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

HW #6 (due at Canvas midnight on Friday, Oct 20, ET)

(There are 6 questions. After you spent at least 30 minutes per question, please look at the hints on the last two pages.)

1. Suppose $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ are independent and identically distributed (iid) with a Uniform$[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ distribution for some unknown $-\infty < \theta < \infty$, i.e., the X_i's have density

$$f_\theta(x) = \begin{cases} 1, & \text{if } \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

It is desired to guess the value of θ under the loss function $L(\theta, d) = (\theta - d)^2$ based on the observed data $\mathbf{X} = (X_1, \ldots, X_n)$. The purpose of this question is to show that the sample mean is inadmissible.

(a) Specify S, Ω, D, and L (i.e., the sample space, the set of all possible distribution functions, the decision space, and the loss function).

(b) Find the risk function of the procedure $\delta_0(\mathbf{X}) = \bar{X}_n = (X_1 + \ldots + X_n)/n$, which is the so-called method of moment estimator.

(c) **Prove that** $T = (X_{(1)}, X_{(n)})$ is a sufficient statistic for θ, where $X_{(1)} = \min(X_1, \ldots, X_n)$ and $X_{(n)} = \max(X_1, \ldots, X_n)$ are the sample minimum and maximum.

(d) While $T = (X_{(1)}, X_{(n)})$ gives all the information about θ, the T itself is not a statistical procedure for estimating θ, since a point estimator of θ must take on real values. To produce point estimators from sufficient statistic T, let us consider a family of procedures of the form

check cov(U(1),U(n))

$$\delta_{a,b}(\mathbf{X}) = aX_{(1)} + (1-a)X_{(n)} + b$$

for some real-valued constants a, b. Show that the risk function of $\delta_{a,b}(\mathbf{X})$ is given by

$$R_{\delta_{a,b}}(\theta) = \Big[a\frac{1}{n+1} + (1-a)\frac{n}{n+1} + b - \frac{1}{2}\Big]^2 + \frac{a^2 n + (1-a)^2 n + 2a(1-a)}{(n+1)^2(n+2)},$$

which is minimized at $b = \frac{1}{2} - \frac{a + n(1-a)}{n+1}$ for any given constant a.

(e) *Among all procedures* $\delta_{a,b}(\mathbf{X})$ in part (d), **show that** the choice $a = \frac{1}{2}$ and $b = \frac{1}{2} - \frac{a+n(1-a)}{n+1} = 0$, i.e., $\delta^*(\mathbf{X}) = (X_{(1)} + X_{(n)})/2$, gives uniformly smallest risk function.

(f) **Prove that** when $n \geq 3$, the procedure $\delta^*(\mathbf{X}) = (X_{(1)} + X_{(n)})/2$ in part (c)(iv) is better than $\delta_0(\mathbf{X}) = \bar{X}_n$, and conclude that $\delta_0(\mathbf{X}) = \bar{X}_n$ is inadmissible when $n \geq 3$.

2. **(Modified from 7.19(a))** Suppose that the random variables $Y_1, \ldots, Y_n (n \geq 2)$ satisfy

$$Y_i = \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n,$$

where $\epsilon_1, \ldots, \epsilon_n$ are iid $N(0, \sigma^2)$, and both β and σ^2 are unknown.

(a) Assume $x_1, \ldots, x_n$ are fixed known constants, and we observe $Y_1 = y_1, \cdots, Y_n = y_n$, e.g., the observed data $\mathbf{Y} = (y_1, \ldots, y_n)$. **Find** a two-dim sufficient statistic of $\mathbf{Y} = (Y_1, \cdots, Y_n)$ for (β, σ^2).

(b) Assume now that $x_1, \ldots, x_n$ are random variables with a known joint distribution $m(x_1, \ldots, x_n)$, and the x_i's are independent of ϵ_i's (it is traditional in the linear regression to use lower case for independent variables x_i's). In this case, the observed data $(\mathbf{Y}, \mathbf{x}) = \{(Y_i, x_i)\}_{i=1,\ldots,n}$. **Find** a three-dimensional sufficient statistic of $(\mathbf{Y}, \mathbf{x})$ for (β, σ^2).

3. **(Modified from 6.5)**. Let $X_1, \cdots, X_n (n \geq 2)$ be independent random variables with pdfs

$$f(x_i|\theta) = \begin{cases} \frac{1}{3i\theta}, & \text{if } -i(\theta - 1) < x_i < i(2\theta + 1); \\ 0, & \text{otherwise,} \end{cases}$$

for $i = 1, 2, \cdots, n$, where $\theta > 0$.

(a) **Show that** $T_a(\mathbf{X}) = (\min_{1\le i\le n}(X_i/i), \max_{1\le i\le n}(X_i/i))$ is a two-dim sufficient statistic for θ.

(b) **Find** a minimal sufficient statistic for θ. Hints: the minimal sufficient statistic is one-dimensional.

4. **(6.25) (b) and (d)**. We have seen a number of theorems concerning sufficiency and related concepts for exponential families. Let $X_1, \dots, X_n (n \ge 2)$ be a random sample for each of the following distribution families, and establish the following results.

   (b) The statistic $T(\mathbf{X}) = \sum_{i=1}^n X_i^2$ is minimal sufficient in the $N(\mu, \mu)$ family.

   (d) The statistic $T(\mathbf{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is minimal sufficient for $\theta = (\mu, \sigma^2)$ in the $N(\mu, \sigma^2)$ family.

5. **(6.9)(a)(b)(d)(e).** For each of the following distribution let $X_1, \dots, X_n (n \ge 2)$ be a random sample. **Find** a minimal sufficient statistic for θ.

   (a) $f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}, \quad -\infty < x < \infty, -\infty < \theta < \infty$ (normal)

   (b) $f(x|\theta) = e^{-(x-\theta)}, \quad \theta < x < \infty, -\infty < \theta < \infty$ (location exponential)

   (d) $f(x|\theta) = \frac{1}{\pi[1+(x-\theta)^2]}, \quad -\infty < x < \infty, -\infty < \theta < \infty$ (Cauchy)

   (e) $f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty, -\infty < \theta < \infty$ (double exponential)

   [In class we will discuss part (c) $f(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}, -\infty < x < \infty, -\infty < \theta < \infty$(logistic).]

6. **(6.12).** A natural ancillary statistic in most problems in the *sample size*. For example, let N be an integer-valued random variable taking values $1, 2, \cdots$ with known probabilities $p_1, p_2, \cdots$, where $\sum_{i=1}^\infty p_i = 1$. Having observed $N = n$, perform n Bernoulli trials with success probability θ, getting X successes.

   (a) Prove that the pair (X, N) is minimal sufficient and N is ancillary for θ.
   (Note that the similarity to some of the hierarchical models discussed in Section 4.4.)

   (b) Prove that the estimator X/N is unbiased for θ and has variance $\theta(1-\theta)\mathbf{E}(1/N)$. In other words, prove that $\mathbf{E}_\theta(X/N) = \theta$ and $Var_\theta(X/N) = \theta(1-\theta)\mathbf{E}(1/N)$.

Hints of Problem 1 (d): To compute its risk function, it is useful to split in the following steps.

(i) Note that if we let $U_i = X_i - \theta + 1/2$, then $X_{(1)} = U_{(1)} + \theta - 1/2$ and $X_{(n)} = U_{(n)} + \theta - 1/2$. Hence we first need to investigate the properties of $U_{(1)} = \min(U_1, \ldots, U_n)$ and $U_{(n)} = \max(U_1, \ldots, U_n)$ when $U_1, \ldots, U_n$ are iid with Uniform$[0, 1]$. Using the fact $\mathbf{P}(u \leq U_{(1)} \leq U_{(n)} \leq v) = \mathbf{P}(u \leq U_i \leq v \text{ for all } i = 1, \ldots, n) = \prod_{i=1}^{n} \mathbf{P}(u \leq U_i \leq v)$ for any u and v, show that the joint density of $U_{(1)}$ and $U_{(n)}$ is

$$f_{U_{(1)},U_{(n)}}(u, v) = \begin{cases} n(n-1)(v-u)^{n-2}, & \text{if } 0 \leq u \leq v \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

whereas the respective (marginal) densities of $U_{(1)}$ and $U_{(n)}$ are

$$f_{U_{(1)}}(u) = \begin{cases} n(1-u)^{n-1}, & \text{if } 0 \leq u \leq 1; \\ 0, & \text{otherwise.} \end{cases} \quad \text{and} \quad f_{U_{(n)}}(v) = \begin{cases} nv^{n-1}, & \text{if } 0 \leq v \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

(ii) Show that $\mathbf{E}(U_{(1)}) = \frac{1}{n+1}$, $\mathbf{E}(U_{(n)}) = \frac{n}{n+1}$, $Var(U_{(1)}) = Var(U_{(n)}) = \frac{n}{(n+1)^2(n+2)}$ and $Cov(U_{(1)}, U_{(n)}) = \frac{1}{(n+1)^2(n+2)}$.

(iii) Use the fact of $\mathbf{E}(Y^2) = [\mathbf{E}(Y)]^2 + Var(Y)$ to show that the risk function of $\delta_{a,b}(\mathbf{X})$ is

$$R_{\delta_{a,b}}(\theta) \quad = \quad \mathbf{E}\Big(aU_{(1)} + (1-a)U_{(n)} + b - 1/2\Big)^2.$$

Hints of Problem 2: Let $\theta = (\beta, \sigma^2)$.

(a) The sample is $\mathbf{Y} = (Y_1, \ldots, Y_n)$, and the joint density function of $\mathbf{Y}$ is

$$f_\theta(\mathbf{Y}) = \prod_{i=1}^{n} f(Y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(Y_i - \beta x_i)^2}{2\sigma^2}\Big).$$

How to factor this joint pdf into two parts? The part that depends on $\theta = (\beta, \sigma^2)$ depends on the sample $\mathbf{Y} = (Y_1, \ldots, Y_n)$ only through which kind of two-dimensional function $T(\mathbf{Y})$? Note that the x_i's are treated as known constants here.

(b) When $x_1, \ldots, x_n$ are random variables with a known joint distribution $m(x_1, \ldots, x_n)$, and the x_i's are independent of ϵ_i's, the joint density of the data $(\mathbf{Y}, \mathbf{X}) = \{(Y_i, x_i)\}_{i=1,\ldots,n}$ is

$$f_\theta(\mathbf{Y}, \mathbf{X}) = m(\mathbf{x}) f_\theta(\mathbf{Y}|\mathbf{X}) = m(x_1, \ldots, x_n) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(Y_i - \beta x_i)^2}{2\sigma^2}\Big).$$

Can you factor this joint pdf into two parts? The part that depends on $\theta = (\beta, \sigma^2)$ depends on the sample $(\mathbf{Y}, \mathbf{X}) = \{(Y_i, x_i)\}_{i=1,\ldots,n}$ only through which kind of three-dimensional function $T(\mathbf{Y}, \mathbf{X})$?

Hints of Problem 3: It is important to focus on the domain of θ in the joint pdf of $\mathbf{X} = (X_1, \ldots, X_n)$. You can write $a(\theta) < x_i < b(\theta)$ for $i = 1, \cdots, n$, into two separate inequalities: $a(\theta) < x_i$ for all i and $x_i < b(\theta)$ for all i. From this, we can conclude that $a(\theta) < \min_i x_i$ and $\max_i x_i < b(\theta)$, and then solve for θ, respectively. To be more specific, the joint density is

$$\begin{aligned} f_\theta(\mathbf{x}) &= \prod_{i=1}^{n} f_{X_i}(x_i|\theta) = \prod_{i=1}^{n} \Big[\frac{1}{3i\theta} I(-i(\theta-1) < x_i < i(2\theta+1)\Big] \\ &= \frac{1}{3^n n! \theta^n} I\Big(-(\theta-1) < \frac{x_i}{i} < 2\theta + 1 \quad \text{for all } i = 1, \ldots, n\Big) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{3^n n! \theta^n} \times I\Big(-(\theta-1) < \frac{x_i}{i} \quad \text{for all } i=1,\dots,n\Big) \times I\Big(\frac{x_i}{i} < 2\theta+1 \quad \text{for all } i=1,\dots,n\Big) \\
&= \frac{1}{3^n n! \theta^n} \times I\Big(-(\theta-1) < \min_{1\le i\le n} \frac{x_i}{i}\Big) \times I\Big(\max_{1\le i\le n} \frac{x_i}{i} < 2\theta+1\Big) \\
&= \frac{1}{3^n n! \theta^n} \times I\Big(\theta > 1 - (\min_{1\le i\le n} \frac{x_i}{i})\Big) \times I\Big(\theta > \frac{1}{2}[(\max_{1\le i\le n} \frac{x_i}{i}) - 1]\Big)
\end{aligned}
$$

Part (a) follows from this immediately. To find the minimal sufficient statistic in part (b), using the fact that $I(\theta > u)I(\theta > v) = I(\theta > \max(u,v))$, you can further simplify the above density function as a function of one-dimensional statistic. Hint: how about us defining

$$T(\mathbf{X}) = \max\Big\{1 - \min_{1\le i\le n} \frac{x_i}{i}, \quad \frac{1}{2}(\max_{1\le i\le n} \frac{x_i}{i} - 1)\Big\}.$$

Also we do not need to simplify T here and it is okay to leave it as is.

Hints of Problem 5(d): The key observation is that

$$
\begin{aligned}
\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} \text{ is constant in } \theta \quad &\Longleftrightarrow \quad \frac{f(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta=0)} = \frac{f(\mathbf{y}|\theta)}{f(\mathbf{y}|\theta=0)} \quad \text{for all } \theta \\
&\Longleftrightarrow \quad \prod_{k=1}^{n} \frac{1+(x_k-\theta)^2}{1+x_k^2} = \prod_{k=1}^{n} \frac{1+(y_k-\theta)^2}{1+y_k^2} \quad \text{for all } \theta.
\end{aligned}
$$

Now both sides are polynomial of θ of degree $2n$, comparing the coefficient of θ^{2n} yields that $\prod(1+x_k^2) = \prod(1+y_k^2)$, and thus

$$\prod_{k=1}^{n}[1+(x_k-\theta)^2] = \prod_{k=1}^{n}[1+(y_k-\theta)^2].$$

Setting these two polynomials to 0 and solving the complex root for θ, the left-hand side polynomial has $2n$ complex roots, $\widehat{\theta} = x_k \pm \sqrt{-1}$, for $k = 1,\dots,n$, whereas the right-hand polynomial leads to another set of $2n$ complex roots, $\widehat{\theta} = y_k \pm \sqrt{-1}$, for $k=1,\dots,n$. Of course these two polynomials in θ will have the same (complex) roots, and thus $x_{(k)} = y_{(k)}$ for $k = 1,\dots,n$. What does this mean?

Hints of Problem 5(e): In this case, the order statistic is also a minimal sufficient statistic. the main difficulty is to show that if $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$ does not depend on θ, then $x_{(i)} = y_{(i)}$ for all $i=1,\dots,n$.
First, let us prove $x_{(1)} = y_{(1)}$. Assume $x_{(1)} \neq y_{(1)}$, and without loss of generality, assume $x_{(1)} < y_{(1)}$. For convenience of notation, define $x_{(0)} = y_{(0)} = -\infty$ and define $x_{(n+1)} = y_{(n+1)} = \infty$. Now let r be the largest $i \ge 1$ such that $x_{(i)} < y_{(1)}$. In other words, $x_{(1)} \le x_{(r)} < y_{(1)} \le x_{(r+1)}$ for some $1 \le r \le n$. Consider the interval $x_{(r)} < \theta < y_{(1)}$, and show that $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$ depends on $\theta \in (x_{(r)}, y_{(1)})$ since $1 \le r \le n$. This is a contradiction that $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$ is a constant of θ. Thus the assumption that $x_{(1)} \neq y_{(1)}$ is wrong, and hence we must have $x_{(1)} = y_{(1)}$.
The above arguments can be easily extended to show that $x_{(i)} = y_{(i)}$ for all $i = 1,\dots,n$. Assume this is not true, and let k be the smallest i such that $x_{(i)} \neq y_{(i)}$, say $x_{(k)} < y_{(k)}$. As above, let r be the largest $i \ge k$ such that $x_{(i)} < y_{(k)}$. Then

$$x_{(1)} = y_{(1)} \le x_{(2)} = y_{(2)} \le \cdots \le x_{(k-1)} = y_{(k-1)} \le x_{(k)} \le x_{(r)} < y_{(k)}$$

for some $k \le r \le n$. Then consider the interval $x_{(r)} < \theta < y_{(k)}$, and see what happens?

Hints of Problem 6(b): Use the facts that $\mathbf{E}(U) = \mathbf{E}(\mathbf{E}(U|V))$ and $Var(U) = \mathbf{E}(Var(U|V)) + Var(\mathbf{E}(U|V))$ for $U = X/N$ and $V = N$. See Theorems 4.4.3 and 4.4.7 on page 164-167 of our text for the proofs of these two useful facts which will be used later.