Midterm Exam

Instructions: This is a closed book exam, but you may refer to one sheet of notes. You have 80 minutes for the exam. Answer as many questions as possible. Partial answers get partial credit. Please write legibly. *Good luck!*

Problem 1 (10 points). For *two* of the three statements below, determine whether or not the statement is correct, and give a *brief* (e.g., a bluebook page or less) justification for your answer.

(a) Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with pdf $f(\cdot|\theta)$ (where $\theta \in \Theta \subseteq \mathbb{R}$ is unknown) and let $t : \mathbb{R} \to \mathbb{R}$ be a function. The statistic $\sum_{i=1}^n t(X_i)$ is sufficient (for θ) if and only if

$$f(x|\theta) = c(\theta) h(x) \exp[w(\theta) t(x)] \qquad \forall (x, \theta) \in \mathbb{R} \times \nleq$$

for some functions $c : \Theta \to \mathbb{R}_{++}$, $h : \mathbb{R} \to \mathbb{R}_+$, and $c : \Theta \to \mathbb{R}$.

The statement is incorrect. The "if" part is correct (by the factorization criterion), but the "only if" part is not. For instance, if $n = 1$, then $X_1 = \sum_{i=1}^n X_i$ is sufficient whether or not $\{f(\cdot|\theta) : \theta \in \Theta\}$ is an exponential family.

(b) Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with pdf $f(\cdot|\theta)$ (where $\theta \in \Theta$ is unknown) and let $T = T(X_1, \ldots, X_n)$ be a statistic. If $\hat{\theta}$ is an unbiased estimator of θ, then $E_\theta\left(\hat{\theta}|T\right)$ is independent of (i.e., not a function of) θ if and only if T is sufficient (for θ).

The statement is incorrect. The "if" part is correct, but the "only if" part is not. For instance, if $X_i \sim$ i.i.d. $\mathcal{N}(\theta, 1)$ and $\hat{\theta} = T = X_1$, then $\hat{\theta}$ is unbiased and $E_\theta\left(\hat{\theta}|T\right) = T$, but T is not sufficient unless $n = 1$.

(c) If (X, Y) is a bivariate random vector, then

$$P\left[|Y - E(Y|X)| \geq r\right] \leq \frac{E\left[Var(Y|X)\right]}{r^2} \leq \frac{Var(Y)}{r^2} \qquad \forall r > 0.$$

The statement is correct. The first inequality holds because it follows from the Markov inequality that

$$P\left[|Y - E(Y|X)| \geq r\right] = P\left[|Y - E(Y|X)|^2 \geq r^2\right] \leq \frac{E\left[|Y - E(Y|X)|^2\right]}{r^2} \qquad \forall r > 0$$

where, using the law of iterated expectations,

$$E\left[|Y - E(Y|X)|^2\right] = E\left(E\left[|Y - E(Y|X)|^2 |X\right]\right) = E\left[Var(Y|X)\right].$$

The second inequality holds because

$$Var(Y) = E\left[Var(Y|X)\right] + \underbrace{Var\left[E(Y|X)\right]}_{\geq 0} \geq E\left[Var(Y|X)\right],$$

where the equality is the conditional variance identity.

Problem 2 (40 points, each part receives equal weight). Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with pdf

$$f_X(x|\theta) = c(\theta)\, x^{-1} \exp\left[-\frac{1}{\theta}(\log x)^2\right] 1(x > 0),$$

where $\theta \in \Theta = (0, \infty)$ is an unknown parameter, $c(\cdot)$ is a function (with argument θ), and $1(\cdot)$ is the indicator function.

(a) Show that $c(\theta) = 1/\sqrt{\pi\theta}$.

Because $\int_{-\infty}^{\infty} f_X(x|\theta)\, dx = 1$, $c(\theta)$ *is given by*

$$c(\theta) = \left(\int_0^{\infty} x^{-1} \exp\left[-\frac{1}{\theta}(\log x)^2\right] dx\right)^{-1}.$$

Using change of variables $(y = \log x)$ *and the fact (about the normal distribution) that*

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\theta}} \exp\left(-\frac{1}{\theta} y^2\right) dy = 1,$$

we have:

$$\int_0^{\infty} x^{-1} \exp\left[-\frac{1}{\theta}(\log x)^2\right] dx = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{\theta} y^2\right) dy = \sqrt{\pi\theta} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\theta}} \exp\left(-\frac{1}{\theta} y^2\right) dy}_{=1} = \sqrt{\pi\theta},$$

from which the stated result follows.

(b) Let $Y_i = \log(X_i)$. Find $f_Y(\cdot|\theta)$, “the” pdf of Y. What is the distribution of Y?

It follows from the answer to (a) that

$$\begin{aligned} f_Y(y|\theta) &= c(\theta) \exp\left(-\frac{1}{\theta} y^2\right) = \frac{1}{\sqrt{\pi\theta}} \exp\left(-\frac{1}{\theta} y^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2(\theta)}} \exp\left(-\frac{1}{2\sigma^2(\theta)} y^2\right), \qquad \sigma^2(\theta) = \theta/2. \end{aligned}$$

In other words, $Y \sim \mathcal{N}(0, \theta/2)$.

(c) Show that $E_\theta(X_i) = \exp(\theta/4)$.

Using change of variables $(y = \log x)$, *the identity*

$$y - \frac{1}{\theta}y^2 = -\frac{1}{\theta}\left(y - \frac{\theta}{2}\right)^2 + \frac{\theta}{4},$$

and the fact (about the $\mathcal{N}(\theta/2, \theta/2)$ *distribution) that*

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\theta}} \exp\left[-\frac{1}{\theta}\left(y - \frac{\theta}{2}\right)^2\right] dy = 1,$$

we have:

$$\begin{aligned} E_\theta(X_i) &= \int_0^\infty x \frac{1}{\sqrt{\pi\theta}} x^{-1} \exp\left[-\frac{1}{\theta}(\log x)^2\right] dx = \int_0^\infty \frac{1}{\sqrt{\pi\theta}} \exp\left[-\frac{1}{\theta}(\log x)^2\right] dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\theta}} \exp(y) \exp\left(-\frac{1}{\theta}y^2\right) dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\theta}} \exp\left(y - \frac{1}{\theta}y^2\right) dy \\ &= \exp(\theta/4) \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\theta}} \exp\left[-\frac{1}{\theta}\left(y - \frac{\theta}{2}\right)^2\right] dy}_{=1} = \exp(\theta/4). \end{aligned}$$

(d) Use the fact that $E_\theta(X_i) = \exp(\theta/4)$ to derive a method moments estimator $\hat{\theta}_{MM}$ of θ. Is $\hat{\theta}_{MM}$ an unbiased estimator of θ?

Because $\theta = 4 \log E_\theta(X_i)$, *a method moments estimator of* θ *is given by*

$$\hat{\theta}_{MM} = 4 \log \bar{X}.$$

The function $\log(\cdot)$ *is strictly concave, so it follows from Jensen's inequality that*

$$E_\theta\left(\hat{\theta}_{MM}\right) = 4E_\theta\left(\log \bar{X}\right) > 4 \log E_\theta\left(\bar{X}\right) = \theta,$$

implying in particular that $\hat{\theta}_{MM}$ *is a biased estimator of* θ.

(e) Find the log likelihood function and show that the maximum likelihood estimator of θ is given by

$$\hat{\theta}_{ML} = \frac{2}{n}\sum_{i=1}^{n}(\log X_i)^2 .$$

The log likelihood function is

$$\begin{aligned}\ell(\theta|X_1,\ldots,X_n) &= \sum_{i=1}^{n}\log f_X(X_i|\theta) = \sum_{i=1}^{n}\log\left[\frac{1}{\sqrt{\pi\theta}}X_i^{-1}\exp\left[-\frac{1}{\theta}(\log X_i)^2\right]\underbrace{1(X_i>0)}_{\equiv 1}\right] \\ &= -\frac{n}{2}\log(\pi\theta) - \sum_{i=1}^{n}\log X_i - \frac{1}{\theta}\sum_{i=1}^{n}(\log X_i)^2 .\end{aligned}$$

Because

$$\begin{aligned}\frac{\partial}{\partial\theta}\ell(\theta|X_1,\ldots,X_n) &= -\frac{n}{2\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}(\log X_i)^2 , \\ \frac{\partial^2}{\partial\theta^2}\ell(\theta|X_1,\ldots,X_n) &= \frac{n}{2\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n}(\log X_i)^2 = -\frac{1}{\theta}\frac{\partial}{\partial\theta}\ell(\theta|X_1,\ldots,X_n) - \underbrace{\frac{1}{\theta^3}\sum_{i=1}^{n}(\log X_i)^2}_{>0},\end{aligned}$$

$\hat{\theta}_{ML} = n^{-1}\sum_{i=1}^{n}(\log X_i)^2$ *is the unique solution to the equation* $\partial\ell(\theta|X_1,\ldots,X_n)/\partial\theta = 0$ *and furthermore satisfies* $\partial^2\ell\left(\hat{\theta}_{ML}|X_1,\ldots,X_n\right)/\partial\theta < 0$. *As a consequence,* $\hat{\theta}_{ML}$ *is the maximum likelihood estimator of* θ.

(f) Show that $\hat{\theta}_{ML}$ is a complete, sufficient statistic for θ.

The family $\{f_X(\cdot|\theta) : \theta \in \Theta\}$ *is an exponential family. Because*

$$f_X(x|\theta) = c(\theta)\,x^{-1}\exp\left[-\frac{n}{2\theta}\frac{2(\log x)^2}{n}\right]1(x>0),$$

it follows from the factorization criterion that $\hat{\theta}_{ML} = \sum_{i=1}^{n}\left[2(\log X_i)^2/n\right]$ *is sufficient. Moreover, because*

$$\left\{-\frac{n}{2\theta} : \theta \in \Theta\right\} = \left\{-\frac{n}{2\theta} : 0 < \theta < \infty\right\} = (-\infty, 0)$$

contains an open set, it follows from the properties of exponential families that $\hat{\theta}_{ML}$ *is complete.*

(g) Find a uniform minimum variance unbiased estimator of θ and compute its variance.
[Hint: Use the fact (about the normal distribution) that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} r^4 \exp\left(-\frac{1}{2\sigma^2} r^2\right) dr = 3\sigma^4$$

for any $\sigma^2 > 0$.]

To show that $\hat{\theta}_{ML} = n^{-1} \sum_{i=1}^{n} \left(2Y_i^2\right)$ is a uniform minimum variance unbiased estimator of θ, it suffices to show that $\hat{\theta}_{ML}$ is unbiased. In turn, this result follows from the fact that

$$E_\theta\left(2Y_i^2\right) = \theta,$$

which itself is a consequence of the distributional result $Y_i \sim \mathcal{N}(0, \theta/2)$ established in (b).

Using the hint (with $\sigma^2 = \theta/2$), we have:

$$Var_\theta\left(Y_i^2\right) = E_\theta\left(Y_i^4\right) - E_\theta\left(Y_i^2\right)^2 = 3\left(\frac{\theta}{2}\right)^2 - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{2}.$$

As a consequence, the variance of $\hat{\theta}_{ML}$ is given by

$$Var_\theta\left(\hat{\theta}_{ML}\right) = Var_\theta\left[\frac{2}{n}\sum_{i=1}^{n}(\log X_i)^2\right] = \frac{4}{n} Var_\theta\left(Y_i^2\right) = \frac{2\theta^2}{n}.$$

(h) Compute the Cramér-Rao bound (on the variance of unbiased estimators of θ). Is this bound attained by the estimator from (g)?

The Fisher information is given by

$$\mathcal{I}(\theta) = E_\theta\left(\left[\frac{\partial}{\partial\theta}\ell(\theta|X_1,\ldots,X_n)\right]^2\right) = Var\left[\frac{1}{\theta^2}\sum_{i=1}^{n}(\log X_i)^2\right] = \frac{n}{\theta^4} Var\left[(\log X_i)^2\right] = \frac{n}{2\theta^2},$$

where the last equality uses the fact (established in (g)) that

$$Var\left[(\log X_i)^2\right] = Var\left(Y_i^2\right) = \theta^2/2.$$

The Cramér-Rao bound is $\mathcal{I}(\theta)^{-1} = 2\theta^2/n$, which is attained by $\hat{\theta}_{ML}$.