STAT 30100 : Mathematical Statistics-1

Winter 2020

•

Homework 8

•

Solutions by

Jinhong Du

12243476

STAT 30100, Homework 8

1. (Casella and Berger Problem 7.38) For each of the following distributions, let $X_1, \dots, X_n$ be a random sample. Is there a function of θ, say $g(\theta)$, for which there exists an unbiased estimator whose variance attains the Cramér-Rao Lower Bound? If so, find it. If not, show why not.

   (a) $f(x|\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$.

Proof. The parameter space $\Theta = \mathbb{R}^+$ is open.

$$\begin{aligned}
\log f(x|\theta) &= [\log(\theta) + (\theta-1)\log(x)]\mathbb{1}_{(0,1)} \\
\frac{\partial \log f(x|\theta)}{\partial\theta} &= \left(\frac{1}{\theta} + \log(x)\right)\mathbb{1}_{(0,1)} \\
\frac{\partial^2 \log f(x|\theta)}{\partial\theta^2} &= -\frac{1}{\theta^2}\mathbb{1}_{(0,1)}
\end{aligned}$$

As $f(x|\theta)$ is in the exponential family, we can pass the derivative under the integral sign to get the Fisher information matrix

$$\mathcal{I}_1(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2 \log f(X_1|\theta)}{\partial\theta^2}\right) = \frac{1}{\theta^2},$$

which is nonsingular for $\theta \in \Theta$. Since the joint density of $\boldsymbol{X}$ is $f_\theta(\boldsymbol{x}) = \theta^n \prod_{i=1}^n x_i^{\theta-1}\mathbb{1}_{(0,1)}(x_i) = \mathbb{1}_{0<x_{(1)}<x_{(n)}<1} \cdot \theta^n \cdot e^{(\theta-1)\sum_{i=1}^n \log(x_i)}$ which is in the exponential family. So the Cramér-Rao Lower Bound is achieved by $w(\boldsymbol{X}) = \sum_{i=1}^n \log(X_i)$.

$$\begin{aligned}
\mathbb{E}_\theta[w(\boldsymbol{X})] &= \sum_{i=1}^n \mathbb{E}_\theta[\log(X_i)] \\
&= n\int_0^1 \log(x)\theta x^{\theta-1}\mathrm{d}x \\
&= n\log(x)x^\theta\big|_0^1 - n\int_0^1 x^{\theta-1}\mathrm{d}x \\
&= -\frac{n}{\theta}x^\theta\big|_0^1 \\
&= -\frac{n}{\theta}.
\end{aligned}$$

So for $g(\theta) = -\frac{n}{\theta}$, there exists a unbiased estimator $w(\boldsymbol{X})$ whose variance attains the Cramér-Rao Lower Bound. □

   (b) $f(x|\theta) = \frac{\log\theta}{\theta-1}\theta^x$, $0 < x < 1$, $\theta > 1$.

Proof. The parameter space $\Theta = (1, \infty)$ is open.

$$\begin{aligned}
\log f(x|\theta) &= [\log[\log(\theta)] - \log(\theta-1) + x\log(\theta)]\mathbb{1}_{(0,1)} \\
\frac{\partial \log f(x|\theta)}{\partial\theta} &= \left(\frac{1}{\theta\log(\theta)} - \frac{1}{\theta-1} + \frac{x}{\theta}\right)\mathbb{1}_{(0,1)} \\
\frac{\partial^2 \log f(x|\theta)}{\partial\theta^2} &= \left(-\frac{\log(\theta)+1}{\theta^2\log^2(\theta)} + \frac{1}{(\theta-1)^2} - \frac{x}{\theta^2}\right)\mathbb{1}_{(0,1)}
\end{aligned}$$

Solution (cont.)

$$\begin{aligned}\mathbb{E}_\theta(X_1) &= \int_0^1 x\frac{\log\theta}{\theta-1}\theta^x \mathrm{d}x \\ &= \frac{1}{\theta-1}\theta^x x\Big|_0^1 - \int_0^1 \frac{\theta^x}{\theta-1}\mathrm{d}x \\ &= \frac{\theta}{\theta-1} - \frac{1}{\log(\theta)}.\end{aligned}$$

As $f(x|\theta)$ is in the exponential family, we can pass the derivative under the integral sign to get the Fisher information matrix

$$\begin{aligned}\mathcal{I}_1(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2 \log f(X_1|\theta)}{\partial\theta^2}\right) &= \frac{\log(\theta)+1}{\theta^2\log^2(\theta)} - \frac{1}{(\theta-1)^2} + \frac{1}{\theta(\theta-1)} - \frac{1}{\theta^2\log(\theta)} \\ &= \frac{1}{\theta^2\log^2(\theta)} - \frac{1}{\theta(\theta-1)^2}\end{aligned}$$

which is nonsingular for $\theta \in \Theta$. Since the joint density of $\boldsymbol{X}$ is $f_\theta(\boldsymbol{x}) = \frac{\log^n(\theta)}{(\theta-1)^n}\theta^{\sum_{i=1}^n x_i}\mathbb{1}_{0<x_{(1)}<x_{(n)}<1} = \mathbb{1}_{0<x_{(1)}<x_{(n)}<1}\cdot\frac{\log^n(\theta)}{(\theta-1)^n}\cdot e^{\ln(\theta)\sum_{i=1}^n x_i}$ which is in the exponential family. So the Cramér-Rao Lower Bound is achieved by $w(\boldsymbol{X}) = \sum_{i=1}^n X_i$.

$$\begin{aligned}\mathbb{E}_\theta[w(\boldsymbol{X})] &= \sum_{i=1}^n \mathbb{E}_\theta(X_i) \\ &= n\left(\frac{\theta}{\theta-1} - \frac{1}{\log(\theta)}\right)\end{aligned}$$

So for $g(\theta) = n\left(\frac{\theta}{\theta-1} - \frac{1}{\log(\theta)}\right)$, there exists a unbiased estimator $w(\boldsymbol{X})$ whose variance attains the Cramér-Rao Lower Bound. $\square$

2. (Casella and Berger Problem 7.46) Let X_1, X_2, and X_3 be a random sample of size three from a Uniform$(\theta, 2\theta)$ distribution, where $\theta > 0$.

(a) Find the method of moments estimator of θ.

Since $\mathbb{E}(X_i) = \int_\theta^{2\theta} x\frac{1}{\theta}\mathrm{d}x = \frac{3\theta}{2}$, we have the moments estimator of θ is $\frac{1}{3}\sum_{i=1}^3 \frac{2}{3}X_i = \frac{2}{9}\sum_{i=1}^3 X_i$.

(b) Find the MLE, $\hat{\theta}$, and find a constant k such that $\mathbb{E}_\theta(k\hat{\theta}) = \theta$.

The likelihood function is

$$\begin{aligned}L(\theta;\boldsymbol{x}) &= \frac{1}{\theta^3}\mathbb{1}_{(\theta,2\theta)}(x_1)\mathbb{1}_{(\theta,2\theta)}(x_2)\mathbb{1}_{(\theta,2\theta)}(x_3) \\ &= \frac{1}{\theta^3}\mathbb{1}_{\theta<x_{(1)}<x_{(3)}<2\theta} \\ \frac{\mathrm{d}L(\theta;\boldsymbol{x})}{\mathrm{d}\theta} &= -\frac{3}{\theta^4} < 0, \qquad \theta < x_{(1)} < x_{(3)} < 2\theta.\end{aligned}$$

So $L(\theta;\boldsymbol{x})$ is decreasing and the MLE of θ is $\hat{\theta} = \frac{1}{2}X_{(3)}$. The density of $X_{(3)}$ is

$$f_{X_3}(x) = 3\left[\frac{1}{\theta}(x-\theta)\right]^2\frac{1}{\theta}\mathbb{1}_{(\theta,2\theta)}(x),$$

Solution (cont.)

so

$$\mathbb{E}_\theta(\hat{\theta}) = \frac{1}{2}\mathbb{E}_\theta(X_{(3)}) = \frac{1}{2}\int_\theta^{2\theta} \frac{3}{\theta}\left[\frac{1}{\theta}(x-\theta)\right]^2 x\mathrm{d}x = \frac{1}{2\theta^3}x(x-\theta)^3\Big|_\theta^{2\theta} - \frac{1}{2\theta^3}\int_\theta^{2\theta}(x-\theta)^3\mathrm{d}x = \frac{7}{8}\theta$$

which impies that $k = \frac{8}{7}$.

(c) Which of the two estimators can be improved by using sufficiency? How?

From Exercise 6.23, a minimal sufficient statistic for θ is $T(\boldsymbol{X}) = (X_{(1)}, X_{(n)})$. $\frac{2}{9}\sum_{i=1}^3 X_i$ is not a function of this minimal sufficient statistic. So by the Rao-Blackwell Theorem, $\mathbb{E}(\frac{2}{9}\sum_{i=1}^3 X_i | T(\boldsymbol{X}))$ is an unbiased estimator of θ with smaller variance than $\frac{2}{9}\sum_{i=1}^3 X_i$. The MLE is a function of $T(\boldsymbol{X})$, so it can not be improved with the Rao-Blackwell Theorem.

(d) Find the method of moments estimate and the MLE of θ based on the data 1.29, .86, 1.33, three observations of average berry sizes (in centimeters) of wine grapes.

The moments estimate is $\frac{2}{9}(1.29 + 0.86 + 1.33) \approx 0.773$. The MLE of θ is $\frac{1}{2} \times 1.33 = 0.665$.

3. (Casella and Berger Problem 7.48) Suppose that X_i, $i = 1, \ldots, n$, are iid Bernoulli(p).

(a) Show that the variance of the MLE of p attains the Cramer-Rao Lower Bound.

Proof. The parameter space $\Theta = (0, 1)$ is open. The log-likelihood is

$$\begin{aligned} l(p; \boldsymbol{x}) &= \sum_{i=1}^n [x_i \log(p) + (1-x_i)\log(1-p)] = \log(p)\sum_{i=1}^n x_i + \log(1-p)\sum_{i=1}^n (1-x_i), \\ \frac{\mathrm{d}l(p;\boldsymbol{x})}{\mathrm{d}p} &= \frac{1}{p}\sum_{i=1}^n x_i - \frac{1}{1-p}\sum_{i=1}^n (1-x_i) = 0, \end{aligned}$$

which yields $p = \overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$. So the MLE of p is $\hat{p} = \overline{X}$.

$$\frac{\mathrm{d}^2 l(p;\boldsymbol{x})}{\mathrm{d}p^2} = -\frac{1}{p^2}\sum_{i=1}^n x_i - \frac{1}{(1-p)^2}\sum_{i=1}^n (1-x_i),$$

which is continuous in $p \in (0, 1)$

Since the joint density

$$f_p(\boldsymbol{x}) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} = (1-p)^n e^{n\ln\left(\frac{p}{1-p}\right)\overline{x}}$$

is in the exponential family, we can pass the derivative under the integral sign to get the Fisher information matrix,

$$\mathcal{I}_1(p) = -\mathbb{E}_p\left[\frac{\mathrm{d}^2 l(p;\boldsymbol{x})}{\mathrm{d}p^2}\right] = -\frac{n}{p} - \frac{n}{1-p}$$

which is nonsingular for $p \in \Theta$. Since the joint density is in the exponential family with natural sufficient statistic $w(\boldsymbol{X}) = \overline{X}$, the Cramér-Rao Lower Bound is achieved by $w(\boldsymbol{X}) = \hat{p}$. □

(b) For $n \geq 4$, show that the product $X_1X_2X_3X_4$ is an unbiased estimator of p^4, and use this fact to find the best unbiased estimator of p^4.

Since $X_1, \dots, X_4$ are independent, we have $\mathbb{E}_p(X_1X_2X_3X_4) = \prod_{i=1}^4 \mathbb{E}_p(X_i) = p^4$. So $X_1X_2X_3X_4$ is an unbiased estimator of p^4. Since $\sum_{i=1}^n X_i$ is a complete sufficient statistic, then $\mathbb{E}_p\left(X_1X_2X_3X_4 | \sum_{i=1}^n X_i = t\right)$ if the best unbiased estimator of p^4.
For $t \geq 4$,

$$\begin{aligned}
\mathbb{E}_p\left(X_1X_2X_3X_4 | \sum_{i=1}^n X_i = t\right) &= \frac{\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, \sum_{i=5}^n X_i = t-4)}{\mathbb{P}(\sum_{i=1}^n X_i = t)} \\
&= \frac{p^4\binom{n-4}{t-4}p^{t-4}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} \\
&= \frac{\binom{n-4}{t-4}}{\binom{n}{t}}
\end{aligned}$$

For $t < 4$, at least one of $X_1, \dots, X_4$ is 0 and therefore $\mathbb{E}_p\left(X_1X_2X_3X_4 | \sum_{i=1}^n X_i = t\right) = 0$.

4. (Casella and Berger Problem 7.52) Let $X_1, \dots, X_n$ be iid Poisson(λ), and let $\overline{X}$ and S^2 denote the sample mean and variance, respectively. We now complete Example 7.3.8 in a different way. There we used the Cramer-Rao Bound; now we use completeness.

(a) Prove that $\overline{X}$ is the best unbiased estimator of λ without using the Cramer-Rao Theorem.

Proof. Since $\mathbb{E}(\overline{X}) = \mathbb{E}(X_1) = \lambda$, so $\overline{X}$ is an unbiased estimator of λ.
The joint density of $\boldsymbol{X}$ is

$$f_\lambda(\boldsymbol{x}) = \left(\prod_{i=1}^n \frac{1}{x_i!}\right) \cdot e^{-n\lambda} \cdot e^{\lambda \sum_{i=1}^n x_i},$$

so $\sum_{i=1}^n X_i$ is the complete sufficient statistic of λ. Since $\overline{X}$ is a function of $\sum_{i=1}^n X_i$ and it is unbiased, it is also the best unbiased estimator of λ.

□

(b) Prove the rather remarkable identity $\mathbb{E}(S^2|\overline{X}) = \overline{X}$, and use it to explicitly demonstrate that $\mathrm{Var}(S^2) > \mathrm{Var}(\overline{X})$.

Since $\mathbb{E}(S^2) = \frac{1}{n-1}\sum_{i=1}^n \mathbb{E}_\lambda[X_i^2 - 2X_i\overline{X} + \overline{X}^2] = \frac{n}{n-1}\left[\lambda^2 + \lambda - \frac{2}{n}\left(\lambda^2 + \lambda + (n-1)\lambda^2\right) + \lambda^2 + \frac{1}{n}\lambda\right] = \lambda$, S^2 is an unbiased estimator of λ. Since $\overline{X}$ is a one-to-one function of $\sum_{i=1}^n X_i$, it is also a complete sufficient statistic of λ. Then $\mathbb{E}(S^2|\overline{X})$ is the unique best unbiased estimator of λ. So $\mathbb{E}(S^2|\overline{X}) = \overline{X}$.
Therefore, $\mathrm{Var}(S^2) = \mathrm{Var}[\mathbb{E}(S^2|\overline{X})] + \mathbb{E}[\mathrm{Var}(S^2|\overline{X})] = \mathrm{Var}(\overline{X}) + \mathbb{E}[\mathrm{Var}(S^2|\overline{X})] > \mathrm{Var}(\overline{X})$.

(c) Using completeness, can a general theorem be formulated for which the identity in part (b) is a special case?

If $T(\boldsymbol{X})$ is a complete sufficient statistic for θ, and $\phi(T(\boldsymbol{X})$ is an integrable function of T, then given any other statistic S such that $\mathbb{E}_\theta(S) = \mathbb{E}_\theta[\phi(T)]$ for all θ, we have $\mathbb{E}[S|T] = \phi(T)$ with probability 1 for all θ.

Solution (cont.)

Proof. Let $g(T) = \mathbb{E}(S|T) - \phi(T)$. Then $\mathbb{E}_\theta[g(T)] = 0$ for all θ. Since T is complete, $\mathbb{P}(g(T) = 0) = 1$. $\square$