Stat 513
Emerson, Winter 04
Midterm #1 Key
February 12, 2004

Instructions:

- **This exam is closed book, closed notes. No use of calculators is permitted.**
- **Write answers to the following questions on the front side of separate sheets of paper, starting each problem at the top of a new page. Be sure to write your name on the top of each page.**
- **In order to receive full credit, you must make clear how you derived the answers to the problems.**
- **If any of the problems seem unsolvable with the information provided, clearly state reasonable assumptions that would allow solution, and use those assumptions to solve the problem. If no such reasonable assumptions are obvious, simply state that and proceed to other problems.**
- **You are allowed 50 minutes for this exam. At the end of the in-class exam, you may complete or correct the solution to any problem as a take-home exam subject to the conditions specified on the last page of the exam.**
- **Write out and sign the following pledge for the in-class exam:**

  "On my honor, I have neither given nor received any unauthorized aid on this in-class examination."

  If for any reason you can not honestly sign the pledge, please discuss this with me on Wednesday.

In addition to the theorems covered in class, you may use the following facts without proof:

- If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$E[X^3] = \mu^3 + 3\mu\sigma^2$$
$$E[X^4] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

1. (30 points) Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. uniform random variables having distribution $X_i \sim \mathcal{U}(0, \theta)$ with unknown $\theta \in \Theta = (0, \infty)$.

   a. Using the definition of sufficiency, show that the maximum likelihood estimator $\hat{\theta} = X_{(n)}$ (the sample maximum) is sufficient for θ.

Ans: To show sufficiency for statistic $\hat{\theta} = X_{(n)}$, we want to show that the distribution of $\vec{X} \mid \hat{\theta} = t$ is independent of θ. The conditional density is

$$f_{\vec{X}|\hat{\theta}}(\vec{x}|\hat{\theta} = t) = \frac{f_{\vec{X},\hat{\theta}}(\vec{x}, \hat{\theta} = t)}{f_{\hat{\theta}}(t)}.$$

Now the joint density of $\vec{X} = \vec{x}$ and $\hat{\theta} = t$ is merely the joint density of the X_i's if the sample maximum is $\hat{\theta} = t$, and zero otherwise:

$$\begin{aligned}
f_{\vec{X},\hat{\theta}}(\vec{x}, \hat{\theta} = t) &= \left(\prod_{i=1}^{n} \frac{1}{\theta} \mathbf{1}_{[0<x_i<\theta]} \right) \mathbf{1}_{[x_{(n)}=t]} \\
&= \frac{1}{\theta^n} \left(\prod_{i=1}^{n} \mathbf{1}_{[0<x_i \leq t]} \right) \mathbf{1}_{[0<x_{(n)}<\theta]} \mathbf{1}_{[x_{(n)}=t]}
\end{aligned}$$

The density for $\hat{\theta} = X_{(n)}$ is easily found from the cumulative distribution function:

$$\begin{aligned}
F_{\hat{\theta}}(t) &= Pr(X_{(n)} \leq t) \\
&= \prod_{i=1}^{n} Pr(X_i \leq t) \\
&= \left(\frac{t}{\theta} \right)^n \mathbf{1}_{[0<t<\theta]} + \mathbf{1}_{[\theta \leq t < \infty]} \\
f_{\hat{\theta}}(t) &= \frac{d}{dt} F_{\hat{\theta}}(t) \\
&= n \frac{t^{n-1}}{\theta^n} \mathbf{1}_{[0<t<\theta]}.
\end{aligned}$$

Hence, the conditional density is

$$\begin{aligned}
f_{\vec{X}|\hat{\theta}}(\vec{x}|\hat{\theta} = t) &= \frac{\frac{1}{\theta^n} \left(\prod_{i=1}^{n} \mathbf{1}_{[0<x_i \leq t]} \right) \mathbf{1}_{[x_{(n)}=t]} \mathbf{1}_{[0<t<\theta]}}{n \frac{t^{n-1}}{\theta^n} \mathbf{1}_{[0<t<\theta]}} \\
&= \frac{\theta^n}{n t^{n-1}} \mathbf{1}_{[x_{(n)}=t]},
\end{aligned}$$

which is independent of θ. (Note that because we are conditioning on an observed $\hat{\theta} = t$, we know that $0 < t < \theta$.)

b. What does the Cramér-Rao Theorem say about the variance of $\hat{\theta}$?

Ans: Because the distribution of the X_i's lack common support across all values of θ, this is not a regular problem, and the Cramér-Rao Theorem provides no guidance about the variance of any estimator.

c. Let $\tilde{\theta}$ be an unbiased estimator of θ based on the MLE. Find $\tilde{\theta}$ and compare its mean squared error to the mean squared error of $\hat{\theta}$.

Ans: We can find the expected value of $\hat{\theta}$ by integrating the density

$$\begin{aligned} E[\hat{\theta}] &= \int_{-\infty}^{\infty} t f_{\hat{\theta}}(t)\, dt \\ &= \int_0^{\theta} n \frac{t^n}{\theta^n}\, dt \\ &= \frac{n}{n+1}\theta. \end{aligned}$$

From this, we quickly deduce that $\tilde{\theta} = (n+1)\hat{\theta}/n$ would be unbiased for θ. Furthermore, we can see that $Var(\tilde{\theta}) = (n+1)^2 Var(\hat{\theta})/n^2$. Since $MSE(T) = Var(T) + bias^2(T)$, we know that

$$MSE(\tilde{\theta}) = \frac{(n+1)^2}{n^2} Var(\hat{\theta})$$

and

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + \frac{1}{(n+1)^2}\theta^2.$$

Now we find $Var(\hat{\theta})$ using the usual approach: $Var(\hat{\theta}) = E[\hat{\theta}^2] - E^2[\hat{\theta}]$ with

$$\begin{aligned} E[\hat{\theta}^2] &= \int_{-\infty}^{\infty} t^2 f_{\hat{\theta}}(t)\, dt \\ &= \int_0^{\theta} n \frac{t^(n+1)}{\theta^n}\, dt \\ &= \frac{n}{n+2}\theta^2 \\ Var(\hat{\theta}^2) &= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 \\ &= \frac{n}{(n+2)(n+1)^2}\theta^2. \end{aligned}$$

This then dictates that

$$\begin{aligned} MSE(\tilde{\theta}) &= \frac{1}{n(n+2)}\theta^2 \\ MSE(\hat{\theta}) &= \left(\frac{n}{(n+2)(n+1)^2} + \frac{1}{(n+1)^2} \right)\theta^2 = \frac{2}{(n+2)(n+1)}\theta^2 \end{aligned}$$

Taking the ratio

$$\frac{MSE(\tilde{\theta})}{MSE(\hat{\theta})} = \frac{n+1}{2n},$$

we find that $\tilde{\theta}$ has lower MSE whenever $n > 1$ (when $n = 1$, the MSE for the two estimators are equal).

2. (20 points) Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. normal random variables having distribution $X_i \sim \mathcal{N}(\mu, 1)$ with unknown $\mu \in (-\infty, \infty)$. Suppose we wish to estimate $\theta = \mu^2$.

   a. Find the Cramér-Rao lower bound for the variance of an unbiased estimator of θ.

Ans: We want to estimate $g(\mu) = \mu^2$. Because $\sigma^2 = 1$, the likelihood for μ is

$$L(\mu \,|\, \vec{X}) = f_{\vec{X}}(\vec{x}; \mu) = \left(\frac{1}{\sqrt{(2\pi)}} \right)^n \exp\left(-\frac{\sum_{i=1}^n (X_i - \mu)^2}{2} \right),$$

yielding log likelihood, score function, and Fisher's information of

$$\begin{aligned} \mathcal{L}(\mu) &= -\frac{n}{2}\log(2\pi) - \frac{\sum_{i=1}^n (X_i - \mu)^2}{2} \\ U(\mu) &= \frac{\partial}{\partial\mu}\mathcal{L}(\mu) = \sum_{i=1}^n (X_i - \mu) \\ I(\mu) &= -E\left(\frac{\partial}{\partial\mu} U(\mu) \right) = n \end{aligned}$$

The Cramér-Rao lower bound then dictates that for this regular problem of estimating $g(\mu) = \mu^2$ (so $g'(\mu) = 2\mu$), an estimator T which is unbiased (so $b(T, g(\mu)) = 0$ and $b'(T, g(\mu)) = 0$) must have

$$Var(T) \geq \frac{[g'(\mu) + b'(T, g(\mu))]^2}{I(\mu)} = \frac{4\mu^2}{n}.$$

   b. Show that the maximum likelihood estimator $\hat{\theta}$ is biased for θ, and compare the mean squared error of that estimator to the Cramér-Rao lower bound for the MSE of a similarly biased estimator. Does the MLE meet that lower bound in small samples? How does the MSE of the MLE behave asymptotically with respect to that lower bound?

Ans: We find the MLE $\hat{\mu}$ for μ from the equation $U(\hat{\mu}) = 0$, which yields $\hat{\mu} = \overline{X}$. By the invariance of maximum likelihood estimates, the MLE of $\theta = g(\mu)$ will be $\hat{\theta} = g(\hat{\mu}) = \overline{X}^2$. Due to the normality of the X_i's, we know that $\overline{X} \sim \mathcal{N}(\mu, 1/n)$, and we can then find the bias of $\hat{\theta}$ as

$$E[\hat{\theta}] = E[\overline{X}^2] = Var(\overline{X}) + E^2[\overline{X}] = \frac{1}{n} + \mu^2.$$

So $\hat{\theta}$ is biased with bias function $b(\hat{\theta}, \theta) = 1/n$.

We are asked to find the Cramér-Rao lower bound for the MSE of an estimator with similar bias. From the formula given in the class notes, the Cramér-Rao lower bound dictates that

$$MSE(\hat{\theta}) \geq \frac{\left[g'(\mu) + b'(\hat{\theta}, g(\mu))\right]^2}{I(\mu)} + [b(\hat{\theta}, g(\mu))]^2 = \frac{4\mu^2}{n} + \frac{1}{n^2}.$$

We can find the actual MSE for $\hat{\theta}$ as

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [b(\hat{\theta}, g(\mu))]^2.$$

Because $\overline{X}$ is normally distributed, we can use the formulas for the fourth non-central moment of a normal random variable (as given in the instructions to the exam) to find

$$\begin{aligned} Var(\hat{\theta}) &= Var(\overline{X}^2) = E[\overline{X}^4] - E^2[\overline{X}^2] \\ &= \mu^4 + 6\frac{mu^2}{n} + \frac{3}{n^2} - \left(\mu^2 + \frac{1}{n}\right)^2 \\ &= \frac{4\mu^2}{n} + \frac{2}{n^2} \\ MSE(\hat{\theta}) &= \frac{4\mu^2}{n} + \frac{3}{n^2}. \end{aligned}$$

Clearly, $\hat{\theta}$ does not meet the Cramér-Rao lower bound for MSE of an estimator having bias $1/n$.

Of course, as a MLE in a regular problem, we know that the $\hat{\theta}$ is consistent for θ, and its asymptotic variance will correspond to that of an unbiased estimator with variance achieving the Cramér-Rao lower bound. This could also be seen by looking at the limit of the ratio of the Cramér-Rao lower bound for the MSE and the actual MSE:

$$\lim_{n\to\infty} \frac{CRLB}{MSE(\hat{\theta})} = \lim_{n\to\infty} \frac{\frac{4\mu^2}{n} + \frac{3}{n^2}}{\frac{4\mu^2}{n} + \frac{1}{n^2}} = 1.$$

Note that it is <u>not</u> sufficient to just argue that the MSE and the CRLB each converge to 0 as $n \to \infty$, because they might do so at different rates.

3. (75 points) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having density function

$$f_X(x) = (\theta + 1)x^\theta \mathbf{1}_{[0<x<1]}$$

for some unknown $\theta > -1$.

a. Find a sufficient statistic for θ. (More credit will be given according to the extent of data reduction in the sufficient statistic.)

Ans: I use the Neyman factorization theorem:

$$\begin{aligned} f_{\vec{X}}(\vec{x};\theta) &= \prod_{i=1}^{n} \left((\theta+1)x_i^{\theta}\mathbf{1}_{[0<x_i<1]}\right) \\ &= (\theta+1)^n \left(\prod_{i=1}^{n} x_i\right)^{\theta} \mathbf{1}_{[0<x_{(1)}]}\mathbf{1}_{[x_{(n)}<1]} \\ &= h(\vec{x})g(\prod_{i=1}^{n} x_i,\theta), \end{aligned}$$

so $T(\vec{X}) = \prod_{i=1}^{n} X_i$ is a sufficient statistic. Note that we can show that this is a minimal sufficient statistic, because if $\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} y_i$,

$$\frac{f_{\vec{X}}(\vec{x};\theta)}{f_{\vec{X}}(\vec{y};\theta)} = \frac{\mathbf{1}_{[0<x_{(1)}]}\mathbf{1}_{[x_{(n)}<1]}}{\mathbf{1}_{[0<y_{(1)}]}\mathbf{1}_{[y_{(n)}<1]}}$$

is independent of θ. (Equivalent minimal sufficient statistics could be based on the geometric mean or the log of the geometric mean (the arithmetic mean of log transformed data).)

b. Find the Cramér-Rao lower bound for the variance of an unbiased estimator of θ.

Ans: In a regular problem, the Cramér-Rao lower bound for an estimator $T(\vec{X})$ of $g(\theta)$ having bias function $b(\theta)$ is

$$Var(T(\vec{X})) \geq \frac{(g'(\theta)+b'(\theta))^2}{I(\theta)}.$$

As we are considering unbiased estimates of $g(\theta)=\theta$, $g'(\theta)=1$ and $b'(\theta)=0$. As we are considering a sampling situation in which we have i.i.d. random variables, $I(\theta) = nI_1(\theta)$. Now in this regular problem

$$\begin{aligned} L_i(\theta|X_i) &= (\theta+1)X_i^{\theta} \\ \mathcal{L}_i(\theta) &= \log(L_i(\theta)) = \log(\theta+1) + \theta\log(X_i) \\ \mathcal{U}_i(\theta) &= \frac{\partial}{\partial\theta}\mathcal{L}_i(\theta) = \frac{1}{\theta+1} + \log(X_i) \\ I_i(\theta) &= -E\left(\frac{\partial}{\partial\theta}\mathcal{U}_i(\theta)\right) = \frac{1}{(\theta+1)^2} \end{aligned}$$

so the Cramér-Rao lower bound for the variance of an unbiased estimator of θ is

$$Var(T(\vec{X})) \geq \frac{(g'(\theta)+b'(\theta))^2}{I(\theta)} = \frac{(\theta+1)^2}{n}.$$

c. For what functions of θ does an unbiased estimator exist which meets the Cramér-Rao lower bound?

Ans: We know that a best regular unbiased estimator (BRUE) exists only for linear transformations of $g(\theta)$ when the score function can be written as

$$\mathcal{U}(\theta) = A(\theta)\left(T(\vec{X}) - g(\theta)\right).$$

From part b, we know the score function is

$$\mathcal{U}(\theta) = \sum_{i=1}^{n}\left(\frac{1}{\theta+1} + \log(X_i)\right) = -n\left(\frac{1}{n}\sum_{i=1}^{n}\log(X_i) - \frac{1}{\theta+1}\right),$$

so $T(\vec{X}) = -\sum_{i=1}^{n}\log(X_i)/n$ is the BRUE for $g(\theta) = 1/(\theta+1)$, and linear transformations of $T(\vec{X})$ will be BRUE for linear transformations of $g(\theta)$. No other BRUE will exist for this problem.

d. Compute the efficiency of the sample mean as an estimator of the expected value of X.

Ans We first find the mean and variance of X_i by the usual approach

$$\begin{aligned}
E[X_i] &= \int_{-\infty}^{\infty} x f_X(x;\theta)\,dx = \int_0^1 (\theta+1)x^{\theta+1}\,dx = \left[\frac{\theta+1}{\theta+2}x^{\theta+2}\right]_0^1 = \frac{\theta+1}{\theta+2}\\
E[X_i^2] &= \int_{-\infty}^{\infty} x^2 f_X(x;\theta)\,dx = \int_0^1 (\theta+1)x^{\theta+2}\,dx = \left[\frac{\theta+1}{\theta+3}x^{\theta+3}\right]_0^1 = \frac{\theta+1}{\theta+3}\\
Var(X_i) &= E[X_i^2] - E^2[X_i] = \frac{\theta+1}{\theta+3} - \frac{(\theta+1)^2}{(\theta+2)^2} = \frac{\theta+1}{(\theta+3)(\theta+2)^2}.
\end{aligned}$$

Now, the sample mean is always an unbiased estimator of the population mean, and its mean and variance are thus

$$\overline{X}_n \sim \left(\frac{\theta+1}{\theta+2}, \frac{\theta+1}{n(\theta+3)(\theta+2)^2}\right).$$

To find its efficiency, we merely compare its variance to the Cramér-Rao lower bound for an unbiased estimator of $g(\theta) = (\theta+1)/(\theta+2)$. We thus have $g'(\theta) = 1/(\theta+2)^2$, and Cramér-Rao lower bound

$$\frac{(g'(\theta)+b'(\theta))^2}{I(\theta)} = \frac{(\theta+1)^2}{n(\theta+2)^4}.$$

The efficiency of the sample mean is thus

$$\left(\frac{(\theta+1)^2}{n(\theta+2)^4}\right)\left(\frac{\theta+1}{n(\theta+3)(\theta+2)^2}\right)^{-1} = \frac{(\theta+1)(\theta+3)}{(\theta+2)^2}.$$

By noting that $(\theta+1)(\theta+3) = (\theta+2-1)(\theta+2+1) = (\theta+2)^2 - 1$, we see that the sample mean is always inefficient, though for large θ, it is nearly efficient.

e. Find the asymptotic distribution of the maximum likelihood estimate of the expected value of X.

Ans: The really fast way to solve this problem is to note that in regular problems, MLE's asymptotically are consistent and attain the Cramér-Rao lower bound. By the invariance property of maximum likelihood estimation, the MLE of $g(\theta) = (\theta+1)/(\theta+2)$ is $g(\hat{\theta}) = (\hat{\theta}+1)/(\hat{\theta}+2)$. Thus we know that

$$(\hat{\theta}+1)/(\hat{\theta}+2) \dot{\sim} \mathcal{N}\left((\theta+1)/(\theta+2), \frac{(\theta+1)^2}{n(\theta+2)^4}\right).$$

This is of course the answer that you would get by using the delta method with the asymptotic distribution of the MLE $\hat{\theta} \dot{\sim} \mathcal{N}(\theta, I^{-1}(\theta))$. From the results pesented in parts b and c, we easily see that $\hat{\theta} = -(1 + n/\sum \log(X_i))$, though this was not at all necessary for solving this problem.

4. (45 points) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having density function

$$f_X(x) = ((2x-1)\theta + 1)\mathbf{1}_{[0<x<1]}$$

for some unknown $\theta \in (-1, 1)$.

a. Find a sufficient statistic for θ. (More credit will be given according to the extent of data reduction in the sufficient statistic.)

Ans: The joint density of the data is

$$f_{\vec{X}}(\vec{x};\theta) = \prod_{i=1}^{n}((2x_i-1)\theta+1)\mathbf{1}_{[0<x_i<1]},$$

which can not be factored further into functions of any statistic beyond the order statistics. We see that the minimal sufficient statistic is indeed the order statistics $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ by examining when the ratio $f_{\vec{X}}(\vec{x};\theta)/f_{\vec{X}}(\vec{y};\theta)$ is independent of θ:

$$\frac{f_{\vec{X}}(\vec{x};\theta)}{f_{\vec{X}}(\vec{y};\theta)} = \frac{\prod_{i=1}^{n}((2x_i-1)\theta+1)\mathbf{1}_{[0<x_i<1]}}{\prod_{i=1}^{n}((2y_i-1)\theta+1)\mathbf{1}_{[0<y_i<1]}},$$

which happens only when $\vec{x}$ and $\vec{y}$ are permutations of each other, i.e., when they have the same order statistics.

b. For what functions of θ does an unbiased estimator exist which meets the Cramér-Rao lower bound?

Ans: We examine whether the score function can be put in the desired form for a

BRUE to exist for some function of θ:

$$L(\theta|X_i) = (\prod_{i=1}^{n}((2x_i - 1)\theta + 1)\mathbf{1}_{[0<x_i<1]}$$

$$\mathcal{L}(\theta) = \log(L(\theta)) = \sum_{i=1}^{n} \log((2x_i - 1)\theta + 1)$$

$$\mathcal{U}(\theta) = \frac{\partial}{\partial\theta}\mathcal{L}(\theta) = \sum_{i=1}^{n} \frac{2x_1 - 1}{(2x_i - 1)\theta + 1}$$

There is no way to factor the score function into the form

$$\mathcal{U}(\theta) = A(\theta)\left(T(\vec{X}) - g(\theta)\right)$$

for any $g(\theta)$.

c. Derive expressions for the asymptotic distribution of the maximum likelihood estimate of the expected value of X.

Ans: We first find the mean by the usual approach

$$E[X_i] = \int_0^1 x((2x - 1)\theta + 1)\,dx = \left[\frac{2}{3}\theta x^3 + \frac{1-\theta}{2}x^2\right]_0^1 = \frac{\theta + 3}{6}$$

Thus we want to estimate $g(\theta) = (\theta+3)/6$. By the invariance of MLEs, the MLE of $g(\theta)$ will be $g(\hat{\theta}) = (\hat{\theta} + 3)/6$. The maximum likelihood estimate $\hat{\theta}$ of θ is that value which satisfies $\mathcal{U}(\hat{\theta}) = 0$, so it is specified by the implicit function

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

which has no closed form solution as an explicit function. The asymptotic distribution of $\hat{\theta}$ in this regular problem will be $\hat{\theta}\dot{\sim}\mathcal{N}\left(\theta, I^{-1}(\theta)\right)$, where $I(\theta) = nI_1(\theta)$ in this problem based on i.i.d. data.

$$\begin{aligned}
I_1(\theta) &= -E\left(\frac{\partial}{\partial\theta}\mathcal{U}_i(\theta)\right) \\
&= -E\left(\frac{(2X_i - 1)^2}{((2X_i - 1)\theta + 1)^2}\right) \\
&= \int_0^1 \frac{(2x - 1)^2}{((2x - 1)\theta + 1)^2}((2x - 1)\theta + 1)\,dx = \int_0^1 \frac{(2x - 1)^2}{((2x - 1)\theta + 1)}\,dx
\end{aligned}$$

This integral can be solved in closed form by making the substitution $u = (2x - 1)\theta + 1$ to obtain

$$I_1(\theta) = \frac{1}{2\theta^3}\int_{1-\theta}^{1+\theta} \frac{(u - 1)^2}{u}\,du = \frac{1}{2\theta^3}\log\left(\frac{1+\theta}{1-\theta}\right) - \frac{1}{\theta^2}$$

Then, using the delta method, Mann-Wald, or Slutsky's, we find

$$\frac{(\hat{\theta} + 3)}{6}\dot{\sim}\mathcal{N}\left(\frac{\theta + 3)}{6}, \frac{1}{36nI_1(\theta)}\right).$$