ISyE 6412A Theoretical Statistics Fall 2023

HW #10 (due at Canvas midnight on Friday, Dec 01, 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.

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 $\widehat{\theta}$.

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.

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 $\delta^* = \frac{(T+1)^2}{n(n-1)}$.

<u>*Remark:*</u> 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).

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 $i \ge 1$ 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 $Y_1, \ldots, Y_n (n \ge 4)$ 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.