# COLUMBIA UNIVERSITY

## DEPARTMENT OF BIOSTATISTICS

## P8109 – STATISTICAL INFERENCE

### *Exercise Sheet 2 (Model Answers)*

---

**Question 1 (2 MARKS)**

We have

$$L_{\mathbf{X}}(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!} = \frac{e^{-n\lambda}\lambda^{\sum_i X_i}}{\prod_i X_i!}$$

$$\therefore \quad \log L_{\mathbf{X}} = -n\lambda + \textstyle\sum_i X_i \log\lambda - \log\left(\prod_i X_i!\right)$$

$$\frac{d}{d\lambda}\log L_{\mathbf{X}} = -n + \frac{\sum_i X_i}{\lambda}$$

Therefore,

$$-n + \frac{\sum_i X_i}{\hat{\lambda}} = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{n}\textstyle\sum_i X_i = \bar{X}\ .$$

---

**Question 2 (2+2=4 MARKS)**

(i) We have

$$L_{\mathbf{X}}(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\mu}} e^{(X_i-\mu)^2/2\mu} = \left(\frac{1}{\sqrt{2\pi\mu}}\right)^n \exp\left\{-\frac{1}{2\mu}\sum_{i=1}^{n}(X_i-\mu)^2\right\}$$

Therefore,

$$\begin{aligned}\log L_{\mathbf{x}}(\mu) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\mu - \frac{1}{2\mu}\sum_i (X_i - \mu)^2 \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\mu - \frac{1}{2\mu}\left(\sum_i X_i^2 - 2\mu\sum_i X_i + n\mu^2\right) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\mu - \frac{\sum_i X_i^2}{2\mu} + \sum_i X_i - \frac{n\mu}{2}\end{aligned}$$

and

$$\frac{d}{d\mu}\log L_{\mathbf{x}} = -\frac{n}{2\mu} + \frac{\sum_i X_i^2}{2\mu^2} - \frac{n}{2}$$

$$\therefore \quad -\frac{n}{2\hat{\mu}} + \frac{\sum_i X_i^2}{2\hat{\mu}^2} - \frac{n}{2} = 0$$

$$\hat{\mu}^2 + \hat{\mu} - m_2' = 0$$

$$\hat{\mu} = \frac{-1 \pm \sqrt{1+4m_2'}}{2}$$

We take the positive root since the variance $\sigma^2 = \mu > 0$, i.e. $\hat{\mu} = (-1 + \sqrt{1+4m_2'})/2$.

(ii) We have

$$L_{\mathbf{x}}(\mu) = \prod_{i=1}^{n} \frac{1}{\mu\sqrt{2\pi}} e^{(X_i-\mu)^2/2\mu^2} = \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\mu^2}\sum_{i=1}^{n}(X_i-\mu)^2\right\}$$

Therefore,

$$\begin{aligned}\log L_{\mathbf{x}}(\mu) &= -\frac{n}{2}\log(2\pi) - n\log\mu - \frac{1}{2\mu^2}\sum_i (X_i - \mu)^2 \\ &= -\frac{n}{2}\log(2\pi) - n\log\mu - \frac{1}{2\mu^2}\left(\sum_i X_i^2 - 2\mu\sum_i X_i + n\mu^2\right) \\ &= -\frac{n}{2}\log(2\pi) - n\log\mu - \frac{\sum_i X_i^2}{2\mu^2} + \frac{\sum_i X_i}{\mu} - \frac{n}{2}\end{aligned}$$

and

$$\frac{d}{d\mu}\log L_{\mathrm{x}} = -\frac{n}{\mu} + \frac{\sum_i X_i^2}{\mu^3} - \frac{n\bar{X}}{\mu^2}$$

$$\therefore \quad -\frac{n}{\hat{\mu}} + \frac{\sum_i X_i^2}{\hat{\mu}^3} - \frac{n\bar{X}}{\hat{\mu}^2} = 0$$

$$-n\hat{\mu}^2 + \sum_i X_i^2 - n\hat{\mu}\bar{X} = 0$$

$$\hat{\mu}^2 + \hat{\mu}\bar{X} - m_2' = 0$$

$$\hat{\mu} = \frac{-\bar{X} \pm \sqrt{\bar{X}^2 + 4m_2'}}{2}$$

We take the positive root since the standard deviation $\sigma = \mu > 0$, i.e.

$\hat{\mu} = (-\bar{X} + \sqrt{\bar{X}^2 + 4m_2'})/2$.

---

**Question 3 (2+1+1=4 MARKS)**

(i) We have

$$\mathcal{E}\left(\frac{\sum_i X_i^2}{n} - 1\right) = \frac{1}{n}\sum_i \mathcal{E}X_i^2 - 1$$

$$= \frac{1}{n}\sum_i \left(\mathrm{var}\, X_i + \mathcal{E}^2 X_i\right) - 1$$

$$= \frac{1}{n}\sum_i \left(1 + \mu^2\right) - 1$$

$$= \frac{1}{n}.n\left(1 + \mu^2\right) - 1$$

$$= \mu^2.$$

Hence, $T^2$ is an unbiased estimator for $\mu^2$.

(ii) $T^2$ is not sensible because it can be negative although $\mu^2$ is always positive.

(iii)Since $T^2$ is unbiased, we have $\mathcal{E}T^2 = \mu^2$. Now

$$\operatorname{var} T = \mathcal{E}T^2 - \mathcal{E}^2 T > 0$$
$$\therefore \quad \mathcal{E}^2 T < \mu^2 \quad \Rightarrow \quad \mathcal{E}T \neq \mu$$

Hence $T$ is biased for $\mu$.

---

**Question 4 (1 +2 = 3 MARKS)**

(a) $T_n$ is a consistent estimator of a parameter $\theta$ if and only if $T_n \xrightarrow{P} \theta$, i.e. for any $\varepsilon > 0$, $\Pr\{|T_n - \theta| < \varepsilon\} \to 1$ as $n \to \infty$

(b) The consistency condition can be written as $\Pr\{|T_n - \theta| \geq \varepsilon\} \to 0$ as $n \to \infty$.

Using Chebychev's inequality,

$$\begin{aligned}\Pr\{|T_n - \theta| \geq \varepsilon\} &\leq \frac{\mathcal{E}\left(T_n - \theta\right)^2}{\varepsilon^2} \\ &= \frac{MSE_\theta\left(T_n\right)}{\varepsilon^2} \\ &= \frac{\operatorname{var}_\theta T + \operatorname{bias}^2\left(T\right)}{\varepsilon^2}\end{aligned}$$

We want $\Pr\{|T_n - \theta| \geq \varepsilon\} \to 0$ as $n \to \infty$. For this to occur, we see from the above that it is sufficient that $\operatorname{bias}(T_n) \to 0$ and $\operatorname{var} T_n \to 0$ as $n \to \infty$.

---

**Question 5 (2+2+1+2+2=9 MARKS)**

(a) We have

$$L_{\mathbf{x}}\left(\theta\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}\left(X_i - \theta\right)^2}{2}\right\}$$

$$\log L_{\mathbf{x}}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2$$

$$S = \frac{\partial}{\partial\theta}\log L_{\mathbf{x}}(\theta) = \sum_{i=1}^{n}(X_i-\theta) = \sum_{i=1}^{n}X_i - n\theta$$

$$\therefore \quad I_{\mathbf{x}} = \operatorname{var} S = n \operatorname{var} X_i = n$$

Hence,

$$CRLB = \frac{\{\tau'(\theta)\}^2}{I_{\mathbf{x}}} = \frac{(1)^2}{n} = \frac{1}{n}.$$

(b) We write the score function as

$$S = \sum_i X_i - n\theta = \underbrace{n}_{g_1(\theta)}\left(\underbrace{\frac{\sum_i X_i}{n}}_{T} - \underbrace{\theta}_{\tau}\right).$$

Hence the MVBU estimator of $\theta$ is $T = (\sum_i X_i)/n = \bar{X}$ .

(c) Any MVBU estimator is an UMVUE. Hence the UMVUE is $T = (\sum_i X_i)/n = \bar{X}$ .

(d) As in (a) above,

$$L_{\mathbf{x}}(\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-\theta)^2}{2}\right\}$$

$$\log L_{\mathbf{x}}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2$$

$$S = \frac{\partial}{\partial\theta}\log L_{\mathbf{x}}(\theta) = \sum_{i=1}^{n}(X_i-\theta) = \left(\sum_{i=1}^{n}X_i - n\theta\right)$$

$$\therefore \quad I_{\mathbf{x}} = \operatorname{var} S = n \operatorname{var} X_i = n$$

But this time, $\tau(\theta) = \theta^2$ and

$$CRLB = \frac{\{\tau'(\theta)\}^2}{I_{\mathbf{x}}} = \frac{(2\theta)^2}{n} = \frac{4\theta^2}{n}.$$

(e) The score function is

$$S = \sum_i X_i - n\theta = \frac{n}{\theta}\left(\frac{\theta}{n}\sum_i X_i - \theta^2\right)$$

This cannot be written in the form $S = g_1(\theta)(T - \tau)$ and there is no MVBU estimator of $\theta^2$.

---

**Question 6 (1+2+1+2+2=8 MARKS)**

(a) We have

$$\Pr\{X_j = x\} = \frac{e^{-\lambda}\lambda^x}{x!} I\left(x \in \{0,\ 1,\ 2,\ ...\}\right)$$
$$\therefore \quad p = \Pr\left\{X_j = 0\right\} = e^{-\lambda}$$

(b) We have

$$L_{\mathbf{x}}(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!} = \frac{e^{-n\lambda}\lambda^{\sum_i X_i}}{\prod_i X_i!}$$
$$\therefore \quad \log L_{\mathbf{x}} = -n\lambda + \sum_i X_i \log\lambda - \log\left(\prod_i X_i!\right)$$
$$S = \frac{d}{d\lambda}\log L_{\mathbf{x}} = -n + \frac{\sum_i X_i}{\lambda}$$
$$I_{\mathbf{x}} = \operatorname{var} S = \frac{1}{\lambda^2}(n\lambda) = \frac{n}{\lambda}$$

We have $p = \tau(\lambda) = e^{-\lambda}$, so

$$CRLB = \frac{\left\{\tau'(\lambda)\right\}^2}{I_X} = \frac{\left(-e^{-\lambda}\right)^2}{n/\lambda} = \frac{\lambda e^{-2\lambda}}{n}.$$

Since $\lambda = -\log p$, we have for any unbiased estimator of $p$,

$$CRLB = \frac{\lambda e^{-2\lambda}}{n} = \frac{(-\log p)e^{-2(-\log p)}}{n} = \frac{(-\log p)\,p^2}{n}.$$

(c) We have

$$\mathcal{E}Y = 1 \times \Pr\{X_j = 0\} + 0 \times \Pr\{X_j \neq 0\} = p$$

Therefore, an unbiased estimator of $p$ is $T = \overline{Y}$ .

(d) We have

$$\begin{aligned} \operatorname{var} T &= \frac{\operatorname{var} Y}{n} \\ &= \frac{\mathrm{E}Y^2 - \mathrm{E}^2 Y}{n} \\ &= \frac{1^2 \times \Pr\{X = 0\} - p^2}{n} \\ &= \frac{p - p^2}{n}. \end{aligned}$$

(e) We see from the graph below that $(p - p^2) \geq (-\log p) p^2$, so that $\operatorname{var} T \geq CRLB$ . Hence $T$ is not the MVBU estimator.