Q1 [15 points]:

Use the method of maximum likelihood to estimate θ in the pdf

$$f_Y(y \mid \theta) = \frac{\theta}{2\sqrt{y}} e^{-\theta\sqrt{y}}, \quad y \geq 0$$

Evaluate θ_e for the following random sample of size 4 : $Y_1 = 6.2, Y_2 = 7.0, Y_3 = 2.5$, and $Y_4 = 4.2$.

Solution:

$$\ell(\theta) = \prod_{i=1}^{4} \frac{\theta}{2\sqrt{y_i}} e^{-\theta\sqrt{y_i}} = \frac{\theta^4}{16\prod_{i=1}^{4}\sqrt{y_i}} e^{-\theta\sum_{i=1}^{4}\sqrt{y_i}}$$

$$\ln \ell(\theta) = 4\ln\theta - \ln\left(16\prod_{i=1}^{4}\sqrt{y_i}\right) - \theta\sum_{i=1}^{4}\sqrt{y_i}$$

$$\frac{d\ln\ell(\theta)}{d\theta} = \frac{4}{\theta} - \sum_{i=1}^{4}\sqrt{y_i}.$$

$$\frac{d\ln\ell(\theta)}{d\theta} = 0 \quad \text{implies} \quad \widehat{\theta} = \frac{4}{\sum_{i=1}^{4}\sqrt{y_i}} = \frac{4}{8.766} = 0.456$$

Q2 [15 points]:
Suppose the random samples are obtained from a two-parameter uniform pdf $Y \sim \mathcal{U}[\theta_1, \theta_2]$. Based on the random sample $Y_1 = 6.3, Y_2 = 1.8$, $Y_3 = 14.2$, and $Y_4 = 7.6$, find the maximum likelihood estimates for θ_1 and θ_2.

$$f_Y(y \mid \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1}, \quad \theta_1 \leq y \leq \theta_2.$$

Solution:

$$\ell(\theta) = \left(\frac{1}{\theta_2 - \theta_1}\right)^n, \text{ if } \theta_1 \leq y_1, y_2, \ldots, y_n \leq \theta_2, \text{ and 0 otherwise}$$

Or, we may write

$$\ell(\theta) = \left(\frac{1}{\theta_2 - \theta_1}\right)^n, \text{ if } \min\{y_i\}_{i=1}^n > \theta_1, \text{ and } \max\{y_i\}_{i=1}^n < \theta_2, \text{ and it is zero otherwise.}$$

Using step function $u(y - y_0)$

$$u(y - y_0) = \begin{cases} 0 & y < y_0 \\ 1 & y > y_0 \end{cases}$$

we may write

$$\ell(\theta) = \left(\frac{1}{\theta_2 - \theta_1}\right)^n u\big(\min\{y_i\} - \theta_1\big)\, u\big(\theta_2 - \max\{y_i\}\big)$$

Clearly, the sufficient statistics for θ_1 and θ_2 are $\min\{y_i\}$ and $\max\{y_i\}$, respectively. So, we obtain

$$\widehat{\theta_1} = \min\{y_i\},$$
$$\widehat{\theta_2} = \max\{y_i\}.$$

From the given data, we obtain $\theta_1 = 1.8$, and $\theta_2 = 14.2$. Note that we have discussed in detail in a problem session that $\widehat{\theta_2} = \max\{y_i\}$ is a biased estimator using order statistics, and how the bias may be removed. Similarly, $\widehat{\theta_1} = \min\{y_i\}$ is also a biased estimator, and the bias may be removed. Please prepare yourself and show your work to me that how may you prove that the estimator $\widehat{\theta_1} = \min\{y_i\}$ is biased in nature.

Q3 [30 points]:

(a) Use the method of moments to estimate θ in the pdf

$$f_Y(y \mid \theta) = \left(\theta^2 + \theta\right) y^{\theta-1}(1-y), \quad 0 \le y \le 1$$

Assume that a random sample of size n has been collected.

(b) Using Taylor's series based method, obtain the bias and variance of the estimator obtained in part **(a)**.

Solution: First of all, we show that

$$E(Y) = \int_0^1 y \left(\theta^2 + \theta\right) y^{\theta-1}(1-y)dy = \left(\theta^2 + \theta\right) \int_0^1 y^{\theta}(1-y)dy = \frac{\theta}{\theta+2}$$

This is sufficient to obtain an estimator of θ from the first-order sample moment of Y. This gives

$$\begin{aligned} \theta &= \frac{2E[Y]}{1-E[Y]} \\ \widehat{\theta} &= \frac{2\widehat{\mu_Y}}{1-\widehat{\mu_Y}} \\ \text{where} \quad \widehat{\mu_Y} &= \frac{1}{n}\sum_{i=1}^{n} Y_i \\ \text{and} \quad E[\widehat{\mu_Y}] &= \frac{\theta}{\theta+2} \end{aligned}$$

For the sake of analysis, we would need to know $E[Y^2]$ for the computation of variance of Y.

$$E(Y^2) = \int_0^1 y^2 \left(\theta^2 + \theta\right) y^{\theta-1}(1-y)dy = \left(\theta^2 + \theta\right) \int_0^1 y^{\theta+1}(1-y)dy = \frac{\theta^2+\theta}{\theta^2+5\theta+6}$$

This gives

$$\mathrm{Var}(Y) = E[Y^2] - (E[Y])^2 = \frac{2\theta}{(\theta+2)^2(\theta+3)}$$

For the ease of analysis, we also define

$$\begin{aligned} \widehat{\mu_Y} &=: T \\ \widehat{\theta} &= \frac{2T}{1-T} =: h(T) \end{aligned}$$

In the sequel, we would need $h'(T)$ and $h''(T)$, so we compute it here

$$h'(T) = \frac{d}{dT}h(T) = \frac{2}{(1-T)^2}$$
$$h''(T) = \frac{d}{dT}h'(T) = \frac{4}{(1-T)^3}$$

We may easily show that the estimator $\widehat{\theta}$ is biased as follows:

$$E[\widehat{\theta}] = E\left[\frac{2\widehat{\mu_Y}}{1-\widehat{\mu_Y}}\right] \neq \frac{2E[\widehat{\mu_Y}]}{1-E[\widehat{\mu_Y}]} = \frac{2\theta/(2+\theta)}{1-\theta/(2+\theta)} = \theta$$

Next we compute the bias and variance of $\widehat{\theta}$. From the lecture 08's slides, we have

> Using Taylor's series based expansion, we may obtain
>
> $$\widehat{\theta} = h(T) = h(\mu_T) + (T-\mu_T)\,h'(\mu_T) + \frac{1}{2}(T-\mu_T)^2\,h''(\mu_T)$$
>
> where the statistic $T = T(Y)$ is the function of given random variables Y, $\mu_T = E[T(Y)]$; next, we take mean of both sides to obtain
>
> $$E\left[\,\widehat{\theta}\,\right] \approx h(\mu_T) + \frac{1}{2}\operatorname{var}(T)h''(\mu_T)$$

and

> $$\operatorname{var}(\widehat{\theta}) = [h'(\mu_T)]^2\operatorname{var}(T)$$
>
> The estimation error is defined as $\widetilde{\theta} = \widehat{\theta} - \theta$, where θ is the true value, constant in nature, to be estimated, therefore the variance of $\widetilde{\theta}$ is same as that of $\widehat{\theta}$, this gives $\operatorname{var}(\widetilde{\theta}) = [h'(\mu_T)]^2\operatorname{var}(T)$.

where $\mu_T = E[T] = E[\widehat{\mu_Y}] = \theta/(2+\theta)$, this gives

$$h(\mu_T) = \frac{2\mu_T}{1-\mu_T} = \theta$$
$$h'(\mu_T) = \frac{2}{(1-\mu_T)^2} = \frac{1}{2}(2+\theta)^2$$
$$h''(\mu_T) = \frac{4}{(1-\mu_T)^3} = \frac{1}{2}(2+\theta)^3$$

We compute $\mathrm{var}(T) = \mathrm{var}(\widehat{\mu_Y}) = \frac{1}{n}\mathrm{var}(Y)$, where $\mathrm{var}(Y)$ has been computed above. Finally, combinung the earlier results, we obtain

$$E[\widehat{\theta}] \approx \theta + \frac{1}{2n}\left(\frac{\theta+2}{\theta+3}\right)\theta$$

$$\mathrm{var}(\widehat{\theta}) \approx \frac{1}{2n}\frac{(2+\theta)^2}{3+\theta}\theta, \qquad \mathrm{var}(\widehat{\theta}) \propto \frac{\theta^2}{n}$$

Estimator is asymptotically consistent because

$$\lim_{n\to\infty} E[\widehat{\theta}] \to \theta$$

$$\lim_{n\to\infty} \mathrm{var}(\widehat{\theta}) \approx 0.$$

Q4 [40 points]:
(a) Let $X_1, X_2, \ldots, X_n$ be a random sample from $f_X(x \mid \theta) = \frac{1}{\theta} e^{-x/\theta}, x > 0$. Find the MLE $\widehat{\theta}$.

By definition, the likelihood function $\ell(\theta)$ is

$$\ell(\theta) = \log\left[\prod_{i=1}^{n} f\big(X_i \mid \theta\big)\right] = \sum_{i=1}^{n} \log\left[f\big(X_i \mid \theta\big)\right]$$

(b) Obtain the bias of the MLE $\widehat{\theta}$.

(c) Obtain the variance of the MLE by computing it explicity as follows:

$$\operatorname{var}\left(\widehat{\theta}\right) = \mathrm{E}\left[\widehat{\theta}^2\right] - \left(\mathrm{E}\left[\widehat{\theta}\right]\right)^2$$

(d) Obtain the asymptotic variance of MLE $\widehat{\theta}$ as follows:

$$\text{asymptotic var}\left(\widehat{\theta}\right) \approx \frac{1}{nI(\theta)}$$
$$\text{where } I(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ell(\theta)\right] = -\mathrm{E}\left[\ell''(\theta)\right]$$

(e) Are the variances obtained in **(c)** and **(d)** equal? Is the ML estimator a best estimator for θ?

Solution: To find the MLE of θ, we first define the likelihood function:

$$\operatorname{lik}(\theta) = f\left(x_1, \ldots, x_n \mid \theta\right) = f\left(x_1 \mid \theta\right) \cdots f\left(x_n \mid \theta\right)$$

Substituting the definition of the density function of X yields

$$\operatorname{lik}(\theta) = \left(\frac{1}{\theta} \cdot e^{-\frac{x_1}{\theta}}\right) \cdot \ldots \cdot \left(\frac{1}{\theta} \cdot e^{-\frac{x_n}{\theta}}\right) = \frac{1}{\theta^n} \cdot e^{-\frac{x_1 + \cdots + x_n}{\theta}}$$

It's easier to work with the natural logarithm of the given expression, so we define

$$l(\theta) = \ln(\operatorname{lik}(\theta)) = -n \cdot \ln(\theta) - \frac{1}{\theta} \cdot \sum_{i=1}^{n} x_i$$

and we need to find its global maximum on the interval $\langle 0, +\infty\rangle$ (where θ can take on values).

The derivative of l is

$$l'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \cdot \sum_{i=1}^{n} x_i.$$

Stationary points are the null points of the above derivative, so

$$l'(\theta) = 0 \Longleftrightarrow -\frac{n}{\theta} + \frac{1}{\theta^2} \cdot \sum_{i=1}^{n} x_i = 0 \Longleftrightarrow \frac{n}{\theta} = \frac{1}{\theta^2} \cdot \sum_{i=1}^{n} x_i \Longleftrightarrow n \cdot \theta = \sum_{i=1}^{n} x_i$$
$$\Longleftrightarrow \theta = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i.$$

Therefore, the MLE of θ is

$$\widehat{\theta} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i = \widehat{\mu_X}.$$

(b) There is no bias. Because $E[\widehat{\theta}] = E[X] = \int_0^\infty x\frac{1}{\theta}e^{-x/\theta}dx = \theta$ (Unbiased).

(c) The variance is computed below:

$$\text{var}(\widehat{\theta}) = \text{var}(\widehat{\mu_X}) = \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \text{var}\,(X_i)$$
$$= \frac{1}{n}\text{var}\,(X) = \frac{1}{n}\left(E[X^2] - (E[X])^2\right) = \frac{1}{n}\left(2\theta^2 - \theta^2\right) = \frac{\theta^2}{n}$$

(d) According to the Cramer-Rao Theorem, no unbiased estimator of θ can have variance less than $\frac{1}{n \cdot I(\theta)}$ (this is the Cramer-Rao lower bound), where

$$I(\theta) = E\left(\left[\frac{\partial}{\partial\theta}\ln f(X \mid \theta)\right]^2\right)$$

(This is called Fisher's information). Remember that $I(\theta)$ can also be calculated as

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\ln f(X \mid \theta)\right).$$

$$\ln f(x \mid \theta) = \ln\left(\frac{1}{\theta} \cdot e^{-\frac{\theta}{\theta}}\right) = -\ln(\theta) - \frac{x}{\theta}.$$

Furthermore,

$$\frac{\partial}{\partial\theta}\ln f(X \mid \theta) = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

from which it follows that

$$\frac{\partial^2}{\partial\theta^2}\ln f(X \mid \theta) = \frac{1}{\theta^2} - \frac{2x}{\theta^3}$$

Since X is an exponential random variable with the parameter $\frac{1}{\theta}$ (this is one of the X_i 's), then its expected value is θ, so

$$E\left(\frac{\partial^2}{\partial\theta^2}\ln f(X \mid \theta)\right) = E\left(\frac{1}{\theta^2} - \frac{2X}{\theta^3}\right) = \frac{1}{\theta^2} - \frac{2 \cdot E(X)}{\theta^3} = \frac{1}{\theta^2} - \frac{2}{\theta^2} = -\frac{1}{\theta^2}$$

from which we can conclude that the value of $I(\theta)$ is

$$I(\theta) = \frac{1}{\theta^2}$$

Finally, the Cramer-Rao lower bound is

$$\operatorname{var}(\widetilde{\theta}) = \frac{1}{n \cdot I(\theta)} = \frac{\theta^2}{n}$$

Notice that this is exactly the variance of $\overline{X} = \widehat{\mu_X}$ (which in this case is our MLE for θ), so we have that the variance of $\widetilde{\theta}$ reaches the Cramer-Rao lower bound, which means that no unbiased estimator for θ can have lower variance than that of $\widetilde{\theta}$ (we say that $\widehat{\theta}$ is an efficient estimator, and also remember that, here, $\operatorname{var}(\widehat{\theta}) = \operatorname{var}(\widetilde{\theta})$).