**ISyE 6412A** **Theoretical Statistics** **Fall 2023**

**HW#9** (due midnight on Tuesday, Nov 21)

There are six questions, and please look at both sides.

1. (**7.44**). Let $X_1, \ldots, X_n$ be iid $N(\theta, 1)$. Show that the best unbiased estimator of $\theta^2$ is $\bar{X}_n^2 - (1/n)$. Calculate its variance, and show that it is greater than the Cramer-Rao Lower Bound.

   *Hints:* When compute the variance $Var(\delta) = \mathbf{E}(\delta^2) - [\mathbf{E}(\delta)]^2$, you can write $\bar{X}_n = a + bZ$ with $Z \sim N(0,1)$ and suitably constants $a, b$ and then use the fact that for $Z \sim N(0,1)$, we have $\mathbf{E}(Z) = 0, \mathbf{E}(Z^2) = 1, \mathbf{E}(Z^3) = 0$ and $\mathbf{E}(Z^4) = 3$.

   **Answer:** It is straightforward from Theorem 6.2.25 that $T = \sum_{i=1}^n X_i$ is a complete and sufficient statistic for $\theta \in \mathbb{R}$. A direct computation shows us that

$$\mathbf{E}_\theta[\bar{X}_n^2] = Var_\theta(\bar{X}_n) + [\mathbf{E}_\theta(\bar{X}_n)]^2 = \frac{1}{n} + \theta^2.$$

   Thus $\mathbf{E}_\theta[\bar{X}_n^2 - \frac{1}{n}] = \theta^2$, i.e, $\bar{X}_n^2 - \frac{1}{n}$ is an unbiased estimator of $\theta^2$; being a function of only of the complete sufficient statistic $T$, it is therefore the best unbiased estimator of $\theta^2$.
   To find the variance of $\bar{X}_n^2 - \frac{1}{n}$, use the hints, we have

$$\begin{aligned} Var_\theta(\bar{X}_n^2 - \frac{1}{n}) &= \mathbf{E}_\theta(\bar{X}_n^2 - \frac{1}{n} - \theta^2)^2 = \mathbf{E}_\theta\Big[(\frac{1}{\sqrt{n}}Z + \theta)^2 - \frac{1}{n} - \theta^2\Big]^2 \\ &= \mathbf{E}[\frac{1}{n}Z^2 + \frac{2\theta}{\sqrt{n}}Z - \frac{1}{n}]^2 \\ &= \mathbf{E}\Big[\frac{1}{n^2}Z^4 + \frac{4\theta}{n\sqrt{n}}Z^3 + (\frac{4\theta^2}{n} - \frac{2}{n^2})Z^2 - \frac{4\theta}{n\sqrt{n}}Z + \frac{1}{n^2}\Big] \\ &= \frac{4\theta^2}{n} + \frac{2}{n^2}. \end{aligned}$$

   Meanwhile, for $N(\theta, 1)$, we have the Fisher information number $I(\theta) = n$ and thus by Cramer-Rao inequality, any unbiased estimator $W$ of $g(\theta) = \theta^2$ must satisfy $Var(W) \geq [g'(\theta)]^2/I(\theta) = \frac{4\theta^2}{n}$. Therefore, the variance of the best unbiased estimator $\delta^* = \bar{X}_n^2 - (1/n)$ is greater than the Cramer-Rao Lower Bound. □

2. (**7.38**). For each of the following distributions, let $X_1, \ldots, X_n$ be a random sample. Is there a function of $\theta$, say $g(\theta)$, for which there exists an unbiased estimator whose variance attains the Cramer-Rao Lower Bound? If so, find it. If not, show why not.

   (a) $f_\theta(x) = \theta x^{\theta - 1}, \quad 0 < x < 1, \quad \theta > 0;$
   (b) $f_\theta(x) = \frac{\log\theta}{\theta - 1}\theta^x, \quad 0 < x < 1, \quad \theta > 1.$

   **Answer:** (a) We have

$$\begin{aligned} \frac{\partial}{\partial\theta}\log L(\theta|\mathbf{X}) &= \frac{\partial}{\partial\theta}\log\prod_{i=1}^n[\theta x_i^{\theta-1}] \\ &= \frac{\partial}{\partial\theta}\sum_{i=1}^n[\log\theta + (\theta - 1)\log x_i] \\ &= \sum_{i=1}^n[\frac{1}{\theta} + \log x_i] \\ &= \frac{n}{\theta} + \sum_{i=1}^n \log x_i \end{aligned}$$

$$
\begin{aligned}
&= c[\frac{1}{c}\sum_{i=1}^{n}\log x_i + \frac{n}{c\theta}] \\
&= a(\theta)[\delta(\mathbf{x}) - \phi(\theta)],
\end{aligned}
$$

where for constant $c \neq 0$, $a(\theta) = c, \delta(\mathbf{X}) = \frac{1}{c}\sum_{i=1}^{n}\log x_i$ and $\phi(\theta) = -\frac{n}{c\theta}$.
Thus for any constant $c$, $\delta(\mathbf{X}) = \frac{1}{c}\sum_{i=1}^{n}\log x_i$ attains the Cramer-Rao lower bound of $\phi(\theta) = -\frac{n}{c\theta}$. In particular, when the constant $c = -n$, we have that $-\sum_{i=1}^{n}\log x_i/n$ is the best unbiased estimator of $1/\theta$.
(b) We have

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{X}) &= \frac{\partial}{\partial\theta}\log\prod_{i=1}^{n}[\frac{\log\theta}{\theta-1}\theta^{x_i}] \\
&= \frac{\partial}{\partial\theta}\sum_{i=1}^{n}[\log\log\theta + x_i\log\theta - \log(\theta-1)] \\
&= \sum_{i=1}^{n}[\frac{1}{\log\theta}\frac{1}{\theta} + \frac{x_i}{\theta} - \frac{1}{\theta-1}] \\
&= \frac{n}{\theta\log\theta} + \frac{\sum_{i=1}^{n}x_i}{\theta} - \frac{n}{\theta-1} \\
&= \frac{c}{\theta}[\frac{\sum_{i=1}^{n}x_i}{c} + \frac{n}{c\log\theta} - \frac{n\theta}{c(\theta-1)}] \\
&= a(\theta)[\delta(\mathbf{x}) - \phi(\theta)],
\end{aligned}
$$

where for constant $c \neq 0$, $a(\theta) = c/\theta, \delta(\mathbf{X}) = \sum_{i=1}^{n}X_i/c$ and $\phi(\theta) = \frac{n}{c}[\frac{\theta}{\theta-1} - \frac{1}{\log\theta}]$.
Therefore, $\delta(\mathbf{X}) = \sum_{i=1}^{n}X_i/c$ is the best unbiased estimator of $\phi(\theta) = \frac{n}{c}[\frac{\theta}{\theta-1} - \frac{1}{\log\theta}]$. In particular, when the constant $c = n$, it means that the sample mean $\delta^*(\mathbf{X}) = \sum_{i=1}^{n}X_i/n$ is the best unbiased estimator of $\phi^*(\theta) = \frac{\theta}{\theta-1} - \frac{1}{\log\theta} = \frac{\theta\log\theta-\theta+1}{(\theta-1)\log\theta}$, which is just the population mean $\mathbf{E}_\theta(X_i)$. □

3. (**Modified by Problem 7.10**). The random variables $X_1, \cdots, X_n$ are iid with probability density function [motivated from a "practical point of view" at the end of this problem]

$$
f_{\theta_1,\theta_2}(x) = \begin{cases} \theta_2^{-\theta_1}\theta_1 x^{\theta_1-1}, & \text{if } 0 < x \leq \theta_2; \\ 0, & \text{otherwise.} \end{cases}
$$

where $\theta_1 > 0, \theta_2 > 0$, and $\Omega$ will be completed specified later.

(a) Assume $\theta_1$ is known (positive) and $\Omega = \{\theta_2 : \theta_2 > 0\}$. Find the MLE of $\theta_2$.
(b) Assume $\theta_2$ is known (positive) and $\Omega = \{\theta_1 : 0 < \theta_1 < \infty\}$. Find the MLE of $\theta_1$.
(c) Show that the estimator in (a) is biased, but in case (b) the MLE of $1/\theta_1$ is unbiased. [*Hints:* $-\int_0^1 x^{\alpha-1}(\log x)dx = \alpha^{-2}$; incidentally, the MLE of $\theta_1$ itself is biased.]
(d) Assume both $\theta_1$ and $\theta_2$ are unknown, and $\Omega = \{(\theta_1, \theta_2) : 0 < \theta_1 < \infty, 0 < \theta_2 < \infty\}$.
   i. Find a two-dimensional sufficient statistic for $(\theta_1, \theta_2)$.
   ii. Find the MLEs of $\theta_1$ and $\theta_2$.
   iii. Find the MLE estimator of $\phi(\theta_1, \theta_2) = \mathbf{P}_{\theta_1,\theta_2}(X_1 > 1)$.
   iv. The length (in millimeters) of cuckoos' eggs found in hedge sparrow nests can be modelled with this distribution. For the data

   22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 21.7, 23.8, 22.8, 23.1, 23.1, 23.5, 23.0, 23.0,

   Compute the value of MLE in parts (ii) and (iii).

[Model that could yield such a problem: There are iid random variables $Y_j$, uniformly distributed from 0 to $\theta_2$. You send an observer out on each of $n$ successive days to observe some $Y_j$'s. He does not record the $Y_j$'s. Instead, knowing that "the maximum of the $Y_j$'s is sufficient and an MLE," he decides to observe a certain number, $\theta_1$, of the $Y_j$'s each day and computes the maximum of these $\theta_1$ observations. He reports you the value of $X_i$, the maximum he computes on the $i$-th day. Unfortunately, he forgets to tell you the $\theta_1$ he used. Then the $X_i$ has the density function stated for this problem, where we have simplified matters by allowing $\theta_1$ to be any positive value instead of restricting it to integers.]

**Answer:** (a) When $\theta_1 > 0$ is known, the likelihood function is a function of $\theta_2$ and is given by

$$\begin{aligned} L(\theta_2) &= \prod_{i=1}^{n} f_{\theta_1,\theta_2}(x_i) = \prod_{i=1}^{n} [\theta_2^{-\theta_1}\theta_1 x_i^{\theta_1-1} I(0 < x_i \le \theta_2) \\ &= [\theta_1^n (x_1 ... x_n)^{\theta_1-1}][\theta_2^{-n\theta_1} I(\theta_2 \ge x_{(n)})], \end{aligned}$$

which is not a continuous function of $\theta_2$. Please plot it as a function of $\theta_2$!!! Clearly, this function is 0 when $\theta_2 < x_{(n)}$ and is positive when $\theta_2 \ge x_{(n)}$. Moreover, when $\theta_2 \ge x_{(n)}$, $\theta_2^{-n\theta_1}$ is a decreasing function of $\theta_2$ (since $\theta_1 > 0$ is known). Hence, the likelihood function $L(\theta_2)$ is maximized at $\theta_2 = x_{(n)}$ and thus the MLE of $\widehat{\theta}_{2,MLE} = X_{(n)}$.

(b) When $\theta_2 > 0$ is known, the likelihood function is a function of $\theta_1$ and is given by

$$\begin{aligned} L(\theta_1) &= \prod_{i=1}^{n} f_{\theta_1,\theta_2}(x_i) = \prod_{i=1}^{n} [\theta_2^{-\theta_1}\theta_1 x_i^{\theta_1-1} I(0 < x_i \le \theta_2) \\ &= [I(\theta_2 \ge x_{(n)})][\theta_2^{-n\theta_1}\theta_1^n (x_1 ... x_n)^{\theta_1-1}], \end{aligned}$$

which is a smooth function of $\theta_1$, and the log-likelihood function is

$$\log L(\theta_1) = C - n\theta_1 \log\theta_2 + n\log\theta_1 + (\theta_1 - 1)\log(x_1 ... x_n),$$

where $C$ does not depend on $\theta_1$. To find $\widehat{\theta}_{1,MLE}$, we set

$$0 = \frac{\partial}{\partial\theta_1}\log L(\theta_1) = -n\log(\theta_2) + \frac{n}{\theta_1} + \sum_{i=1}^{n}\log x_i.$$

Solving this equation yields the solution

$$\widehat{\theta}_1 = \frac{n}{n\log(\theta_2) - \sum_{i=1}^{n}\log x_i} = \frac{1}{\log(\theta_2) - (1/n)\sum_{i=1}^{n}\log x_i}.$$

Next, we need to check this is indeed a global maximum. To see this,

$$\frac{\partial^2}{\partial\theta_1^2}\log L(\theta_1) = \frac{\partial}{\partial\theta_1}\Big( - n\log(\theta_2) + \frac{n}{\theta_1} + \sum_{i=1}^{n}\log x_i\Big) = -\frac{n}{\theta_1^2} < 0$$

and $\widehat{\theta}_1$ is the local maximum. Moreover, we also need to check the boundary values of $\theta_1$ : it is easy to see that $\log L(\theta_1) \to -\infty$ as $\theta_1 \to 0$ or $+\infty$. Thus $\widehat{\theta}_1$ is the global maximum. In other words, the MLE of $\theta_1$ is

$$\widehat{\theta}_{1,MLE} = \frac{n}{n\log(\theta_2) - \sum_{i=1}^{n}\log x_i} = \frac{1}{\log(\theta_2) - (1/n)\sum_{i=1}^{n}\log x_i}.$$

(c) For the MLE in (a), $\widehat{\theta}_{2,MLE} = X_{(n)}$, we need to first derive of its pdf. It is not difficult to show that the pdf of $\widehat{\theta}_{2,MLE} = X_{(n)}$ is

$$f_{X_{(n)}}(x) = n\theta_1\theta_2^{-n\theta_1}x^{n\theta_1-1}I(0 \le x \le \theta_2).$$

Next, the mean of $\widehat{\theta}_{2,MLE} = X_{(n)}$ is

$$\mathbf{E}_{\theta_2}(X_{(n)}) = \int_0^{\theta_2} n\theta_1 \theta_2^{-n\theta_1} x^{n\theta_1} dx = \frac{n\theta_1}{n\theta_1 + 1}\theta_2$$

and

$$\text{Bias}(\widehat{\theta}_{2,MLE}) = E_{\theta_2}(X_{(n)}) - \theta_2 = -\frac{1}{n\theta_1 + 1}\theta_2$$

Thus, it is biased.
In part (b), it is clear that the MLE of $1/\theta_1$ is

$$\frac{1}{\widehat{\theta}_{1,MLE}} = \log(\theta_2) - \frac{1}{n}\sum_{i=1}^{n} \log(X_i).$$

From the hint we have

$$\mathbf{E}\log(X_i) = \int_0^{\theta_2} \theta_2^{-\theta_1}\theta_1 x^{\theta_1 - 1}\log(x)dx = \int_0^1 (\log t + \log\theta_2)\theta_1 t^{\theta_1 - 1}dt = -\frac{1}{\theta_1} + \log(\theta_2).$$

Hence,

$$\mathbf{E}_{\theta_1}\Big(\frac{1}{\widehat{\theta}_{1,MLE}}\Big) = \mathbf{E}_{\theta_1}\Big(log(\theta_2) - \frac{1}{n}\sum_{i=1}^{n}\log(X_i)\Big) = \log(\theta_2) - \log(\theta_2) + \frac{1}{\theta_1} = \frac{1}{\theta_1},$$

and thus the MLE of $1/\theta_1$, $\log(\theta_2) - (1/n)\sum_{i=1}^{n}\log(X_i)$ is unbiased.
(d)(i) From (a) we know the joint pdf is $\theta_2^{-n\theta_1}\theta_1^n \prod_{i=1}^{n}(x_1...x_n)^n I(\theta_2 \geq x_{(n)})$. We set $T_1(\mathbf{x}) = \prod_{i=1}^{n} x_i$ and $T_2(\mathbf{x}) = x_{(n)}$, define $h(\mathbf{x}) = 1$,

$$g(T_1, T_2|\theta_1, \theta_2) = \theta_2^{-n\theta_1}\theta_1^n T_1^{\theta_1 - 1} I(T_2 \leq \theta_2).$$

Then the pdf $f(\mathbf{x}|\theta_1, \theta_2) = h(x)g(T_1, T_2)$. Thus, by the Factorization Theorem, $(T_1, T_2) = (\prod_{i=1}^{n} X_i, X_{(n)})$ is a two-dimensional sufficient statistic for $(\theta_1, \theta_2)$.
(ii) To find the MLEs of $(\theta_1, \theta_2)$, it is equivalent to maximize the likelihood function $L(\theta_1, \theta_2)$. From (a), we know for any $\theta_1 > 0$, the likelihood function is maximized at $\theta_2 = x_{(n)}$. Hence, the MLE of $\theta_2$ is $X_{(n)}$. Then, from (b), we know given $\theta_2 = X_{(n)}$, the likelihood function is maximized at $\theta_1 = \frac{1}{log(X_{(n)}) - (1/n)\sum_{i=1}^{n} log(X_i)} = n/\sum_{i=1}^{n}\log(X_{(n)}/X_i)$. Thus, we get the MLE of $(\theta_1, \theta_2) = (n/\sum_{i=1}^{n}\log(X_{(n)}/X_i), X_{(n)})$
(iii) The key is note that $\phi(\theta_1, \theta_2) = 0$ if $\theta_2 \leq 1$ and $= 1 - \theta_2^{-\theta_1}$ if $\theta_2 > 1$. Hence, this can be re-written as $\phi(\theta_1, \theta_2) = \max(0, 1 - \theta_2^{-\theta_1})$. Hence, the corresponding MLE of $\phi(\theta_1, \theta_2) = \mathbf{P}_{\theta_1,\theta_2}(X_1 > 1)$ is $\phi(\widehat{\theta}_1, \widehat{\theta}_2) = \max(0, 1 - (\widehat{\theta}_2)^{-\widehat{\theta}_1}) = \max(0, 1 - (X_{(n)})^{-n/\sum_{i=1}^{n}\log(X_{(n)}/X_i)})$.
(iv) The values are (in part (ii)) $(\widehat{\theta}_1, \widehat{\theta}_2) = (12.59, 25)$ and (in part (iii)) $\max(0, 1 - 25^{12.59}) = 1$. □

4. Recall that in problem 6.3 of our text (i.e., problem #5 of HW #5, or problem #1 of HW #8), $X_1, \ldots, X_n$ are assumed to be a random sample from the pdf

$$f(x|\mu, \sigma) = \frac{1}{\sigma}e^{-(x-\mu)/\sigma}, \quad \mu \leq x < \infty, \quad 0 < \sigma < \infty.$$

In each of the following three scenarios, estimate the parameter(s) using both the maximum likelihood estimator (MLE) and the best unbiased estimator:

**(a)** Assume that $\sigma$ is known. Find both MLE and the best unbiased estimator of $\mu$.

**(b)** Assume that $\mu$ is known. Find both MLE and the best unbiased estimator of $\sigma$.

**(c)** Assume that both $\mu$ and $\sigma$ are unknown. Find both MLE and the best unbiased estimator of $\mu$ and $\sigma$.

**Answer:** (a) When $\sigma$ is known, the likelihood function is

$$L(\mu) = \prod_{i=1}^{n} \Big( \frac{1}{\sigma} e^{-(x_i-\mu)/\sigma} I(\mu \le x_i) \Big) = \frac{1}{\sigma^n} \exp\Big( -\frac{1}{\sigma} \sum_{i=1}^{n} x_i + \frac{n\mu}{\sigma} \Big) I(\mu \le x_{(1)}).$$

Note that $L(\mu) = 0$ when $\mu > x_{(1)}$ and $L(\mu) > 0$ when $\mu \le x_{(1)}$. Hence for the MLE, it suffices to consider the region $\mu \le x_{(1)}$. Now the function $L(\mu)$ is an increasing function of $\mu$ when $\mu \le x_{(1)}$. Thus, the MLE of $\mu$ is

$$\widehat{\mu}_{MLE} = X_{(1)}.$$

To find the best unbiased estimator of $\mu$, we need to use the complete sufficient statistics. From previous homework, recall that $X_{(1)}$ is complete sufficient for $\mu$, and the distribution of $n(X_{(1)} - \mu)/\sigma$ is the standard exponential distribution which has mean 1. That is,

$$\mathbf{E}(n(X_{(1)} - \mu)/\sigma) = 1,$$

which implies that

$$\mathbf{E}(X_{(1)} - (\sigma/n)) = \mu.$$

Thus the best unbiased estimator of $\mu$ is

$$\delta^* = X_{(1)} - (\sigma/n).$$

(b) When $\mu$ is known, the complete sufficient statistic of $\sigma$ is $T = \sum_{i=1}^{n} X_i$. In this case, we have

$$\mathbf{E}(X_i) = \mu + \sigma$$

(one way is to note that $(X_i - \mu)/\sigma$ is a standard exponential random variable). Hence $\mathbf{E}(T) = n(\mu+\sigma)$ or

$$\mathbf{E}\Big( \frac{1}{n} T - \mu \Big) = \sigma,$$

and thus the best unbiased estimator of $\sigma$ is

$$\delta^* = \frac{1}{n} T - \mu = \frac{1}{n} \sum_{i=1}^{n} X_i - \mu.$$

To find the MLE of $\sigma$ when $\mu$ is known, note that the log-likelihood function of $\sigma$ is

$$\log L(\sigma) = -n \log(\sigma) - \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma} + \text{Constant}.$$

First, we set

$$0 = \frac{\partial}{\partial \sigma} \log L(\sigma) = -\frac{n}{\sigma} + (\sum_{i=1}^{n} X_i - n\mu) \frac{1}{\sigma^2} = 0.$$

Solving this equation yields the candidate solution

$$\widehat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} X_i - \mu.$$

Second, note that

$$\frac{\partial^2}{\partial \sigma^2} \log L(\sigma)|_{\sigma = \widehat{\sigma}} = \frac{n}{\sigma^2} - 2(\sum_{i=1}^{n} X_i - n\mu) \frac{1}{\sigma^3}|_{\sigma=\widehat{\sigma}} = -\frac{n}{\widehat{\sigma}^2} < 0,$$

which means that $\widehat{\sigma}$ is a local maximum. Also as $\sigma \to 0$ or $\infty$, it is clear that $\log L(\sigma) \to -\infty$. Hence, the MLE of $\sigma$ is

$$\widehat{\sigma} = \frac{1}{n}\sum_{i=1}^{n} X_i - \mu.$$

(c) When both $\mu$ and $\sigma$ are unknown, by (a) and (b), the MLE of $(\mu, \sigma)$ is

$$\widehat{\mu}_{MLE} = X_{(1)} \quad \text{and} \quad \widehat{\sigma}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i - X_{(1)}.$$

To find the best unbiased estimator of $(\mu, \sigma)$, note that $T = (T_1, T_2) = (X_{(1)}, \sum_{i=1}^n X_i)$ is a two-dimensional complete sufficient statistics. As shown in (a) and (b), we have

$$\begin{aligned}
\mathbf{E}(T_1) &= \mathbf{E}(X_{(1)}) = \mu + \frac{1}{n}\sigma \\
\mathbf{E}(T_2) &= \mathbf{E}(\sum_{i=1}^{n} X_i) = n\mu + n\sigma
\end{aligned}$$

From these equations, we have

$$\begin{aligned}
\mathbf{E}(\frac{n^2 T_1 - T_2}{n^2 - n}) &= \mu \\
\mathbf{E}(\frac{1}{n-1}(T_2 - nT_1)) &= \sigma
\end{aligned}$$

Thus,

$$\delta_1^* = \frac{n^2 T_1 - T_2}{n^2 - n} = T_1 - \frac{1}{n(n-1)}(T_2 - nT_1) = X_{(1)} - \frac{1}{n(n-1)}\sum_{i=1}^{n}(X_i - X_{(1)})$$

is an unbiased estimator of $\mu$, whereas

$$\delta_2^* = \frac{1}{n-1}(T_2 - nT_1) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{(1)})$$

is an unbiased estimator of $\sigma$. Since both $\delta_1^*$ and $\delta_2^*$ are functions of the complete sufficient statistic $T = (T_1, T_2) = (X_{(1)}, \sum_{i=1}^n X_i)$, we conclude that they are the best unbiased estimators of $(\mu, \sigma)$. □

5. (**Modified from Problem 7.9**). Let $X_1, \ldots, X_n$ be iid with pdf

$$f_\theta(x) = \frac{1}{\theta}, \quad 0 \le x \le \theta, \quad \theta > 0.$$

**(a)** Estimate $\theta$ using both the method of moments and maximum likelihood.

**(b)** Calculate the means and variances of the two estimators in part (a). Which one should be preferred and why?

**(c)** One can improve the MLE $\widehat{\theta}_{MLE}$ to an unbiased estimator of the form $\delta_c = c\widehat{\theta}_{MLE}$. Find a constant $c$ such that $\mathbf{E}_\theta(\delta_c) = \theta$, i.e., $\delta_c = c\widehat{\theta}_{MLE}$ is an unbiased estimator of $\theta$. Is it the best unbiased estimator of $\theta$?

**(d)** The best estimator of the form of $\delta_c = c\widehat{\theta}_{MLE}$ is the one that uniformly minimizes the risk function $\mathcal{R}_{\delta_c}(\theta) = \mathbf{E}_\theta(\delta_c - \theta)^2$. Find such constant $c$.

**Answer:** (a) For the MOM, note that the population mean is $\mathbf{E}_\theta(X_i) = \theta/2$, and equating the sample mean to the population mean leads to

$$\frac{X_1 + \cdots + X_n}{n} = \frac{\theta}{2},$$

and thus the MOM estimator can be defined as

$$\widehat{\theta}_{MOM} = 2\frac{X_1 + \ldots + X_n}{n} = 2\bar{X}.$$

For the MLE, note that the likelihood function is

$$L(\theta) = f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n}[\frac{1}{\theta}I(0 \le x_i \le \theta)] = \theta^n I(\theta \le \max(x_i)) = \theta^n I(\theta \ge x_{(n)}).$$

Plot $L(\theta)$ as a function of $\theta$! Then we can see that $\widehat{\theta}_{MLE} = X_{(n)}$.

(b) Since $\mathbf{E}_\theta(X_i) = \theta/2$, and $Var_\theta(X_i) = \theta^2/12$, it is easy to see that $\mathbf{E}_\theta(\widehat{\theta}_{MOM}) = \theta$ and $Var_\theta(\widehat{\theta}_{MOM}) = \frac{4}{n}Var_\theta(X_i) = \frac{\theta^2}{3n}$.

We have derived the distribution of $X_{(n)}$ before (in class for complete sufficient statistics), and it is not difficult to show that $\mathbf{E}_\theta(\widehat{\theta}_{MLE}) = \frac{n}{n+1}\theta$ and $Var_\theta(\widehat{\theta}_{MLE}) = \frac{n\theta^2}{(n+2)(n+1)^2}$.

The MLE is preferable (when $n \ge 3$ under the squared error loss function). Why???

Let us consider the risk functions. For any given $\theta$,

$$\begin{aligned}
R_{\widehat{\theta}_{MOM}}(\theta) &= \mathbf{E}_\theta\big(\widehat{\theta}_{MOM} - \theta\big)^2 = Var_\theta\big(\widehat{\theta}_{MOM}\big) + \big(\mathbf{E}_\theta(\widehat{\theta}_{MOM}) - \theta\big)^2 \\
&= \frac{1}{3n}\theta^2, \\
R_{\widehat{\theta}_{MLE}}(\theta) &= \mathbf{E}_\theta\big(\widehat{\theta}_{MLE} - \theta\big)^2 = Var_\theta\big(\widehat{\theta}_{MLE}\big) + \big(\mathbf{E}_\theta(\widehat{\theta}_{MLE}) - \theta\big)^2 \\
&= \frac{n\theta^2}{(n+2)(n+1)^2} + (\frac{n}{n+1}\theta - \theta)^2 \\
&= \frac{2}{(n+1)(n+2)}\theta^2.
\end{aligned}$$

When $n \ge 3$, $\frac{1}{3n} > \frac{2}{(n+1)(n+2)}$, and thus $R_{\widehat{\theta}_{MLE}}(\theta) < R_{\widehat{\theta}_{MOM}}(\theta)$ for all $\theta > 0$. So the MLE is preferable when $n \ge 3$.

(c)(d) We can improve $\widehat{\theta}_{MLE}$ by considering $\delta_c = c\widehat{\theta}_{MLE} = cX_{(n)}$ as follows:

- **Unbiased**. Let $c = \frac{n+1}{n}$ and we have shown in class that $\delta^* = \frac{n+1}{n}X_{(n)}$ is the best unbiased estimator, and $R_{\delta^*}(\theta) = \frac{1}{n(n+2)}\theta^2$, which is smaller than $R_{\widehat{\theta}_{MLE}}(\theta)$ when $n \ge 2$.
- **More interestingly**, when $c = \frac{n+2}{n+1}$ and the procedure $\delta^{**} = \frac{n+2}{n+1}X_{(n)}$ is the best estimator within the family of $\delta_c = c\widehat{\theta}_{MLE} = cX_{(n)}$ in the sense of minimizing the risk function $R_{\delta_c}(\theta)$ at each $\theta$. To see this, note that

$$\begin{aligned}
R_{\delta_c}(\theta) &= \frac{nc^2}{(n+2)(n+1)^2}\theta^2 + (\frac{n}{n+1}c\theta - \theta)^2 = [\frac{n}{(n+2)(n+1)^2}c^2 + (\frac{n}{n+1}c - 1)^2]\theta^2 \\
&= [\frac{n}{n+2}c^2 - \frac{2n}{n+1}c + 1]\theta^2 \\
&= \Big[\frac{n}{n+2}\big(c - \frac{n+2}{n+1}\big)^2 + \frac{1}{(n+1)^2}\Big]\theta^2,
\end{aligned}$$

which is minimized at $c = \frac{n+2}{n+1}$ and the corresponding minimum value is $\frac{1}{(n+1)^2}\theta^2$.

In summary, we prefer $\widehat{\theta}_{MLE} = X_{(n)}$ over $\widehat{\theta}_{MOM} = 2\bar{X}$, the estimator $\delta^* = \frac{n+1}{n}X_{(n)}$ is better than $\widehat{\theta}_{MLE} = X_{(n)}$, and the best estimator is $\delta^{**} = \frac{n+2}{n+1}X_{(n)}$.

$\square$

6. (**This is to show that sometimes MLE has poor performance**). Suppose that $X_1, \ldots, X_n$ are iid with density

$$f_\theta(x) = \begin{cases} \frac{2\theta^2}{(x+\theta)^3}, & \text{if } x > 0; \\ 0, & \text{if } x \le 0. \end{cases}$$

where $\Omega = \{\theta : \theta > 0\}$.

**(a)** If $n = 1$, show that an MLE estimator of $\theta$ is $\widehat{\theta}_a = 2X_1$.

**(b)** Show that $\widehat{\theta}_a$ in part (a) is not an unbiased estimator of $\theta$.
[Verify or believe: $\int_0^\infty \frac{x}{(x+1)^3} dx = \int_1^\infty \frac{u-1}{u^3} du = \frac{1}{2}$ with $u = x + 1$.]

**(c)** Under the squared error loss function $L(\theta, d) = (\theta - d)^2$, show that $\widehat{\theta}_a$ in part (a) is much worse than the constant estimator $\widehat{\theta}^* \equiv 17$. [*Hints*: $\int_0^\infty \frac{x^2}{(1+x)^3} dx = +\infty$.]

**(d)** If $n = 2$, show that an MLE of $\theta$ is $\widehat{\theta}_b = \frac{1}{4}[X_1 + X_2 + \sqrt{X_1^2 + 34X_1X_2 + X_2^2}]$.

[If you want, you can consider the general $n$ by yourself. For general $n$, describe the computation of the MLE in terms of solving a polynomial equation of some degree, checking whether a local maximum is a global maximum, etc.]

**Answer:** (a) when $n = 1$, the likelihood function $L(\theta) = f_\theta(x_1)$, and thus the log-likelihood function is

$$\log L(\theta) = \ell(\theta) \quad = \ln(2\theta^2) - \ln(x_1 + \theta)^3 = \ln(2) + 2\ln\theta - 3\ln(x_1 + \theta)$$

for $x > 0$. Since the domain of $f_\theta(x_1)$ is $\{x_1 > 0\}$, which does not depend on $\theta$. Taking derivatives and setting it to 0, we have

$$0 = \frac{\partial}{\partial\theta} \log L(\theta) \quad = \tfrac{2}{\theta} - \tfrac{3}{x_1+\theta}.$$

Solving $\theta$ from this equation yields that $\widehat{\theta}_a = 2X_1$. This shows that $\widehat{\theta}_a = 2X_1$ is a candidate for the MLE, and now we need to check that it is indeed a global maximum.
For that purpose, we will first check that $\widehat{\theta}_a = 2X_1$ is a local maximum, or equivalently, the concavity of $\log L(\theta)$. This can be done by looking the second-order derivative of $\log L(\theta)$ :

$$\frac{\partial^2}{\partial\theta^2} \log L(\theta)\big|_{\theta=\widehat{\theta}_a} = \Big(-\frac{2}{\theta^2} + \frac{3}{(x_1+\theta)^2}\Big)\big|_{\theta=\widehat{\theta}_a=2x_1} = -\frac{2}{4x_1^2} + \frac{3}{9x_1^2} = -\frac{1}{6x_1^2} < 0.$$

This shows that $\theta_a = 2X_1$ is indeed a local maximum. Next, we also need to check the boundary behavior of $\log L(\theta)$ when $\theta$ goes to 0 or $\infty$. It is clear that for any given observation $x_1$, the log-likelihood function $\log L(\theta) = \ln(2) + 2\ln\theta - 3\ln(x_1 + \theta) \to -\infty$ as $\theta$ goes to 0 or $\infty$. Thus, $\theta_a = 2X_1$ is an MLE estimator.
(b) Using the density function

$$\begin{aligned} \mathbf{E}_\theta[\widehat{\theta}_a] &= \mathbf{E}_\theta(2X_1) = \int_0^\infty 2x \frac{2\theta^2}{(x+\theta)^3} dx \\ &= 4\theta^2 \int_0^\infty \frac{x}{(x+\theta)^3} dx \qquad (\text{let } x = \theta t) \\ &= 4\theta^2 \int_0^\infty \frac{\theta t}{\theta^3 (t+1)^3} \theta dx \end{aligned}$$

$$
\begin{aligned}
&= 4\theta \int_0^\infty \frac{t}{(t+1)^3} dt \qquad (\text{let } t = u-1) \\
&= 4\theta \int_1^\infty \frac{u-1}{u^3} du \\
&= 4\theta(\frac{1}{2}) = 2\theta \neq \theta \qquad (\text{as } \theta > 0).
\end{aligned}
$$

Thus $\theta_a = 2X_1$ is not an unbiased estimator of $\theta$.
(c) By (a) and (b), the risk function of $\widehat{\theta}_a = 2X_1$ is

$$
R_{\widehat{\theta}_a}(\theta) = \mathbf{E}_\theta[(\theta - 2X_1)^2] = \theta^2 - 4\theta\mathbf{E}_\theta[X_1] + 4\mathbf{E}_\theta[X_1^2] = 4\mathbf{E}_\theta[X_1^2] - 3\theta^2.
$$

However,

$$
\begin{aligned}
\mathbf{E}_\theta[X_1^2] &= \int_0^\infty x^2 \frac{2\theta^2}{(x+\theta)^3} dx \qquad (\text{let } x = \theta t) \\
&= 2\theta^2 \int_0^\infty \frac{\theta^2 t^2}{\theta^3 (t+1)^3} \theta dt \\
&= 2 \int_0^\infty \frac{t^2}{(t+1)^3} dt = +\infty.
\end{aligned}
$$

In other words, $\mathbf{E}_\theta(X_1^2) = +\infty$ or equivalently, $Var_\theta(X_1) = +\infty$. Thus, the risk function of the MLE $\widehat{\theta}_a = 2X_1$ is $R_{\widehat{\theta}_a}(\theta) = +\infty$ for each $\theta > 0$. On the other hand, for the constant estimator $\widehat{\theta}^* \equiv 17$, its risk function is $R_{\widehat{\theta}^*}(\theta) = (\theta - 17)^2$, which is finite for each finite $\theta > 0$. Thus, the constant estimator $\widehat{\theta}^* \equiv 17$ is better than the MLE $\widehat{\theta}_a = 2X_1$ (in the sense of uniformly smaller risk function).
(d) The part (d) is very similar to part (a). In this case, we have $n = 2$ and the log-likelihood function

$$
\begin{aligned}
\log L(\theta) &= \ln f_\theta(x_1, x_2) = \ln\Big(\frac{4\theta^4}{(x_1+\theta)^3(x_2+\theta)^3}\Big) \\
&= \ln 4 + 4\ln\theta - 3\ln(x_1+\theta) - 3\ln(x_2+\theta).
\end{aligned}
$$

Taking derivative with respect to $\theta$ and setting it to 0,

$$
0 = \frac{\partial}{\partial\theta} \log L(\theta) = \frac{4}{\theta} - \frac{3}{x_1+\theta} - \frac{3}{x_2+\theta}.
$$

So

$$
\begin{aligned}
0 &= 4(x_1+\theta)(x_2+\theta) - 3\theta(x_2+\theta) - 3\theta(x_1+\theta) \\
&= -2\theta^2 + (x_1+x_2)\theta + 4x_1x_2,
\end{aligned}
$$

and

$$
\Big(\theta - \frac{x_1+x_2}{4}\Big)^2 = \Big(\frac{x_1+x_2}{4}\Big)^2 + 2x_1x_2 = \frac{x_1^2 + 34x_1x_2 + x_2^2}{16}.
$$

Solve $\widehat{\theta}$ from this equation, and recall that $\theta > 0$, we have

$$
\widehat{\theta} = \frac{x_1 + x_2 + \sqrt{x_1^2 + 34x_1x_2 + x_2^2}}{4},
$$

which is a candidate of MLE.
Now for this question, it is a little more challenging to prove that this candidate is a global maximum,

and thus the remaining part is not required. We will present two different approaches.
The first approach is to check the second-order derivatives

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\log L(\theta) &= -\frac{4}{\theta^2}+\frac{3}{(x_1+\theta)^2}+\frac{3}{(x_2+\theta)^2} \\ &= \frac{2\theta^4-2(x_1+x_2)\theta^3-(x_1^2+16x_1x_2+x_2^2)\theta^2-8x_1x_2(x_1+x_2)\theta-4x_1^2x_2^2}{\theta^2(x_1+\theta)^2(x_2+\theta)^2}.\end{aligned}$$

We can prove that this is negative at $\widehat{\theta}=\frac{x_1+x_2+\sqrt{x_1^2+34x_1x_2+x_2^2}}{4}$. To see this, note that $x_1>0, x_2>0$ and $\widehat{\theta}>0$, it suffices to show that

$$2\widehat{\theta}^4-2(x_1+x_2)\widehat{\theta}^3<0, \qquad \text{or equivalently,} \qquad \widehat{\theta}<x_1+x_2.$$

Now $\widehat{\theta}=\frac{x_1+x_2+\sqrt{x_1^2+34x_1x_2+x_2^2}}{4}<x_1+x_2$ if and only if

$$\begin{aligned}&\sqrt{x_1^2+34x_1x_2+x_2^2}<3(x_1+x_2) \\ \Longleftrightarrow\ & x_1^2+34x_1x_2+x_2^2<9(x_1^2+2x_1x_2+x_2^2) \\ \Longleftrightarrow\ & 8x_1^2-16x_1x_2+8x_2^2=8(x_1-x_2)^2\geq 0.\end{aligned}$$

This shows that $\frac{\partial^2}{\partial\theta^2}\log L(\theta)|_{\theta=\widehat{\theta}}<0$, and thus $\widehat{\theta}$ is a local maximum.
Clearly, as $\theta\to 0$ or $\infty$, the log-likelihood function $\log L(\theta)=\ln 4+4\ln\theta-3\ln(x_1+\theta)-3\ln(x_2+\theta)$ goes to $-\infty$. Thus $\widehat{\theta}$ reaches the global maximum, and thus it is the MLE.

The second approach to prove that $\widehat{\theta}$ is a global maximum is to take a further look at the first-order derivative:

$$\frac{\partial}{\partial\theta}\log L(\theta)=-\frac{2}{\theta(x_1+\theta)(x_2+\theta)}\Big[\big(\theta-\frac{x_1+x_2}{4}\big)^2-\frac{x_1^2+34x_1x_2+x_2^2}{16}\Big].$$

It is clear that when $0<\theta<\widehat{\theta}=\frac{x_1+x_2+\sqrt{x_1^2+34x_1x_2+x_2^2}}{4}$, we have $\frac{\partial}{\partial\theta}\log L(\theta)<0$, and thus $\log L(\theta)$ is an decreasing function of $\theta$ over the interval $(0,\widehat{\theta})$. Meanwhile, when $\theta>\widehat{\theta}$, we have $\frac{\partial}{\partial\theta}\log L(\theta)>0$, and thus $\log L(\theta)$ is an increasing function of $\theta$ over the interval $(\widehat{\theta},\infty)$. Combining these two facts together, $\log L(\theta)$ is maximized at $\widehat{\theta}$, which completes the proof of the MLE. □