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: 4 January 2016, 14:00 - 18:00, M building

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

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

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

with the restriction that $r$ must be an integer.

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

   There are many practial effects of the CTL. 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).

4. (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

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

5. (a) A confidence interval for $\mu$ is given by

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

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

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

in other words, [13.417, 18.961].

6. (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) $\mathbf{E}(\overline{X}) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}(X_i) = \frac{1}{n}\sum_{i=1}^n p = p$ and $\mathbf{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^n \mathbf{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^n p(1-p) = p(1-p)/n$.

(c) By the central limit theorem, $\overline{X}$ 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

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

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

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

and thus

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

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

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

so

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

is a confidence interval with confidence degree 95%.

7. (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 $(x_1, \widehat{y_1} - \epsilon), (x_2, \widehat{y_2} - \epsilon), \ldots, (x_n, \widehat{y_n} - \epsilon)$ has a smaller sum of squares than the original regression line. This is a contradiction, proving that $S$ is indeed zero.

8. • $P_1$ 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.