

## Studyguide Problems Solution Set

1. Let  $X_1, \dots, X_n$  be *i.i.d.* random variables from  $\text{Uniform}(\theta, 2\theta)$  with the pdf

$$f_X(x|\theta) = \frac{1}{\theta} I(\theta \leq x \leq 2\theta), \quad \theta > 0.$$

- (a) Find a minimal sufficient statistic for  $\theta$ .

**Solution:** The joint pdf is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\theta) &= \prod_{i=1}^n \left[ \frac{1}{\theta} I(\theta \leq x_i \leq 2\theta) \right] \\ &= \frac{1}{\theta^n} I(\theta \leq x_{(1)} \leq x_{(n)} \leq 2\theta) \\ &= \frac{1}{\theta^n} I(x_{(n)}/2 \leq \theta \leq x_{(1)}) \\ &= g(T(\mathbf{x})|\theta) h(\mathbf{x}) \end{aligned}$$

if  $g(t_1, t_2|\theta) = \frac{1}{\theta^n} I(t_2/2 \leq \theta \leq t_1)$ .  $T(\mathbf{x}) = (x_{(1)}, x_{(n)})$ ,  $h(\mathbf{x})$ . By Factorization Theorem  $T(\mathbf{x})$  is a sufficient statistic for  $\theta$ .

Let  $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ . Then

$$\begin{aligned} \frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} &= \frac{\frac{1}{\theta^n} I(x_{(n)}/2 \leq \theta \leq x_{(1)})}{\frac{1}{\theta^n} I(y_{(n)}/2 \leq \theta \leq y_{(1)})} \\ &= \frac{I(x_{(n)}/2 \leq \theta \leq x_{(1)})}{I(y_{(n)}/2 \leq \theta \leq y_{(1)})} \end{aligned}$$

is constant to  $\theta$  if and only if  $T(\mathbf{x}) = (x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)}) = T(\mathbf{y})$ . Therefore, by Theorem 6.2.13,  $T(\mathbf{x})$  is a minimal sufficient statistic.

- (b) Is the minimal sufficient statistic obtained in (a) also a complete statistic? Justify your answer.

**Solution:** The distribution of  $X$  is a scale family of  $\text{Uniform}(1, 2)$ . Therefore,  $R(T(\mathbf{X})) = X_{(n)}/X_{(1)}$ , which is a function of minimal sufficient statistic, is an ancillary statistic for  $\theta$ . Because any function of complete sufficient statistic cannot be ancillary,  $T$  is not a complete statistic.

(c)

Find a maximum likelihood estimator for  $\theta(\mathbf{X})$ .

**Solution:** The likelihood function is

$$L(\theta|\mathbf{x}) = \frac{1}{\theta^n} I(x_{(n)}/2 \leq \theta \leq x_{(1)})$$

This function is a decreasing function of  $\theta$  when  $x_{(n)}/2 \leq \theta \leq x_{(1)}$  and zero otherwise, so the likelihood is maximized when  $\theta = \frac{x_{(n)}}{2}$ . Therefore, the maximum likelihood estimator is  $\hat{\theta} = \frac{x_{(n)}}{2}$ .

2. Let  $X_1, \dots, X_n$  be *i.i.d.* random variables from  $\text{Gamma}\left(2, \frac{1}{\theta}\right)$  with the pdf

$$f_X(x|\theta) = \theta^2 x e^{-\theta x}, \quad x \geq 0, \quad \theta > 0.$$

- (a) Show that this pdf belongs to the exponential family. Identify  $h(\cdot), c(\cdot), w_j(\cdot)$ , and  $t_j(\cdot)$  terms in your proof.

**Solution:** The pdf can be written as

$$f_X(x|\theta) = \theta^2 x \exp[-\theta x] = c(\theta)h(x) \exp[w(\theta)t(x)]$$

when  $c(\theta) = \theta^2$ ,  $h(x) = x$ ,  $w(\theta) = -\theta$ ,  $t(x) = x$ . Thus, by definition, it belongs to the exponential family

- (b) Show that  $\sum_{i=1}^n X_i$  is a complete sufficient statistic for parameter  $\theta$ .

**Solution:** By Theorem 6.2.10,  $\sum_{i=1}^n t(X_i) = \sum_{i=1}^n X_i$  is a sufficient statistic for  $\theta$ . Because  $\Theta = \{w(\theta) : \theta > 0\} = (-\infty, 0)$  contains an open set in  $R$ , by Theorem 6.2.25,  $\sum_{i=1}^n X_i$  is also a complete statistic for  $\theta$ . Therefore, it is a complete sufficient statistic.

- (c) Find the maximum likelihood estimator for  $\text{Var}(X)$ .

**Solution:** The likelihood, log-likelihood, its derivate and roots are

$$\begin{aligned}
L(\mathbf{x}|\theta) &= \prod_{i=1}^n [\theta^2 x_i e^{-\theta x_i}] \\
l(\mathbf{x}|\theta) &= \sum_{i=1}^n \log[\theta^2 x_i e^{-\theta x_i}] \\
&= \sum_{i=1}^n [2 \log \theta + \log x_i - \theta x_i] \\
&= 2n \log \theta + \sum_{i=1}^n \log x_i - \theta \sum_{i=1}^n x_i \\
\frac{\partial l(\mathbf{x}|\theta)}{\partial \theta} &= \frac{2n}{\theta} - \sum_{i=1}^n x_i = 0 \\
\hat{\theta} &= \frac{2n}{\sum_{i=1}^n x_i}
\end{aligned}$$

The second derivative of log-likelihood function is

$$\frac{\partial^2 l(\mathbf{x}|\theta)}{\partial \theta^2} = -\frac{2n}{\theta^2} < 0$$

always concave, so the unique interior maximum  $\hat{\theta}$  is also a global maximum. Therefore, the MLE of  $\theta$  is  $\hat{\theta} = \frac{2n}{\sum_{i=1}^n X_i}$ .

Because  $\text{Var}(X) = \frac{2}{\theta^2}$ , by invariance property, the MLE is  $\frac{2}{\hat{\theta}^2} = \frac{[\sum_{i=1}^n X_i]^2}{2n^2}$ .

3. Let  $X_1, \dots, X_n$  be *i.i.d.* random variables from a  $\text{Binomial}(k, p)$  distribution with pmf

$$\Pr(X = x | k, p) = \binom{k}{x} p^x (1-p)^{k-x}, \quad k \in \{1, 2, \dots\}, \quad x \in \{0, 1, 2, \dots, k\}, \quad 0 \leq p \leq 1$$

Assuming  $p$  is known, define

$$W(\mathbf{X}) = \frac{\sum_{i=1}^n X_i}{np}$$

as an estimator for the unknown parameter  $k$ .

- (a) Calculate the bias and MSE of  $W(\mathbf{X})$ .

**Solution:**  $\sum X_i$  follows Binomial( $nk, p$ ). So the expectation and variance of  $W$  is

$$\begin{aligned} \text{EW} &= \text{E} \left[ \frac{\sum_i X_i}{np} \right] = \frac{npk}{np} = k \\ \text{Var}W &= \text{Var} \left[ \frac{\sum_i X_i}{np} \right] = \frac{nkp(1-p)}{n^2 p^2} = \frac{k(1-p)}{np} \end{aligned}$$

So  $W$  is an unbiased estimator of  $k$  (i.e. bias is zero), and the MSE is equal to  $\text{Var}W = \frac{k(1-p)}{np}$ .

- (b) As an estimator for  $k$ , what is the most glaring disadvantage of  $W$ ?

**Solution:**  $k$  is an integer, but  $W$  is not necessarily integer, so the estimates does not belong to valid parameter space.

4. Let  $X_1, \dots, X_n$  be *i.i.d.* random variables from  $N(0, \theta)$  with pdf

$$f_X(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ -\frac{x^2}{2\theta} \right\}, \quad x \in R, \quad \theta > 0.$$

- (a) Using Corollary 7.3.15 (Attainment), find  $\tau(\theta)$  for which the best unbiased estimator (UMVUE) attains its Cramer-Rao lower bound.

**Solution:** The log-likelihood and score function is

$$\begin{aligned} l(\theta|\mathbf{x}) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \theta - \frac{\sum_{i=1}^n x_i^2}{2\theta} \\ \frac{\partial l(\theta|\mathbf{x})}{\partial \theta} &= -\frac{n}{2\theta} + \frac{\sum_{i=1}^n x_i^2}{2\theta^2} \\ &= \frac{n}{2\theta^2} \left[ \frac{1}{n} \sum_{i=1}^n x_i^2 - \theta \right] \\ &= a(\theta)[W(\mathbf{x}) - \theta] \end{aligned}$$

Therefore,  $\tau(\theta) = \theta$  and  $W(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i^2$  is the best unbiased estimator for  $\theta$ .

- (b) Find the best unbiased estimator (UMVUE) for  $\theta^2$ .

**Solution:** First note that the distribution of  $X$  belongs to an exponential family because

$$f_X(x|\theta) = h(x)c(\theta)\exp[w(\theta)t(x)]$$

if  $h(x) = 1/\sqrt{2\pi}$ ,  $c(\theta) = \frac{1}{\sqrt{\theta}}$ ,  $t(x) = x^2$  and  $w(\theta) = -\frac{1}{\theta}$ .

Because  $\Theta = \{w(\theta) : \theta > 0\} = (-\infty, 0)$  contains an open set in  $\mathbb{R}$ , by Theorem 6.2.10 and 6.2.25,  $T(\mathbf{X}) = \sum_{i=1}^n X_i^2$  is a complete sufficient statistic.

Now  $T/\theta == \sum_{i=1}^n X_i^2/\theta$  follows a  $\chi_n^2$  distribution. It follows that

$$ET^2 = \text{Var}(T) + (ET)^2 = 2n\theta^2 + n^2\theta^2 = (2n + n^2)\theta^2$$

Because  $T$  is a complete sufficient statistic,  $\frac{T^2}{2n+n^2} = \frac{(\sum_{i=1}^n X_i^2)^2}{n(n+2)}$  is the best unbiased estimator of its expected value  $\theta^2$  by Theorem 7.3.23.

- (c) Calculate the Cramer-Rao lower bound of the variance of unbiased estimators for  $\theta^2$ , and justify whether the variance of UMVUE obtained in (b) attains the bound or not.

**Solution:** Based on a single observation,

$$\begin{aligned}\log f(x|\theta) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \frac{x^2}{2\theta} \\ \frac{\partial \log f(x|\theta)}{\partial \theta} &= -\frac{1}{2\theta} + \frac{x^2}{2\theta^2} \\ \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} &= \frac{1}{2\theta^2} - \frac{x^2}{\theta^3}\end{aligned}$$

Since  $X^2/\theta \sim \chi_1^2$ , we have  $E(X^2) = \theta$  and hence

$$\begin{aligned}I(\theta) &= -E\left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2}\right] \\ &= -\frac{1}{2\theta^2} + \frac{\theta}{\theta^3} \\ &= \frac{1}{2\theta^2}\end{aligned}$$

Then  $I_n(\theta) = nI(\theta) = n/(2\theta^2)$  is the Fisher Information Number. Since  $\tau(\theta) = \theta^2$ , the CRLB for variances of unbiased estimators for  $\theta^2$  equals

$$CRLB = \frac{(\tau'(\theta))^2}{I_n(\theta)} = \frac{4\theta^2}{n/(2\theta^2)} = \frac{8\theta^4}{n}.$$

The variance of UMVUE obtained in (b) does not attain the bound since part (a) identifies  $\tau(\theta) = \theta$  to be the only function of  $\theta$  for which attainment occurs.