UNIVERSITY OF NEW SOUTH WALES
DEPARTMENT OF STATISTICS

MATH5856 Introduction to Statistics and Statistical Computations
Tutorial Problems week 10, Solutions

1. First note that

$$\begin{aligned} E(\widehat{\mu}_1) &= \frac{E(X_1)+E(X_2)+E(X_3)+E(X_4)+E(X_5)}{5} \\ &= \frac{\mu+\mu+\mu+\mu+\mu}{5} = \mu \end{aligned}$$

so

$$\mathsf{bias}(\widehat{\mu}_1) = E(\widehat{\mu}_1) - \mu = 0.$$

Similarly,

$$E(\widehat{\mu}_2) = \frac{\mu+2\mu+3\mu+4\mu+5\mu}{15} = \mu$$

so $\mathsf{bias}(\widehat{\mu}_2) = 0$. Next,

$$\begin{aligned} \mathrm{Var}(\widehat{\mu}_1) &= \frac{\mathrm{Var}(X_1)+\mathrm{Var}(X_2)+\mathrm{Var}(X_3)+\mathrm{Var}(X_4)+\mathrm{Var}(X_5)}{25} \\ &= \frac{\sigma^2+\sigma^2+\sigma^2+\sigma^2+\sigma^2}{25} = \frac{\sigma^2}{5} \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(\widehat{\mu}_2) &= \frac{\mathrm{Var}(X_1)+4\mathrm{Var}(X_2)+9\mathrm{Var}(X_3)+16\mathrm{Var}(X_4)+25\mathrm{Var}(X_5)}{225} \\ &= \frac{\sigma^2+4\sigma^2+9\sigma^2+16\sigma^2+25\sigma^2}{225} = \frac{11\sigma^2}{45}. \end{aligned}$$

It follows immediately that

$$\textsc{mse}(\widehat{\mu}_1) = \frac{\sigma^2}{5} = 0.2\sigma^2 \quad \text{while} \quad \textsc{mse}(\widehat{\mu}_2) = \frac{11\sigma^2}{45} \simeq 0.244\sigma^2.$$

Since MSE is a quality measure for estimators, $\widehat{\mu}_1$ is better. Intitu- ively, $\widehat{\mu}_2$ should be inferior since instead of a simple average with equal weighting to all members of the random sample, it applies a weighted average with more weight given to higher indexed members of the sample. Since all members of the sample are on ‘equal footing’ with one another $\widehat{\mu}_2$ seems to be making improper (perhaps inefficient) use of the sample.

2. First note that $E(X_i) = \mathrm{Var}(X_i) = \lambda$ for all $1 \leq i \leq n$. Then

$$\begin{aligned} \mathsf{bias}(\widehat{\lambda}) &= E(\widehat{\lambda}) - \lambda \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) - \lambda \\ &= \frac{1}{n}\sum_{i=1}^{n} \lambda - \lambda = \lambda - \lambda = 0. \end{aligned}$$

Next,

$$\mathrm{Var}(\widehat{\lambda}) = \mathrm{Var}(\widehat{\lambda}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \lambda = \frac{\lambda}{n}$$

Hence,

$$\mathsf{se}(\widehat{\lambda}) = \sqrt{\mathrm{Var}(\widehat{\lambda})} = \sqrt{\frac{\lambda}{n}}.$$

Finally,

$$\text{MSE}(\widehat{\lambda}) = \mathsf{bias}^2(\widehat{\lambda}) + \mathrm{Var}(\widehat{\lambda}) = \frac{\lambda}{n}.$$

3. (a)

$$\mathsf{bias}(\widehat{\mu}) = E(\overline{X}) - \mu = \mu - \mu = 0.$$

$$\mathrm{Var}(\widehat{\mu}) = \mathrm{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

so

$$\mathsf{se}(\widehat{\mu}) = \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad \text{MSE}(\widehat{\mu}) = \mathsf{bias}^2(\widehat{\mu}) + \mathrm{Var}(\widehat{\mu}) = \frac{\sigma^2}{n}.$$

(b) Note that

$$\widehat{\sigma}^2 = \frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{(n-1)S^2}{n}.$$

Then, from results on the distribution of S^2,

$$E(S^2) = \sigma^2 \quad \text{and} \quad \mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}.$$

Hence,

$$\mathsf{bias}(\widehat{\sigma}^2) = E\left(\frac{(n-1)S^2}{n}\right) - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = \frac{-\sigma^2}{n}$$

and

$$\mathrm{Var}(\widehat{\sigma}^2) = \mathrm{Var}\left(\frac{(n-1)S^2}{n}\right) = \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2}.$$

Thus, $\mathsf{se}(\widehat{\sigma}^2) = \frac{\sigma^2\sqrt{2(n-1)}}{n}$ and

$$\mathrm{MSE}(\widehat{\sigma}^2) = \left(\frac{-\sigma^2}{n}\right)^2 + \frac{2(n-1)\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}.$$

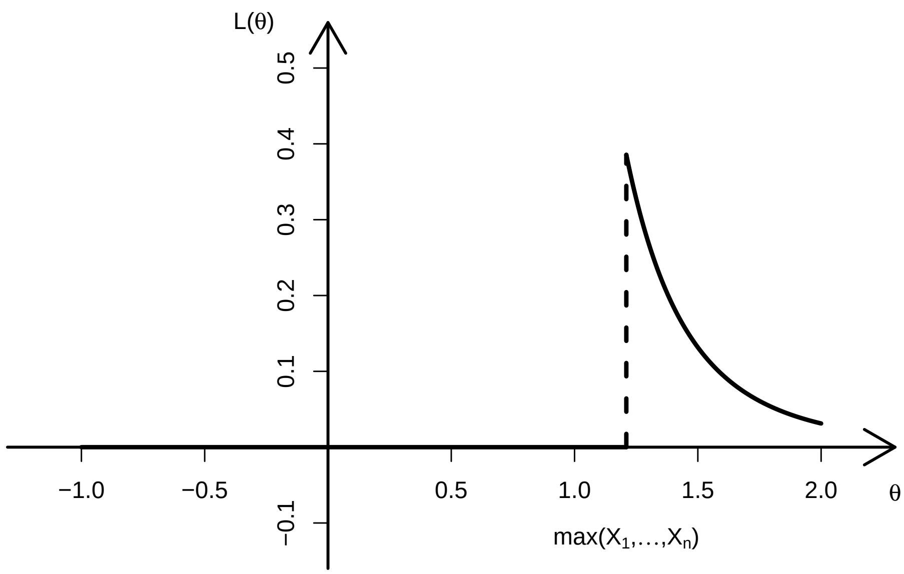
4. (a) The common density function of $X_1, \ldots, X_n$ may be written

$$f_X(x;\theta) = \left(\frac{1}{\theta}\right)\mathcal{I}(0 < x < \theta)$$

where $\mathcal{I}(\mathcal{P})$ is the indicator function of the condition $\mathcal{P}$. The likelihood function of θ is then

$$\begin{aligned}
\mathcal{L}(\theta) &= \prod_{i=1}^{n} f_X(X_i;\theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n}\mathcal{I}(0 < X_i < \theta) \\
&= \left(\frac{1}{\theta}\right)^n \mathcal{I}(\theta \geq X_1 \& \ldots \& \theta \geq X_n) \\
&= \left(\frac{1}{\theta}\right)^n \mathcal{I}(\theta \geq \max(X_1, \ldots, X_n))
\end{aligned}$$

As shown in the accompanying figure, $\mathcal{L}(\theta)$, attains a unique maximum at $\theta = \max(X_1, \ldots, X_n)$.



Therefore the maximum likelihood estimator for θ is

$$\widehat{\theta} = \max(X_1, \ldots, X_n).$$

(b) For $0 < x < \theta$, the cumulative distribution function of $\widehat{\theta}$ is

$$\begin{aligned} F_{\widehat{\theta}}(x) &= P(\widehat{\theta} \leq x) = P(\max(X_1, \ldots, X_n) \leq x) \\ &= P(X_1 \leq x \,\&\ldots\&X_n \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) = (x/\theta) \cdots (x/\theta) \\ &= (x/\theta)^n \end{aligned}$$

For all other x, $F_{\widehat{\theta}}(x)$ is constant so the density function of $\widehat{\theta}$ is

$$f_{\widehat{\theta}}(x) = \frac{d}{dx} F_{\widehat{\theta}}(x) = (n/\theta^n)x^{n-1}, \quad 0 < x < \theta.$$

The bias of $\widehat{\theta}$ is

$$\begin{aligned}\mathsf{bias}(\widehat{\theta}) &= E(\widehat{\theta}) - \theta = \int_0^\theta x(n/\theta^n)x^{n-1}\,dx - \theta \\ &= \theta\left(\frac{n}{n+1}\right) - \theta \to 0 \text{ as } n \to \infty.\end{aligned}$$

Next note that

$$E(\widehat{\theta}^2) = \int_0^\theta x^2(n/\theta^n)x^{n-1}\,dx = \frac{n\theta^2}{n+2}$$

so that

$$\mathrm{Var}(\widehat{\theta}) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2.$$

which also has a limit of zero as $n \to \infty$. Hence,

$$\lim_{n\to\infty} \mathrm{MSE}(\widehat{\theta}) = 0$$

and $\widehat{\theta}$ is consistent.

5. The logarithm of the likelihood function is

$$\begin{aligned}\ln \mathcal{L}(\theta) &= \ln \prod_{i=1}^n f(X_i;\theta) \\ &= \ln \prod_{i=1}^n \{2\theta X_i e^{-\theta X_i^2}\} \\ &= n\ln(2) + n\ln(\theta) + \sum_{i=1}^n \ln(X_i) - \theta \sum_{i=1}^n X_i^2\end{aligned}$$

The first derivative of $\ln \mathcal{L}(\theta)$ is

$$(\partial/\partial\theta)\ln \mathcal{L}(\theta) = n/\theta - \sum_{i=1}^n X_i^2$$

and the second derivative of $\ln \mathcal{L}(\theta)$ is

$$(\partial^2/\partial\theta^2)\ln \mathcal{L}(\theta) = -n/\theta^2 < 0 \quad \text{for all } \theta > 0.$$

Hence $\ln \mathcal{L}(\theta)$ is concave downwards over its domain and is therefore maximised at a value for which $(\partial/\partial\theta)\ln\mathcal{L}(\theta) = 0$ for $\theta = 0$. Solving this for θ we obtain

$$\theta = n \Big/ \sum_{i=1}^{n} X_i^2 > 0$$

so the maximum likelihood estimate of θ is

$$\widehat{\theta} = n \Big/ \sum_{i=1}^{n} X_i^2$$

6. The logarithm of the likelihood function is

$$\begin{aligned}
\ln \mathcal{L}(\mu, \sigma^2) &= \ln \prod_{i=1}^{n} f(X_i; \mu, \sigma^2) \\
&= \ln \prod_{i=1}^{n} \{(2\pi\sigma^2)^{-1/2} e^{-(X_i-\mu)^2/(2\sigma^2)}\} \\
&= -(n/2)\ln(2\pi) - (n/2)\ln(\sigma^2) - (2\sigma^2)^{-1} \sum_{i=1}^{n} (X_i - \mu)^2
\end{aligned}$$

Thus,

$$(\partial/\partial\mu) \ln \mathcal{L}(\mu, \sigma^2) = (\sigma^2)^{-1} \sum_{i=1}^{n} (X_i - \mu)$$

and

$$(\partial/\partial\sigma^2) \ln \mathcal{L}(\mu, \sigma^2) = -(n/2)(\sigma^2)^{-1} + \tfrac{1}{2}(\sigma^2)^{-3} \sum_{i=1}^{n} (X_i - \mu)^2$$

Clearly

$$(\partial/\partial\mu) \ln \mathcal{L}(\mu, \sigma^2) = 0 \Longleftrightarrow \mu = \overline{X}$$

for all $\sigma^2 > 0$. Hence, both partial derivatives equal zero if and only if

$$\mu = \overline{X} \quad \text{and} \quad -(n/2)(\sigma^2)^{-1} + \tfrac{1}{2}(\sigma^2)^{-3} \sum_{i=1}^{n} (X_i - \overline{X})^2 = 0.$$

This is equivalent to

$$\mu = \overline{X} \quad \text{and} \quad \sigma^2 = \tfrac{1}{n} \textstyle\sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Assuming that this does correspond to a global maximum of $\ln \mathcal{L}(\mu, \sigma^2)$ (the checking of which is a bit beyond the level of the course), the maximum likelihood estimates of μ and σ^2 are

$$\widehat{\mu} = \overline{X} \quad \text{and} \quad \widehat{\sigma}^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

7.

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_{X_i}(X_i;\theta) = \begin{cases} e^{n\theta - \sum_{i=1}^{n} X_i} 1 & \min(X_1, \ldots, X_n) \geq \theta \\ 0 & \text{otherwise.} \end{cases}$$

A graph of $\mathcal{L}(\theta)$ shows that the maximum of the function occurs at $\widehat{\theta} = \min(X_1, \ldots, X_n)$.

8. (a) The joint probability function of the X_i's and Y_i's is

$$\begin{aligned} f_{X_1,\ldots,X_{100},Y_1,\ldots,Y_{100}}(x_1, \ldots, x_{100}, y_1, \ldots, y_{100}) &= \prod_{i=1}^{100} \frac{\lambda_X^{x_i} e^{-\lambda_X}}{x_i!} \times \prod_{i=1}^{100} \frac{\lambda_Y^{y_i} e^{-\lambda_Y}}{y_i!} \\ &= \prod_{i=1}^{100} \frac{\lambda_X^{x_i} \lambda_Y^{y_i} e^{-\lambda_X - \lambda_Y}}{x_i! y_i!}. \end{aligned}$$

(b) The log-likelihood of λ_X and λ_Y is

$$\begin{aligned} \ell(\lambda_X, \lambda_Y) &= \ln f_{X_1,\ldots,X_{100},Y_1,\ldots,Y_{100}}(X_1, \ldots, X_{100}, X_1, \ldots, X_{100}) \\ &= \sum_{i=1}^{100} \{X_i \ln(\lambda_X) + Y_i \ln(\lambda_Y) - \lambda_X - \lambda_Y - \ln(X_i!) - \ln(Y_i!)\} \end{aligned}$$

The partial derivatives with respect to λ_X and λ_Y are

$$\frac{\partial}{\partial \lambda_X} \ell(\lambda_X, \lambda_Y) = (1/\lambda_X) \sum_{i=1}^{100} X_i - 100$$

and

$$\frac{\partial}{\partial \lambda_Y} \ell(\lambda_X, \lambda_Y) = (1/\lambda_Y) \sum_{i=1}^{100} Y_i - 100$$

Setting these to zero leads to the unique stationary point

$$\widehat{\lambda}_X = \overline{X}, \quad \widehat{\lambda}_Y = \overline{Y}.$$

Second derivative analysis can be used to show that these corresponds to the unique global maximiser of $\ell(\lambda_X, \lambda_Y)$, so these are the maximum likelihood estimators of λ_X and λ_Y.

(c) Under the stated hypothesis the likelihood function is

$$\mathcal{L}(\lambda) = \prod_{i=1}^{100} \frac{\lambda^{X_i+Y_i} e^{-2\lambda}}{X_i! Y_i!}$$

(d) The log-likelihood function is

$$\ell(\lambda) = \sum_{i=1}^{100} \{X_i \ln(\lambda) + Y_i \ln(\lambda) - 2\lambda - \ln(X_i!) - \ln(Y_i!)\}$$

The partial derivative with respect to λ is

$$\frac{\partial}{\partial \lambda}\ell(\lambda) = (1/\lambda) \sum_{i=1}^{100} (X_i + Y_i) - 200$$

which is zero if and only if

$$\lambda = \frac{1}{200} \sum_{i=1}^{100} (X_i + Y_i) = \tfrac{1}{2}(\overline{X} + \overline{Y}).$$

$$\frac{\partial^2}{\partial \lambda^2}\ell(\lambda) = (-1/\lambda^2) \sum_{i=1}^{100} (X_i + Y_i) \leq 0$$

so the $\widehat{\lambda} = \frac{1}{2}(\overline{X} + \overline{Y})$ is the unique maximiser of $\ell(\lambda)$ and hence corresponds to the maximum likelihood estimator.