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 θ^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 $\mathbf{E}(Z^4) = 3$.

2. (**7.38**). For each of the following distributions, let $X_1, \ldots, X_n$ be a random sample. Is there a function of θ, 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) $f_\theta(x) = \theta x^{\theta-1}, \qquad 0 < x < 1, \quad \theta > 0;$
   (b) $f_\theta(x) = \frac{\log\theta}{\theta-1}\theta^x, \qquad 0 < x < 1, \quad \theta > 1.$

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]

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

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

   (a) Assume θ_1 is known (positive) and $\Omega = \{\theta_2 : \theta_2 > 0\}$. Find the MLE of θ_2.

   (b) Assume θ_2 is known (positive) and $\Omega = \{\theta_1 : 0 < \theta_1 < \infty\}$. Find the MLE of θ_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 θ_1 itself is biased.]

   (d) Assume both θ_1 and θ_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 (θ_1, θ_2).
      ii. Find the MLEs of θ_1 and θ_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 θ_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, θ_1, of the Y_j's each day and computes the maximum of these θ_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 θ_1 he used. Then the X_i has the density function stated for this problem, where we have simplified matters by allowing θ_1 to be any positive value instead of restricting it to integers.]

4. Recall that in problem 6.3 of our text (i.e., problem #6 of HW #7 with special cases in problem #5 of HW #7 and problem #2 of HW #8), $X_1, \ldots, X_n$ are assumed to be a random sample from the pdf

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

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 σ is known. Find both MLE and the best unbiased estimator of μ.

(b) Assume that μ is known. Find both MLE and the best unbiased estimator of σ.

(c) Assume that both μ and σ are unknown. Find both MLE and the best unbiased estimator of μ and σ.

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 θ 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 θ. Is it the best unbiased estimator of θ?

(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.

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 θ is $\widehat{\theta}_a = 2X_1$.

(b) Show that $\widehat{\theta}_a$ in part (a) is not an unbiased estimator of θ.
[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 θ 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.]