ISyE 6412A Theoretical Statistics Fall 2023

HW #10 (due at Canvas midnight on Friday, Dec 1, 2023)

Please turn in your solutions to the first four questions. Question #5 includes a list of some extra questions from our text — no credits & not required.

1. (**Modified from 10.1**). This question illustrates that the Method of Moments (MOM) estimators are typically consistent). A random sample $X_1, \ldots, X_n$ is drawn from a population with pdf

$$f_\theta(x) = \frac{1}{2}(1+\theta x), \qquad -1 < x < 1, \quad -1 < \theta < 1.$$

(a) Find the population mean, $\mathbf{E}_\theta(X_1)$, and population variance, $Var_\theta(X_1)$.

(b) Find the MOM estimator $\widehat{\theta}_{MOM}$ of θ, by equating the sample mean to the population mean.

(c) Compute the bias and variance of $\widehat{\theta}_{MOM}$, and show that $\widehat{\theta}_{MOM}$ is a consistent estimator of θ.

Answer: (a) Note that the population mean is

$$\mathbf{E}_\theta(X) = \int_{-1}^{1}[x\frac{1}{2}(1+\theta x)]dx = \int_{-1}^{1}(\frac{1}{2}x + \frac{\theta}{2}x^2)dx = (\frac{1}{4}x^2 + \frac{\theta}{2\times 3}x^3)|_{x=-1}^{x=1} = \theta/3.$$

To calculate its variance, note that

$$\mathbf{E}_\theta(X^2) = \int_{-1}^{1}[x^2\frac{1}{2}(1+\theta x)]dx = \int_{-1}^{1}(\frac{1}{2}x^2 + \frac{\theta}{2}x^3)dx = 1/3,$$

and thus

$$Var_\theta(X) = \mathbf{E}_\theta(X^2) - [\mathbf{E}_\theta(X)]^2 = \frac{1}{3} - (\frac{\theta}{3})^2 = \frac{3-\theta^2}{9}.$$

(b) Equating the sample mean to the population mean yields that the MOM estimator is $\widehat{\theta}_{MOM} = 3\bar{X}$.
(c) It is clear that $\widehat{\theta}_{MOM}$ is an unbiased estimator of θ, i.e., $Bias = 0$. Moreover,

$$Var_\theta(\widehat{\theta}_{MOM}) = 9Var_\theta(\bar{X}) = 9Var_\theta(X)/n = \frac{3-\theta^2}{n},$$

which converges to 0 as $\to \infty$. By Theorem 10.1.3, $\widehat{\theta}_{MOM} = 3\bar{X}$ is consistent. □

2. (**Modified from 10.3**). A random sample $X_1, \ldots, X_n$ ($n \geq 2$) is drawn from $N(\theta, \theta)$, where $\theta > 0$.

(a) Show that the MLE of θ, $\widehat{\theta}$, is a root of the quadratic equation $\theta^2 + \theta - W = 0$, where $W = (1/n)\sum_{i=1}^n X_i^2$, and determine which root equals the MLE.

(b) Compute Fisher information numbers $I_n(\theta)$ and $I_1(\theta)$, and find the asymptotic distribution of $\widehat{\theta}$.

(c) Find the approximate variance of $\widehat{\theta}$ (using the techniques of Section 10.1.3 or other methods).

(d) Suppose that we observe the following $n = 10$ observations:

2.84, 0.93, 4.73, 5.69, 2.11, 2.88, 2.01, 1.17, 2.82, 4.49

For your convenience, the sample mean $\bar{x} = 2.9670$, the sample variance $s^2 = 2.4352$, and the sample second-moment $\sum_{i=1}^n x_i^2 = 109.9475$. Since $\bar{x}$ and s^2 are close, it is reasonable to assume that this data set is a random sample from $N(\theta, \theta)$ distribution. For this dataset, **calculate** the MLE $\widehat{\theta}$ and **find** its approximate variance.

Remark: If interested, you can further compute a so-called 95% confidence interval on the true θ, which is given $[\widehat{\theta} - 1.96\sqrt{\widehat{V}}, \widehat{\theta} + 1.96\sqrt{\widehat{V}}]$, where $\widehat{V}$ is the approximate variance of MLE $\widehat{\theta}$.

Answer: (a) The log-likelihood function is

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

Differentiating this and setting it equal to zero, we have

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

which becomes $\theta^2 + \theta - W = 0$ with $W = (1/n)\sum_{i=1}^n X_i^2$. The roots of this equation are $(-1 \pm \sqrt{1+4W})/2$, and since the MLE has to be non-negative (as the variance $\theta \geq 0$), a candidate for the MLE is

$$\widehat{\theta} = (-1 + \sqrt{1+4W})/2.$$

Note that $W > 0$ with probability 1, and thus $\widehat{\theta} > 0$ with probability 1.
To be more rigours, we also need to check whether this is indeed a global maximum or not. To see this, first check the second-order derivatives:

$$l''(\theta) = \frac{n}{2\theta^2} - \frac{\sum_{i=1}^n x_i^2}{\theta^3} = \frac{n}{2\theta^2} - \frac{nW}{\theta^3} = \frac{n(\theta - 2W)}{\theta^3}.$$

At the point $\theta = \widehat{\theta} = (-1+\sqrt{1+4W})/2$, we have $\widehat{\theta}^2 + \widehat{\theta} - W = 0$ and thus $\widehat{\theta} - 2W = -\widehat{\theta}^2 - W < 0$ (since $W > 0$ with probability 1). Combining this with the fact that $\widehat{\theta} > 0$ yields that $l''(\widehat{\theta}) < 0$, and thus $\widehat{\theta}$ is a local maximum. In addition, it is clear that $l(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\theta) - \frac{\sum_{i=1}^n x_i^2}{2\theta} + \sum_{i=1}^n x_i - \frac{n\theta}{2}$ goes to $-\infty$ as θ goes to 0 or ∞ (check that the coefficients of $\frac{1}{\theta}$ and θ are negative). Therefore, $\widehat{\theta} = (-1 + \sqrt{1+4W})/2$ is indeed the global maximum, and thus it is the MLE of θ.
(b) Replace the x_i by the corresponding random variable X_i and note that $\mathbf{E}_\theta(X_i^2) = Var_\theta(X_i + (E_\theta(X_i))^2 = \theta + \theta^2$, we have that the Fisher Information Number is

$$I_n(\theta) = -\mathbf{E}_\theta(\frac{n}{2\theta^2} - \frac{\sum_{i=1}^n X_i^2}{\theta^3}) = -\frac{n}{2\theta^2} + \frac{n\mathbf{E}_\theta(X_i^2)}{\theta^3} = -\frac{n}{2\theta^2} + \frac{n(\theta+\theta^2)}{\theta^3} = \frac{n(2\theta+1)}{2\theta^2}.$$

and

$$I_1(\theta) = \frac{2\theta+1}{2\theta^2}.$$

By the asymptotic theory for the MLE, the asymptotic variance of $\widehat{\theta}$ is $\frac{1}{I_n(\theta)} = \frac{2\theta^2}{n(2\theta+1)}$, and thus

$$\widehat{\theta} \sim N(\theta, \frac{2\theta^2}{n(2\theta+1)})$$

or equivalently,

$$\sqrt{n}(\widehat{\theta} - \theta) \sim N(0, \frac{1}{I_1(\theta)}) = N(0, \frac{2\theta^2}{2\theta+1}).$$

(c) The above result is fine from the theoretical viewpoint, however, the true θ is unknown, and thus we cannot compute the variance $\frac{2\theta^2}{n(2\theta+1)}$ directly from the data. Hence, it is better to approximate $Var(\widehat{\theta})$ using the "observed" Fisher Information.

$$\mathrm{Var}_\theta(\widehat{\theta}) \approx \frac{1}{-(\frac{n}{2\theta^2} - \frac{\sum_{i=1}^n X_i^2}{\theta^3})} \approx \frac{1}{-\frac{n}{2\widehat{\theta}^2} + \frac{\sum_{i=1}^n X_i^2}{\widehat{\theta}^3}} = \frac{2\widehat{\theta}^3}{2nW - n\widehat{\theta}} = \frac{2\widehat{\theta}^2}{n(2\widehat{\theta}+1)}.$$

(d) By (a) and (c), $W = \sum_{i=1}^n X_i^2/n = 10.99475$ and

$$\widehat{\theta} = (-1 + \sqrt{1+4W})/2 = (-1 + \sqrt{1 + 4 \times 10.99475})/2 = 2.8533$$

and

$$\widehat{V} = Var(\widehat{\theta}) \approx \frac{2\widehat{\theta}^2}{n(2\widehat{\theta}+1)} = \frac{2 \times 2.8533^2}{10(2 \times 2.8533 + 1)} = 0.2428.$$

A 95% confidence interval on θ is

$$\widehat{\theta} \pm 1.96\sqrt{\widehat{V}} = 2.8533 \pm 1.96\sqrt{0.2428} = [1.8875, 3.8191].$$

Did you how I get those data? I used R to randomly generate these ten observations from $N(\theta, \theta)$ when $\theta = 2$!!! So the 95% confidence interval looks pretty good in this case. You can write some compute code to randomly generate other datasets to see what happen. □

3. (**Modified from 10.9**). Assume that $X_1, \ldots, X_n$ are iid Poisson(λ). As in Midterm II, please feel free to use the fact that $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a complete sufficient statistic for λ. Suppose now that we are interested in estimating $\phi(\lambda) = \lambda e^{-\lambda}$, the probability that $X = 1$.

   (a) Find the best unbiased estimator of $\phi(\lambda) = \lambda e^{-\lambda}$, the probability that $X = 1$.

   (b) Calculate the Fisher Information numbers $I_n(\lambda)$ and $I_1(\lambda)$.

   (c) Derive the Cramer-Rao lower bound for any unbiased estimator of $\phi(\lambda) = \lambda e^{-\lambda}$.

   (d) Find the MLE of $\phi(\lambda) = \lambda e^{-\lambda}$, and derive its asymptotic distribution.

   (e) A preliminary test of a possible carcinogenic compound can be performed by measuring the mutation rate of microorganisms exposed to the compound. An experimenter places the compound in 15 petri dishes and records the following number of mutant colonies:

   $$10,\ 7,\ 8,\ 13,\ 8,\ 9,\ 5,\ 7,\ 6,\ 8,\ 3,\ 6,\ 6,\ 3,\ 5.$$

   For this data set, **calculate** both the best unbiased estimator and the MLE of $\phi(\lambda) = \lambda e^{-\lambda}$, the probability that one mutant colony will emerge. In addition, **find** the approximate variance of the MLE of $\phi(\lambda) = \lambda e^{-\lambda}$ (using the techniques of Section 10.1.3 or other methods).

   Remark: all your answers in part (e) should be numerical values.

Answer: (a) For the best unbiased estimator, note that $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a complete sufficient statistic for λ (can you prove it by yourself? Also see problem 7.52 on page 365 of our text). To find the best unbiased estimator of $h(\lambda) = \mathbf{P}_\lambda(X = 1) = \lambda e^{-\lambda}$, Let

$$\delta = \delta(X_1, \cdots, X_n) = \begin{cases} 1, & \text{if } X_1 = 1; \\ 0, & \text{if } X_1 \neq 1. \end{cases}$$

It is easy to see that $\mathbf{E}_\lambda(\delta) = \mathbf{P}_\lambda(\delta = 1) = \mathbf{P}_\lambda(X_1 = 1) = \lambda e^{-\lambda}$. So δ is an unbiased estimator of $h(\lambda) = \mathbf{P}_\lambda(X = 1) = \lambda e^{-\lambda}$. By the Rao-Blackwell theorem, the best unbiased estimator is given by $\delta^* = \mathbf{E}(\delta|T)$.

By the definition, δ is either 1 or 0, and thus, when $T(\mathbf{X}) = b$ for some non-negative integer b,

$$\begin{aligned}
\delta^*(b) &= \mathbf{E}(\delta|T = b) \\
&= \sum_c c\mathbf{P}_\lambda(\delta(\mathbf{X}) = c|T(\mathbf{X}) = b) \\
&= \mathbf{P}_\lambda(\delta(\mathbf{X}) = 1|T(\mathbf{X}) = b) \qquad \text{(since } c = 0 \text{ or } 1) \\
&= \frac{\mathbf{P}_\lambda(\delta(\mathbf{X}) = 1 \text{ and } T(\mathbf{X}) = b)}{\mathbf{P}_\lambda(T(\mathbf{X}) = b)} \\
&= \frac{\mathbf{P}_\lambda(X_1 = 1 \text{ and } \sum_{i=2}^n X_i = b - 1)}{\mathbf{P}_\lambda(\sum_{i=1}^n X_i = b)}
\end{aligned}$$

Recall that the sum of independent Poisson RVs is still Poisson distributed, and thus $\sum_{i=1}^n X_i$ and $\sum_{i=2}^n X_i$ are Poisson distributed with means $n\lambda$ and $(n-1)\lambda$, respectively. Thus,

$$\begin{aligned}
\delta^*(b) &= \begin{cases} 0, & \text{if } b = 0; \\ \frac{\{\lambda e^{-\lambda}\}\{[(n-1)\lambda]^{b-1}e^{-[(n-1)\lambda]}/(b-1)!\}}{(n\lambda)^b e^{-(n\lambda)}/b!)} = b\frac{(n-1)^{b-1}}{n^b}, & \text{if } b \geq 1. \end{cases} \\
&= \frac{b}{n}\Big(1 - \frac{1}{n}\Big)^{b-1}.
\end{aligned}$$

Replacing b by $T = \sum_{i=1}^n X_i$ will lead to the corresponding best unbiased estimator of $\lambda e^{-\lambda}$:

$$\delta^* = \frac{T}{n}\Big(1 - \frac{1}{n}\Big)^{T-1} \qquad \text{with} \quad T = \sum_{i=1}^n X_i.$$

(b) Note that

$$\begin{aligned}
\ell(\lambda) &= \log \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = (\sum_{i=1}^n x_i)\log\lambda - n\lambda - \log(\prod x_i!) \\
S(\lambda) &= \frac{\partial}{\partial\lambda}\ell(\lambda) = (\sum_{i=1}^n x_i)/\lambda - n \\
\ell''(\lambda) &= \frac{\partial^2}{\partial\lambda^2}\ell(\lambda) = \frac{\partial}{\partial\lambda}S(\lambda) = -(\sum_{i=1}^n x_i)/\lambda^2 \\
I_n(\lambda) &= -\mathbf{E}(\ell''(\lambda)) = \frac{1}{\lambda^2}\mathbf{E}_\lambda(\sum_{i=1}^n X_i) = \frac{1}{\lambda^2}(n\lambda) = n/\lambda.
\end{aligned}$$

Thus the Fisher Information number is $I_n(\theta) = \frac{n}{\lambda}$ and $I_1(\theta) = \frac{1}{\lambda}$. Also see Example 10.1.7 on page 476 of our text.
(c) The Cramer-Rao lower bound for any unbiased estimator of $\phi(\lambda) = \lambda e^{-\lambda}$ is

$$\frac{[\phi'(\lambda)]^2}{I_n(\lambda)} = \frac{(1-\lambda)^2 e^{-2\lambda}}{n/\lambda} = \frac{1}{n}\lambda(1-\lambda)^2 e^{-2\lambda}.$$

(d) For the MLE, first note that the log-likelihood function is

$$\log L(\lambda) = \log \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = (\sum_{i=1}^n x_i)\log\lambda - n\lambda + \prod_{i=1}^n \log(x_i!)$$

and

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\log L(\lambda) &= \frac{\sum_{i=1}^n x_i}{\lambda} - n \\
\frac{\partial^2}{\partial\lambda^2}\log L(\lambda) &= -\frac{\sum_{i=1}^n x_i}{\lambda^2}.
\end{aligned}$$

From this, it is not difficult to prove that the MLE of λ is $\widehat{\lambda} = \sum_{i=1}^n X_i/n = \bar{X}$. Hence, the MLE of $\mathbf{P}(X = 1) = \lambda e^{-\lambda}$ is $\bar{X}e^{-\bar{X}}$.
By Theorem 10.1.12,

$$\sqrt{n}(\tau(\widehat{\lambda}) - \tau(\lambda)) \to N(0, V) \text{ in distribution} \qquad \text{with } V = \frac{[\tau'(\lambda)]^2}{1/\lambda} = \lambda[\tau'(\lambda)]^2.$$

In particular, $\sqrt{n}(\widehat{\lambda}e^{-\widehat{\lambda}} - \lambda e^{-\lambda})$ converges to $N(0, \lambda(1-\lambda)^2 e^{-2\lambda})$ in distribution.
(e) Here $n = 15, T = \sum_{i=1}^n X_i = 104$ and $\widehat{\lambda} = \bar{X} = T/n = 6.9333$.

The best unbiased estimator $\delta^* = 0.0056850$ and the MLE is $\widehat{\lambda}e^{-\widehat{\lambda}} = 0.0067582$.
To approximate the variance of the MLE of $\phi(\lambda) = \lambda e^{-\lambda}$, using the techniques in Section 10.1.3, we have

$$\begin{aligned}
\widehat{\mathrm{Var}}[\phi(\widehat{\lambda})] &= \frac{[\phi'(\lambda)]^2|_{\lambda=\widehat{\lambda}}}{-\frac{\partial^2}{\partial^2\lambda^2}\ell(\lambda)|_{\lambda=\widehat{\lambda}}} = \frac{(1-\lambda)^2 e^{-2\lambda}|_{\lambda=\widehat{\lambda}}}{(\sum_{i=1}^n X_i)/\lambda^2|_{\lambda=\widehat{\lambda}}} \\
&= \frac{(1-\widehat{\lambda})^2 e^{-2\widehat{\lambda}}}{(\sum_{i=1}^n X_i)/\widehat{\lambda}^2} \\
&= \frac{1}{n}\widehat{\lambda}(1-\widehat{\lambda})^2 e^{-2\widehat{\lambda}} \qquad (\text{since } \sum_{i=1}^n X_i = n\widehat{\lambda} \\
&= \frac{1}{15} \times 6.9333 \times (1-6.9333)^2 \times e^{-2\times 6.9333} = 1.5462 \times 10^{-5}.
\end{aligned}$$

□

4. (MLE and other topics). Assume that $X_1, \ldots, X_n$ are iid with density

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}e^{-x/\theta}, & \text{if } x > 0; \\ 0, & \text{if } x \le 0. \end{cases}$$

and we want to estimate $\phi(\theta) = \theta^2$ under the squared error loss. Here $\Omega = \{\theta : \theta > 0\}$, $D = \{: d > 0\}$, and $L(\theta, d) = (\phi(\theta) - d)^2$. Recall that $T = \sum_{i=1}^n X_i$ is a complete sufficient statistic for θ, and has a Gamma(n, θ) distribution and $\mathbf{E}_\theta(T^k) = \theta^k(n+k-1)!/(n-1)!$ for k an integer ($= \infty$ if $n+k-1 < 0$.)

(a) Find that the MLE estimator of $\phi(\theta) = \theta^2$.

(b) Show that the estimator of θ^2 of the form $\delta_c = cT^2$ (with c constant) has risk function

$$R_{\delta_c}(\theta) = \mathbf{E}_\theta(\delta_c - \theta^2)^2 = \theta^4\{c^2 n(n+1)(n+2)(n+3) - 2cn(n+1) + 1\}.$$

In particular, the MLE estimator of $\phi(\theta) = \theta^2$ corresponds to $\delta_c = cT^2$ with $c = \frac{1}{n^2}$ and show that the MLE in (a) yields risk function $\theta^4\left\{\frac{4n^2+11n+6}{n^3}\right\}$.

(c) One can improve the MLE to an unbiased estimator of the form $\delta_c = cT^2$. Show that the best unbiased estimator of θ^2 corresponds to $c = \frac{1}{n(n+1)}$ and yields risk function $\theta^4\left\{\frac{4n+6}{n(n+1)}\right\}$.

(d) The best estimator of the form of $\delta_c = cT^2$ is the one that uniformly minimizes the risk function. For the estimator of the form $\delta_c = cT^2$, show that the "best" choice of c is $c = \frac{1}{(n+2)(n+3)}$ and that the resulting estimator has the smallest risk function $\theta^4\left\{\frac{4n+6}{(n+2)(n+3)}\right\}$.

(e) **Compute** the Fisher information $I_n(\theta)$ and $I_1(\theta)$, and **show that** the Cramer-Rao lower bound for any unbiased estimator of θ^2 is $H_n(\theta) = 4\theta^4/n$. Consequently, assuming that biased estimators also cannot have risk much smaller than this bound when n is large, we define a sequence of estimators $\delta(X_1, \ldots, X_n) = \delta_n$ to be *asymptotically efficient* if $R_{\delta_n}(\theta)/H_n(\theta) \to 1$ as $n \to \infty$, for each θ in Ω. Here $R_{\delta_n}(\theta)$ is the risk function of the estimator δ_n. **Show that** under this definition, each of the sequences of estimators in (a), (c), (d) is asymptotically efficient.

(f) Another type of estimators can be obtained by a Bayes procedure relative to some prior density $\pi(\theta)$. One possible choice of $\pi(\theta)$ is $\pi(\theta) = \begin{cases} \theta^{-2}e^{-1/\theta}, & \text{if } \theta > 0; \\ 0, & \text{otherwise.} \end{cases}$
First **show that** this is indeed a probability density function by integrating it, using the transformation $u = \theta^{-1}$. Next, recall our lectures on Bayes procedure, and **show that** for $n \ge 2$, the Bayes procedure (under the squared error loss $L(\theta, d) = (\theta^2 - d)^2$) is $\delta^* = \frac{(T+1)^2}{n(n-1)}$.

Remark: Note that, for fixed "true" $\theta > 0$, T/n is very likely to be close to θ when n is large, so that the Bayes estimator δ^* is likely to be close in values to the estimators of (b), (c), (d).

Answer: (a) The log-likelihood function is

$$\ell(\theta) = \log f_\theta(X_1, \ldots, X_n) = \log \prod_{i=1}^n \frac{1}{\theta} \exp\left(-\frac{X_i}{\theta}\right) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^n X_i.$$

Taking the derivative with respect to θ, and setting it to 0 yields that

$$0 = \frac{\partial \ell(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n X_i.$$

Solving this equation, we have that $\widehat{\theta} = \sum_{i=1}^n X_i/n = T/n$ is a candidate of MLE. To check whether it is a global maximum, we first look at the second-order derivative

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n X_i = \frac{n}{\theta^2}\left(1 - 2\frac{T/n}{\theta}\right).$$

and clearly

$$\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=T/n} = \frac{n^3}{T^2}\left(1 - 2\frac{T/n}{T/n}\right) = -\frac{n^3}{T^2} < 0,$$

and thus $\widehat{\theta} = \sum_{i=1}^n X_i/n = T/n$ is a local maximum. Moreover, as $\theta \to 0$ or ∞, it is trivial to see that $\ell(\theta) \to -\infty$. Hence, $\widehat{\theta}$ is the MLE of θ, and the MLE of $\phi(\theta) = \theta^2$ is $(T/n)^2 = T^2/n^2$.

(b) Note that

$$R_{\delta_c}(\theta) = \mathbf{E}_\theta[(\theta^2 - \delta_c)^2] = Var_\theta(\delta_c) + (\theta^2 - \mathbf{E}_\theta[\delta_c])^2.$$

Since $\mathbf{E}_\theta[T^k] = \theta^k (n+k-1)(n+k-2)\cdots(n+1)n$, we have

$$\begin{aligned}
\mathbf{E}_\theta[\delta_c] &= \mathbf{E}_\theta[cT^2] = c\theta^2 n(n+1), \\
(\theta^2 - \mathbf{E}_\theta[\delta_c])^2 &= (\theta^2 - c\theta^2 n(n+1))^2 = \theta^4(c^2 n^2 (n+1)^2 - 2cn(n+1) + 1), \\
Var_\theta(\delta_c) &= c^2 Var_\theta(T^2) \\
&= c^2(\mathbf{E}_\theta[T^4] - (\mathbf{E}_\theta[T^2])^2) \\
&= c^2\theta^4\left((n+3)(n+2)(n+1)n - (n+1)^2 n^2\right).
\end{aligned}$$

Thus, the risk function of $\delta_c = cT^2$ is

$$\begin{aligned}
R_{\delta_c}(\theta) = &\; c^2\theta^4\left((n+3)(n+2)(n+1)n - (n+1)^2 n^2\right) \\
&+ \theta^4(c^2 n^2 (n+1)^2 - 2cn(n+1) + 1) \\
= &\; \theta^4(c^2(n+3)(n+2)(n+1)n - 2cn(n+1) + 1) \\
= &\; \theta^4\left((n+1)n\left(c - \tfrac{1}{(n+3)(n+2)}\right)^2 + \text{const.}\right),
\end{aligned}$$

In particular, by (a), the MLE is of the form $\delta_c = cT^2$ with $c = 1/n^2$. Plugging $c = 1/n^2$ into the previous result, we have

$$\begin{aligned}
R_{\delta_c}(\theta) &= \theta^4\left(\frac{1}{n^4}(n+3)(n+2)(n+1)n - 2\frac{1}{n^2}n(n+1) + 1\right) \\
&= \frac{\theta^4}{n^3}\left((n+3)(n+2)(n+1) - 2n^2(n+1) + n^3\right) \\
&= \frac{\theta^4(4n^2 + 11n + 6)}{n^3}.
\end{aligned}$$

(c) When $c = 1/(n(n+1))$, the procedure $\delta_c = cT^2$ satisfies

$$\begin{aligned}
\mathbf{E}_\theta[\delta_c] &= \frac{1}{n(n+1)}\mathbf{E}_\theta[T^2] \\
&= \frac{1}{n(n+1)}\theta^2(n+1)n = \theta^2,
\end{aligned}$$

and thus $\delta_c = \frac{1}{n(n+1)}T^2$ is the unbiased estimator of θ^2. Since this unbiased estimator is the function of T, the complete sufficient statistic of θ, we can conclude that this is also the best unbiased estimator of θ^2.

Plugging $c = 1/(n(n+1))$ into the result in (a), the risk function of the best unbiased estimator is

$$\begin{aligned} R_{\delta_c}(\theta) &= \theta^4\left(\frac{1}{n^2(n+1)^2}(n+3)(n+2)(n+1)n - 2\frac{1}{n(n+1)}n(n+1)+1\right) \\ &= \theta^4\left(\frac{(n+2)(n+3)}{n(n+1)} - 2 + 1\right) = \frac{\theta^4(4n+6)}{n(n+1)}. \end{aligned}$$

(d) From the result in (b), $R_{\delta_c}(\theta)$ is minimized when $c = 1/((n+2)(n+3))$, and the corresponding risk function becomes

$$\begin{aligned} R_{\delta_c}(\theta) &= \theta^4\left(\frac{(n+3)(n+2)(n+1)n}{(n+2)^2(n+3)^2} - 2\frac{n(n+1)}{(n+2)(n+3)} + 1\right) \\ &= \frac{\theta^4}{(n+2)(n+3)}(n(n+1) - 2n(n+1) + (n+2)(n+3)) \\ &= \frac{\theta^4(4n+6)}{(n+2)(n+3)}. \end{aligned}$$

(e) We have $I_n(\theta) = n/\theta^2$ and $I_1(\theta) = 1/\theta^2$. The proof (or computation) is straightforward, and see the following for the details.

$$\begin{aligned} I_n(\theta) &= \mathbf{E}_\theta\left[-\frac{\partial^2 \ell(\theta)}{\partial\theta^2}\right] = \mathbf{E}_\theta\left[-\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^n X_i\right] \\ &= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\mathbf{E}_\theta[T] = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\theta n = \frac{n}{\theta^2}, \\ I_1(\theta) &= \frac{1}{\theta^2}, \\ H_n(\theta) &= \frac{(2\theta)^2}{I_n(\theta)} = \frac{4\theta^2}{n/\theta^2} = \frac{4\theta^4}{n}. \end{aligned}$$

The sequence of estimators in (a), (c), and (d) are asymptotically efficient since as $n \to \infty$,

$$\begin{aligned} (\mathrm{a}): \quad \frac{R_\delta(\theta)}{H_n(\theta)} &= \frac{4n^2+11n+6}{4n^2} \to 1, \\ (\mathrm{c}): \quad \frac{R_\delta(\theta)}{H_n(\theta)} &= \frac{4n+6}{4(n+1)} \to 1, \\ (\mathrm{d}): \quad \frac{R_\delta(\theta)}{H_n(\theta)} &= \frac{n(4n+6)}{4(n+2)(n+3)} \to 1. \end{aligned}$$

(f) In part (f), note that, for each $X = (x_1, \ldots, x_n)$, a Bayes procedure shall minimize

$$\begin{aligned} h^*_\pi(X,d) &= \int_{-\infty}^{\infty} L(\theta,d) f_\theta(x_1,\cdots,x_n)\pi(\theta)d\theta \\ &= \int_0^\infty (\theta^2-d)^2[\theta^{-n}e^{-T/\theta}][\theta^{-2}e^{-1/\theta}]d\theta \\ &= \int_0^\infty (\theta^4 - 2\theta^2 d + d^2)[\theta^{-(2+n)}e^{-(T+1)/\theta}]d\theta \\ &= Ad^2 - 2Bd + C, \end{aligned}$$

where

$$A = \int_0^\infty [\theta^{-(2+n)}e^{-(T+1)/\theta}]d\theta$$

$$
\begin{aligned}
&= \int_0^\infty [u^{2+n} e^{-(T+1)u}] u^{-2} du \qquad \text{(let } u = 1/t) \\
&= \int_0^\infty u^n e^{-(T+1)u} du \\
&= \frac{1}{(T+1)^{n+1}} \int_0^\infty t^n e^{-t} dt \qquad \text{(let } t = (T+1)u) \\
&= \frac{\Gamma(n+1)}{(T+1)^{n+1}} = \frac{n!}{(T+1)^{n+1}} \\
B &= \int_0^\infty \theta^2 [\theta^{-(2+n)} e^{-(T+1)/\theta}] d\theta \\
&= \int_0^\infty u^{-2} [u^{2+n} e^{-(T+1)u}] u^{-2} du \qquad \text{(let } u = 1/t) \\
&= \int_0^\infty u^{n-2} e^{-(T+1)u} du \\
&= \frac{1}{(T+1)^{n-1}} \int_0^\infty t^{n-2} e^{-t} dt \qquad \text{(let } t = (T+1)u) \\
&= \frac{\Gamma(n-1)}{(T+1)^{n-1}} = \frac{(n-2)!}{(T+1)^{n-1}} \\
C &= \int_0^\infty \theta^4 [\theta^{-(2+n)} e^{-(T+1)/\theta}] d\theta \\
&= \int_0^\infty u^{-4} [u^{2+n} e^{-(T+1)u}] u^{-2} du \qquad \text{(let } u = 1/t) \\
&= \int_0^\infty u^{n-4} e^{-(T+1)u} du \\
&= \frac{1}{(T+1)^{n-3}} \int_0^\infty t^{n-4} e^{-t} dt \qquad \text{(let } t = (T+1)u) \\
&= \frac{\Gamma(n-3)}{(T+1)^{n-3}} = \frac{(n-4)!}{(T+1)^{n-3}}
\end{aligned}
$$

and thus the integral $h_\pi^*(X, d)$ is minimized at

$$\widehat{d} = \frac{B}{A} = \frac{(n-2)!}{(T+1)^{n-1}} \times \frac{(T+1)^{n+1}}{n!} = \frac{(T+1)^2}{n(n-1)},$$

which is exactly δ^*. Thus, δ^* minimizes $h_\pi^*(X, d)$ over D for all X. Hence, it is Bayes relative to the prior density $\pi(\theta)$. □

5* (**Optional, Extra Questions: NOT Required and No credits**, as the TA will not be able to grade these optional questions). To enhance your understanding, you may work many other exercises/questions from our text such as: 7.15 & 7.16 (MLE), 7.46, 7.47 & 7.48 (the best unbiased estimator), 7.49 (estimation via sufficient statistic), 7.50 & 7.51 (how to improve an estimator), 7.58 (Unbiased), 7.62 (Bayes), 7.66 (Jackknife), 10.7 & 10.8 (proofs for MLE asymptotic), 10.10 & 10.11, 10.17 (a)(d)(e), 10.22, 10.23 (Asymptotic properties).

Below is a research type problem for those students who are interested in research.

(a) (**Modified from 7.19**). Suppose that we observe the real-valued random variables $Y_1, \cdots, Y_n$ that satisfies the AR(1) model: $Y_i = \beta Y_{i-1} + \epsilon_i$ for $i \geq 1$ and $Y_0 = 0$, where the ϵ_i's are iid $N(0, \sigma^2)$, σ^2 unknown. Find the MLE of β, and derive the asymptotic distribution of $\widehat{\beta}_{MLE}$.
Remark: The asymptotic distribution will depend on the true value of $\beta : |\beta| < 1, = 1$ or > 1.

(b) In practice, the initial value Y_0 is generally unobservable, how would you modify your answers in (a)? That is, when we are not sure whether $Y_0 = 0$ or not, how to estimate β based on $Y_1, Y_2, \cdots, Y_n$, and what is the asymptotic distribution of your estimator?

(c) Another extension is the AR(2) model: assume that the observed data $Y_1, \ldots, Y_n (n \geq 4)$ are from the AR(2) model: $Y_i = \beta_1 Y_{i-1} + \beta_2 Y_{i-2} + \epsilon_i$, where the ϵ_i's are iid $N(0, \sigma^2)$, σ^2 unknown. Find a good estimator of $\theta = (\beta_1, \beta_2)$, and derive its asymptotic distribution.