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

MVE055 / MSG810 Matematisk statistik och diskret matematik

Exam 6 April 2018, 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. We know $\mathbb{E}[X_i] = \mu,\ Var(X_i) = \sigma^2\ \forall i = 1, ..., n$.

   (a) E_1 is an unbiased estimators of μ. In fact,

   $$\mathbb{E}[E_1] = \frac{1}{2}\left[\mathbb{E}[X_1] + \frac{3}{4}\mathbb{E}[X_2] + \frac{1}{4}\mathbb{E}[X_3]\right] = \frac{1}{2}\left[\mu + \frac{3}{4}\mu + \frac{1}{4}\mu\right] = \mu.$$

   E_2 is instead a biased estimators of μ.

   $$\mathbb{E}[E_2] = \mathbb{E}[X_1^2 - X_1] = \mathbb{E}[X_1]^2 + Var(X_1)^2 - \mathbb{E}[X_1] = \mu^2 + \sigma^2 - \mu.$$

   E_3 is an unbiased estimator for μ as

   $$\mathbb{E}[E_3] = \frac{1}{6}\sum_{i=1}^{3} i\mathbb{E}[X_i] = \frac{\mu}{6}\sum_{i=1}^{3} i = \frac{\mu}{6}\cdot(1+2+3) = \mu.$$

   (b) Using the independence property

   $$Var(E_1) = Var\left(\frac{1}{2}X_1 + \frac{3}{8}X_2 + \frac{1}{8}X_3\right) = \sigma^2\left(\frac{1}{4} + \frac{9}{64} + \frac{1}{64}\right) = \frac{13}{32}\sigma^2$$

   $$Var(E_3) = \frac{1}{36}(Var(X_1 + 2X_2 + 3X_3) = \frac{1}{36}(\sigma^2 + 4\sigma^2 + 9\sigma^2) = \frac{7}{18}\sigma^2.$$

   We conclude that E_3 is the unbiased estimator with minimum variance among the unbiased estimators presented above.

2. (a) The random variables $X_1, ..., X_{10}$ cannot be independent as we have $X_1 = 500 - \sum_{i=2}^{10} X_i$, which implies that the correlation coefficient $\rho(X, Y) = -1$ as there exists a linear relationship between X and Y. If $X_1, ..., X_{10}$ were independent then we would have $\rho(X, Y) = 0$. As this is not the case we conclude they are dependent.

(b) The distribution of Z is not binomial. In particular, it is impossible that $Z = 10$ as we would not have a total weight of 500 grams. Thus, Z cannot have a binomial distribution.

3. Define the events: CH = email received on Chalmers account, GM = email received on gmail account, GU = email received on GU account, S = mail is spam.

   (a) $P(S) = P(S|GM)P(GM) + P(S|CH)P(CH) + P(S|GU)P(GU) = 0.4 \cdot 0.02 + 0.35 \cdot 0.01 + 0.25 \cdot 0.05 = 0.024$.

   (b) $P(GM|S) = \frac{P(S|GM)P(GM)}{P(S)} = \frac{0.4 \cdot 0.02}{0.024} = 0.333$.

4. Define the generating function $f(x) = \sum_{n=0}^{\infty} a_n x^n$. By multiplying the equation describing a_n by x_n and summing over all such equations we obtain

$$f(x) = 3xf(x) + \frac{2x}{1-x} + 5$$

   which can be solved for $f(x)$ giving

$$f(x) = \frac{5-3x}{(1-x)(1-3x)} = \frac{6}{1-3x} - \frac{1}{1-x}.$$

   Thus, $a_n = 6 \cdot 3^n - 1$.

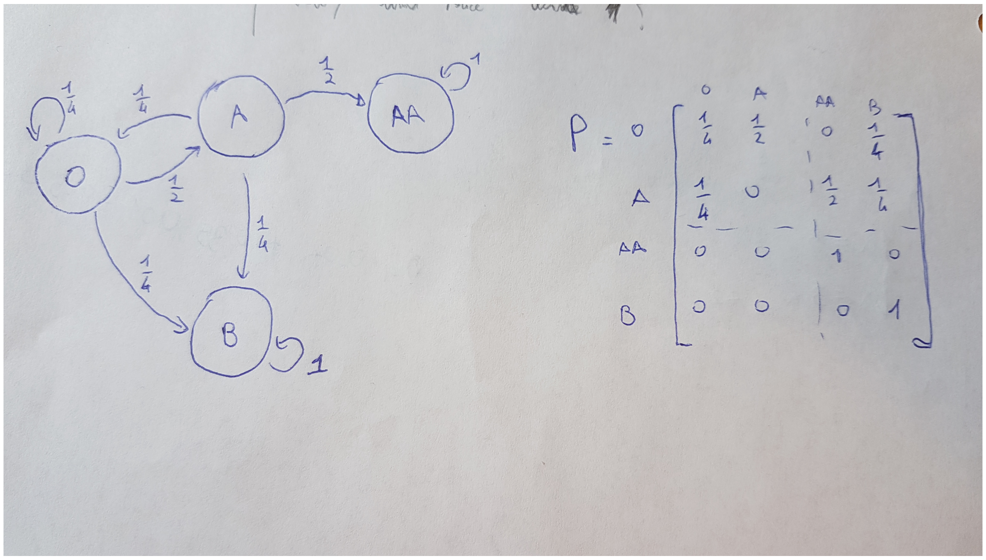
   (a) $\hat{p} = \frac{120}{541} = 0.2218$ is the observed proportion of obese people in the sample of size $n = 541$ and by p the true proportion in the population. Consider the one sided test H_0: $p = p_0 = 0.2$, $H_1 : p > 0.2$. Under the null hypothesis we have $np = 541 \cdot 0.2 = 108.2 \geq 5$ and hence we can use the normal approximation as the sample size is large enough. Consider the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

   We reject the null hypothesis if $Z \geq z_{0.05} = 1.645$. Since the observed value of Z is 1.27 we fail to reject the null hypothesis at a 5% level.

   (b) Type I error: we conclude that more than 20% of the people in the population are obese when in fact the true percentage is 20%. Type II error: we conclude that the percentage of obese people is not more than 20% when in fact it is higher. Since we fail to reject the null hypothesis, we are subject to type II errors.

5. Consider the following Markov Chain with states as in the Penney's game: 0 =START, A=the latest card picked was a diamonds or hearts, AA=the two last cards picked were diamonds or hearts (Alice wins), B=the latest card picked was a Spades (Bob wins). See the picture below for the transition matrix P for this markov chain in canonical form and the corresponding representation



The fundamental matrix N is then given by

$$N = \begin{bmatrix} \frac{8}{5} & \frac{4}{5} \\ \frac{2}{5} & \frac{6}{5} \end{bmatrix}.$$

Hence,

$$NR = \begin{bmatrix} \frac{8}{5} & \frac{4}{5} \\ \frac{2}{5} & \frac{6}{5} \end{bmatrix} \cdot \begin{bmatrix} 0 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{2}{5} \end{bmatrix}.$$

The probability that Alice will win is then $\frac{2}{5} = 0.4$.