and T_n is consistent. ■

A final important result: although MLEs are sometimes biased, they are always consistent.

2.6 CRAMÉR-RAO INEQUALITY

Let $\mathbf{X} = (X_1, ..., X_n)$ be a vector r.v. with likelihood function $L_{\mathbf{X}}(\theta)$, where θ is a univariate parameter. Assume that: (i) the p.d.f. of each X_i has a range that does not depend on θ; (ii) $L_{\mathbf{X}}(\theta)$ is differentiable w.r.t. θ; (iii) derivatives w.r.t. θ can be moved inside and outside integrals involving $L_{\mathbf{X}}(\theta)$.

We now define two important quantities associated with families of distributions:

$$S = S(\mathbf{X};\theta) = \frac{\partial}{\partial\theta}\log L_{\mathbf{X}}(\theta),$$
$$I_{\mathbf{X}}(\theta) = \operatorname{var} S(\mathbf{X};\theta).$$

$S(\mathbf{X};\theta)$ is called the <u>score function</u>. $I_{\mathbf{X}}(\theta)$ is called the <u>Fisher information</u> in $\mathbf{X}$ (and is used to measure the amount of information about θ in the *n* observations).

<u>Theorem 2.6A</u>

(a) $\mathrm{E}S(\mathbf{X};\theta) = 0$;

(b) $I_{\mathbf{X}}(\theta) = \mathrm{E}S(\mathbf{X};\theta)^2 = -\mathrm{E}\frac{\partial}{\partial\theta}S(\mathbf{X};\theta) = -\mathrm{E}\frac{\partial^2}{\partial\theta^2}\log L_{\mathbf{X}}(\theta)$

<u>Proof:</u> (a) We have

$$\begin{aligned}
\mathrm{E}S(\mathbf{X};\theta) &= \mathrm{E}\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{X};\theta) \\
&= \int\left\{\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\} f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} \\
&= \int\frac{\frac{\partial}{\partial\theta}f_{\mathrm{X}}(\mathbf{x};\theta)}{f_{\mathrm{X}}(\mathbf{x};\theta)} f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} \\
&= \frac{\partial}{\partial\theta}\int f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} \\
&= \frac{\partial}{\partial\theta}(1) \\
&= 0.
\end{aligned}$$

(b) For the first equality,

$$\begin{aligned}
I_{\mathrm{X}}(\theta) &= \operatorname{var} S(\mathbf{X};\theta) \\
&= \mathrm{E}S(\mathbf{X};\theta)^2 - \mathrm{E}^2 S(\mathbf{X};\theta) \\
&= \mathrm{E}S(\mathbf{X};\theta)^2 \qquad \text{[by result (a) above]}
\end{aligned}$$

For the second and third equality, from (a)

$$0 = \mathrm{E}S(\mathbf{X};\theta) = \int\left\{\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\} f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x}$$

By differentiating on both sides w.r.t. θ,

$$\begin{aligned}
0 &= \int\left\{\frac{\partial^2}{\partial\theta^2}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\} f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} + \int\left\{\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\}\frac{\partial f_{\mathrm{X}}(\mathbf{x};\theta)}{\partial\theta}\,d\mathbf{x} \\
&= \int\left\{\frac{\partial^2}{\partial\theta^2}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\} f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} + \int\left\{\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\}\frac{\frac{\partial}{\partial\theta}f_{\mathrm{X}}(\mathbf{x};\theta)}{f_{\mathrm{X}}(\mathbf{x};\theta)} f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} \\
&= \int\left\{\frac{\partial^2}{\partial\theta^2}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\} f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} + \int\left\{\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\}^2 f_{\mathrm{X}}(\mathbf{x};\theta)\,d\mathbf{x} \\
&= \mathrm{E}\frac{\partial^2}{\partial\theta^2}\log f_{\mathrm{X}}(\mathbf{x};\theta) + \mathrm{E}\left\{\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\}^2
\end{aligned}$$

Hence, $I_{\mathrm{X}}(\theta) = \mathrm{E}\left\{\frac{\partial}{\partial\theta}\log f_{\mathrm{X}}(\mathbf{x};\theta)\right\}^2 = -\mathrm{E}\frac{\partial^2}{\partial\theta^2}\log f_{\mathrm{X}}(\mathbf{x};\theta) = -\mathrm{E}\frac{\partial}{\partial\theta}S(\mathbf{X};\theta)$. ■

We now consider the Cramer-Rao Inequality. This is a remarkable result that shows that the variance of any unbiased estimator has a lower bound depending on the sample size and the family of the distribution under consideration.

Theorem 2.6B (Cramér-Rao Inequality)

Let $T = T(\mathbf{X})$ be an unbiased estimator of $\tau(\theta)$. Then, under the regularity conditions stated at the start of Sec. 2.5,

$$\operatorname{var} T \geq \frac{\left\{\tau'(\theta)\right\}^2}{I_{\mathbf{X}}(\theta)} \tag{2-7}$$

with equality iff

$$L_{\mathbf{X}}(\theta) = \exp\left\{h(\theta)T(\mathbf{X}) + j(\theta) + u(\mathbf{X})\right\} \tag{2-8}$$

for some functions $h(\theta), j(\theta)$ and $u(\mathbf{X})$.■

Remarks (i) The form of L in (2-8) is the one-parameter exponential family. From (1-13),

$$\begin{aligned} & f_X(x_i;\theta) = a(\theta)b(x_i)\exp\left\{c(\theta)d(x_i)\right\} \\ \Rightarrow \quad & L_{\mathbf{X}}(\theta) = a^n(\theta)\prod_i b(X_i)\exp\left\{c(\theta)\sum_i d(X_i)\right\} \\ & \qquad = \exp\left\{h(\theta)T(\mathbf{X}) + j(\theta) + u(\mathbf{X})\right\}, \end{aligned}$$

where $a^n(\theta) = \exp\{j(\theta)\}$, $\prod_i b(X_i) = \exp\{u(\mathbf{X})\}$, $h(\theta) = c(\theta)$, and $T(\mathbf{X}) = \sum_i d(X_i)$.

(ii) If T is an unbiased estimator of $\boldsymbol{\theta}$ itself, then $\tau(\theta) = \theta \Rightarrow \tau'(\theta) = 1$ so that the Cramer-Rao inequality becomes

$$\operatorname{var} T \geq \frac{1}{I_{\mathbf{X}}(\theta)} \tag{2-9}$$

(iii)The right side of (2-7) is called the Cramer-Rao Lower Bound (CRLB). Estimators which achieve the CRLB are said to be efficient or minimum variance bound unbiased (MVBU). Not all estimators achieve the CRLB: some UMVUE's (see Sec 2.4) have variances which exceed the CRLB. In other words,

CRLB = theoretical minimum variance

var(UMVUE) = actual minimum variance

In general, var(UMVUE) $\geq$ CRLB

If var(UMVUE) = CRLB, then UMVUE is MVBU or efficient.

Thus, some UMVUE's do not achieve the CRLB and are not MVBUE's's. On the other hand, all MVBUE's are UMVUE's.

Example 2.6A

Let $X_1,...,X_n$ be i.i.d. $N(\mu,\sigma^2)$ r.v.'s The following estimators are all unbiased for μ.

$$T_1 = X_1,\quad T_2 = \frac{X_1+2X_2}{3},\quad T_3 = \frac{X_1+2X_2+3X_3}{6},\quad T_4 = \frac{X_1+...+X_n}{n}$$

Obtain the CRLB for *any* unbiased estimator of μ.

Solution We have

$$L_{\mathbf{x}}(\mu,\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}\right\}$$

$$\log L_{\mathbf{x}}(\mu,\sigma) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{n}{2}\log(2\pi)$$

$$S = \frac{\partial}{\partial\mu}\log L_{\mathbf{x}}(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}X_i - n\mu\right)$$

$$\frac{\partial S}{\partial\mu} = -\frac{n}{\sigma^2}$$

$$\therefore \quad I_{\mathbf{x}} = -\mathrm{E}\frac{\partial S}{\partial\mu} = \frac{n}{\sigma^2}$$

Also, $\tau(\mu)=\mu$. Hence,

$$CRLB = \frac{\left\{\tau'\left(\mu\right)\right\}^2}{I_{\mathbf{X}}} = \frac{\left(1\right)^2}{n/\sigma^2} = \frac{\sigma^2}{n}.$$

Note that $\operatorname{var} T_1 = \sigma^2$: thus, although it is unbiased, it does not attain the CRLB. Similarly for T_2 and T_3. On the other hand, $\operatorname{var} T_4 = \sigma^2 / n = CRLB$. Hence T_4 is the MVBUE while the others are not.

Proof Theorem 2.6B (Cramér-Rao Inequality)

Since T is unbiased,

$$\tau\left(\theta\right) = \mathrm{E}T = \int T\left(\mathbf{x}\right) f_{\mathbf{X}}\left(\mathbf{x};\theta\right) d\mathbf{x}$$

Differentiating both sides w.r.t. θ,

$$\begin{aligned}\tau'\left(\theta\right) &= \int T\left(\mathbf{x}\right)\frac{\partial}{\partial\theta} f_{\mathbf{X}}\left(\mathbf{x};\theta\right) d\mathbf{x} = \int T\left(\mathbf{x}\right)\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}\left(\mathbf{x};\theta\right)\right\} f_{\mathbf{X}}\left(\mathbf{x};\theta\right) d\mathbf{x} \\ &= \mathrm{E}\left\{TS\left(\mathbf{X};\theta\right)\right\} \\ &= \operatorname{cov}\left\{T, S\left(\mathbf{X};\theta\right)\right\} \qquad [\text{since} \quad \mathrm{E}S\left(\mathbf{X};\theta\right) = 0]\end{aligned}$$

Now, from the Cauchy-Schwartz inequality $\operatorname{cov}^2\{T, S(\mathbf{X};\theta)\} \le (\operatorname{var} T)\{\operatorname{var} S(\mathbf{X};\theta)\}$, so that

$$\left\{\tau'\left(\theta\right)\right\}^2 \le \left(\operatorname{var} T\right) I_{\mathbf{X}}$$

Since $I_{\mathbf{X}} = \operatorname{var} S(\mathbf{X};\theta) > 0$, we obtain Cramér-Rao inequality.

From the Cauchy-Schwartz inequality above, it is seen that equality holds when $S(\mathbf{X};\theta)$ is a linear combination of T (the correlation between them is then unity), i.e. when

$$S(\mathbf{X};\theta) = g_1(\theta)T + g_2(\theta), \tag{2-10}$$

where g_1 and g_1 are functions of θ. The condition for equality is thus

$$\log L(\mathbf{X};\theta) = \int g_1(\theta)T(\mathbf{X})d\theta + \int g_2(\theta)d\theta + g_3(\mathbf{X}) = T(\mathbf{X})h(\theta) + j(\theta) + u(\mathbf{X}). \blacksquare$$

For the exponential family, the MVBU estimator can be obtained by using (2-10):

$$
\begin{aligned}
S(\mathbf{X};\theta) &= g_1(\theta)T + g_2(\theta) \\
\mathrm{E}S(\mathbf{X};\theta) &= g_1(\theta)\mathrm{E}T + g_2(\theta) \\
0 &= g_1(\theta)\tau(\theta) + g_2(\theta) \\
\therefore \quad g_2(\theta) &= -g_1(\theta)\tau(\theta)
\end{aligned}
$$

Hence, (2-10) can be written in the form

$$S(\mathbf{X};\theta) = g_1(\theta)T - g_1(\theta)\tau = g_1(\theta)(T-\tau) \qquad \text{(2-11)}$$

from which the MVBUE can be extracted.

Remarks: (i) Eq. (2-11) is the same as (2-10) and is often more useful to verify if the MVBUE exists.

(ii) For the exponential family, the MVBUE exists for the *particular* function $\tau(\theta) = -g_2(\theta)/g_1(\theta) = -j'(\theta)/h'(\theta)$ and not for *any* function $\tau(\theta)$

Example 2.6B

Let $X_1,\ldots,X_n$ be i.i.d. $N(0,\sigma^2)$ r.v.'s. Obtain the $\tau(\sigma)$ for which an MVBU estimator for exists and write down the corresponding estimator.

Solution. We have

$$L_{\mathbf{X}}(\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^{n} X_i^2}{2\sigma^2}\right)$$

$$\log L_{\mathbf{X}}(\sigma) = -n\log\sigma - \frac{n}{2}\log(2\pi) - \frac{\sum_{i=1}^{n} X_i^2}{2\sigma^2}$$

$$
\begin{aligned}
S &= \frac{\partial}{\partial\sigma}\log L_{\mathbf{x}}(\mu,\sigma) \\
&= -\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n} X_i^2 \\
&= \frac{\sum_{i=1}^{n} X_i^2 - n\sigma^2}{\sigma^3} \\
&= \underbrace{\left(\frac{n}{\sigma^3}\right)}_{g_1(\theta)}\left(\underbrace{\frac{\sum_{i=1}^{n} X_i^2}{n}}_{T} - \underbrace{\sigma^2}_{\tau}\right)
\end{aligned}
$$

Hence, from Eq. (2-11) $\tau(\sigma)=\sigma^2$ and the MVBU estimator of σ^2 is $T=(\sum_i X_i^2)/n$.

Remark: From, the above, we see while there is an MVBU estimator for σ^2, there is none for (say) σ and σ^3.

Example 2.6C

Let $X_1,...,X_n$ be i.i.d. Cauchy r.v.'s where each X_i has density

$$
f_X(x;\theta)=\frac{1}{\pi}.\frac{1}{1+(x-\theta)^2}, \qquad -\infty < x < \infty
$$

Verify if the CRLB is attained (i.e. if an MVBU estimator exists for θ).

Solution.

We have

$$
L_{\mathbf{x}}(\theta)=\prod_{i=1}^{n} f_X(x_i;\theta)=\left(\frac{1}{\pi}\right)^n\prod_{i=1}^{n}\left\{\frac{1}{1+(x_i-\theta)^2}\right\}
$$

$$
\log L_{\mathbf{x}}(\theta)=-n\log\pi-\sum_{i=1}^{n}\log\left\{1+(x_i-\theta)^2\right\}
$$

$$
\frac{\partial}{\partial\theta}\log L_{\mathbf{x}}(\theta)=2\sum_{i=1}^{n}\frac{(x_i-\theta)}{1+(x_i-\theta)^2}
$$

Since the above cannot be written in the form (2-11), the CRLB cannot be attained by any unbiased estimator and no MBVU estimator exists.

Remark: An alternative solution is to show that the Cauchy distribution does not belong to the exponential family.

We conclude with an interesting connection between MVBU estimators and MLE's: not all MLE's are MBVU estimators, but all MVBU estimators are MLEs.

2.7 ASYMPTOTIC PROPERTIES OF MLEs

Suppose we wish to estimate $\tau(\theta)$. Let $\hat{\theta}_n$ be the MLE of θ, and $\hat{\tau}_n = \tau(\hat{\theta})$ be the MLE of $\tau(\theta)$. Then

$$\hat{\tau}_n \mathrel{\dot{\sim}} N\left(\tau, \frac{\left\{\tau^{/}(\theta)\right\}^2}{I_{\mathbf{X}}(\theta)}\right) \quad \text{or} \quad \sqrt{I_{\mathbf{X}}(\theta)}\left(\hat{\tau}_n - \tau\right) \xrightarrow{D} N\left(0, \left\{\tau^{/}(\theta)\right\}^2\right), \tag{2-12}$$

where $I_n(\theta)$ is the Fisher information in the *n* observations. This means that:

(i) $\hat{\tau}_n$ is asymptotically normal;

(ii) $\hat{\tau}_n$ is asymptotically unbiased;

(iii) $\hat{\tau}_n$ is asymptotically MVBU (i.e. efficient).

If we further assume that $X_1, \ldots, X_n$ are i.i.d. then $I_{\mathbf{X}}(\theta) = nI_{X_i}(\theta)$, where $I_{X_i}(\theta)$ is the Fisher information in any one observation. Also, let the parameter being estimated be θ itself (i.e. let $\tau(\theta) = \theta$). Then Eq. (2-12) can be written as

$$\hat{\theta} \mathrel{\dot{\sim}} N\left(\theta, \frac{1}{nI_{X_1}(\theta)}\right) \quad \text{or} \quad \sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{D} N\left(0, \frac{1}{I_{X_1}(\theta)}\right). \tag{2-13}$$

Therefore, for i.i.d. r.v.'s, any MLE $\hat{\theta}$ is consistent and MVBU asymptotically (the same is true for any function of θ, see Theorem 2.5A

2.8 SUFFICIENCY

Consider a random vector $\mathbf{X} = (X_1, ..., X_n)$ with density $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$. A statistic $\mathbf{T} = \mathbf{T}(\mathbf{X})$ is said to be sufficient for $\boldsymbol{\theta}$ if the conditional density of $\mathbf{X}$ given $\mathbf{T}$ does not depend on the unknown parameter $\boldsymbol{\theta}$, i.e.

$$f(\mathbf{x}|\mathbf{T}) \text{ does not depend on } \boldsymbol{\theta}$$

This means that if $\mathbf{T}$ is sufficient for $\boldsymbol{\theta}$, then $\mathbf{T}$ gives as much information about $\boldsymbol{\theta}$ as does the whole sample $\mathbf{X}$.

Remark: The sample $\mathbf{X} = (X_1, ..., X_n)$ is always sufficient for $\boldsymbol{\theta}$. But we are interested in more condensed statistics.

Example 2.8A

Consider a sequence of i.i.d. Bernoulli r.v.s $Z_1, ..., Z_n$ such that

$$Z_i = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

By using conditional distributions, show that $T = Z_1 + ... + Z_n$ is sufficient for *p*.

Solution.

$T \sim \text{binomial}(n, p)$. Therefore,

$$\begin{aligned} p(z_1, ..., z_n | T) &= \frac{\Pr\{(Z_1 = z_1, ...Z_n = z_n) \cap (T = t)\}}{\Pr\{T = t\}} \\ &= \frac{p^t (1-p)^{n-t} I(t \in \{0, 1, ..., n\})}{\binom{n}{t} p^t (1-p)^{n-t} I(t \in \{0, 1, ..., n\})} \\ &= \binom{n}{t}^{-1}, \end{aligned}$$

which does not depend on *p*. Hence, *T* is sufficient for *p*.

It is often very difficult to obtain conditional distributions. In such cases, the following theorem helps in identifying sufficient statistics.

Theorem 2.8A (Fisher-Neyman Factorization Theorem)

$\mathbf{T} = \mathbf{T}(\mathbf{X})$ is sufficient for $\boldsymbol{\theta}$ iff the likelihood $L_{\mathbf{X}}(\boldsymbol{\theta})$ can be factored as

$$L_{\mathbf{X}}(\boldsymbol{\theta}) = g_{\boldsymbol{\theta}}\left\{\mathbf{T}(\mathbf{X})\right\} h(\mathbf{X}), \qquad (2\text{-}14)$$

where $g_{\boldsymbol{\theta}}$ is a function which depends on $\boldsymbol{\theta}$ and on $\mathbf{X}$ only through $\mathbf{T}$.■

Example 2.8A (Revisited)

Consider a sequence of i.i.d. Bernoulli r.v.s $Z_1, ..., Z_n$ such that

$$Z_i = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

By the factorization theorem, find a sufficient statistic for *p*.

Solution.

$$\begin{aligned} L_{\mathbf{Z}}(p) &= p^{Z_1}(1-p)^{1-Z_1} \times ... \times p^{Z_n}(1-p)^{1-Z_n}\, I\left(Z_1, ..., Z_n \in \{0, 1\}\right) \\ &= \underbrace{I\left(Z_1, ..., Z_n \in \{0, 1\}\right)}_{h(\mathbf{Z})} \underbrace{p^{Z_1+...+Z_n}(1-p)^{n-(Z_1+...+Z_n)}}_{g_p(T)} \end{aligned}$$

By the factorization theorem, $T = \sum_i Z_i$ is sufficient for *p*.

Example 2.8B

Let $X_1, ..., X_n$ be i.i.d. r.v.'s such that $X_i \sim \text{Poisson}(\lambda)$. Obtain a sufficient statistic for λ.

Solution

We have

$$L_{\mathbf{X}}(\lambda) = e^{-\lambda}\frac{\lambda^{x_1}}{X_1!}......e^{-\lambda}\frac{\lambda^{x_n}}{X_n!}I\left(X_1,...,X_n \in \{0, 1, ...\}\right)$$
$$= \underbrace{\frac{I\left(X_1,...,X_n \in \{0, 1, ...\}\right)}{X_1!...X_n!}}_{h(\mathbf{X})}\underbrace{e^{-n\lambda}\lambda^{X_1+...+X_n}}_{g_\lambda(T)}$$

By the factorization theorem, $T = \sum_i X_i$ is sufficient for λ

Proof of Theorem 2.8A (Fisher-Neyman Factorization Theorem)

We will consider the discrete case only. The continuous case is similar.

Sufficiency ⇒ Factorization: Suppose $\mathbf{T} = \mathbf{T}(\mathbf{X})$ is sufficient. Then $\mathbf{X} = \mathbf{x}$ if and only if $\mathbf{X} = \mathbf{x}$ and $\mathbf{T}(\mathbf{X}) = \mathbf{T}(\mathbf{x})$. Therefore, the probability mass function of $\mathbf{X}$ is

$$p(\mathbf{x};\boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}\{\mathbf{X}=\mathbf{x}\} = \Pr_{\boldsymbol{\theta}}\{\mathbf{X}=\mathbf{x}, \mathbf{T}=\mathbf{T}(\mathbf{x})\}$$
$$= \Pr_{\boldsymbol{\theta}}\{\mathbf{X}=\mathbf{x} | \mathbf{T}=\mathbf{T}(\mathbf{x})\}\Pr_{\boldsymbol{\theta}}\{\mathbf{T}=\mathbf{T}(\mathbf{x})\}$$
$$= h(\mathbf{x})g_{\boldsymbol{\theta}}\{\mathbf{T}(\mathbf{x})\}$$

Factorization ⇒ Sufficiency: Assume factorization holds. Then, the conditional mass function of $\mathbf{X}$ given $\mathbf{T}$ is

$$
\begin{aligned}
\Pr_{\boldsymbol{\theta}}\left\{\mathbf{X}=\mathbf{x}|\mathbf{T}=\mathbf{t}\right\} &= \frac{\Pr_{\boldsymbol{\theta}}\left\{\mathbf{X}=\mathbf{x},\mathbf{T}=\mathbf{t}\right\}}{\Pr_{\boldsymbol{\theta}}\left\{\mathbf{T}=\mathbf{t}\right\}} \\
&= \frac{\Pr_{\boldsymbol{\theta}}\left\{\mathbf{X}=\mathbf{x}\right\}}{\Pr_{\boldsymbol{\theta}}\left\{\mathbf{T}=\mathbf{t}\right\}} \\
&= \frac{p\left(\mathbf{x};\theta\right)}{\sum_{\mathbf{z}:\mathbf{T}=\mathbf{t}} p\left(\mathbf{z};\theta\right)} \\
&= \frac{h\left(\mathbf{x}\right) g_{\boldsymbol{\theta}}\left\{\mathbf{T}(\mathbf{x})\right\}}{\sum_{\mathbf{z}\in A_t} h\left(\mathbf{z}\right) g_{\boldsymbol{\theta}}\left\{\mathbf{T}(\mathbf{z})\right\}} \\
&= \frac{h\left(\mathbf{x}\right) g_{\boldsymbol{\theta}}\left\{\mathbf{T}(\mathbf{x})\right\}}{g_{\boldsymbol{\theta}}\left\{\mathbf{T}(\mathbf{x})\right\} \sum_{\mathbf{z}\in A_t} h\left(\mathbf{z}\right)} \\
&= \frac{h\left(\mathbf{x}\right)}{\sum_{\mathbf{z}\in A_t} h\left(\mathbf{z}\right)}
\end{aligned}
$$

which is independent of $\boldsymbol{\theta}$. Hence $\mathbf{T}$ is sufficient for $\boldsymbol{\theta}$ ■

Notice that the exponential family in (2-8) can be written as

$$
L_{\mathrm{X}}\left(\theta\right) = \underbrace{u^{*}\left(\mathbf{X}\right)}_{h(\mathbf{X})} \underbrace{\exp\left\{h\left(\theta\right) T\left(\mathbf{X}\right) + k\left(\theta\right)\right\}}_{g_{\theta}(T)} \tag{2-15}
$$

which is factorizable in the Fisher-Neyman form. <u>Hence exponential families admit sufficient statistics</u>. On the other hand, distributions admitting sufficient statistics do not always belong to the exponential family.

The following theorem concerns the sufficiency of functions of sufficient statistics.

<u>Theorem 2.8B</u>

Let $\mathbf{T} = \mathbf{T}(\mathbf{X})$ be a sufficient statistic for $\boldsymbol{\theta}$, and let $\mathbf{S} = \mathbf{S}(\mathbf{X})$ be an invertible function of $\mathbf{T}$. Then $\mathbf{S}$ is also sufficient for $\boldsymbol{\theta}$.

<u>Proof</u>

Since $\mathbf{S}$ is an invertible function of $\mathbf{T}$, there exists a function R such that $\mathbf{T} = R(\mathbf{S})$. Since $\mathbf{T}$ is sufficient, the factorization theorem yields

$$\begin{aligned} L_{\mathbf{X}}(\boldsymbol{\theta}) &= g_{\boldsymbol{\theta}}\left\{\mathbf{T}(\mathbf{X})\right\} h(\mathbf{X}) \\ &= g_{\boldsymbol{\theta}}\left\{R\left(\mathbf{S}(\mathbf{X})\right)\right\} h(\mathbf{X}) \end{aligned}$$

By the factorization theorem, $\mathbf{S}$ is also sufficient. ■

Example 2.8C

Let $X_1,...,X_n$ be i.i.d. r.v.'s such that $X_i \sim \mathrm{N}(\mu,\sigma^2)$. Obtain sufficient statistics for μ and σ^2.

Solution

Let $\boldsymbol{\theta} = (\mu,\sigma^2)$. Then

$$\begin{aligned} L_{\mathbf{X}}\left(\mu,\sigma^2\right) &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{2\sigma^2}\right\} \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} X_i^2 - 2\mu\sum_{i=1}^{n} X_i + n\mu^2\right)\right\} \\ &= h(\mathbf{X})\, g_{\boldsymbol{\theta}}(\mathbf{T}) \end{aligned}$$

where

$$h(\mathbf{X}) = 1, \qquad g_{\boldsymbol{\theta}}\left(T_1,T_2\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(\overbrace{\sum_{i=1}^{n} X_i^2}^{T_2} - 2\mu \overbrace{\sum_{i=1}^{n} X_i}^{T_1} + n\mu^2\right)\right\}$$

By the factorization theorem, $(T_1,T_2) = (\sum_i X_i, \sum_i X_i^2)$ is jointly sufficient for (μ,σ^2).

Remarks: (i) We have $\bar{X} = T_1/n$ and $S^2 = (T_2 - T_1^2/n)/(n-1)$. By Theorem 2.8B, since $(\bar{X},S^2)$ is an invertible function of (T_1,T_2), $(\bar{X},S^2)$ is also jointly sufficient for (μ,σ^2).

(ii)The fact that $(\bar{X}, S^2)$ is jointly sufficient for (μ, σ^2) <u>does not</u> mean that $\bar{X}$ is sufficient for μ and S^2 is sufficient for σ^2. In fact:

- $\bar{X}$ is sufficient for μ,
- $(\bar{X}, S^2)$ is sufficient for σ^2,
- $(\bar{X}, S^2)$ is sufficient for (μ, σ^2).

Example 2.8D

Let $X_1, ..., X_n$ be i.i.d. r.v.'s such that $X_i \sim \mathrm{N}(\mu, \sigma_0^2)$, where σ_0^2 is known. Obtain a sufficient statistic for μ.

Solution

We have

$$
\begin{aligned}
L_{\mathbf{x}}(\mu) &= \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma_0^2}\right\} \\
&= \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^{n} X_i^2 - 2\mu\sum_{i=1}^{n} X_i + n\mu^2\right)\right\} \\
&= \underbrace{\left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^{n} X_i^2}{2\sigma_0^2}\right)}_{h(\mathbf{X})} \underbrace{\exp\left\{\frac{1}{2\sigma_0^2}\left(-2\mu\sum_{i=1}^{n} X_i + n\mu^2\right)\right\}}_{g_\mu(\mathbf{T})}
\end{aligned}
$$

By the factorization theorem, $T = \sum_i X_i$ is sufficient for μ.

Example 2.8E

Let $X_1, ..., X_n$ be i.i.d. r.v.'s such that $X_i \sim \mathrm{N}(\mu_0, \sigma^2)$, where μ_0 is known. Obtain a sufficient statistic for σ^2.

Solution

We have

$$L_{\mathbf{X}}\left(\sigma^2\right)=\underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}\left(X_i-\mu_0\right)^2}{2\sigma^2}\right\}}_{g_{\sigma^2}(\mathbf{T})}$$

By the factorization theorem, the statistic $T=\sum_i (X_i-\mu_0)^2$ is sufficient for σ^2 (thus the statistic $\sum_i (X_i-\mu_0)^2/n$ is used in this situation to estimate σ^2)

Example 2.8F

Let $X_1,\ldots,X_n$ be i.i.d. r.v.'s from a $\text{uniform}(0,\theta)$ distribution. Obtain a sufficient statistic for θ.

Solution

The p.d.f. of each X_i is $f(x_i)=1/\theta$ for $0<x_i<\theta$, i.e. $f(x_i)=\theta^{-1}I(0<x_i<\theta)$. The likelihood function is

$$\begin{aligned}
L_{\mathbf{X}}(\theta) &= \frac{1}{\theta^n} I\left(0<X_1<\theta\right) I\left(0<X_2<\theta\right)\ldots. I\left(0<X_n<\theta\right) \\
&= \frac{1}{\theta^n} I\left(0<X_{(1)}\right) I\left(X_{(n)}<\theta\right) \\
&= \underbrace{I\left(0<X_{(1)}\right)}_{h(\mathbf{X})} \underbrace{\frac{1}{\theta^n} I\left(X_{(n)}<\theta\right)}_{g_\theta(\mathbf{T})}
\end{aligned}$$

By the factorization theorem, $T=X_{(n)}$ is sufficient.

The following theorem shows why sufficient statistics are very important in statistical inference.

<u>**Theorem 2.8C (Rao-Blackwell)**</u>

Let $\mathbf{T}=\mathbf{T}(\mathbf{X})$ be a sufficient statistic for $\boldsymbol{\theta}$, and let $d(\mathbf{X})$ be an unbiased estimator of $\tau(\boldsymbol{\theta})$. If

$$d^*\left(\mathbf{T}\right)=\mathrm{E}\left\{d\left(\mathbf{X}\right)|\mathbf{T}\right\}, \tag{2-16}$$

then:

(i) $d^*(\mathbf{T})$ is a statistic;

(ii) $d^*(\mathbf{T})$ is an unbiased estimator of $\tau(\boldsymbol{\theta})$;

(iii) $\operatorname{var} d^*(\mathbf{T}) \le \operatorname{var} d(\mathbf{T})$, with equality only if $d(\mathbf{X}) = d^*(\mathbf{T})$. ■

Remarks: (i) The theorem tells us that, given an initial unbiased estimator of a parameter, by conditioning on a sufficient statistic we can find a second unbiased estimator which is as good as or better than the first one. Thus, in our search for UMVUE's, the Rao-Blackwell theorem tells us to look for functions of sufficient statistics.

(ii)The new estimator $d^*(\mathbf{T})$ is called the Rao-Blackwellized estimator.

Example 2.8G

Let $X_1,...,X_n$ be i.i.d. r.v.'s such that $X_i \sim \mathrm{N}(\mu,\sigma_0^2)$, where σ_0^2 is known. We showed in the previous example that $T = \sum_i X_i$ is sufficient for μ. Suppose we now use $d = X_1$ (which is unbiased) as an estimator of μ. Obtain the Rao-Blackwellized estimator.

Solution

We have

$$\begin{aligned} d^* &= \mathrm{E}(X_1|T) = \mathrm{E}\{T - (X_2 + ... + X_n)|T\} \\ &= T - \mathrm{E}\{(X_2 + ... + X_n)|T\} \\ &= T - \mathrm{E}(X_2|T) - ... - \mathrm{E}(X_n|T) \\ &= T - (n-1)d^* \end{aligned}$$

Therefore, $d^* = T_n / n = \bar{X}$ is the Rao-Blackwellized estimator.

Proof of Theorem 2.8C (Rao-Blackwell)

(i) By the definition of sufficiency, $d^*(\mathbf{T})$ does not depend on $\boldsymbol{\theta}$ and is therefore a statistic.

(ii) $\mathrm{E}d^*(\mathbf{T}) = \mathrm{E}\left[\mathrm{E}\{d(\mathbf{X})|\mathbf{T}\}\right] = \mathrm{E}d(\mathbf{X})$

(iii) We have

$$\begin{aligned}\operatorname{var} d(\mathbf{X}) &= \operatorname{var}\mathrm{E}\{d(\mathbf{X})|\mathbf{T}\} + \mathrm{E}\operatorname{var}\{d(\mathbf{X})|\mathbf{T}\}\\ &= \operatorname{var} d^*(\mathbf{T}) + \mathrm{E}\operatorname{var}\{d(\mathbf{X})|\mathbf{T}\}\end{aligned}$$

Now,

$$\operatorname{var}\{d(\mathbf{X})|\mathbf{T}\} \geq 0 \qquad \Rightarrow \qquad \mathrm{E}\operatorname{var}\{d(\mathbf{X})|\mathbf{T}\} \geq 0.$$

Therefore,

$$\operatorname{var} d^*(\mathbf{T}) \leq \operatorname{var} d(\mathbf{X}).$$

For equality to hold, we need

$$\begin{aligned}\operatorname{var}\{d(\mathbf{X})|\mathbf{T}\} &= 0\\ \mathrm{E}\left[\{d(\mathbf{X}) - \mathrm{E}d(\mathbf{X})\}|\mathbf{T}\right]^2 &= 0\\ \{d(\mathbf{X}) - d^*(\mathbf{T})\}^2 &= 0\\ \text{i.e.} \qquad d(\mathbf{X}) &= d^*(\mathbf{T})\end{aligned}$$

■

There is a very interesting connection between sufficient statistics and MLE's, as the following theorem shows.

<u>Theorem 2.8D</u>

Any MLE is a function of a sufficient statistic or a set of jointly sufficient statistics.

<u>Proof</u>

Suppose $\mathbf{T} = (T_1, \ldots, T_k)$ is jointly sufficient for a parameter $\boldsymbol{\theta}$. Then by the factorization theorem (Theorem 2.8A),

$$L_{\mathbf{X}}(\boldsymbol{\theta}) = g_{\boldsymbol{\theta}}(T_1, \ldots, T_k)\, h(\mathbf{X}).$$

As a function of $\boldsymbol{\theta}$, $L_{\mathbf{X}}(\boldsymbol{\theta})$ will attain its maximum at the same place as $g_{\boldsymbol{\theta}}(T_1,...,T_k)$ has its maximum. Now, $g_{\boldsymbol{\theta}}(T_1,...,T_k)$ attains a maximum at values of $\boldsymbol{\theta}$ which depend on X only through $T_1,...,T_k$. Hence MLEs are functions of sufficient statistics. ■

We conclude by referring to the remark we made at the start of this section, namely that the sample $\mathbf{X}=(X_1,...,X_n)$ is always sufficient for $\boldsymbol{\theta}$ and that we are interested in more condensed sufficient statistics. Thus, in Example 2.8C, both of the following are sufficient statistics for (μ,σ^2):

$$\mathbf{T}_1=(X_1,...,X_n) \quad \text{and} \quad \mathbf{T}_2=(\textstyle\sum_i X_i, \sum_i X_i^2)$$

It is therefore natural to ask if further condensation is possible. The most condensed form of a sufficient statistic is called a <u>minimally sufficient statistic.</u> A statistic is said to be minimally sufficient if:

(i) it is sufficient;

(ii) it is a function of every other sufficient statistic.

Using this definition, it is seen that $(\sum_i X_i, \sum_i X_i^2)$ is minimally sufficient for (μ,σ^2) and that no further condensation is possible.

Lehmann and Scheffe have provided a useful method to identify minimally sufficient statistics. Consider a likelihood function $L_{\mathbf{X}}(\boldsymbol{\theta})$. Choose a particular $\mathbf{X}=\mathbf{X}_0$ such that

$$\frac{L_{\mathbf{X}}(\boldsymbol{\theta})}{L_{\mathbf{X}_0}(\boldsymbol{\theta})}=k\left(\mathbf{X},\mathbf{X}_0\right) \tag{2-17}$$

is independent of $\boldsymbol{\theta}$. From the Fisher-Neyman factorization,

$$\frac{L_{\mathbf{X}}(\boldsymbol{\theta})}{L_{\mathbf{X}_0}(\boldsymbol{\theta})}=\frac{g_\theta\left(T\right)}{g_\theta\left(T_0\right)}\cdot\frac{h\left(\mathbf{X}\right)}{h\left(\mathbf{X}_0\right)}, \tag{2-18}$$

where $T=T(\mathbf{X})$ is a sufficient statistic. Now Eq. (**2-18**) shows that Eq. (**2-17**) holds iff $T=T_0$. Eq. (**2-18**) enables the identification of sufficient statistics, and in particular minimal sufficient statistics.

Example 2.8G

Consider the i.i.d. r.v.'s $X_1,..., X_n$, where $X_i \sim N(\mu,\sigma^2)$. Use the Lehmann-Scheffe procedure to obtain a minimal sufficient statistic for (μ,σ^2)

Solution. Eq. (2-17) can be written as

$$\frac{L_{\mathbf{X}}(\mu,\sigma)}{L_{\mathbf{X}_0}(\mu,\sigma)} = \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}\right\}}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_{0i}-\mu)^2}{2\sigma^2}\right\}} = \exp\left\{\frac{\left(\sum_{i=1}^{n}X_{0i}^2-\sum_{i=1}^{n}X_i^2\right)-2\mu\left(\sum_{i=1}^{n}X_{0i}-\sum_{i=1}^{n}X_i\right)}{2\sigma^2}\right\}$$

The above is independent of (μ,σ^2) if, for example, each $X_i = X_{0i}$, i.e. $(X_1,..., X_n)$ is sufficient. The most compact condition of the likelihood ratio to be independent of (μ,σ^2) is for $\sum_i X_{0i}^2 = \sum_i X_i^2$ and $\sum_i X_{0i} = \sum_i X_i$. Hence, a minimally sufficient statistic for (μ,σ^2) is $\left(\sum_i X_i^2, \sum_i X_i\right)$.

In general, the sufficient statistic $T(\mathbf{X})$ from the exponential family (see Eq. (2-15))

$$L_{\mathbf{X}}(\theta) = \underbrace{u^*(\mathbf{X})}_{h(\mathbf{X})}\underbrace{\exp\{h(\theta)T(\mathbf{X})+k(\theta)\}}_{g_\theta(T)}$$

is also minimally sufficient.

2.9 SUFFICIENCY AND COMPLETENESS

N.B. If $g(x)$ is a function of *x*, then the bilateral (or two-sided) Laplace transform of $g(x)$ is

$$G(s) = \int_{x=-\infty}^{\infty} g(x)e^{-sx}dx.$$

A uniqueness property of the bilateral Laplace transform is that $G(s) \equiv 0$ implies $g(x) \equiv 0$.

As useful as the Rao-Blackwell theorem is, it does not address one issue: assuming we have improved an initial unbiased estimator by conditioning on a sufficient statistic, could further improvements be made by further conditioning on other sufficient statistics? That is, when can we know that we have reached the UMVUE? To answer this question, we need the concept of completeness.

Let $\mathbf{X} = (X_1, ..., X_n)$ be a random sample such that each X_i has p.d.f. $f(x_i;\theta)$ and $T = T(\mathbf{X})$ is a statistic. The family of densities of T is defined to be complete if and only if

$$\forall\theta: \quad \mathrm{E}_\theta h(T) \equiv 0 \quad \Rightarrow \quad h(T) \equiv 0 \text{ a.s.} \tag{2-19}$$

The statistic T is said to be complete if and only if its family of densities is complete. Completeness means that the only unbiased estimator of zero is zero itself.

Example 2.9A

Consider the random sequence $Z_1, ..., Z_n$ where each $Z_i \sim \text{Bernoulli}(p)$. Verify if the following two statistics are complete:

(i) $T_1 = Z_1 - Z_2$;

(ii) $T_2 = \sum_{i=1}^n Z_i$.

Solution

(i) $\mathrm{E}T_1 = p - p = 0$ is true for all values of T_1. Therefore, if take $h(T_1) = T_1$ we have

$\mathrm{E}h(T_1) \equiv 0 \nRightarrow h(T_1) \equiv 0$. Hence, T_1 is not complete.

(ii) We have

$$\mathrm{E}h(T_2) = \sum_{u=0}^{n} h(u)\binom{n}{u} p^u (1-p)^{n-u} = (1-p)^n \sum_{u=0}^{n} h(u)\binom{n}{u}\left(\frac{p}{1-p}\right)^u$$

$$\therefore \quad \mathrm{E}h(T_2) \equiv 0 \quad \Rightarrow \quad \sum_{u=0}^{n} h(u)\binom{n}{u}\left(\frac{p}{1-p}\right)^u \equiv 0$$

The above can be viewed as a polynomial in $\alpha = p/(1-p)$. That this polynomial is identically equal to zero means that its coefficients must all be zero, i.e. $h(u) \equiv 0$. Hence, T_2 is complete.

Example 2.9B

Let X be a r.v. from a $N(\mu, \sigma_0^2)$ distribution, where μ ($\mu \neq 0$) is unknown and σ_0^2 is known. Verify if X belongs to a complete family.

Solution

We set $\mathrm{E}h(X) \equiv 0$. Therefore,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma_0\sqrt{2\pi}} h(x) \exp\left\{-\frac{(x-\mu)^2}{2\sigma_0^2}\right\} dx = 0$$

$$\int_{-\infty}^{\infty} h(x) e^{-\frac{(x^2-2\mu x)}{2\sigma_0^2}} dx = 0$$

$$\int_{-\infty}^{\infty} \underbrace{h(x) e^{-\frac{x^2}{2\sigma_0^2}}}_{g(x)} . e^{-\overbrace{\left(-\frac{\mu}{\sigma_0^2}\right)}^{s} x} dx = 0$$

For $\mu \neq 0$, the above is the bilateral Laplace transform of $g(x) = h(x) e^{-x^2/(2\sigma_0^2)}$. By the uniqueness property of the bilateral transform, we have $h(x) e^{-\frac{x^2}{2\sigma_0^2}} = 0$ so that $h(x) = 0$. Hence X belongs to a complete family.

Note that if $\mu = 0$, $\mathrm{E}h(X) \equiv 0$ would imply

$$\int_{-\infty}^{\infty} h(x) e^{-\frac{x^2}{2\sigma_0^2}} dx = 0$$

The above is zero not only when $h(x) = 0$ but also when $h(x) = x^{2n+1}, n = 1, 2, \ldots$ Hence the $N(0, \sigma_0^2)$ family is not complete.

Note that if a sufficient statistic is complete, it is also minimally sufficient. On the other hand, minimally sufficient statistics are not necessarily complete unless they belong to the exponential family (see Eq. (2-8)).

The importance of completeness lies in its ability to establish uniqueness, as the following theorem shows.

Theorem 2.9A (Lehmann-Scheffé)

Suppose T is a complete sufficient statistic. If $g(T)$ is an unbiased estimator of $\tau(\theta)$, then $g(T)$ is both unique and best.

Proof

Suppose $g_1(T)$ another unbiased estimator. Then,

$$
\begin{aligned}
&\mathrm{E}g(T)=\tau(\theta),\\
&\mathrm{E}g_1(T)=\tau(\theta),\\
\Rightarrow\quad &\mathrm{E}\underbrace{\{g(T)-g_1(T)\}}_{h(T)}=0
\end{aligned}
$$

Since the family is complete, $g(T)-g_1(T)=0$ a.s., i.e. $g(T)$ is unique. Furthermore, by the Rao-Blackwell theorem, $g(T)$ is best (i.e. has minimum variance) ■

Example 2.9C

Let $X_1,...,X_n$ be i.i.d. r.v.'s from a $\text{uniform}(0,\theta)$ distribution.

(a) Show that $X_{(n)}$ is sufficient;

(b) Show that $X_{(n)}$ is complete;

(c) Obtain the UMVUE of θ.

Solution

(a) See Example 2.8F

(b) We first need to determine the p.d.f. of $X_{(n)}$:

$$\begin{aligned}
\Pr\left\{X_{(n)} \le u\right\} &= \Pr\left\{X_1 \le u, \ldots, X_n \le u\right\} \\
&= \left\{\frac{1}{\theta}\left[x\right]_0^u\right\}^n \\
&= \frac{u^n}{\theta^n}
\end{aligned}$$

$$\therefore \qquad f(u) = \frac{nu^{n-1}}{\theta^n} \qquad 0 \le u \le \theta$$

We now check if $U = X_{(n)}$ is complete. $\mathrm{E}h(U) \equiv 0$ implies

$$\frac{n}{\theta^n}\int_0^\theta h(u)u^{n-1}du = 0.$$

Differentiating the above w.r.t. *u*, we get from the Fundamental Theorem of Calculus,

$$h(u)u^{n-1} = 0 \quad \Rightarrow \quad h(u) = 0$$

Hence, $U = X_{(n)}$ is complete

(c)From the Lehmann-Scheffé theorem, we need to find a function $g(X_{(n)})$ which is unbiased. Such a function is then the UMVUE. We have

$$
\begin{aligned}
\mathrm{E}X_{(n)} &= \frac{n}{\theta^n}\int_{u=0}^{\theta} u^n du \\
&= \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1} \\
&= \frac{n}{n+1}\theta
\end{aligned}
$$

$$
\therefore \qquad \mathrm{E}\frac{n+1}{n}X_{(n)} = \theta
$$

Hence, the UMVUE of θ is $(n+1)X_{(n)}/n$.

---

3 CONFIDENCE LEVELS AND TESTS

3.1 CONFIDENCE INTERVALS

In the last chapter, we studied <u>point</u> estimators. However, we are often interested in an <u>interval</u> estimator, i.e. an interval $(L(\mathbf{X}), R(\mathbf{X}))$ for which we are fairly confident that

$$
L(\mathbf{X}) \le \tau(\boldsymbol{\theta}) \le R(\mathbf{X}).
$$

The random interval $(L(\mathbf{X}), R(\mathbf{X}))$ is called a $100(1-\alpha)\%$ <u>confidence interval</u> (CI) for $\tau(\boldsymbol{\theta})$ if

$$
\Pr\{L(\mathbf{X}) \le \tau(\boldsymbol{\theta}) \le R(\mathbf{X})\} = 1-\alpha. \qquad (3\text{-}1)
$$

Example 3.1A

Consider the i.i.d. r.v.'s $X_1,\ldots,X_n$, where $X_i \sim N(\mu,\sigma^2)$ and σ is unknown. Obtain a $100(1-\alpha)\%$ CI for $\tau(\mu,\sigma) = \mu$.

Solution

From (1-11), we have

$$
\frac{\sqrt{n}\left(\bar{X}-\mu\right)}{S} \sim t_{n-1}.
$$