Marco Longfils and Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE055 / MSG810 Matematisk statistik och diskret matematik

Exam 13 January 2017, 8:30 - 12:30

Allowed aids: Chalmers-approved calculator
and one (two-sided) A4 sheet of paper with your own notes.
Total number of points: 30. To pass, at least 12 points are needed.
Note: All answers should be motivated.

1 Solutions

1. Let us rename the states of the chain to $Milan = 1, Gothenburg = 2, Stockholm = 3, Tromso = 4$.

   (a) The chain $X_0, X_1, X_2, \dots$ is a Markov chain, because given the value of X_n, the distribution for the value of X_{n+1} does not depend on the previous values $X_0, \dots, X_{n-1}$ of the chain. The transition matrix is given by

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

   (b) Let $t = (t_1, t_2, t_3, t_4)$ be the vector where t_i is the expected number of days it takes to reach Tromsø starting from state $i, i = 1, 2, 3, 4$. t satisfies the following system of equations:

$$\begin{cases} t_4 = 0 \\ t_3 = \frac{1}{3}(t_1 + t_2) + 1 \\ t_2 = \frac{1}{3}(t_1 + t_3) + 1 \\ t_1 = \frac{1}{2}(t_2 + t_3) + 1 \end{cases}$$

   which has solution $t = (5, 4, 4, 0)$ and we conclude that the expected number of days to reach Tromsø from Milan is $t_1 = 5$. (Note: It is also possible to arrive at the answer using matrix calculations; this involves inverting a 3×3 matrix. A smart trick is to look at the Markov chain where Gothenburg and Stockholm are joined into one state, and realize that the question can be answered using this simplified chain. Then, only a 2×2 matrix needs to be inverted.)

   (c) Y_n is not a Markov chain, as the probabilities for the length of the next trip depends on not only how long Marco has flown up to this point, but also on where he has arrived, and values of the chain before the last step carry information about where he is. For example, consider $Y_2 = 7$. After two flights, Marco could be either in Stockholm (so that $Y_1 = 5$) or in Tromsø (so that $Y_1 = 4$). What happens next (i.e., the value of Y_3) depends not only on Y_2, but also on Y_1.

2. (a) X does not have a binomial distribution because the attempts have different probabilities of success, while in a binomial distribution the probability of success remains the same from trial to trial.

   (b) Since in the last trial we always obtain a success as $p_{10} = 1$, it means that the only way we obtain exactly one success is when all the first nine trials end up in a failure. We have a failure in trial i with probability $q_i = \frac{10-i}{10}$. Thus,

$$P(X = 1) = \prod_{i=1}^{9} q_i = \frac{9}{10}\frac{8}{10}\frac{7}{10}\frac{6}{10}\frac{5}{10}\frac{4}{10}\frac{3}{10}\frac{2}{10}\frac{1}{10} = 0.000363$$

   (c) X has a binomial distribution with parameters $n = 10$ trials and probability of success $p = \frac{1}{10}$. This is due to the fact that we consider a fixed number of experiments, each with outcome "success" or "failure", where the trials are independent and identical, and X counts the number of success obtained.

   (d) Y does not have a binomial distribution since it has hypergeometric distribution, as the experiments is equivalent to selecting objects from a group of items, of which only some of them have a trait ("successes").

3. Denote by X the random variable that models the number of attempts needed to pass the exam.

   (a) The probability that a student will pass the exam in less than 3 attempts is given by

$$P(\text{less than 3 attempts}) = 1 - P(\text{at least three attempts}) = 1 - 0.3^2 = 1 - 0.09 = 0.91$$

   (b) X has a geometric distribution with parameter $p = 0.7$ as we are counting the number of trials needed to pass the exam (i.e. the first success).

   (c) Since $X + Y = 200$ as the total number of students is just the sum of those who pass and those who fail the exam, we have that $Y = 200 - X$, so the correlation coefficient between X and Y is -1, as there exist a linear relationship between them and the slope of the linear relationship is negative.

4. (a) The number of cases X in one year follows a Poisson distribution with parameter 8. Thus, the probability of three or fewer cases is given by

$$P(X \leq 3) = \sum_{i=0}^{3} e^{-8}\frac{8^i}{i!} = 0.0424$$

(b) The number of cases Y in one month follows a Poisson distribution with parameter $\frac{8}{12}$. Thus

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-\frac{8}{12}} = 0.4866$$

(c) As cases occur independently, the number of cases Z in the next 2.5 months follows a Poisson distribution with parameter $\frac{8 \cdot 2.5}{12}$. So

$$P(Z = 0) = e^{-\frac{5}{3}} = 0.1889$$

(d) The number of cases W in 10 years follows a Poisson distribution with parameters 80. Thus W can be approximated by a normal distribution V with mean 80 and variance 80.

$$P(W \geq 103) \approx P(V \geq 103) = P(N(0,1) \geq \frac{103 - 80}{\sqrt{80}}) = P(N(0,1) \geq 2.57) = 0.0051$$

5. $n_X = 43, p_X = \frac{37}{43}, n_Y = 35, p_Y = \frac{21}{35}$.

(a) An estimate of how many more percent of the workers are satisfied with their job at X compared to at Y is given by

$$p_X - p_Y = 26\%.$$

(b) A 90% confidence interval for the estimate defined above is given by

$$(p_X - p_Y) \pm z_{\alpha/2} \sqrt{\frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}}$$

where $z_{\alpha/2} = 1.645$. Thus the confidence interval is given by

$$[0.0989, 0.4221]$$

(c) The length of the confidence interval is in general

$$2z_{\alpha/2} \sqrt{\frac{p_X(1-p_X) + p_Y(1-p_Y)}{n}} \leq z_{\alpha/2} \sqrt{\frac{2}{n}}$$

as we always have $p_X(1-p_X) \leq 1/4$ and $p_Y(1-p_Y) \leq 1/4$. Solving $z_{\alpha/2}\sqrt{2/n} \leq 0.1$ gives

$$n \geq 2\left(\frac{z_2}{0.1}\right)^2 = 2\left(\frac{1.645}{0.1}\right)^2 = 541.205$$

so n has to be at least 542.

6. (a)
- $P(A) < P(A|B)$: sometimes true. It cannot be always true as if A and B are independent then we would have an equality. It cannot be never as if we consider A such that $P(A) < 1$ and $B = A$ the inequality is satisfied for example.

- $P(A) = P(A|B)$: true if and only if the two events are independent. Thus, sometimes true.

(b) Suppose now that $P(A|B) > P(A)$.

- $P(A \cap B) = 0$: never true as we would have $0 = \frac{P(A \cap B)}{P(B)} = P(A|B) > P(A) > 0$ which is absurd.
- $P(B|A) > P(B)$: always true as by assumption $\frac{P(A|B)}{P(A)} > 1$ implies $P(B|A) = \frac{P(A|B)}{P(A)} P(B) > P(B)$