

Petter Mostad
Mathematical Statistics
Chalmers and GU

Solution to
MVE055/MSG810 Mathematical statistics and discrete mathematics
MVE055/MSG810 Matematisk statistik och diskret matematik

Re-exam: 24 August 2016, 14:00 - 18:00

1. (a) One can use the Hypergeometric distribution:

$$\frac{\binom{3}{3}\binom{3}{1}}{\binom{6}{4}} = \frac{1}{5} = 0.2.$$

Alternatively, the probability can be computed by using that it is the same as sequentially choosing two non-defect items out of the 6, i.e.,

$$\frac{3}{6} \cdot \frac{2}{5} = \frac{1}{5} = 0.2.$$

- (b) One can use the Hypergeometric distribution repeatedly. Alternatively, one can compute the probability of the particular sequence of events described in the question. It is:

$$\frac{3}{6} \cdot \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{20} = 0.05.$$

2. A confidence interval for a proportion p with significance level α is

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1 - \hat{p})/n}$$

where \hat{p} is the observed frequency, n is the sample size, and $z_{\alpha/2}$ is the corresponding quantile of the standard normal distribution. In our case, $\hat{p} = 9/23$, $n = 23$, and $z_{\alpha/2} = z_{0.025} = 1.96$ so we get

$$9/23 \pm 1.96 \sqrt{9/23(1 - 9/23)/23} = 0.3913 \pm 0.1995,$$

i.e., [0.1918, 0.5908].

3. (a) A 95% confidence interval:

$$\begin{aligned} & [7.43 - z_{0.025} \sqrt{0.041/7}, 7.43 + z_{0.025} \sqrt{0.041/7}] \\ & = [7.43 - 1.96 \cdot 0.07653, 7.43 + 1.96 \cdot 0.07653] = [7.28, 7.58] \end{aligned}$$

and a 99% confidence interval:

$$\begin{aligned} & [7.43 - z_{0.005} \sqrt{0.041/7}, 7.43 + z_{0.005} \sqrt{0.041/7}] \\ & = [7.43 - 2.58 \cdot 0.07653, 7.43 + 2.58 \cdot 0.07653] = [7.23, 7.63] \end{aligned}$$

(b) A 95% confidence interval:

$$\begin{aligned} & [7.43 - t_{6,0.025} \sqrt{0.039/7}, 7.43 + t_{6,0.025} \sqrt{0.039/7}] \\ & = [7.43 - 2.45 \cdot 0.07464, 7.43 + 2.45 \cdot 0.07464] = [7.24, 7.61] \end{aligned}$$

4. (a) The distribution can be computed as

$$\begin{bmatrix} 0.1 \\ 0.3 \\ 0 \\ 0.6 \end{bmatrix}' \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix}^2 = \begin{bmatrix} 0.31 \\ 0.11 \\ 0.15 \\ 0.61 \end{bmatrix}' \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.057 \\ 0.059 \\ 0.182 \\ 0.702 \end{bmatrix}'$$

(b) We get

$$\Pr(X_0 = 1) \Pr(X_1 = 2 | X_0 = 1) \Pr(X_2 = 3 | X_1 = 2) = 0.1 \cdot 0.2 \cdot 0.2 = 0.004.$$

(c) The chain is neither ergodic nor absorbing.

(d) We get

$$\Pr(X_0 = 2 | X_1 = 1) = \frac{\Pr(X_1 = 1 | X_0 = 2) \Pr(X_0 = 2)}{\Pr(X_1 = 1)} = \frac{0.4 \cdot 0.3}{0.31} = 0.387.$$

5. The likelihood for λ given data x_1, \dots, x_n is

$$L(\lambda) = e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \cdot \dots \cdot e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} = e^{-n\lambda} \frac{\lambda^{x_1 + \dots + x_n}}{x_1! \dots x_n!} = C \exp(-n\lambda) \lambda^S$$

where $C = x_1! \dots x_n!$ and $S = x_1 + \dots + x_n$. So the derivative of the likelihood is

$$L'(\lambda) = C \left(\exp(-n\lambda) S \lambda^{S-1} - n \exp(-n\lambda) \lambda^S \right) = C \exp(-n\lambda) \lambda^{S-1} (S - n\lambda)$$

Setting $L'(\lambda) = 0$ gives $\lambda = S/n$, so the maximum likelihood estimator for λ is S/n . We get that

$$\mathbf{E}\left(\frac{S}{n}\right) = \frac{1}{n} \mathbf{E}(S) = \frac{1}{n} \cdot n \mathbf{E}(x_i) = \lambda$$

so the ML estimator is unbiased.

6. (a) We have that

$$\begin{aligned} 1 &= \int_0^\theta \frac{ax}{\theta} dx + \int_\theta^1 \frac{a(1-x)}{1-\theta} dx = \frac{a}{\theta} \left[\frac{x^2}{2} \right]_0^\theta + \frac{a}{1-\theta} \left[x - \frac{x^2}{2} \right]_\theta^1 \\ &= \frac{a\theta^2}{2\theta} + \frac{a}{1-\theta} \left(\frac{1}{2} - \theta + \frac{\theta^2}{2} \right) = \frac{a\theta}{2} + \frac{a(1-\theta)^2}{2(1-\theta)} = \frac{a}{2} \end{aligned}$$

which shows that $a = 2$. Alternatively, we can see that the density function looks like two triangles, with the same height a and with total length of baselike equal to 1. Thus their area is $a/2$, and for this area to be 1, we need to have $a = 2$.

(b) We have

$$\begin{aligned} \mathbf{E}(X) &= \int_0^\theta x \frac{ax}{\theta} dx + \int_\theta^1 x \frac{a(1-x)}{1-\theta} dx = \frac{a}{\theta} \left[\frac{x^3}{3} \right]_0^\theta + \frac{a}{1-\theta} \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_\theta^1 \\ &= \frac{a\theta^3}{3\theta} + \frac{a}{1-\theta} \left(\frac{1}{2} - \frac{1}{3} - \frac{\theta^2}{2} + \frac{\theta^3}{3} \right) = \frac{a}{3}\theta^2 + \frac{a}{6(1-\theta)}(1 - 3\theta^2 + 2\theta^3) \\ &= \frac{a}{3}\theta^2 + \frac{a}{6}(1 + \theta - 2\theta^2) = \frac{a}{6}(1 + \theta) \end{aligned}$$

7. We get

$$\begin{aligned} \mathbf{E}(\cos X) &= \int_0^\pi \cos x \frac{1}{\pi} dx = \frac{1}{\pi} [\sin x]_0^\pi = 0 \\ \mathbf{Var}(\cos X) &= \mathbf{E}(\cos^2 X) - \mathbf{E}(\cos X)^2 = \int_0^\pi \cos^2 x \frac{1}{\pi} dx = \frac{1}{\pi} \int_0^\pi \left(\frac{1}{2} + \frac{1}{2} \cos(2x) \right) dx \\ &= \frac{1}{\pi} \left[\frac{1}{2}x + \frac{1}{4} \sin(2x) \right]_0^\pi = \frac{\pi}{2\pi} = \frac{1}{2} \\ \mathbf{E}(\sin X) &= \int_0^\pi \sin x \frac{1}{\pi} dx = \frac{1}{\pi} [-\cos x]_0^\pi = \frac{2}{\pi} \\ \mathbf{Var}(\sin X) &= \mathbf{E}(\sin^2 X) - \mathbf{E}(\sin X)^2 = \int_0^\pi \sin^2 x \frac{1}{\pi} dx - \frac{4}{\pi^2} \\ &= \frac{1}{\pi} \int_0^\pi \left(\frac{1}{2} - \frac{1}{2} \cos(2x) \right) dx - \frac{4}{\pi^2} \\ &= \frac{1}{\pi} \left[\frac{1}{2}x - \frac{1}{4} \sin(2x) \right]_0^\pi - \frac{4}{\pi^2} = \frac{\pi}{2\pi} - \frac{4}{\pi^2} = \frac{1}{2} - \frac{4}{\pi^2} \end{aligned}$$

8. (a) When x increases, $x/(1+x^2)$ goes to zero at the rate of $1/x$: Specifically, we prove that, for $x > 1$,

$$\frac{x}{1+x^2} > \frac{1}{2x}$$

by noting that $2x^2 > 1+x^2$ when $x > 1$. Thus we get

$$\int_0^\infty x \frac{1}{\pi(1+x^2)} dx = \frac{1}{\pi} \int_0^\infty \frac{x}{1+x^2} dx > \frac{1}{\pi} \int_0^\infty \frac{1}{2x} dx = \frac{1}{\pi} \left[\frac{1}{2} \log(2x) \right]_0^\infty = \infty$$

so the expectation does not exist.

- (b) If we had started with a Beta distribution or a uniform distribution, the histograms would have been bell-shaped, i.e., they would look like samples from normal distributions. The reason is the Central Limit Theorem. However, for the Central Limit Theorem to apply, the distribution we start with must have an existing expectation and an existing variance. Thus, the CLT does not apply to the Cauchy distribution.

9. (a) $\mathbf{P}(X > 1) = \mathbf{P}(X = 2) + \mathbf{P}(X = 3) = 0.2 + 0.1 = 0.3$
 (b) $\mathbf{P}(XY > 0) = \mathbf{P}(X > 0 \text{ and } Y > 0) = \mathbf{P}(X > 0) \mathbf{P}(Y > 0) = (1 - 0.4) \cdot 0.3 = 0.18$
 (c)

$$\begin{aligned} \mathbf{P}(X + Y = 1) &= \mathbf{P}(X = 0 \text{ and } Y = 1) + \mathbf{P}(X = 1 \text{ and } Y = 0) + \mathbf{P}(X = 2 \text{ and } Y = -1) \\ &= \mathbf{P}(X = 0) \mathbf{P}(Y = 1) + \mathbf{P}(X = 1) \mathbf{P}(Y = 0) + \mathbf{P}(X = 2) \mathbf{P}(Y = -1) \\ &= 0.4 \cdot 0.3 + 0.3 \cdot 0.4 + 0.2 \cdot 0.3 = 0.3 \end{aligned}$$

(d) $\mathbf{E}(X) = 0 \cdot 0.4 + 1 \cdot 0.3 + 2 \cdot 0.2 + 3 \cdot 0.1 = 1$

(e) $\mathbf{E}(XY) = \mathbf{E}(X) \mathbf{E}(Y) = 0$ as $\mathbf{E}(Y) = 0$.

(f) $\mathbf{Var}(Y) = \mathbf{E}(Y^2) - \mathbf{E}(Y)^2 = \mathbf{E}(Y^2) = 0^2 \cdot 0.4 + 1^2 \cdot 0.3 + (-1)^2 \cdot 0.3 = 0.6$

10. Assume X_1, X_2, \dots, X_n is a random sample from some probability distribution, and assume this distribution is from a family of distributions parametrized by parameters $\theta_1, \dots, \theta_k$. The purpose of the Method of Moments is to construct functions $\hat{\theta}_1, \dots, \hat{\theta}_k$ of the random sample that can work as estimators for the parameters $\theta_1, \dots, \theta_k$. The idea is the following: If M_1, \dots, M_s denote the first s moments of a distribution in the parametric family, then these depend on the parameters $\theta_1, \dots, \theta_k$, and one can obtain formulas expressing relating M_1, \dots, M_s to $\theta_1, \dots, \theta_k$. In these formulas, one may make the replacement

$$M_j \approx \frac{1}{n} \sum_{i=1}^n X_i^j$$

and solve for the parameters $\theta_1, \dots, \theta_k$ in order to obtain estimators.

As an example, consider the Negative Binomial distribution with parameters r and p , where r is a positive integer and $p \in (0, 1)$. The expressions for its expectation and variance gives us

$$\begin{aligned} M_1 &= r/p \\ M_2 - M_1^2 &= r(1 - p)/p^2 \end{aligned}$$

Solving for the parameters and making the substitutions, one obtain formulas for example on the form

$$\begin{aligned} p &= \frac{M_1}{M_2 - M_1^2 + M_1} = \frac{\bar{X}}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 + \bar{X}} \\ r &= \frac{M_1^2}{M_2 - M_1^2 + M_1} = \frac{\bar{X}^2}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 + \bar{X}} \end{aligned}$$

with the restriction that r must be an integer.

11. The statement of the Central Limit Theorem (CLT) given in Milton and Arnold is: Let X_1, X_2, \dots, X_n be a random sample of size n from a distribution with mean μ and variance σ^2 . Then for large n , \bar{X} is approximately normal with mean μ and variance σ^2/n . Furthermore, for large n , the random variable $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is approximately standard normal.

There are many practical effects of the CLT. One very fundamental is that many variables measured in practice will tend to have a normal distribution, as their values can be modelled as the sum of many small variables that are more or less independent. (Example: Weight of a bag of chips supposed to weigh 200 grams).

12. (a) $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$.
 (b)

$$\begin{aligned} m(t) &= \mathbf{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 - 2tx)\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 - 2tx + t^2) + \frac{1}{2}t^2\right) dx \\ &= e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x - t)^2\right) dx \\ &= e^{\frac{1}{2}t^2} \end{aligned}$$

(c)

$$\begin{aligned} m'(t) &= te^{\frac{1}{2}t^2} \\ m''(t) &= e^{\frac{1}{2}t^2} + t^2 e^{\frac{1}{2}t^2} = (1 + t^2)e^{\frac{1}{2}t^2} \\ m'''(t) &= 2te^{\frac{1}{2}t^2} + (1 + t^2)te^{\frac{1}{2}t^2} = (3t + t^3)e^{\frac{1}{2}t^2} \end{aligned}$$

yields

$$\begin{aligned} M_1 &= m'(0) = 0 \\ M_2 &= m''(0) = 1 \\ M_3 &= m'''(0) = 0 \end{aligned}$$

13. (a) A confidence interval for μ is given by

$$\bar{x} \pm z_{0.025}\sigma/\sqrt{n} = 16.189 \pm 1.96 \cdot 3.7/\sqrt{9} = 16.189 \pm 2.417,$$

in other words, [13.772, 18.606].

(b) The length of the interval will be

$$2 \cdot z_{0.025}\sigma / \sqrt{41 + 9} = 2 \cdot 1.96 \cdot 3.7 / \sqrt{50} = 2.051$$

(c) We get

$$\begin{aligned} 2 \cdot z_{0.005}\sigma / \sqrt{n} &= 1.1 \\ 2 \cdot 2.58 \cdot 3.7 / 1.1 &= \sqrt{n} \\ 17.356^2 &= n \\ n &= 301 \end{aligned}$$

so one should sample a total of 301 values.

(d) The sample standard deviation for the numbers is $s = 3.606$. The 95% confidence interval becomes

$$\bar{x} \pm t_{8,0.025}s / \sqrt{n} = 16.189 \pm 2.306 \cdot 3.606 / \sqrt{9} = 16.189 \pm 2.772,$$

in other words, [13.417, 18.961].

14. (a) $\mathbf{E}(X_i) = p \cdot 1 + (1 - p) \cdot 0 = p$ and $\mathbf{Var}(X_i) = \mathbf{E}(X_i^2) - \mathbf{E}(X_i)^2 = p \cdot 1^2 + (1 - p) \cdot 0^2 - p^2 = p(1 - p)$.
- (b) $\mathbf{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}(X_i) = \frac{1}{n} \sum_{i=1}^n p = p$ and $\mathbf{Var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \mathbf{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n p(1 - p) = p(1 - p)/n$.
- (c) By the central limit theorem, \bar{X} has an approximately normal distribution, and by the above, it is then approximately distributed as a normal distribution with expectation p and variance $p(1 - p)/n$.
- (d) From the above, we get that

$$\mathbf{P}\left(p - z_{\alpha/2} \sqrt{p(1 - p)/n} \leq \bar{x} \leq p + z_{\alpha/2} \sqrt{p(1 - p)/n}\right) \approx 1 - \alpha.$$

As $p \approx \bar{x}$, we substitute some p 's with \bar{x} and get

$$\mathbf{P}\left(p - z_{\alpha/2} \sqrt{\bar{x}(1 - \bar{x})/n} \leq \bar{x} \leq p + z_{\alpha/2} \sqrt{\bar{x}(1 - \bar{x})/n}\right) \approx 1 - \alpha,$$

and thus

$$\mathbf{P}\left(\bar{x} - z_{\alpha/2} \sqrt{\bar{x}(1 - \bar{x})/n} \leq p \leq \bar{x} + z_{\alpha/2} \sqrt{\bar{x}(1 - \bar{x})/n}\right) \approx 1 - \alpha.$$

In particular, we get, for $\alpha = 0.05$,

$$\mathbf{P}\left(\bar{x} - 1.96 \sqrt{\bar{x}(1 - \bar{x})/n} \leq p \leq \bar{x} + 1.96 \sqrt{\bar{x}(1 - \bar{x})/n}\right) \approx 1 - \alpha.$$

so

$$\bar{x} \pm 1.96 \sqrt{\bar{x}(1 - \bar{x})/n}$$

is a confidence interval with confidence degree 95%.

15. (a) A simple description of simple linear regression is that one tries to fit a straight line to a set of data points in the plane. More precisely the best-fitting line is considered to be the line such that the sum of the squares of the vertical distances between the points and the line is minimized. Such a line represents the least squares solution.
- (b) The definition of S shows that it is the sum of the residuals $y_i - \widehat{y}_i$ of the regression. One may remember that this sum is always zero. However, one may also show directly that $S = 0$: Assume S is not zero. Then there exists an $\epsilon \neq 0$ such that $\sum_{i=1}^n (y_i - \widehat{y}_i + \epsilon) = 0$. But then

$$\begin{aligned}
 \sum_{i=1}^n (y_i - \widehat{y}_i)^2 &= \sum_{i=1}^n (y_i - \widehat{y}_i + \epsilon - \epsilon)^2 \\
 &= \sum_{i=1}^n \left[(y_i - \widehat{y}_i + \epsilon)^2 - 2(y_i - \widehat{y}_i + \epsilon)\epsilon + \epsilon^2 \right] \\
 &= \sum_{i=1}^n (y_i - \widehat{y}_i + \epsilon)^2 - 2\epsilon \sum_{i=1}^n (y_i - \widehat{y}_i + \epsilon) + n\epsilon^2 \\
 &= \sum_{i=1}^n (y_i - \widehat{y}_i + \epsilon)^2 + n\epsilon^2.
 \end{aligned}$$

Thus the line going through the points $(x_1, \widehat{y}_1 - \epsilon), (x_2, \widehat{y}_2 - \epsilon), \dots, (x_n, \widehat{y}_n - \epsilon)$ has a smaller sum of squares than the original regression line. This is a contradiction, proving that S is indeed zero.

16. • P_1 is a quadratic matrix of non-negative numbers with rows summing to 1, so it is a transition matrix for a Markov chain. It is not absorbing as it does not have any absorbing states. It is ergodic, and also regular, as it has only positive entries.
- P_2 has some negative values, so it is not a transition matrix.
- P_3 is a quadratic matrix of non-negative numbers with rows summing to 1, so it is a transition matrix. It is an absorbing chain, as it has an absorbing state and all states has a positive probability ending up in the absorbing state. It is not ergodic and not regular.
- P_4 is a quadratic matrix of non-negative numbers with rows summing to 1, so it is a transition matrix. It has an absorbing state, but it is not absorbing as the other states have zero probability of entering the absorbing state. It is not ergodic, and not regular.
- P_5 is a quadratic matrix of non-negative numbers with rows summing to 1, so it is a transition matrix. it is not absorbing. It is however ergodic. It is also regular, as one can go from any state to any other state in at most two steps.
- P_6 does not have rows summing to 1, so it is not a transition matrix of a Markov chain.