Practice Midterm II Solution

STA 200B
Winter 2020
University of California, Davis

Exam Rules: This exam is closed book and closed notes. You may bring one page of notes, double-sided. Use of calculators, cell phones or any other electronic or communication devices is not allowed. You must show all of your work to receive credit. You will have 50 minutes to complete the exam. This exam has 4 pages, make sure you have all four pages.

Note: You do not need to show that the second derivative is negative when deriving MLEs. If needed, you may use that for the Beta(α,β) distribution we have $EX = \alpha/(\alpha+\beta)$, $\quad \text{var}(X) = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$ and for the Gamma(α,β) distribution $EX = \alpha/\beta$, $\quad \text{var}(X) = \alpha/\beta^2$.

Name : ____________________

ID : ____________________

Signature : ____________________

1. Let $X_1, \ldots, X_n$ be a random sample from a distribution with p.d.f.

$$f(x|\theta_1,\theta_2) = \frac{1}{\theta_2} e^{-(x-\theta_1)/\theta_2},$$

for $x \geq \theta_1$, $-\infty < \theta_1 < \infty$, and $\theta_2 > 0$.

(a) Find jointly sufficient statistics (T_1, T_2) when θ_1 and θ_2 are both unknown.

<u>**solution**</u> The joint p.d.f is

$$f_n(\mathbf{x}|\theta_1,\theta_2) = \prod_{i=1}^n \frac{1}{\theta_2^n} e^{-\frac{(X_i-\theta_1)}{\theta_2}} \mathbf{1}_{\{X_i \geq \theta_1\}} = \frac{1}{\theta_2^n} e^{-\frac{\sum_{i=1}^n X_i - n\theta_1}{\theta_2}} \mathbf{1}_{\{X_{(1)} \geq \theta_1\}}.$$

$(T_1, T_2) = (\sum_{i=1}^n X_i, X_{(1)})$ is jointly sufficient for (θ_1, θ_2) by factorization theorem with $u(\mathbf{x}) = 1$ and $v((T_1,T_2),(\theta_1,\theta_2)) = \frac{1}{\theta_2^n} e^{-\frac{T_1 - n\theta_1}{\theta_2}} \mathbf{1}_{\{T_2 \geq \theta_1\}}$.

(b) Find a method of moments estimator for for $\theta_1+\theta_2$ and obtain the MSE for this estimator.

solution

$$\begin{aligned} E(X) &= \int_{\theta_1}^{\infty} x\frac{1}{\theta_2}e^{-(x-\theta_1)/\theta_2}dx \\ &= \int_{\theta_1}^{\infty} -xde^{-(x-\theta_1)/\theta_2} \\ &= -xe^{-(x-\theta_1)/\theta_2}\Big|_{\theta_1}^{\infty} + \int_{\theta_1}^{\infty} e^{-(x-\theta_1)/\theta_2}dx \\ &= \theta_1+\theta_2, \end{aligned}$$

so $\mu_1 = \frac{\sum_{i=1}^n X_i}{n}$ is a method of moments estimator of $\theta_1+\theta_2$. It is unbiased by the calculation above, so the MSE can be calculated by

$$\text{MSE}(\frac{\sum_{i=1}^n X_i}{n}) = \text{Var}(\frac{\sum_{i=1}^n X_i}{n}) = \frac{\text{Var}(X_i)}{n} = \frac{\theta_2^2}{n}.$$

(c) If θ_1 is known, find the UMVUE for θ_2.

solution

$$f(x|\theta_2) = \mathbf{1}_{\{x\geq\theta_1\}}\frac{1}{\theta_2}e^{\frac{\theta_1}{\theta_2}}e^{-\frac{x}{\theta_2}}$$

belongs to the exponential family with $a(\theta_2) = \frac{1}{\theta_2}e^{\frac{\theta_1}{\theta_2}}$, $b(x) = \mathbf{1}_{\{x\geq\theta_1\}}$, $c(\theta_2) = \frac{1}{\theta_2}$ and $d(x) = x$. The parameter space $\theta_2 > 0$ contains an open set. So by theorem on complete statistics in the exponential family $\sum_{i=1}^n X_i$ is complete for θ_2.
By (b) we have $E(\sum_{i=1}^n X_i) = n(\theta_1+\theta_2)$, so $\frac{\sum_{i=1}^n X_i}{n} - \theta_1$ is the UMVUE for θ_2.

(d) If θ_2 is known, find the Fisher information in the sample.

solution The Fisher information for θ_1 does not exist because the support $[\theta_1,\infty)$ of the distribution of X_i depends of θ_1.

2. Suppose $X_1,\ldots,X_n$ form a random sample from a Poisson distribution with parameter $\theta > 0$.

(a) Obtain the UMVUE by first showing efficiency of a suitable statistic.

solution The likelihood function is

$$f_n(\mathbf{x}|\theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}$$

and derivative of log-likelihood

$$\frac{\partial \log f_n(\mathbf{x}|\theta)}{\partial\theta} = -n + \frac{\sum_{i=1}^n X_i}{\theta}.$$

By Cramér-Rao bound $\frac{\sum_{i=1}^n X_i}{n} = u(\theta)\frac{\partial \log f_n(\mathbf{x}|\theta)}{\partial\theta} + v(\theta)$ with $u(\theta) = \frac{\theta}{n}$ and $v(\theta) = \theta$, $\frac{\sum_{i=1}^n X_i}{n}$ is an efficient estimator. Then by $E(\frac{\sum_{i=1}^n X_i}{n}) = \theta$ it is the UMVUE of θ.

(b) Obtain the UMVUE without using Fisher information or the log-density derivative.

solution The pmf $\frac{e^{-\theta}e^{x\log(\theta)}}{x!}$ belongs to the exponential family with $a(\theta) = e^{-\theta}$, $b(x) = \frac{1}{x!}$, $c(\theta) = \log(\theta)$ and $d(x) = x$. The parameter space $\theta > 0$ contains an open set. So by theorem on complete statistics in the exponential family $\sum_{i=1}^n X_i$ is a complete statistics of θ. From (a) we know $\frac{\sum_{i=1}^n X_i}{n}$ is unbiased and this means it is the UMVUE since it is a function of $\sum_{i=1}^n X_i$.

(c) Obtain method of moments estimators based on first and second moments. Which of these is preferred? State your reasons.

solution By $E(X_i) = \theta$, $\hat{\theta}_1 = \mu_1 = \frac{\sum_{i=1}^n X_i}{n}$ is the method of moments estimator based on the first moment. The second moment is $E(X_i^2) = \theta + \theta^2$. Then solving $\theta + \theta^2 = \mu_2$ with constraint $\theta > 0$, we have $\hat{\theta}_2 = \frac{-1+\sqrt{1+4\mu_2}}{2}$ is the method of moment estimator based on the second moment, where $\mu_2 = \frac{\sum_{i=1}^n X_i^2}{n}$.
Since $\hat{\theta}_2$ is not a function of complete statistics, it is inadmissible and therefore not a UMVUE. $\hat{\theta}_1$ is preferred.

(d) Show that there is no unbiased estimator of $1/\theta$.

solution Let $n = 1$, if there exists an unbiased estimator $\delta(X)$ we have $E(\delta(X)) = \frac{1}{\theta}$. Write the expectation out we get

$$\sum_{x=0}^{\infty} \delta(x)\frac{e^{-\theta}\theta^x}{x!} = \frac{1}{\theta}$$

and therefore

$$\sum_{x=0}^{\infty} \delta(x)\frac{\theta^{x+1}}{x!} = e^{\theta} = \sum_{x=0}^{\infty} \frac{\theta^x}{x!}.$$

Matching coefficients of θ^x with $x = 0$ we get $\frac{1}{x!} = 0$, which should not happen. So such δ does not exist and there is no unbiased estimator of $\frac{1}{\theta}$.

3. Suppose we draw a sample $X_1, \ldots, X_n$ of size n from the distribution $N(\mu_1, \sigma_1^2)$ and a sample $Y_1, \ldots, Y_m$ of size m from the distribution $N(\mu_2, \sigma_2^2)$. Assume $\sigma_1^2 = 4\sigma_2^2$ and $\mu_1 = \mu_2 =: \mu$. We aim to estimate $\theta = \mu$ and use the estimator $\theta_\alpha = \alpha \bar{X}_n + (1-\alpha)\bar{Y}_m$, where $\bar{X}$ denotes the sample means.

   (a) For what value of α is the MSE minimized? What is the value of the MSE at the minimum?

   solution We first observe $E(\theta_\alpha) = \alpha E(\bar{X}_n) + (1-\alpha)E(\bar{Y}_m) = \alpha\mu + (1-\alpha)\mu = \mu$ is unbiased. Then

$$
\begin{aligned}
\mathrm{MSE}(\theta_\alpha) &= \mathrm{Var}(\alpha \bar{X}_n + (1-\alpha)\bar{Y}_m) \\
&= \frac{\alpha^2}{n}\mathrm{Var}(X_i) + \frac{(1-\alpha)^2}{m}\mathrm{Var}(Y_i) \\
&= \frac{4\alpha^2\sigma_2^2}{n} + \frac{(1-\alpha)^2\sigma_2^2}{m}.
\end{aligned}
$$

   Denote this by $g(\theta)$ and take the derivative $\frac{dg(\theta)}{d\theta} = \frac{8\alpha\sigma_2^2}{n} + \frac{2(\alpha-1)\sigma_2^2}{m}$. Set this to 0 we get $\alpha = \frac{n}{4m+n}$ minimize the MSE. (The second derivative $\frac{8\sigma_2^2}{n} + \frac{2\sigma_2^2}{m} > 0$ means it is convex.)

   (b) How does this MSE compare to that of the estimator that is obtained when you pool the two samples into one and take the sample average as estimator?

   solution This pooled estimator is $\frac{n\bar{X}_n + m\bar{Y}_m}{n+m} = \theta_\alpha$ with $\alpha = \frac{n}{m+n}$. By the optimization in (a) it clearly has a larger MSE compared to the optimal one.

4. Suppose $X_1, \ldots, X_n$ form a random sample from a distribution with p.d.f. $f(x|\theta) = \theta x^{\theta-1}$, for $0 < x < 1$ and $\theta > 0$.

   (a) Find a minimal sufficient statistics for θ.

   solution
   Method 1: By $f(x|\theta) = \theta x^{\theta-1} = \theta e^{\log(x)(\theta-1)}$, it belongs to the exponential family with $a(\theta) = \theta$, $b(x) = 1$, $c(\theta) = \theta - 1$ and $d(x) = \log(x)$. The parameter space $\theta > 0$ contains an open set. So by theorem on complete statistics of exponential family $\sum_{i=1}^n \log(X_i)$ is complete and therefore minimal sufficient.
   Method 2:

$$f_n(\mathbf{x}|\theta) = \theta^n (\prod_{i=1}^n X_i)^{\theta-1}$$

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

$$\frac{dL}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^n \log(X_i)$$

$$\frac{d^2L}{d\theta^2} = -\frac{n}{\theta} < 0.$$

So MLE of θ is $-\frac{n}{\sum_{i=1}^n \log(X_i)}$. $T = \prod_{i=1}^n X_i$ is sufficient by factorization theorem with $u(\mathbf{x}) = 1$ and $v(T,\theta) = \theta^n T^{\theta-1}$. Then MLE is sufficient and therefore minimal sufficient.
Method 3:

$$\frac{f_n(\mathbf{x}|\theta)}{f_n(\mathbf{y}|\theta)} = (\frac{\prod_{i=1}^n X_i}{\prod_{i=1}^n Y_i})^{\theta-1}$$

is constant if and only if $\prod_{i=1}^n X_i = \prod_{i=1}^n Y_i$, so $\prod_{i=1}^n X_i$ is minimal sufficient.

(b) Is the sample mean admissible for estimating θ under square error loss? Provide your reasoning.

solution
The sample mean $\frac{\sum_{i=1}^n X_i}{n}$ is not a function of minimal sufficient statistics and by Rao-Blackwell theorem it is inadmissible.

(c) Use that fact that $E(-\log X_i) = \frac{1}{\theta}$ to find an UMVUE of $\frac{1}{\theta}$.

solution
In (a) method 1 we have shown $\sum_{i=1}^n \log X_i$ is a complete statistics, so $-\frac{\sum_{i=1}^n \log X_i}{n}$ is the UMVUE of $\frac{1}{\theta}$.