, independently from the others. There is a long line of children waiting for face paint, so the face paint artist works all the time.

   (a) Model the face paint artist's activity with a discrete time Markov chain. What are the states? Compute the transition probability matrix.

   (b) What is the ratio of time she spends with painting tigers?

   (c) The cost of a tiger is 1200 HUF and the cost of a butterfly is 1800 HUF. Calculate the long term average income of the face paint artist per hour.

   Solution 1. The states are 10 minute intervals. In each interval, the face paint artist is either painting a tiger, or the first half of a butterfly, or the second half of a butterfly. Accordingly, the states are T, B1, B2, and

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

   The stationary distribution $v_{\text{st}} = (x_1\, x_2\, x_3)$ can be computed from

$$2/3x_1 + 2/3x_3 = x_1, \quad 1/3x_1 + 1/3x_3 = x_2, \quad x_2 = x_3, \quad x_1 + x_2 + x_3 = 1,$$

   whose solution is $v_{\text{st}} = (1/2\,1/4\,1/4)$, so she is spending 1/2 of her time painting tigers.

   To calculate her long term average income, we use the ergodic theorem. We need to cut the cost of a butterfly in 2 parts; cutting it into 900+900 gives that the long term average income is

$$1/2 \cdot 1200 + \frac{1}{/}4 \cdot 900 + \frac{1}{/}4 \cdot 900 = 1050$$

   HUF per 10 minute interval, which is $6 \cdot 1050 = 6300$ (HUF per hour). Due to $x_2 = x_3$, it doesn't matter how the cost of a butterly is cut in half, the result will be the same.

   Solution 2. The states could also be the individual paintings; with this setup, we have two states, T and B, and

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

   (This actually corresponds to an iid sequence.) The stationary distribution is then $v_{\text{st}} = (2/3\,1/3)$, which is the ratio of tiger paintings and butterfly paintings. We also have to consider that tiger and butterfly paintings take different length of time; accordingly, the ratio of time spent with painting tigers is

$$\frac{2/3 \cdot 10}{2/3 \cdot 10 + 1/3 \cdot 20} = \frac{1}{2}.$$

   Since she spends half her time painting tigers, in a one hour interval, she will spend 30 minutes on average painting tigers, which is 3 tigers (on average), and 30 minutes painting butterflies, which is 1.5 butterflies on average. Her income from 3 tigers and 1.5 butterflies is

$$3 \cdot 1200 + 1.5 \cdot 1800 = 6300,$$

which is thus her long term average income per hour.

2. The parking lot of a small store has room for 2 cars. A customer arrives by car every 10 minutes on average. If the parking lot has room for a car, they park and enter the store. If the parking lot is full, they leave immediately, without entering the parking lot or the store. Each customer who enters the store spends on average 5 minutes inside, then leaves.

   (a) Let X_t denote the number of cars in the parking lot at time t. What assumptions do we need to make so that the Markov property holds for X_t? Calculate the generator.

   (b) What is the probability that at a random time, we find the parking lot empty?

   (c) What is the long term average ratio of customers lost due to a full parking lot?

   Solution. In order for the Markov property to hold, we need to assume that the interarrival times and the time spent in the store by the customers are independent and both have exponential distribution.

   The states are $0, 1, 2$, and the generator is

$$Q = \begin{pmatrix} -1/10 & 1/10 & 0 \\ 1/5 & -3/10 & 1/10 \\ 0 & 2/5 & -2/5 \end{pmatrix}$$

   since the expectation of the interarrival times should be 10, which corresponds to an arrival rate of $1/10$. For the service rate, each customer leaves with rate $1/5$, so the rate of $1 \to 0$ transition is $1/5$, but the rate of $2 \to 1$ transition is $2/5$ since each customer leaves with rate $1/5$, and there are 2 customers inside.

   This is a Markov queue; the stationary distribution $v_{\mathrm{st}} = (x_0\, x_1\, x_2)$ can be computed from the balance equations

$$1/10 x_0 = 1/5 x_1, \quad 1/10 x_1 = 2/5 x_2, \quad x_1 + x_2 + x_3 = 1,$$

   from which $v_{\mathrm{st}} = \left(\frac{8}{13}\, \frac{4}{13}\, \frac{1}{13}\right)$.

   From the stationary distribution, the probability that at a random time, we find the parking lot empty is $8/13$.

   The long term average ratio of customers lost due to a full parking lot is equal to the ratio of time when the parking lot is full, which is $1/13$ (once again from the stationary distribution).

3. A translator can translate 10 pages of text per day. She receives a job offer every 4 days on average. If she is idle, she takes the first incoming offer, but she declines all incoming offers while she is working on a job. Each job is translating a text; the average length of the text is 80 pages.

   (a) Model the translator's activity with a continuous time Markov chain. State the assumptions we need to make to ensure that the Markov property holds. Calculate the generator.

   (b) Right now, she is idle. Estimate the probability that 12 hours from now, she will be working on a job.

(c) She charges 30 euros per page for translation. Calculate her long-term average daily income.

Solution. There are 2 states: idle or working. If the job offers arrive according to a Poisson process, and the length of a working period has exponential distribution, then the Markov property holds for the process. We use 1 day for the unit of time, then the generator is

$$Q = \begin{pmatrix} -1/4 & 1/4 \\ 1/8 & -1/8 \end{pmatrix}$$

from the expectation of the interarrival times of job offers and the length of a text, considering that 80 pages takes 8 days to translate.

Assuming right now she is idle, we have $v(0) = (1\,0)$, and the state vector 12 hours from now can be estimated using the short term approximation. 12 hours is 0.5 days, so

$$v(0.5) = v(0)e^{0.5Q} \approx v(0)(I+0.5Q) = (1\,0)\cdot\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + 0.5\begin{pmatrix} -1/4 & 1/4 \\ 1/8 & -1/8 \end{pmatrix}\right) = \begin{pmatrix} \frac{7}{8} & \frac{1}{8} \end{pmatrix},$$

so the probability that 12 hours from now, she will be working on a job is approximately 1/8.

The stationary vector is $v_{\mathrm{st}} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix}$, and from the ergodic theorem, her long term average daily income is

$$\frac{1}{3}\cdot 0 + \frac{2}{3}\cdot 10\cdot 30 = 300$$

euros, considering that she translates 10 pages per day and charges 30 euros per page.

4. Each student passes a certain exam with probability p, but p is unknown. In each of the last 5 semesters, 10 students have taken the exam, and the number of students passing the exam was 5, 7, 5, 9, 6 for each semester respectively. Based on this sample, calculate either the moment estimate for p or the maximum likelihood estimate for p (one of them is enough, you may choose).

Solution 1. The background distribution is BIN(10,p) since in each semester, 10 students take the exam and each passes with probability p. For the moment estimate, we need the function

$$g(p) = E_p(X_1) = 10p,$$

so $g^{-1}(x) = x/10$, and the moment estimate is

$$\hat{p} = g^{-1}(\bar{x}) = \frac{6.4}{10} = 0.64,$$

where $\bar{x}$ is the sample mean.

Solution 2. The background distribution is BIN(10,p) once again, and the likelihood and log-likelihood function of the sample are

$$\begin{aligned} L(p) &= \binom{10}{5}p^5(1-p)^5\binom{10}{7}p^7(1-p)^3\binom{10}{5}p^5(1-p)^5\binom{10}{9}p^9(1-p)^1\binom{10}{6}p^6(1-p)^4 = \\ &= Cp^{32}(1-p)^{18}, \\ l(p) &= \log C + 32\log p + 18\log(1-p), \end{aligned}$$

where C denotes a positive constant (whose precise value is not relevant). We solve

$$0 = \frac{\mathrm{d}l(p)}{\mathrm{d}t} = \frac{32}{p} - \frac{18}{1-p},$$

whose only solution is

$$\hat{p} = \frac{32}{50} = 0.64,$$

which is the maximum likelihood estimate.

Solution 3. We actually don't need the detailed information about individual semesters, we can add all of the semesters up to obtain a sample of size 1: $X_1 = 32$ from the background distribution BIN(50,p). Then either the ML estimate or the moment estimate can be carried out for this sample and will give the same estimate $\hat{p} = 0.64$ as before.

5. The box of a board game states that the average game length is 60 minutes. We play 4 games and the length of the games turns out to be 42, 54, 60, 48 minutes each. Based on this sample, test the hypothesis that the average game length is 60 minutes versus the hypothesis that the average game length is not equal to 60 minutes on a 95% confidence level. Also state the type of test used explicitly.

   Solution. We need to test the unknown mean against a given value. The deviation is not known; we will do a 2-sided, 1-sample t-test.

   - H_0: the mean length of a game is $m = 60$ minutes.
   - H_1: $m \neq 60$ minutes.

   The sample mean is

$$\bar{x} = \frac{42 + 54 + 60 + 48}{4} = 51,$$

   and the corrected empirical variance is

$$s_n^{*2} = \frac{1}{3}((42 - 51)^2 + (54 - 51)^2 + (60 - 51)^2 + (48 - 51)^2) = 60,$$

   and $s_n^* = \sqrt{60} \approx 7.75$.

   We compute the statistic

$$t = \frac{\bar{x} - \mu}{s_n^*}\sqrt{n} = \frac{51 - 60}{\sqrt{60}}\sqrt{4} \approx -2.32,$$

   the percentile

$$t_{\varepsilon/2} = 3.182$$

   from the table of the t-distribution with degree of freedom $4 - 1 = 3$, then the comparison

$$t = -2.32 \in [-3.182, 3.182] = [-t_{\varepsilon/2}, t_{\varepsilon/2}]$$

   holds, so we accept H_0.