Tutorial Letter 201/2/2014

Applied Statistics II

STA2601

Semester 2

Department of Statistics

Solutions to Assignment 1

BAR CODE

QUESTION 1

(a) (i) Let $P(4) = x$, then $P(1) = 2x$

Now

$$\begin{aligned}
\sum_{x \in A} f_X(x) &= 1 \\
P(0) + P(1) + P(2) + P(3) + P(4) &= 1 \\
0.1 + 2x + 0.4 + 0.2 + x &= 1 \\
3x + 0.7 &= 1 \\
3x &= 1 - 0.7 \\
3x &= 0.3 \\
\implies x &= 0.1
\end{aligned}$$

(ii) The probability distribution of X is

X	0	1	2	3	4
$P(x)$	0.1	0.2	0.4	0.2	0.1

(iii) The mean value of X is calculated as:

$$\begin{aligned}
\mu &= E\left(x\right) = \sum_x xP(X = x) \\
&= 0\left(0.1\right) + 1\left(0.2\right) + 2\left(0.4\right) + 3\left(0.2\right) + 4\left(0.1\right) \\
&= 0 + 0.2 + 0.8 + 0.6 + 0.4 \\
&= 2
\end{aligned}$$

(iv) Standard deviation, $\sigma = \sqrt{\sigma^2} = \sqrt{\text{variance}}$.

Now the *variance* of X is calculated as:

$$\begin{aligned}
\sigma^2 &= \sum_{x \in A} \left(x - \mu\right)^2 f_X\left(x\right) \\
&= \left(0 - 2\right)^2 \times 0.1 + \left(1 - 2\right)^2 \times 0.2 + \left(2 - 2\right)^2 \times 0.4 + \left(3 - 2\right)^2 \times 0.2 + \left(4 - 2\right)^2 \times 0.1 \\
&= \left(-2\right)^2 \times 0.1 + \left(-1\right)^2 \times 0.2 + \left(0\right)^2 \times 0.4 + \left(1\right)^2 \times 0.2 + \left(2\right)^2 \times 0.1 \\
&= 0.4 + 0.2 + 0 + 0.2 + 0.4 \\
&= 1.2
\end{aligned}$$

OR

$$\begin{aligned}
\sigma^2 &= E\left(X^2\right) - (E(X))^2 \\
&= \sum_x x^2 \, f_X(x) - \mu^2 \\
&= 0^2(0.1) + 1^2(0.2) + 2^2(0.4) + 3^2(0.2) + 4^2(0.1) - 2^2 \\
&= 0 + 0.2 + 1.6 + 1.8 + 1.6 - 4 \\
&= 5.2 - 4 \\
&= 1.2
\end{aligned}$$

$\Longrightarrow$Standard deviation, $\sigma = \sqrt{1.2} \approx 1.0954$

(v) The coefficient of skewness of X is calculated as $\beta_1 = \frac{\mu_3}{\sigma^3}$. Now $\mu_3 =$ the *third central moment* of X is calculated as:

$$\begin{aligned}
\mu_3 &= \sum_{x \in A} (x-\mu)^3 \, P(X=x) \\
&= (0-2)^3 \times 0.1 + (1-2)^3 \times 0.2 + (2-2)^3 \times 0.4 + (3-2)^3 \times 0.2 + (4-2)^3 \times 0.1 \\
&= (-2)^3 \times 0.1 + (-1)^3 \times 0.2 + (0)^3 \times 0.4 + (1)^3 \times 0.2 + (2)^3 \times 0.1 \\
&= -0.8 - 0.2 + 0 + 0.2 + 0.8 \\
&= 0
\end{aligned}$$

Thus,

$$\begin{aligned}
\beta_1 &= \frac{\mu_3}{\sigma^3} \\
&= \frac{0}{\left(\sqrt{(1.2)}\right)^3} \\
&= \frac{0}{1.314534138} \\
&= 0
\end{aligned}$$

Data is symmetrical.

(b) $X_i \sim n(\mu; \sigma^2)$ for $i = 1;\ 2;\ 3$ and $4.$

$\therefore E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for $i = 1;\ 2;\ 3$ and $4.$

(i)

$$\begin{aligned} E(T_1) &= E\left[\frac{X_1 + X_2 + X_3 + X_4}{4}\right] \\ &= \tfrac{1}{4}\left[E(X_1) + E(X_2) + E(X_3) + E(X_4)\right] \\ &= \tfrac{1}{4}[\mu + \mu + \mu + \mu] \\ &= \tfrac{1}{4}[4\mu] \\ &= \mu \\ E(T_2) &= E\left[\frac{X_1 + 2X_2 + 3X_3 - 2X_4}{4}\right] \\ &= \tfrac{1}{4}\left[E(X_1) + 2E(X_2) + 3E(X_3) - 2E(X_4)\right] \\ &= \tfrac{1}{4}[\mu + 2\mu + 3\mu - 2\mu] \\ &= \tfrac{1}{4}[4\mu] \\ &= \mu \end{aligned}$$

T_1 and T_2 are all unbiased estimators of μ.

(ii) If X and Y are stochastically independent variables, then $Var(aX + bY) = a^2Var(X) + b^2Var(Y)$

Thus
$$\begin{aligned} Var(T_1) &= Var\left[\frac{X_1 + X_2 + X_3 + X_4}{4}\right] \\ &= \left[\frac{1}{4}\right]^2 Var(X_1) + \left[\frac{1}{4}\right]^2 Var(X_2) + \left[\frac{1}{4}\right]^2 Var\left(X_3\right) + \left[\frac{1}{4}\right]^2 Var\left(X_4\right) \\ &= \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 \\ &= \frac{1}{4}\sigma^2 \\ &= 0.25\sigma^2. \end{aligned}$$

$$
\begin{aligned}
\text{Similarly } Var(T_2) &= Var\left[\frac{X_1 + 2X_2 + 3X_3 - 2X_4}{4}\right] \\
&= \left[\frac{1}{4}\right]^2 Var(X_1) + \left[\frac{2}{4}\right]^2 Var(X_2) + \left[\frac{3}{4}\right]^2 Var\left(X_3\right) + \left[\frac{-2}{4}\right]^2 Var\left(X_4\right) \\
&= \frac{1}{16}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{9}{16}\sigma^2 + \frac{1}{4}\sigma^2 \\
&= \frac{9\sigma^2}{8} \\
&= 1.1252\sigma^2.
\end{aligned}
$$

$\therefore \quad Var(T_1) < Var(T_2) \Longrightarrow T_1$ is a more efficient estimator for μ.

QUESTION 2

(a) True. The variance of any distribution is the sum of the squares of the deviation.

(b) False. There are some highly sophisticated method that can also be used and the method of least squares is used especially in problems where the unknown function are linear functions of known constants.

(c) False. A type I error is committed when H_0 is rejected when in actual fact it is true.

(d) False. There is a positive correlation between the length (in cm) and the mass (in kg) of a child.

(e) True. A correlation coefficient of zero, i.e., $r = 0$ indicates absence of a linear relationship.

(f) False. Correlation does not mean causation, i.e., correlation does not imply causality. Thus, one can not draw cause and effect conclusions based on correlation.

(g) False. The assumptions underlying a one-way analysis of variance are

- observations are independent (given).
- data comes from a normal population.
- equal population variances.

QUESTION 3

(a) Yes. $f_{X_2}(x_2) = f_{X_5}(x_5)$

(b)

$$\begin{aligned} f_{X_4}(x_4) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_4-\mu}{\sigma}\right)^2} \\ &= \frac{1}{5\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_4-60}{5}\right)^2} \qquad -\infty < x_4 < \infty \end{aligned}$$

(c)

$$\begin{aligned} P(X_3 < 65) &= P\left(\frac{X_3-\mu}{\sigma} < \frac{65-\mu}{\sigma}\right) \\ &= P\left(\frac{X_3-\mu}{\sigma} < \frac{65-60}{5}\right) \\ &= P\left(Z < \frac{5}{5}\right) \\ &= P(Z < 1) \\ &= 0.8413 \end{aligned}$$

(d)

$$\begin{aligned} P(55 < X_3 < 72) &= P\left(\frac{55-\mu}{\sigma} < \frac{X_3-\mu}{\sigma} < \frac{72-\mu}{\sigma}\right) \\ &= P\left(\frac{55-60}{5} < Z < \frac{72-60}{5}\right) \\ &= P\left(\frac{-5}{5} < Z < \frac{12}{5}\right) \\ &= P(-1 < Z < 2.4) \\ &= P(Z < 2.4) - P(Z < -1) \\ &= P(Z < 2.4) - P(Z > 1) \\ &= P(Z < 2.4) - [1 - P(Z \leq 1)] \\ &= 0.9918 - [1 - 0.8413] \\ &= 0.9918 - 0.1587 \\ &= 0.8331 \end{aligned}$$

(e) $V_3 = \sum_{i=1}^{8} \frac{[X_i - 60]^2}{25}$ is defined as the sum of 8 independent squared $n(0;\ 1)$ variates. Using **result 1.2 in our study guide (page 29),** $V_3 \sim \chi_n^2 \Longrightarrow V_3 \sim \chi_8^2$. Since $V_3 \sim \chi_8^2$, it follows from the properties of the chi-square distribution that $E(V_3) = 8$ using **result 1.1 in our study guide (page 28).**

(f) $Y = \sum_{i=1}^{10} \left[\frac{X_i - \overline{X}}{\sigma} \right]^2$ then Y is a $\chi_{n-1}^2 \Longrightarrow$ a χ_9^2 variate and from the properties of the chi-square distribution in the study guide, it follows that $Var(Y) = 2d = 2 \times 9 = 18.$

(g) $T = \frac{\sqrt{10}\left(\overline{X} - \mu\right)}{\sqrt{\frac{\Sigma\left(X_i - \overline{X}\right)^2}{9}}} \sim t_9$ $\quad P(T \geq 3.250) = 0.005$ from Table III (Stoker) with $\nu = 9$ and $p = 0.005$

(h) Since $V_1 \sim \chi_3^2$ and $V_2 \sim \chi_7^2$, then $U = \frac{V_1/3}{V_2/7} \sim F_{3;\,7}$ using definition 1.21.

(i) Using result 1.4 (page 32), if $U \sim F_{3;7}$, then $\frac{1}{U} \sim F_{7;3}$.

(j) The table gives $P(U > a) = \alpha$. $P(U < a) = 1 - P(U > a) = 1 - \alpha$.

Since $U \sim F_{3;\,7} \Longrightarrow P(U < a) = 1 - P(U > a) = 1 - 0.05$

and from Table V (Stoker) we find that $F_{0.05;3;7} = 4.35$ and $P(U > 4.35) = 0.05.$

Hence $P(U < 4.35) = 0.95$ and thus $a = 4.35$

(k) Since $\frac{1}{U} \sim F_{7;\,3} \Longrightarrow P\left(\frac{1}{U} > a\right) = \alpha$ and from Table V (Stoker) we find that $F_{0.05;7;3} = 8.89.$

Thus $P\left(\frac{1}{U} > 8.89\right) = 0.05 \Longrightarrow a = 8.89.$

QUESTION 4

(a) $E(X_i) = \theta_1 + C_i\theta_2$

The least square estimator is

$$\begin{aligned} Q(\theta_1;\theta_2) &= \sum_{i=1}^{n} (X_i - E(X_i))^2 \\ \Longrightarrow Q(\theta_1;\theta_2) &= \sum_{i=1}^{n} (X_i - (\theta_1 + C_i\theta_2))^2 \\ &= \sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2)^2 \end{aligned}$$

$$\begin{aligned} \frac{dQ(\theta_1;\theta_2)}{d\theta_1} &= 2\sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2) \times -1 \\ &= -2\sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2) \end{aligned}$$

If we set $\frac{dQ(\theta_1;\theta_2)}{d\theta_1} = 0$. Now

$$\begin{aligned} 0 &= -2\sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2) \\ 0 &= \sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2) \\ 0 &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \theta_1 - \theta_2 \sum_{i=1}^{n} C_i \\ 0 &= \sum_{i=1}^{n} X_i - n\theta_1 - \theta_2 \sum_{i=1}^{n} C_i \end{aligned}$$

Making θ_1 subject of the formula

$$\begin{aligned} n\theta_1 &= \sum_{i=1}^{n} X_i - \theta_2 \sum_{i=1}^{n} C_i \\ \theta_1 &= \frac{\sum_{i=1}^{n} X_i - \theta_2 \sum_{i=1}^{n} C_i}{n} \quad \text{..........................(1)} \end{aligned}$$

Making θ_2 subject of the formula

$$\begin{aligned}\theta_2\sum_{i=1}^{n}C_i &= \sum_{i=1}^{n}X_i - n\theta_1\\ \theta_2 &= \frac{\sum_{i=1}^{n}X_i - n\theta_1}{\sum_{i=1}^{n}C_i}\text{..........................(2)}\end{aligned}$$

Now

$$\begin{aligned}\frac{dQ\left(\theta_1;\theta_2\right)}{d\theta_2} &= 2\sum_{i=1}^{n}\left(X_i-\theta_1-C_i\theta_2\right)\times -C_i\\ &= -2\sum_{i=1}^{n}\left(X_iC_i-\theta_1C_i-C_i^2\theta_2\right)\end{aligned}$$

If we set $\frac{dQ\left(\theta_1;\theta_2\right)}{d\theta_2}=0$. Now

$$\begin{aligned}0 &= -2\sum_{i=1}^{n}\left(X_iC_i-\theta_1C_i-C_i^2\theta_2\right)\\ 0 &= \sum_{i=1}^{n}\left(X_iC_i-\theta_1C_i-C_i^2\theta_2\right)\\ 0 &= \sum_{i=1}^{n}X_iC_i-\theta_1\sum_{i=1}^{n}C_i-\theta_2\sum_{i=1}^{n}C_i^2\end{aligned}$$

Making θ_1 subject of the formula

$$\begin{aligned}\theta_1\sum_{i=1}^{n}C_i &= \sum_{i=1}^{n}X_iC_i-\theta_2\sum_{i=1}^{n}C_i^2\\ \theta_1 &= \frac{\sum_{i=1}^{n}X_iC_i-\theta_2\sum_{i=1}^{n}C_i^2}{\sum_{i=1}^{n}C_i}\text{..........................(3)}\end{aligned}$$

Making θ_2 subject of the formula

$$\begin{aligned}
\theta_2 \sum_{i=1}^{n} C_i^2 &= \sum_{i=1}^{n} X_i C_i - \theta_1 \sum_{i=1}^{n} C_i \\
\theta_2 &= \frac{\sum_{i=1}^{n} C_i X_i - \theta_1 \sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} C_i^2} \quad \text{..........................(4)}
\end{aligned}$$

Finding θ_1 by equating equations 2 and 4.

$$\begin{aligned}
\frac{\sum_{i=1}^{n} X_i - n\theta_1}{\sum_{i=1}^{n} C_i} &= \frac{\sum_{i=1}^{n} X_i C_i - \theta_1 \sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} C_i^2} \\
\sum_{i=1}^{n} C_i^2 \left(\sum_{i=1}^{n} X_i - n\theta_1 \right) &= \sum_{i=1}^{n} C_i \left(\sum_{i=1}^{n} X_i C_i - \theta_1 \sum_{i=1}^{n} C_i \right) \\
\sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i - n\theta_1 \sum_{i=1}^{n} C_i^2 &= \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i C_i - \theta_1 \left(\sum_{i=1}^{n} C_i \right)^2 \\
\theta_1 \left(\sum_{i=1}^{n} C_i \right)^2 - n\theta_1 \sum_{i=1}^{n} C_i^2 &= \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i C_i - \sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i \\
\theta_1 \left(\left(\sum_{i=1}^{n} C_i \right)^2 - n \sum_{i=1}^{n} C_i^2 \right) &= \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i C_i - \sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i \\
&\Longrightarrow \widehat{\theta}_1 = \frac{\sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i C_i - \sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i}{\left(\sum_{i=1}^{n} C_i \right)^2 - n \sum_{i=1}^{n} C_i^2}
\end{aligned}$$

Finding θ_2 by equating equations 1 and 3.

$$\begin{aligned}
\frac{\sum_{i=1}^{n} X_i - \theta_2 \sum_{i=1}^{n} C_i}{n} &= \frac{\sum_{i=1}^{n} X_i C_i - \theta_2 \sum_{i=1}^{n} C_i^2}{\sum_{i=1}^{n} C_i} \\
\sum_{i=1}^{n} C_i \left(\sum_{i=1}^{n} X_i - \theta_2 \sum_{i=1}^{n} C_i \right) &= n \left(\sum_{i=1}^{n} X_i C_i - \theta_2 \sum_{i=1}^{n} C_i^2 \right) \\
\sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i - \theta_2 \left(\sum_{i=1}^{n} C_i \right)^2 &= n \sum_{i=1}^{n} X_i C_i - n\theta_2 \sum_{i=1}^{n} C_i^2 \\
n\theta_2 \sum_{i=1}^{n} C_i^2 - \theta_2 \left(\sum_{i=1}^{n} C_i \right)^2 &= n \sum_{i=1}^{n} X_i C_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i \\
\theta_2 \left(n \sum_{i=1}^{n} C_i^2 - \left(\sum_{i=1}^{n} C_i \right)^2 \right) &= n \sum_{i=1}^{n} X_i C_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i \\
\widehat{\theta}_2 &= \frac{n \sum_{i=1}^{n} X_i C_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i}{n \sum_{i=1}^{n} C_i^2 - \left(\sum_{i=1}^{n} C_i \right)^2}
\end{aligned}$$

(b) $f_X(x) = cx^{c-1}$ for $x > 1$. The maximum likelihood is

$$\begin{aligned}
L(c) &= \prod_{i=1}^{n} f(x_i;\ c) \qquad \text{(see definition 2.5)} \\
&= \prod_{i=1}^{n} c x_i^{c-1} \\
&= c x_1^{c-1} \times c x_2^{c-1} \times \cdots \times c x_n^{c-1} \\
&= c^n \prod_{i=1}^{n} x_i^{c-1}
\end{aligned}$$

$\therefore$

$$\begin{aligned}
\log L(c) &= n \log c + (c-1) \log \prod_{i=1}^{n} x_i \\
&= n \log c + c \log \prod_{i=1}^{n} x_i - \log \prod_{i=1}^{n} x_i \\
\frac{d \log L(c)}{dc} &= \frac{n}{c} + \log \prod_{i=1}^{n} x_i
\end{aligned}$$

Setting $\dfrac{d \log L(c)}{dc} = 0$

$$
\begin{aligned}
\therefore \qquad \frac{n}{c} + \log \prod_{i=1}^{n} x_i &= 0 \\
\frac{n}{c} &= -\log \prod_{i=1}^{n} x_i \\
n &= -c \log \prod_{i=1}^{n} x_i \\
c \log \prod_{i=1}^{n} x_i &= -n \\
\widehat{c} &= \frac{-n}{\log \prod_{i=1}^{n} x_i}
\end{aligned}
$$