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

MVE055 / MSG810 Matematisk statistik och diskret matematik

Exam 24 October 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. (a) Consider X = "number of men in the group", then X follows an hypergeometric distribution with parameters $N = 13, n = 4, r = 5$. The answer is then

$$\Pr(X = 2) = \frac{\binom{5}{2}\binom{8}{2}}{\binom{13}{4}} = 0.3916 \tag{1}$$

(b) We need to compute

$$\Pr(X = 0) = \frac{\binom{5}{0}\binom{8}{4}}{\binom{13}{4}} = 0.0979 \tag{2}$$

(c) Denote by (F, F, M, M) the sequence of female,female,male,male extracted. The first female is picked randomly from a set of 13 people with 8 women. So the probability that the sequence starts with female is $\frac{8}{13}$. The second female has to be picked from a set of 12 people, as we have removed the first woman selected in the group, of which 7 are female. Similarly, the probabilities for picking the two men are respectively $\frac{5}{11}$ and $\frac{4}{10}$. Thus, the probability of the sequence (F, F, M, M) is given by

$$\frac{8}{13}\frac{7}{12}\frac{5}{11}\frac{4}{10} = 0.0653$$

2. (a) The distribution of $\overline{X}$ is Normal$\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$, thus the probability distribution is given by

$$f_{\overline{X}}(x) = \frac{1}{\sqrt{2\pi\frac{\sigma_x^2}{n}}} \exp\left(-\frac{(x-\mu_x)^2}{2\frac{\sigma_x^2}{n}}\right) \tag{3}$$

(b) The distribution of $\overline{X} - \overline{Y}$ is Normal$\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right)$, thus the probability distribution is given by

$$f_{\overline{X}-\overline{Y}}(z) = \frac{1}{\sqrt{2\pi\left(\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right)}} \exp\left(-\frac{(z - \mu_x + \mu_y)^2}{2\left(\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right)}\right) \tag{4}$$

(c) From the previous point, and given that if $Z \sim$ Normal(μ, σ) then $\frac{Z-\mu}{\sigma} \sim$ Normal$(0, 1)$, we conclude that the denominator should be the standard deviation of $\overline{X} - \overline{Y}$, so the random variable is

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

(d) We know that

$$\Pr\left(-1.96 \le \frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \le 1.96\right) = 0.95 \tag{5}$$

which means

$$\Pr\left(-1.96\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \le \overline{X} - \overline{Y} - (\mu_x - \mu_y) \le 1.96\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right) = 0.95. \tag{6}$$

Hence,

$$\Pr\left(\overline{X} - \overline{Y} - 1.96\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \le (\mu_x - \mu_y) \le \overline{X} - \overline{Y} + 1.96\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right) = 0.95 \tag{7}$$

and finally we find

$$L_1 = \overline{X} - \overline{Y} - 1.96\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}, \quad L_2 = \overline{X} - \overline{Y} + 1.96\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

3. Define the following events:

- HD=the person has the disease, $\Pr(HD) = 0.01$, $\Pr(HD^c) = 0.99$ (HD^c means healthy);
- TP=the test is positive, $\Pr(TP|HD) = 0.8$, $\Pr(TP|HD^c) = 0.05$

(a)

$$\Pr(TP) = \Pr(TP \cap HD) + \Pr(TP \cap HD^c) = \\ \Pr(TP|HD)\Pr(HD) + \Pr(TP|HD^c)\Pr(HD^c) = 0.8 * 0.01 + 0.05 * 0.99 = 0.0575; \tag{8}$$

(b)

$$\Pr(HD|TP) = \frac{\Pr(TP \cap HD)}{\Pr(TP)} = \frac{\Pr(TP|HD)\Pr(HD)}{\Pr(TP)} = \frac{0.8 * 0.01}{0.0575} = 0.1391 \tag{9}$$

4. (a) The transition matrix is

$$P = \begin{bmatrix} 1 & 0 & \dots & & 0 \\ 0.75 & 0 & 0.25 & \dots & 0 \\ \vdots & \ddots & & & \\ 0 & \dots & 0.75 & 0 & 0.25 \\ 0 & \dots & 0 & 0.75 & 0.25 \end{bmatrix}$$

(b) For $N = 2$,

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0.75 & 0 & 0.25 \\ 0 & 0.75 & 0.25 \end{bmatrix}$$

thus

$$N = \begin{bmatrix} \frac{4}{3} & \frac{4}{9} \\ \frac{4}{3} & \frac{16}{9} \end{bmatrix}$$

and $\dot{N}[1,1]^T = [\frac{16}{9}, \frac{28}{9}]^T$. Thus, the expected time to extiction if the initial state is $X_0 = 2$ is $\frac{28}{9}$.

(c) Consider $Y_n = \max\{X_0, ..., X_n\}$. Y_n is not a Markov chain as the future is not independent of the past. Consider X_n as above with $X_0 = 2$ and $N = 3$, and take $Y_n = 1$ for some integer n, thus the maximum state reached by $X_m, m \leq n$ is 1. Now the probability that at the next step Y_{n+1} is equal to 2, correspond to the probability that $X_{n+1} = 2$. This probability, depends on all the $X_m, m \leq n$. In fact, consider the case in which we have $X_0 = 2, X_1 = 1, X_2 = 0$, i.e. we get extinct as fast as possible. Then the probability that $Y_3 = 2$ is zero, as the population cannot grow if there is no one left. On the other hand, consider $X_0 = 2, X_1 = 1, X_2 = 2$ (one death followed by a birth). Then $Y_3 = 2$ happens if we get a birth, i.e. with probability 0.25. Thus, the past has an effect on the future.

5. (a)

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, ..., n.$$

(b)

$$\begin{aligned}
\Pr[X=n] &= \sum_{k=0}^{n} P(X_A=k, X_B=n-k) = \sum_{k=0}^{n} P(X_A=k)P(X_B=n-k) = \\
&\sum_{k=0}^{n} e^{-\lambda_A}\frac{\lambda_A^k}{k!}e^{-\lambda_B}\frac{\lambda_B^{n-k}}{(n-k)!} = e^{-(\lambda_A+\lambda_B)}\sum_{k=0}^{n}\frac{\lambda_A^k}{k!}\frac{\lambda_B^{n-k}}{(n-k)!} = \\
&e^{-(\lambda_A+\lambda_B)}\sum_{k=0}^{n}\binom{n}{k}\frac{1}{n!}\lambda_A^k\lambda_B^{n-k} = e^{-(\lambda_A+\lambda_B)}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda_A^k\lambda_B^{n-k} = \\
&= e^{-(\lambda_A+\lambda_B)}\frac{(\lambda_A+\lambda_B)^n}{n!}
\end{aligned} \tag{10}$$

where in the last passage we used the hint:

$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{(n-k)} \tag{11}$$

(c)

$$\begin{aligned}
P[X_A=k|X=n] =& \frac{P(X_A=k, X=n)}{P(X=n)} = \frac{P(X_A=k, X_B=n-k)}{P(X=n)} = \\
&= \frac{e^{-\lambda_A}\frac{\lambda_A^k}{k!}e^{-\lambda_B}\frac{\lambda_B^{n-k}}{(n-k)!}}{e^{-(\lambda_A+\lambda_B)}\frac{(\lambda_A+\lambda_B)^n}{n!}} = \binom{n}{k}\frac{\lambda_A^k}{(\lambda_A+\lambda_B)^k}\frac{\lambda_B^{n-k}}{(\lambda_A+\lambda_B)^{n-k}} = \\
&= \binom{n}{k}\left(\frac{\lambda_A}{(\lambda_A+\lambda_B)}\right)^k\left(1-\frac{\lambda_A}{(\lambda_A+\lambda_B)}\right)^{n-k}
\end{aligned} \tag{12}$$

We obtained a binomial distribution with probability of success equal to $\frac{\lambda_A}{(\lambda_A+\lambda_B)}$.

6. Assume the continuous random variable has a probability distribution with expectation μ and variance σ^2, and assume $X_1, \ldots, X_n$ is a random sample from this distribution.

(a)

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2. \tag{13}$$

(b)

$$E\left[(X_1-\overline{X})^2\right] = Var(X_1-\overline{X}) + \left(E[X_1-\overline{X}]\right)^2 \tag{14}$$

where the second term on the right hand side above is

$$E[X_1-\overline{X}] = E[X_1] - \frac{1}{n}\sum_{i=1}^{n}E[X_i] = \mu - \frac{n\mu}{n} = 0 \tag{15}$$

and by using independence we handle the first term

$$Var(X_1 - \overline{X}) = Var\left(\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{i=2}^{n} X_i\right) = Var\left(\frac{n-1}{n}X_1\right) + Var\left(\frac{1}{n}\sum_{i=2}^{n} X_i\right)$$
$$= \left(\frac{n-1}{n}\right)^2 \sigma^2 + \frac{n-1}{n^2}\sigma^2 = \frac{n-1}{n}\sigma^2 \tag{16}$$

(c)

$$E[s^2] = \frac{1}{n-1}\sum_{i=1}^{n} E\left[(X_i - \overline{X})^2\right] = \frac{1}{n-1}n\frac{n-1}{n}\sigma^2 = \sigma^2 \tag{17}$$

and thus is an unbiased estimator.