## Module - IV

## Unbiased Estimation

The risk function

$$R(\theta, \delta) = E_\theta(L(\theta, \delta(x))), \quad \theta \in \Theta$$

measures the preciseness with which $\delta$ estimates $\psi(\theta)$.

- One would like to choose $\delta$ that minimizes the risk for all values of $\theta \in \Theta$.
- For fixed $\theta_0 \in \Theta$, consider estimator $\delta_0(\underline{x}) = \psi(\theta_0)$, $\forall\ \underline{x}$.

$$R(\theta_0, \delta_0) = E_{\theta_0}(L(\theta_0, \delta_0(x)))$$

$$= E_{\theta_0}(L(\theta_0, \psi(\theta_0))) = 0$$

Thus at parametric point $\theta_0 \in \Theta$ $\delta_0$ has the smallest possible risk 0 and no other estimator can have smaller risk than $\delta_0$ at parametric point $\theta_0$.

- Thus there exist no uniformly best estimator, i.e. no estimator exists which minimizes the risk for all values of $\theta \in \Theta$, except in the trivial case that $\psi(\theta)$ is constant.
- One way to avoid this difficulty is to restrict the class of estimators by ruling out estimators which do not possess certain desirable characteristics.
- One such desirable characteristic is unbiasedness.

**Definition** An estimator $\delta(x)$ is said to be an unbiased estimator of $\psi(\theta)$ if

$$E_\theta(\delta(x)) = \psi(\theta), \quad \forall\ \theta \in \Theta.$$

If there exists an unbiased estimator of $\psi$, the estimand will be called U-estimable.

**Example** (Unbiased estimators $\psi(\theta)$ may not exist)

$X \sim Bin(n, \theta)$, $\theta \in \Theta = (0,1)$, $n$ is a known positive integer

$\psi(\theta) = \frac{1}{\theta}$.

Example (Unbiased estimate may be absurd)

$X \sim Po(\theta)$, $\theta \in \Theta = (0, \infty)$, $\psi(\theta) = e^{-3\theta}$

The estimator $\delta(X) = (-2)^X$ is unbiased. However it takes both positive and negative values, whereas $\psi(\theta)$ is always positive.

**Remark**: After a best unbiased estimator (an unbiased estimator having uniformly smallest risk among all unbiased estimators) has been found, its performance should be studied to find out if there exists a slightly biased estimator with uniformly much smaller risk function. Such a possibility exists and should not be ruled out.

Define $\mathcal{D}$ = class of all estimators of $\psi(\theta)$

$\mathcal{D}_0$ = the class of all unbiased estimators of $\psi(\theta)$

$= \{\delta \in \mathcal{D} : E_\theta(\delta(\underline{X})) = \psi(\theta), \ \forall \theta \in \Theta\}$

$\mathcal{U}$ = the class of all unbiased estimators of 0

$= \{U \in \mathcal{D} : E_\theta(U(X)) = 0, \ \forall \theta \in \Theta\}$

**Lemma** Let $\delta_0 \in \mathcal{D}_0$ be fixed. Then

$\mathcal{D}_0 = \{\delta \in \mathcal{D} : \delta(x) = \delta_0(x) - U(x), \ U \in \mathcal{U}\}$

**Proof** Trivial

# UMVU Estimators

Consider the squared error loss (SEL) function

$$L(\theta, d) = (\psi(\theta) - d)^2, \quad d \in \Theta^*, \ \theta \in \Theta$$

Then, for any $\delta \in \mathcal{D}_0$

$$R(\theta, \delta) = E_\theta((\delta(X) - \psi(\theta))^2)$$

$$= E_\theta((\delta(X) - E_\theta(\delta(X)))^2) \quad (\text{since } \delta \in \mathcal{D}_0)$$

$$= V_\theta(\delta), \quad \forall\, \theta \in \Theta,$$

where $V_\theta(\delta)$ denotes the variance of $\delta$.

Thus finding an estimator $\delta^*$ in $\mathcal{D}_0$ that has smallest risk among all estimators in $\mathcal{D}_0$ is equivalent to finding an estimator in $\mathcal{D}_0$ that has the smallest variance.

<u>Note</u>: here we may consider only non-randomized estimators as the SEL function is strictly convex.

<u>Definition</u>. (a) An estimator $\delta^* \in \mathcal{D}_0$ is said to be locally minimum variance unbiased (LMVU) estimator at $\theta_0 \in \Theta$ if for any $\delta \in \mathcal{D}_0$

$$V_{\theta_0}(\delta^*) \le V_{\theta_0}(\delta_0)$$

(b) An estimator $\delta^* \in \mathcal{D}_0$ is said to be uniformly minimum variance unbiased (UMVU) estimator if for any $\delta \in \mathcal{D}_0$

$$V_\theta(\delta^*) \le V_\theta(\delta), \quad \forall\, \theta \in \Theta$$

Let $\delta_0 \in \mathcal{D}_0$ be fixed. Then for any other $\delta \in \mathcal{D}_0$

$$\delta(X) = \delta_0(X) - U(X),$$

for some $U \in \mathcal{U}$, and

$$V_\theta(\delta) = V_\theta(\delta_0 - U) = E_\theta((\delta_0(X) - U(X) - \psi(\theta))^2)$$

$$= E_\theta((\delta_0(X) - U(X))^2) - (\psi(\theta))^2$$

Thus to find LMVU estimator at $\theta_0$ it is enough to find a $U^* \in \mathcal{U}$ such that

$$E_{\theta_0}(\delta_0(X) - U^*(X))^2 \le E_{\theta_0}((\delta_0(X) - U(X))^2), \ \forall\, U \in \mathcal{U}.$$

<u>Example</u> Let $X$ have p.m.f.

$$f_\theta(k) = \begin{cases} \theta, & \text{if } k=-1 \\ (1-\theta)^2\theta^k, & k=0,1,2,\dots \\ 0, & \text{otherwise} \end{cases}, \quad \theta \in \Theta = (0,1).$$

Let $\psi_1(\theta) = \theta$ and $\psi_2(\theta) = \theta^2$. Find UMVU estimators of $\psi_1(\theta)$ and $\psi_2(\theta)$, provided they exist.

<u>Solution</u> It can be seen that $U \in \mathcal{U}_0$ iff

$$U(k) = -k\,U(-1), \quad k=0,1,2,\dots$$

$$\Leftrightarrow \quad U(k) = ak, \quad k=-1,0,1,\dots,$$

for some $a \in \mathbb{R}$.

(I) Estimation of $\psi_1(\theta) = \theta$

It can be seen that $\delta_0(X) = \begin{cases} 1, & \text{if } X=-1 \\ 0, & \text{otherwise} \end{cases} \in \mathcal{D}_0.$

Thus $\delta \in \mathcal{D}_0$ iff

$$\delta(X) = \delta_0(X) - U(X) = \delta_0(X) - aX,$$

for some $a \in \mathbb{R}$.

We desire to minimise, for fixed $\theta_0 \in \Theta$,

$$h_1(a) = E_{\theta_0}((\delta_0(X) - aX)^2)$$

$$= \left[\theta_0 + (1-\theta_0)^2 \sum_{k=1}^{\infty} k^2\theta_0^k\right] a^2 + 2a\theta_0 + \theta_0.$$

Clearly

$$a^* = \frac{-\theta_0}{\theta_0 + (1-\theta_0)^2 \sum_{k=1}^{\infty} k^2 \theta_0^k}$$

minimizes $h_1(a)$ and is the unique minimizer ($h_1''(a^*) > 0$).

Thus $\delta^*(x) = \delta_0 - a^* x$ is the LMVUE of $\psi(\theta)$ at $\theta = \theta_0$.
Since $a^*$ depends on $\theta = \theta_0$ and $a^*$ is the unique minimizer
we conclude that no UMVUE of $\psi_1(\theta)$ exists.

(II) Estimation of $\psi_2(\theta) = (1-\theta)^2$

Let

$$\delta_p(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Then $\delta_1 \in \mathcal{D}_0$. Thus $\delta \in \mathcal{D}_0$ iff

$$\delta(x) = \delta_1(x) - U(x) = \delta_1(x) - ax,$$

for some $a \in \mathbb{R}$.
Consider, for fixed $\theta_0 \in \Theta$,

$$h_2(a) = E_{\theta_0}((\delta_1(x) - ax)^2)$$

$$= [\theta_0 + (1-\theta_0)^2 \sum_{k=1}^{\infty} k^2 \theta_0^k] a^2 + (1-\theta_0)^2.$$

Clearly $b^* = 0$ minimizes $h_2(a)$ and it is the unique minimizer. Moreover $b^*$ does not depend on $\theta_0$.
It follows that $\delta_1(x)$ is the UMVUE of $\psi_2(\theta) = (1-\theta)^2$.

<u>Remark</u> Even restricting to unbiased estimators may not ensure existence of the best estimator

Define

$$\Delta = \{\delta \in \mathcal{D}_0 : E_\theta(\delta^2(x)) < \infty\}.$$

For finding the UMVUE it is enough to restrict to class $\Delta$ of estimators.

**Theorem** Let $\delta \in \Delta$. Then $\delta$ is UMVUE for $\psi(\theta) = E_\theta(\delta(X))$ iff, for every $U \in \mathcal{U}$ with $E_\theta(U^2(X)) < \infty$,

$$Cov_\theta(\delta, U) = E_\theta(\delta U) = 0 \quad \forall\, \theta \in \Theta.$$

**Proof.** First suppose that $\delta$ is UMVUE and $U \in \mathcal{U}$ be s.t. $E_\theta(U^2(X)) < \infty$ $\forall\, \theta \in \Theta$. Consider class of estimators

$$\delta_\lambda(X) = \delta(X) + \lambda U(X),$$

so that each $\delta_\lambda$, $\lambda \in \mathbb{R}$, belongs to $\mathcal{D}_0$.

Also $\delta$ is UMVUE implies that for every $\theta \in \Theta$

$$V_\theta(\delta) \le V_\theta(\delta_\lambda), \quad \forall\, \lambda \in \mathbb{R}$$

$$\Leftrightarrow \quad \lambda^2 V_\theta(U) + 2\lambda\, Cov_\theta(\delta, U) \ge 0 \quad \forall\, \lambda \in \mathbb{R}$$

$$\Rightarrow \quad Cov_\theta(\delta, U) = 0. \qquad (\text{Why ?!})$$

Conversely suppose that $\delta \in \mathcal{D}_0$ ~~be~~ is such that for any $U \in \mathcal{U}$ with $E_\theta(U^2) < \infty$

$$Cov_\theta(\delta, U) = E_\theta(\delta U) = 0 \quad \forall\, \theta \in \Theta$$

Let $\delta_0 \in \Delta$ be any other estimator in $\mathcal{D}_0$. Define

$$U(X) = \delta(X) - \delta_0(X),$$

so that $U \in \mathcal{U}$ and $E_\theta(U^2(X)) < \infty$ $\forall\, \theta \in \Theta$.

Then, $\forall\, \theta \in \Theta$

$$\begin{aligned} 0 &= E_\theta(\delta(X) U(X)) \\ &= Cov_\theta(\delta(X), U(X)) \\ &= Cov_\theta(\delta(X), \delta(X) - \delta_0(X)) \\ &= V_\theta(\delta(X)) - Cov_\theta(\delta(X), \delta_0(X)) \end{aligned}$$

$$\Rightarrow \quad V_\theta(\delta) = Cov_\theta(\delta, \delta_0) \le \sqrt{V_\theta(\delta)\, V_\theta(\delta_0)}$$

$$\Rightarrow \quad V_\theta(\delta) \le V_\theta(\delta_0), \quad \forall\, \theta \in \Theta.$$

**Example** Let $X$ be a r.v. with p.m.f.

$$f_\theta(k) = \begin{cases} \theta, & \text{if } k = -1 \\ (1-\theta)^2 \theta^k, & \text{if } k = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Determine the class of all U-estimable estimands for which UMVUE exist.

**Solution** We know that $\delta$ is UMVUE for $\psi(\theta) = E_\theta(\delta(X))$ iff for every $U \in \mathcal{U}$ with $E_\theta(U^2(X)) < \infty$,

$$E_\theta(\delta U) = 0 \quad \forall\ \theta \in \textcircled{H} \quad \ldots \quad (*)$$

But $U \in \mathcal{U}$ iff

$$U(k) = ak, \quad k = -1, 0, 1, \ldots,$$

for some $a \in \mathbb{R}$.

Thus $(*)$ reduces to

$$E_\theta(X\,\delta(X)) = 0, \quad \forall\ \theta \in \textcircled{H}$$

$$\Leftrightarrow \quad -\delta(-1)\theta + (1-\theta)^2 \sum_{k\ge 1}^{\infty} k\,\delta(k)\,\theta^k = 0, \quad \forall\ 0<\theta<1$$

$$\Rightarrow \quad \delta(k) = \delta(-1), \quad k = 1, 2, 3, \ldots,$$

with $\delta(0)$ being arbitrary. Then let $\delta(0) = b$ and $\delta(-1) = a$.

$$E_\theta(\delta(X)) = a\theta + b(1-\theta)^2 + a \sum_{k\ge 1}^{\infty} (1-\theta)^2 \theta^k$$

$$= b(1-\theta)^2 + a(1-(1-\theta)^2)$$

Thus $\psi(\theta)$ possesses UMVUE with finite variance iff $\psi(\theta) = a + c(1-\theta)^2$, for some $a, c \in \mathbb{R}$.

**Note:** The above example illustrates that not all non-constant U-estimable functions may have UMVUE.

**Example** (No non-constant U-estimable function has UMVUE)

Let $X_1, \dots, X_n$ be i.i.d. r.v.s with

$$P_\theta(X_1 = k) = \begin{cases} \frac{1}{3}, & \text{if } k \in \{\theta-1, \theta, \theta+1\} \\ 0, & \text{otherwise} \end{cases},$$

where $\theta \in \Theta = \{0, \pm 1, \pm 2, \dots\}$. Show that no non-constant estimator is UMVUE.

**Solution** Consider the case $n=1$.

$$u \in \mathcal{U} \iff E_\theta(u(X_1)) = 0, \quad \forall \theta \in \Theta$$

$$\iff u(\theta-1) + u(\theta) + u(\theta+1) = 0 \quad \forall \theta \in \Theta$$

$$\Rightarrow u(\theta-1) + u(\theta) + u(\theta+1) = 0 \quad \forall \theta \in I$$
$$\& \quad u(\theta) + u(\theta+1) + u(\theta+2) = 0$$

$$\Rightarrow u(\theta-1) = u(\theta+2) \quad \forall \theta \in \Theta \qquad \iff \qquad \left.\begin{array}{l} u(\theta) = u(\theta+3) \\ u(\theta)+u(\theta+1)+u(\theta+2)=0 \end{array}\right\} \forall \theta \in$$
$$u(\theta) + u(\theta+1) + u(\theta+2) = 0 \quad \forall \theta \in I$$

Thus $u \in \mathcal{U}$ iff

$$u(0) = a, \quad u(1) = b, \quad u(2) = -(a+b), \quad u(\theta) = u(\theta+3), \quad \theta \in \Theta$$

i.e.

$$\mathcal{U} = \{U(\cdot) : u(0) = a, \ u(1) = b, \ U(2) = -(a+b), \ u(\theta) = u(\theta+3), \ \forall \theta \in \Theta, \text{ for some } a, b \in \mathbb{R}\}$$

$\delta$ is UMVUE of $E_\theta(\delta(X_1))$ iff

$$E_\theta(\delta(X_1) U(X_1)) = 0, \quad \forall \theta \in \Theta \quad \forall u \in \mathcal{U}$$

$$\delta(\theta) u(\theta) + \delta(\theta-1) u(\theta-1) + \delta(\theta+1) u(\theta+1) = 0, \quad \forall \theta \in \Theta$$
$$\iff \delta(\theta) u(\theta) + \delta(\theta+1) u(\theta+1) + \delta(\theta+2) u(\theta+2) = 0 \quad \forall u \in \mathcal{U}$$

$$\Rightarrow a\delta(0) + b\delta(1) - (a+b)\delta(2) = 0, \quad \forall a, b \in \mathbb{R}$$

$$\Rightarrow a(\delta(0) - \delta(2)) + b(\delta(1) - \delta(2)) = 0 \quad \forall a, b \in \mathbb{R}$$

$$\Rightarrow \delta(0) = \delta(2), \ \delta(1) = \delta(2) \Rightarrow \delta(0) = \delta(1) = \delta(2)$$

We also have (as for $U(\cdot) \in \mathcal{U}$)

$$\delta(\theta) U(\theta) = \delta(\theta+3) u(\theta+3), \quad \forall \theta \in \Theta \quad u \in \mathcal{U}$$

$$\Rightarrow \quad \delta(0)U(0) = \delta(3)U(3), \qquad \forall\, U \in \mathcal{U}$$

$$\Rightarrow \quad a\,\delta(0) = a\,\delta(3), \qquad \forall\, a \in \mathbb{R}$$

$$\Rightarrow \quad \delta(0) = \delta(3).$$

By induction it follows that

$$\delta(0) = \delta(\pm 1) = \delta(\pm 2) = \cdots = c, \text{ say, for some } c \in \mathbb{R}. \quad \square$$

Now we prove the following theorem which describes situations where every U-estimable function has the UMVUE.

<u>Theorem (Lehmann-Scheffe Theorem)</u> Let $\psi(\theta)$ be an U-estimable function and let $T(\underline{X})$ be a complete-sufficient statistic for $\theta$. Then $\psi(\theta)$ has one and only one unbiased estimator that is a function of $T(\underline{X})$, which is also the unique UMVUE of $\psi(\theta)$.

<u>Proof</u> Since $\psi(\theta)$ is U-estimable there exists an estimator (in fact non-randomized) such that (How?!)

$$E_\theta(\delta(X)) = \psi(\theta), \quad \forall\, \theta \in \Theta$$

Let $\delta_0$ be any unbiased estimator of $\psi(\theta)$. Define

$$\eta(t) = E_\theta(\delta_0(x) \mid T = t).$$

By virtue of sufficiency of $T$ $\eta(t)$ does not depend on $\theta$ and therefore it is a proper estimator.

Also

$$E_\theta(\eta(T)) = E_\theta(E_\theta(\delta_0(\underline{X}) \mid T))$$

$$= E_\theta(\delta_0(\underline{X})) = \psi(\theta), \quad \forall\, \theta \in \Theta$$

and by Rao-Blackwell Theorem (since the SEL is strictly convex)

$$V_\theta(\delta_0) = E_\theta((\delta_0(x) - \psi(\theta))^2)$$

$$> E_\theta((\eta(T) - \psi(\theta))^2)$$

unless $\delta_0(\underline{X}) = \eta(T)$ w.p. 1.

If $\delta_1$ and $\delta_2$ are two-unbiased estimators of $\psi(\theta)$ which are functions of $T$, then
$$E_\theta(\delta_1(T)-\delta_2(T)) = 0, \quad \forall \theta\in\Theta$$
$$\Rightarrow \delta_1(T)=\delta_2(T), \quad \text{w.p.1},$$
by virtue of completeness of $T$. ▨

The above theorem deals with squared error loss function. However the Rao-Blackwell Theorem applies to any convex loss function and therefore we have the following result.

<u>Theorem</u> Suppose that the loss function $L(\theta, d)$ is a strictly convex function of $d$, for every $\theta\in\Theta$. Let $T(\underline{X})$ be a complete-sufficient statistic for $\theta$.

(a) For every U-estimable function $\psi(\theta)$ there exists an uniformly minimum risk unbiased estimator based on $T(\underline{X})$ (UMRUE: an unbiased estimator that uniformly minimizes the risk). In particular this estimator is UMVUE.

(b) The UMRUE of (i) is the unique unbiased estimator which is a function of $T$ and it is the unique UMRUE provided the risk is finite.

<u>Remark</u> (a) If $\delta$ is a function of complete-sufficient statistic $T(\underline{X})$ then $\delta$ is UMVUE (or UMRUE) of $\psi(\theta)=E_\theta(\delta(T))$ (provided the loss function $L(\theta, d)$ is convex in $d$).

(b) Note that $\delta(X)\equiv 0 \in \mathcal{U}$ and thus $\mathcal{U}\neq\phi$. Note that $U\in\mathcal{U}$ is nothing more than a random noise (as it has, in some sense, no information about $\theta$). It makes sense to estimate 0 by 0 and not with random noise. Therefore if an estimator could be improved by adding random noise to it, the estimator is probably defective (this justifies the condition $Cov(\delta, U)=0$ $\forall$ $U\in\mathcal{U}$ for $\delta$ to be UMVUE of its expectation)

(c) If $\mathcal{U}=\{0\}$, i.e. $\mathcal{U}$ 0 is the only unbiased estimator of 0. Then finding the UMVUE is equivalent to finding an unbiased estimator of $\psi(\theta)$. Recall that property of completeness guarantees such a situation.

$\boxed{10/4}$

# Methods of Finding a UMVUE

(i) If $T(\underline{X})$ is a complete-sufficient statistic, then the UMVUE of any U-estimable function $\psi(\theta)$ is uniquely determined by the set of equations

$$E_\theta(\delta(T)) = \psi(\theta) \quad \forall\ \theta \in \Theta$$

(ii) If $\delta(\underline{X})$ is any unbiased estimator of $\psi(\theta)$ and $T \equiv T(\underline{X})$ is a complete-sufficient statistic then

$$\eta(T) = E_\theta(\delta(\underline{X}) \mid T)$$

is the UMVUE of $\psi(\theta)$.

**Example** Let $X_1, \dots, X_n$ $(n \geq 2)$ be i.i.d. $N(\mu, \sigma^2)$ rvs, where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$. Find the UMVUEs of $\mu$ and $\sigma^2$. Find the UMVUE of $\psi(\underline{\theta}) = \mu + \sigma$

**Example** Let $X_1, \dots, X_n$ be i.i.d. $Bin(1, \theta)$ rvs, where $\theta \in \Theta = (0, 1)$ is an unknown parameter. Find the UMVUE of $\psi(\theta) = V_\theta(X_1)$ using method (i) above.

**Example** Let $X_1, \dots, X_n$ be i.i.d. $U(0, \theta)$ rvs where $\theta \in \Theta = (0, \infty)$ is an unknown parameter. Using method (ii) above, find the UMVUE of $\psi(\theta) = \theta$.

**Remark** For a quite general class of loss functions Basu (1955) proved that no LMVUE exists. Similarly for bounded loss functions no LMVUE exist except in trivial cases. (see the next example)

**Example** Let the loss function $L(\theta, d)$ be bounded (say) $0 \le L(\theta, d) \le M$, $\forall \theta, d$). Also assume that $L(\theta, \psi(\theta)) = 0$ $\forall \theta \in \Theta$. Suppose that $\psi(\theta)$ is U-estimable and $\theta_0$ be an arbitrary point in $\Theta$. Consider estimation of $\psi(\theta)$ based on a single observation $X$. For any unbiased estimator $\delta(\underline{x})$ of $\psi(\theta)$ (such a $\delta$ exists as $\psi(\theta)$ is U-estimable), define

$$\delta_n(x) = \begin{cases} \psi(\theta_0) & \text{w.p. } 1-\frac{1}{n} \\ n[\delta(x) - \psi(\theta_0)] + \psi(\theta_0) & \text{w.p. } \frac{1}{n} \end{cases}, \quad n \ge 1, 2, \ldots$$

Show that $E_\theta(\delta_n(x)) = \psi(\theta)$, $\forall \theta \in \Theta$ and

$$0 \le R(\theta_0, \delta_n) \le \frac{M}{n} \to 0, \text{ as } n \to \infty$$

$$\left(\text{i.e. } \lim_{n\to\infty} R(\theta_0, \delta_n) = 0\right)$$

**Exercise** Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$, and $\sigma > 0$

(a) Find the UMVUE of $\mu$ when $\sigma$ is known;

(b) Find the UMVUE of $\sigma^r$ (for integer $r > -n$) when $\mu$ is known;

(c) Find the UMVUE of $\sigma^r$ (for integer $r > -n+1$) when $\mu$ is unknown;

(d) Find the UMVUE $\xi_p \equiv \xi_p(\underline{\theta})$, where

$$P_\theta(X_1 \le \xi_p(\theta)) = p,$$

and $p \in (0,1)$ is given.

(e) For a fixed $u_0 \in \mathbb{R}$ and taking $\sigma = 1$, find the UMVUE of

$$\psi(\theta) = P_\theta(X_1 \le u)$$

(f) For a fixed $u_0 \in \mathbb{R}$ and taking $\sigma = 1$, find the UMVUE of $\phi(u_0 - \theta)$, the pdf of $X_1$ at point $u_0$.

(g) Same as (e) and (f) when $\sigma$ is unknown.

$\boxed{12/4}$

* ←

**Example** Let $X_1, \dots, X_n$ be a random sample from the exponential distribution $E(\mu, \sigma)$ having the p.d.f.

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}}, \quad x > \mu,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$.

(a) Find the UMVUE of $\mu$ when $\sigma$ is known;

(b) Find the UMVUE of $\sigma$ when $\mu$ is known;

(c) Find the UMVUE of $\mu$ when $\sigma$ is unknown;

(d) Find the UMVUE of $\sigma$ when $\mu$ is unknown.

## Power Series Distribution

**Definition** A r.v. $X$ is said to have power series distribution if it has a p.m.f. belonging to the family $\mathcal{P} = \{ p_\theta : \theta \in \Theta \}$ where $\Theta$ is an interval in $(0, \infty)$ and

$$p_\theta(x) = P_\theta(X = x) = \frac{a(x)\, \theta^x}{c(\theta)}, \quad x = 0, 1, 2, \dots \qquad \dots\dots (*)$$

for some function $a(x)$ for which $\sum_{x=0}^{\infty} a(x)\, \theta^x < \infty$.

**Example** (i) For fixed positive integer $n$, let $X \sim Bin(n, p)$, $p \in (0, 1)$. Let $\theta = p/1-p$, $c(\theta) = (\theta+1)^n$, $a(x) = \begin{cases} \binom{n}{x}, & x = 0, 1, \dots, n \\ 0, & \text{otherwise} \end{cases}$

Then $\mathcal{P} = \{ p_\theta : \theta \in (0, \infty) \}$.

(ii) Let $X \sim NB(m, p)$ (# of failures preceding the m-th success) where $m$ is a fixed positive integer and $p \in (0, 1)$. Let $\theta = 1-p$, $c(\theta) = (1-\theta)^{-m}$,

$$a(x) = \binom{m+x-1}{m-1} = \binom{m+x-1}{x}, \quad x = 0, 1, 2, \dots$$

(iii) Let $X \sim Poisson(\theta)$, $\theta > 0$. Let $c(\theta) = e^{\theta}$, and $a(x) = \frac{1}{x!}$, $x = 0, 1, 2, \dots$

Exercise: (i) Show that the family (*) is Complete.

(ii) If $a(x) > 0$ $\forall x = 0, 1, 2, \ldots$, then, for any positive integer $r$, $\theta^r$ is U-estimable with its unique unbiased estimator given by

$$\delta(x) = \begin{cases} 0 & \text{if } x = 0, 1, \ldots, r-1 \\ \frac{a(x-r)}{a(x)}, & x = r, r+1, \ldots \end{cases}$$

Exercise Let $X_1, \ldots, X_n$ be a random sample from power series family (*).

(a) Show that the distribution of $T = \sum_{i=1}^{n} X_i$ is the power series family

$$f_\theta(t) = P_\theta(T = t) = \frac{A(t, n)\,\theta^t}{[c(\theta)]^n}, \quad t = 0, 1, 2, \ldots$$

Where $A(t, n)$ is the coefficient of $\theta^t$ in the power series expansion of $[c(\theta)]^n$.

(b) Show that $T = \sum_{i=1}^{n} X_i$ is a complete and sufficient statistic.

(c) For any positive integer $r$, the UMVUE of $\psi(\theta) = \theta^r$ is

$$\delta_1(T) = \begin{cases} 0 & \text{if } T = 0, 1, \ldots, r-1 \\ \frac{A(T-r, n)}{A(T, n)}, & \text{if } T = r, r+1, \ldots \end{cases}$$

(d) The UMVUE of $p_\theta(x) = P_\theta(X_1 = x)$ is given by

$$\delta_2(T) = \frac{a(x)\,A(T-x, n-1)}{A(t, n)}; \quad T = x, x+1, \ldots$$

here $n > 1$ and $x \in \{0, 1, \ldots\}$ is fixed.

Exercise Let $\underline{X} = (X_1, \dots, X_k) \sim \text{Mult}(n, \theta_1, \dots, \theta_k)$, where $0 < \theta_i < 1$, $i = 1, \dots, k$, and $\sum_{i=1}^{k} \theta_i = 1$. Find the UMVUE of: (i) $\theta_i$, $i = 1, \dots, k$, (ii) $\theta_i \theta_j$, $i \neq j$, $i, j = 1, \dots, k$.

## Variance Inequalities

Let

$V_L(\theta_0)$: Variance of LMVUE at $\theta_0$.

When UMVUE does not exist, $V_L(\theta)$ can be used as a benchmark for evaluating the performance of any unbiased estimator.

Unfortunately $V_L(\theta)$ is generally difficult to evaluate. It may be useful to derive some lower bounds on $V_L(\theta)$, which are simple to evaluate.

Recall that for any function $\psi(x, \theta)$ ~~[illegible]~~

$$V_\theta(\delta) = \text{Var}_\theta(\delta) \geq \frac{[\text{Cov}_\theta(\delta, \psi)]^2}{\text{Var}(\psi)}, \quad \cdots \text{(A)}$$

provided the moment exist.

The above inequality is not useful for our purposes as the rhs also involves $\delta$.

However when $\text{Cov}_\theta(\delta, \psi)$ depends on $\delta$ only through $g(\theta) = E_\theta(\delta)$, (A) does provide a lower bound for the variance of all unbiased estimators of $g(\theta)$.

**Lemma** A necessary and sufficient condition for $Cov_\theta(\delta, \psi)$ to depend on $\delta$ only through $g(\theta) = E_\theta(\delta)$ is that for every $\theta$

$$Cov_\theta(\psi, U) = 0, \quad \forall\, U \in \mathcal{U},$$

where $\mathcal{U}$ is the class of all statistics such that $E_\theta(\cdot) = 0$, $\forall\, \theta$.

**Proof.** Suppose that

$$Cov_\theta(\psi, U) = 0, \quad \forall\, U \in \mathcal{U}, \ \forall \theta.$$

Let $\delta_1$ and $\delta_2$ be two estimators with $E_\theta(\delta_1) = E_\theta(\delta_2) = g(\theta)$. Then $\delta_1 - \delta_2 \in \mathcal{U}$ and thus

$$Cov_\theta(\psi, \delta_1 - \delta_2) = 0$$

$$\Rightarrow Cov_\theta(\psi, \delta_1) = Cov_\theta(\psi, \delta_2).$$

Conversely suppose that, for every $\theta$, $Cov_\theta(\delta, \psi)$ depends on $\delta$ only through $g(\theta) = E_\theta(\delta)$. Let $U \in \mathcal{U}$. Then $E_\theta(\delta) = E_\theta(\delta + U) = g(\theta)$ and thus

$$Cov_\theta(\delta, \psi) = Cov_\theta(\delta + U, \psi)$$

$$\Rightarrow Cov_\theta(U, \psi) = 0, \quad \forall \theta.$$

**Theorem** (Hammersley–Chapman–Robbins Inequality). Suppose that $X \overset{pdf/pmf}{\sim} f_\theta(\cdot)$ where $\theta \in \Theta$ is unknown. Suppose that $A = \{x : f_\theta(x) > 0\}$ is independent of $\theta \in \Theta$. For any fixed $\theta \in \Theta$, define $B_\theta = \{\Delta \in \Theta : \theta + \Delta \in \Theta \text{ and } g(\theta) \neq g(\theta + \Delta)\}$. Let $\delta$ be such that $E_\theta(\delta) = g(\theta)$, $\forall\, \theta \in \Theta$. Then

$$Var_\theta(\delta) \geq \sup_{\Delta \in B_\theta} \left[ \frac{(g(\theta+\Delta) - g(\theta))^2}{E_\theta\left[\left(\frac{f_{\theta+\Delta}(X)}{f_\theta(X)} - 1\right)^2\right]} \right], \quad \theta \in \Theta.$$

Proof. Fix $\theta \in \Theta$ and $\Delta$ such that $\theta + \Delta \in \Theta$. Define

$$\psi_\Delta(x;\theta) = \frac{f_{\theta+\Delta}(x)}{f_\theta(x)} - 1, \quad x \in A.$$

Then

$$E_\theta(\psi_\Delta(x,\theta)) = \int \left[\frac{f_{\theta+\Delta}(x)}{f_\theta(x)} - 1\right] f_\theta(x)\,dx$$

$$= \int f_{\theta+\Delta}(x)\,dx - \int f_\theta(x)\,dx = 1 - 1 = 0.$$

and for any $U \in \mathcal{U}$

$$Cov_\theta(\psi_\Delta(x,\theta), U(x)) = E_\theta[\psi_\Delta(x,\theta)\, U(x)]$$

$$= \int \left[\frac{f_{\theta+\Delta}(x)}{f_\theta(x)} - 1\right] U(x) f_\theta(x)\,dx$$

$$= \int U(x) f_{\theta+\Delta}(x)\,dx - \int U(x) f_\theta(x)\,dx$$

$$= E_{\theta+\Delta}(U(x)) - E_\theta(U(x))$$

$$= 0 - 0 = 0.$$

Thus $\psi_\Delta$ satisfies the assumptions of last lemma.

Note that

$$Cov_\theta(\delta(x), \psi_\Delta(x,\theta)) = E_\theta[\delta(x)\psi_\Delta(x,\theta)]$$

$$= \int \left[\frac{f_{\theta+\Delta}(x)}{f_\theta(x)} - 1\right] \delta(x) f_\theta(x)\,dx$$

$$= g(\theta+\Delta) - g(\theta).$$

Thus

$$Var_\theta(\delta) \geq \frac{[Cov_\theta(\delta,\psi)]^2}{Var_\theta(\psi)} = \frac{(g(\theta+\Delta) - g(\theta))^2}{E_\theta\left[\left(\frac{f_{\theta+\Delta}(x)}{f_\theta(x)} - 1\right)^2\right]}$$

17/4

$$\Rightarrow \quad Var_\theta(\delta) \geq \sup_{\Delta \in B_\theta} \left[ \frac{(g(\theta+\Delta) - g(\theta))^2}{E_\theta\left[\left(\frac{f_{\theta+\Delta}(x)}{f_\theta(x)} - 1\right)^2\right]} \right]$$

Regularity Conditions $R_1$:

(a) $\Theta$ is an open interval (finite, infinite or semi-infinite).

(b) The distributions $\{f_\theta : \theta \in \Theta\}$ have the common support so that, WLOG, the set $A = \{x : f_\theta(x) > 0\}$ is independent of $\theta$.

(c) For any $x \in A$ and any $\theta \in \Theta$, the derivative $f'_\theta(x) = \frac{\partial}{\partial\theta} f_\theta(x)$ exists, is finite and $\int f_\theta(x)\,dx$ can be differentiated under the integral sign.

Theorem (Rao-Cramer Inequality) Let $\delta$ be an estimator with $E_\theta(\delta) = g(\theta)$, $\forall \theta \in \Theta$. Then, under regularity Conditions $R_1$,

$$V_\theta(\delta) \geq \frac{[g'(\theta)]^2}{V_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta(x)\right)}, \quad \theta \in \Theta,$$

provided for any $U \in \mathcal{U}$

$$\int U(x) f_\theta(x)\,dx$$

and $$\int \delta(x) f_\theta(x)\,dx = g(\theta)$$

can be differentiated under the integral sign.

Proof From Hammersley-Chapman-Robbins inequality we have, for a fixed $\theta$,

$$V_\theta(\delta) \geq \frac{(g(\theta+\Delta) - g(\theta))^2}{E_\theta\left[\left(\frac{f_{\theta+\Delta}(x)}{f_\theta(x)} - 1\right)^2\right]}, \quad \forall \Delta \in B_\theta$$

$$= \frac{\left(\frac{g(\theta+\Delta) - g(\theta)}{\Delta}\right)^2}{E_\theta\left[\frac{1}{f_\theta^2(x)}\left(\frac{f_{\theta+\Delta}(x) - f_\theta(x)}{\Delta}\right)^2\right]}, \quad \forall \Delta \in B_\theta$$

On taking $\Delta \to 0$, we get

$$V_\theta(\delta) \geq \frac{[g'(\theta)]^2}{E_\theta\left[\left(\frac{f'_\theta(x)}{f_\theta(x)}\right)^2\right]}$$

$$= \frac{[g'(\theta)]^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta} \ln f_\theta(x)\right)^2\right]}.$$

Now the result follows on noting that

$$E_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta(x)\right) = \int \frac{f'_\theta(x)}{f_\theta(x)} f_\theta(x)\, dx$$

$$= \int f'_\theta(x)\, dx = \int \frac{\partial}{\partial\theta} f_\theta(x)\, dx$$

$$= \frac{\partial}{\partial\theta} \int f_\theta(x)\, dx = \frac{\partial}{\partial\theta}(1) = 0.$$

<u>Remark</u>: (a) Suppose that assumptions of last theorem are satisfied and for some unbiased estimator $\delta_0$ of $g(\theta)$, $Var_\theta(\delta_0) = \frac{[g'(\theta)]^2}{E_\theta[(\frac{\partial}{\partial\theta} \ln f_\theta(x))^2]}$, then $\delta_0$ is the UMVUE of $g(\theta)$.

(b) $$\psi(\theta) = \frac{\partial}{\partial\theta} \ln f_\theta(x)$$

$$= \frac{1}{f_\theta(x)} \frac{\partial}{\partial\theta} f_\theta(x)$$

$\longrightarrow$ relative rate at which the density $f_\theta(x)$ changes at $x$

$$I(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right)^2\right] = V_\theta\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right)$$

$\rightarrow$ Average rate of change of density

The large value of $I(\theta_0)$ indicates that it is easier to distinguish $\theta_0$ from neighboring values of $\theta$ and therefore more accurately $\theta$ can be estimated at $\theta_0$.

$I(\theta)$: Fisher's Information that $\underline{X}$ contains about the parameter $\theta$.

(c) Under the regularity conditions $R_1$

$$E_\theta\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right) = 0$$

and

$$I(\theta) = V_\theta\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right)^2\right).$$

(d) If in addition to regularity conditions $R_1$, the second derivative w.r.t. $\ln f_\theta(x)$ exist $\forall x$ and $\forall\theta$ and $\int f_\theta(x)\,dx$ can be differentiated twice by differentiating under the integral sign, then

$$\int\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right) f_\theta(x)\,dx = 0$$

$$\Rightarrow \frac{\partial}{\partial\theta}\left(\int\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right) f_\theta(x)\,dx\right) = 0$$

$$\Rightarrow \int\left[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x)\right] f_\theta(x)\,dx + \int\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right) f'_\theta(x)\,dx = 0$$

$$\Rightarrow E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x)\right] = -E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right)^2\right]$$

$$\Rightarrow I(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right)^2\right) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x)\right].$$

**Theorem** Let $X$ have the pdf/pmf

$$f_\theta(x) = e^{\eta(\theta)T(x) - B(\theta)} h(x), \quad x \in \mathcal{X}, \ \theta \in \Theta,$$

belonging to the exponential family with $\theta = E_\theta(T(X))$ (mean value parametrization) and $\Theta$ containing an interval. Then

$$I(\theta) = \frac{1}{Var_\theta(T)}.$$

**Proof.** For exponential families regularity conditions $R_1$ are satisfied. Moreover $\forall x \in \mathcal{X}$, $\ln f_\theta(x)$ is twice differentiable and $\int f_\theta(x)\,dx$ can be differentiated twice under the integral sign. Therefore by differentiating under the integral sign

$$I(\theta) = Var_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta(X)\right) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\ln f_\theta(X)\right)^2\right)$$

$$= -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \ln f_\theta(X)\right), \quad \theta \in \Theta$$

We have

$$\ln f_\theta(x) = \eta(\theta)T(x) - B(\theta) + \ln h(x)$$

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = \eta'(\theta)T(x) - B'(\theta) = \eta'(\theta)\left[T(x) - \frac{B'(\theta)}{\eta'(\theta)}\right]$$

$$\Rightarrow \quad I(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_\theta(X)\right)^2\right] = (\eta'(\theta))^2 E_\theta\left[\left(T(X) - \frac{B'(\theta)}{\eta'(\theta)}\right)^2\right]$$

Also

$$\int f_\theta(x)\,dx = 1 \Rightarrow \int e^{\eta(\theta)T(x) - B(\theta)} dx = 1, \quad \forall \theta \in \Theta \quad \ldots\ldots (*)$$

$$\Rightarrow \int (\eta'(\theta)T(x) - B'(\theta)) f_\theta(x)\,dx = 0$$

$$\Rightarrow \quad E_\theta(T(X)) = \frac{B'(\theta)}{\eta'(\theta)} = \theta$$

$\Rightarrow$ $I(\theta) = (\eta'(\theta))^2 Var_\theta(T(X))$, $\forall \theta \in \Theta$ and

$$\frac{\eta'(\theta)B''(\theta) - \eta''(\theta)B'(\theta)}{(\eta'(\theta))^2} = 1, \quad \forall \theta \in \Theta \quad \ldots\ldots (A)$$

Further, from $(*)$,

$$\int (\eta'(\theta)T(x) - B'(\theta))^2 f_\theta(x)\,dx + \int (\eta''(\theta)T(x) - B''(\theta)) f_\theta(x)\,dx = 0$$

$$\Rightarrow (\eta'(\theta))^2 \int \Big(T(x) - \underbrace{\frac{B'(\theta)}{\eta'(\theta)}}_{=E_\theta(T(x))}\Big)^2 f_\theta(x)\,dx + \eta''(\theta)\, E_\theta(T(x)) - B''(\theta) = 0, \quad \forall\, \theta \in \Theta$$

$$\Rightarrow (\eta'(\theta))^2 \, Var_\theta(T(x)) + \eta''(\theta)\frac{B'(\theta)}{\eta'(\theta)} - B''(\theta) = 0, \quad \forall\, \theta \in \Theta$$

$$\Rightarrow (\eta'(\theta))^2 \, Var_\theta(T(x)) - \frac{\eta'(\theta)B''(\theta) - B'(\theta)\eta''(\theta)}{\eta'(\theta)} = 0$$

$$\Rightarrow (\eta'(\theta))^2 \, Var_\theta(T(x)) - \eta'(\theta) = 0, \qquad (\text{from } (A))$$

$$\Rightarrow Var_\theta(T(x)) = \frac{1}{\eta'(\theta)}$$

$$\Rightarrow I(\theta) = (\eta'(\theta))^2 \, E_\theta\left[\left(T(x) - \frac{B'(\theta)}{\eta'(\theta)}\right)^2\right]$$

$$= (\eta'(\theta))^2 \, Var_\theta(T(x))$$

$$= \frac{1}{Var_\theta(T(x))}, \quad \forall\, \theta \in \Theta$$

<u>Remark</u>: Let $\theta = h(\eta)$ and let the information contained in $X$ about $\eta$ be denoted by $I^*(\eta)$. If $h(\cdot)$ is differentiable, then

$$I^*(\eta) = E_\theta\left(\left(\frac{\partial}{\partial \eta} \ln f_\theta(x)\right)^2\right)$$

$$= E_\theta\left(\left\{\left(\frac{\partial}{\partial \theta} \ln f_\theta(x)\right)\frac{\partial \theta}{\partial \eta}\right\}^2\right)$$

$$= [h'(\eta)]^2 \, I(\theta)$$

↳ Information contained in $X$ about $\theta$.

**Remark** Under the assumptions of last theorem, $T(X)$ is the UMVUE of $E_\theta(T(X)) = \theta$. Therefore $Var_\theta(T(X))$ is a measure of difficulty of estimating $\theta = E_\theta(T(X))$ or equivalently reciprocal of the $Var_\theta(T(X))$ is the measure of ease with which $\theta$ can be estimated and in this case it matches with $I(\theta)$.

**Theorem** Let $X_i$ have pdf/pmf $f_\theta^{(i)}(\cdot)$ and let $I_i(\theta)$ denote the information contained in $X_i$ about $\theta$, $i = 1, \dots, n$. Assume that $X_1, \dots, X_n$ are independent and let $I(\theta)$ denote the information contained in $\underline{X} = (X_1, \dots, X_n)$ about $\theta$. Then, under regularity conditions $R_1$,

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta), \qquad \theta \in \Theta.$$

**Proof.** The joint p.d.f/p.m.f. of $\underline{X} = (X_1, \dots, X_n)$ is

$$f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta^{(i)}(x_i)$$

$$\Rightarrow \quad \ln f_\theta(\underline{x}) = \sum_{i=1}^{n} \ln f_\theta^{(i)}(x_i)$$

$$I(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta} \ln f_\theta(\underline{X})\right)^2\right] = E_\theta\left[\left\{\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_\theta^{(i)}(X_i)\right\}^2\right]$$

$$= E_\theta\left[\sum_{i=1}^{n} \left(\frac{\partial}{\partial\theta} \ln f_\theta^{(i)}(X_i)\right)^2\right] + E_\theta\left[\sum_{i\neq j}\sum \frac{\partial}{\partial\theta} \ln f_\theta^{(i)}(X_i) \frac{\partial}{\partial\theta} \ln f_\theta^{(j)}(X_j)\right]$$

$$= \sum_{i=1}^{n} E_\theta\left[\left(\frac{\partial}{\partial\theta} \ln f_\theta^{(i)}(X_i)\right)^2\right] + \sum_{i\neq j}\sum E_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta^{(i)}(X_i) \frac{\partial}{\partial\theta} \ln f_\theta^{(j)}(X_j)\right).$$

$$= \sum_{i=1}^{n} I_i(\theta) + \sum_{i\neq j}\sum E_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta^{(i)}(X_i)\right) E_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta^{(j)}(X_j)\right)$$

$$= \sum_{i=1}^{n} I_i(\theta)$$

**Corollary** Under the assumptions of last theorem let $X_1, \dots, X_n$ be i.i.d. so that $I_1(\theta) = \dots = I_n(\theta) = i(\theta)$ $\forall \theta \in \Theta$, say. Then $I(\theta) = n\, i(\theta)$, $\theta \in \Theta$

**Theorem (Reformulation of Rao-Cramer Inequality).**

(I) Under the regularity conditions $R_1$ let $I(\theta) > 0$, $\forall \theta \in \Theta$, and let $\delta$ be any statistic with $E_\theta(\delta^2(X)) < \infty$, $\forall \theta \in \Theta$, for which the derivative of $E_\theta(\delta(X)) = \int \delta(x) f_\theta(x)\, dx$ w.r.t. $\theta$ exists and can be obtained by differentiating under the integral sign. Then

$$Var_\theta(\delta(X)) \geq \frac{\left[\frac{\partial}{\partial\theta} E_\theta(\delta(X))\right]^2}{I(\theta)}.$$

(II) Let $X_1, \dots, X_n$ be i.i.d. with p.d.f. / p.m.f. $f_\theta(x)$. Under the regularity conditions $R_1$ let $i(\theta) = Var_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta(X_1)\right) > 0$ and let $\delta(X)$ be any estimator of $\psi(\theta)$ with bias $b(\theta) = E_\theta(\delta(\underline{X})) - \psi(\theta)$, $\theta \in \Theta$. If $E_\theta(\delta(\underline{X})) = \int \delta(\underline{x}) \prod_{i=1}^{n} f_\theta(x_i)\, d\underline{x} = b(\theta) + \psi(\theta)$, $\theta \in \Theta$, is differentiable w.r.t. $\theta$ and the derivative can be obtained by differentiating under the integral sign, then

$$Var_\theta(\delta(X)) \geq \frac{\left[b'(\theta) + \psi'(\theta)\right]^2}{n\, i(\theta)}.$$

**Theorem** Let the r.v. $X$ be distributed with pdf/pmf $f_\theta(\cdot)$, $\theta \in \Theta$. Then

$$Var_\theta(\delta(X)) = \frac{\left[\frac{\partial}{\partial\theta} E_\theta(\delta(X))\right]^2}{I(\theta)}$$

iff $f_\theta(x) = e^{\eta(\theta) T(x) - B(\theta)} h(x)$, $x \in \mathcal{X}$ with $T(x) = \delta(x)$.

Proof. Note that in the Covariance inequality

$$Var_\theta(\delta(x)) \geq \frac{[Cov_\theta(\delta, \psi)]^2}{Var_\theta(\psi)}$$

the equality is attained iff $\delta$ and $\psi$ are linearly related, i.e.,

$$\psi(\underset{\sim}{x}, \theta) = a(\theta)\, \delta(x) - b(\theta)$$

$$\Leftrightarrow \frac{\partial}{\partial\theta} \ln f_\theta(x) = a(\theta)\, \delta(x) - b(\theta)$$

$$\Leftrightarrow \ln f_\theta(x) = \eta(\theta)\, \delta(x) - B(\theta) + c(x) \qquad \left[\begin{array}{l} a(\theta) = \eta'(\theta) \\ b(\theta) = B'(\theta) \end{array}\right]$$

$$\Leftrightarrow f_\theta(x) = e^{\eta(\theta)\delta(x) - B(\theta)}\, h(x), \quad x \in \mathcal{X}.$$

Example Let $X_1, \ldots, X_n$ be i.i.d. Poisson($\theta$), $\theta > 0$. Let the estimand be $\psi(\theta) = \theta$.

Method I $T(\underset{\sim}{X}) = \bar{X}$ is complete-sufficient

$E_\theta(T(\underset{\sim}{X})) = E_\theta(\bar{X}) = \theta, \quad \forall\, \theta > 0$

$\Rightarrow T(\underset{\sim}{X}) = \bar{X}$ is the UMVUE of $\theta$.

Method II By C.R.C. inequality, for any unbiased estimator $\delta$,

$$Var_\theta(\delta) \geq \frac{\psi'(\theta)}{n\, i(\theta)} = \frac{1}{n\, i(\theta)},$$

where

$$i(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f_\theta(x_1)\right)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \ln f_\theta(x_1)\right)$$

$$f_\theta(x_1) = \frac{e^{-\theta}\, \theta^{x_1}}{\lfloor x_1}$$

$$\ln f_\theta(x_1) = -\theta + x_1 \ln\theta - \ln \lfloor x_1$$

$$\frac{\partial}{\partial\theta} \ln f_\theta(x_1) = -1 + \frac{x_1}{\theta}$$

$$\frac{\partial^2}{\partial\theta^2} \ln f_\theta(x_1) = -\frac{x_1}{\theta^2}$$

$$I(\theta) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \ln f_\theta(x_1)\right) = E\left(\frac{x_1}{\theta^2}\right) = \frac{1}{\theta}$$

$\Rightarrow$ For any unbiased estimator $\delta$ of $\theta$

$$Var_\theta(\delta) \geq \frac{\theta}{n} = Var_\theta(T(\underset{\sim}{x}))$$

This is happening since $\underset{\sim}{X}$ has dist belonging to exponential family

$\Rightarrow$ $T(\underset{\sim}{x})$ is the UMVUE of $\theta$.

<u>Example</u> Let $x_1, \dots, x_n$ be i.i.d. $U(0, \theta)$ $\theta \in \Theta = (0, \infty)$

Let the estimand be $\psi(\theta) = \theta$, $\theta > 0$.

$T(\underset{\sim}{x}) = \frac{n+1}{n} X_{(n)}$ is the UMVUE.

$$f_\theta(x_1) = \frac{1}{\theta}, \quad 0 < x_1 < \theta$$

$$\ln f_\theta(x_1) = -\ln\theta$$

$$\frac{\partial}{\partial\theta} \ln f_\theta(x_1) = -\frac{1}{\theta}$$

$$E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f_\theta(x_1)\right)^2\right) = \frac{1}{\theta^2}$$

R.C. Bound for any unbiased estimator

$$= \frac{\theta^2}{n}$$

$$Var_\theta(T(\underset{\sim}{x})) = \left(\frac{n+1}{n}\right)^2 Var_\theta(X_{(n)})$$

$$= \left(\frac{n+1}{n}\right)^2 \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right]\theta^2$$

$$= \frac{\theta^2}{n(n+2)} < \text{C.R. lower bound } \frac{\theta^2}{n}$$

$\Rightarrow$ R.C. Theorem is not applicable here. To see this, consider for any continuous function $h(\cdot)$

$$\frac{d}{d\theta}\int_0^\theta h(x)\, f_\theta(x)\,dx = \frac{d}{d\theta}\int_0^\theta h(x)\,\frac{1}{\theta}\,dx$$

$$= \frac{h(\theta)}{\theta} + \int_0^\theta h(x)\,\frac{\partial}{\partial\theta}\cdot\frac{1}{\theta}\,dx$$

$$\neq \int_0^\theta h(x)\,\frac{\partial}{\partial\theta} f_\theta(x)\,dx,$$

unless $\frac{h(\theta)}{\theta} = 0, \quad \forall\, \theta > 0$.

In general R.C. Theorem is not applicable to distributions whose range of the p.d.f./p.m.f. depends on the parameter.

Remark (i) R.C. Theorem, when applicable, sometimes can be used to provide UMVUE.

(ii) Even if R.C. Theorem is applicable, the lower bound may not be sharp and the UMVUE may have uniformly larger variance than the R.C. lower bound. In such situations the R.C. Theorem can not be used to provide the UMVUE.

Example Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$, where $\theta \in \Theta = \mathbb{R}$. $\longrightarrow$ Exponential family

$\psi(\theta) = \theta$.

Then, for any unbiased estimator $\delta$

$$\mathrm{Var}_\theta(\delta) \geq \frac{\left[\frac{d}{d\theta} E_\theta(\bar{X})\right]^2}{(\mathrm{Var}_\theta(\bar{X}))^{-1}} = \frac{1}{n}$$

Also $\bar{X}$ is unbiased and

$$Var_\theta(\bar{X}) = \frac{1}{n}, \forall \theta = \text{R.C. lower bound}$$

$\Rightarrow$ $\bar{X}$ is the UMVUE of $\theta = E_\theta(\bar{X})$.

<u>Example</u> Let $X_1, \dots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$.

Let the estimand be $\tau(\underline{\theta}) = \sigma^2$.

$$f_{\underline{\theta}}(x_1) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2}$$

Assuming $\mu$ to be fixed (so that class of unbiased estimators of $\tau(\underline{\theta})$ under consideration is large)

$$\frac{\partial}{\partial \sigma^2} \ln f_{\underline{\theta}}(x_1) = \frac{\partial}{\partial \sigma^2}\left[-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}(x_1-\mu)^2\right]$$

$$= -\frac{1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(x_1-\mu)^2$$

$$\frac{\partial^2}{\partial(\sigma^2)^2} \ln f_\theta(x_1) = \frac{1}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}(x_1-\mu)^2$$

$$\Rightarrow \quad -E_{\underline{\theta}}\left[\frac{\partial^2}{\partial(\sigma^2)^2} \ln f_\theta(x_1)\right] = -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} = \frac{1}{2\sigma^4}$$

$\Rightarrow$ R.C. bound for an unbiased estimator of $\tau(\underline{\theta})$

$$= \frac{2\sigma^4}{n} \longrightarrow \text{does not depend on } \mu$$

We know that $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is the UMVUE

$$Var_\theta(S^2) = \frac{2\sigma^4}{n-1} > \text{R.C. lower bound.}$$

Note that for fixed $\mu$ $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$ is the UMVUE and

$$\mathrm{Var}_{\theta}(\hat{S}^2) = \mathrm{Var}\left(\frac{\sigma^2}{n} \chi^2_n\right)$$

$$= \frac{\sigma^4}{n^2} \mathrm{Var}(\chi^2_n)$$

$$= \frac{2\sigma^4}{n} = \text{R.C. lower bound}$$

$\Rightarrow$ $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ is the UMVUE of $\sigma^2$ for fixed $\mu$.

## Generalizations of Rao-Cramer Theorem

$\underline{X} = (X_1, \ldots, X_p)'$ : a $p\times 1$ vector

$\underline{Y} = (Y_1, \ldots, Y_q)'$ : a $q\times 1$ vector

$\mathrm{Var}(\underline{X}) = \Sigma_{11}$, $\mathrm{Var}(\underline{Y}) = \Sigma_{22}$, $\mathrm{Cov}(\underline{X}, \underline{Y}) = \Sigma_{12}$

so that $\Sigma_{11}$ is a $p\times p$ n.n.d. matrix; $\Sigma_{22}$ is a $q\times q$ n.n.d matrix and $\Sigma_{12}$ is a $p\times q$ matrix

For a fixed $\underline{b} \in \mathbb{R}^q - \{0\}$ suppose you want to predict $\underline{b}'\underline{Y}$ by $\underline{a}'\underline{X}$. We consider the correlation coefficient

$$\mathrm{Corr}(\underline{a}'\underline{X}, \underline{b}'\underline{Y})$$

and choose $\underline{a}$ which gives the maximum value of above correlation coefficient.

For a fixed $\underline{b} \in \mathbb{R}^q - \{0\}$, define

$$\rho^* = \sup_{\underline{a}\in\mathbb{R}^p} \mathrm{Corr}(\underline{a}'\underline{X}, \underline{b}'\underline{Y})$$

**Theorem** (a)

$$\rho^* = \frac{\sqrt{\underline{b}'\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\underline{b}}}{\sqrt{\underline{b}'\Sigma_{22}\underline{b}}}$$

(b) The matrix $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ is n.n.d.

Proof. (a) Let

$$\psi(\underline{a}) = \mathrm{Corr}(\underline{a}'\underline{X}, \underline{b}'\underline{Y}) = \frac{\underline{a}'\Sigma_{12}\underline{b}}{\sqrt{\underline{a}'\Sigma_{11}\underline{a}}\sqrt{\underline{b}'\Sigma_{22}\underline{b}}}$$

Note that,

$$\psi(\underline{a}) = \psi\left(\frac{a}{\underline{a}'\Sigma_{22}\underline{a}}\right), \quad \forall\ \underline{a} \in \mathbb{R}^p.$$

Thus for maximization of $\psi(\underline{a})$ WLOG we may assume that $\underline{a}'\Sigma_{22}\underline{a} = 1$. Then maximization of $\psi(\underline{a})$ is equivalent to maximization of

$$\psi_1(\underline{a}) = \underline{a}'\Sigma_{12}\underline{b}$$

subject to $\underline{a}'\Sigma_{22}a = 1$.

For using the method of Lagrange's multiplier Consider the function

$$\psi_2(\underline{a}, \lambda) = \underline{a}'\Sigma_{12}\underline{b} - \lambda(\underline{a}'\Sigma_{22}\underline{a} - 1)$$

$$\frac{\partial \psi_2}{\partial a} = \Sigma_{12}\underline{b} - 2\lambda\Sigma_{11}a = 0$$

$$\Rightarrow 2\lambda\underline{a} = \Sigma_{11}^{-1}\Sigma_{12}b \quad \cdots\cdots \quad (i)$$

$$\frac{\partial \psi_2}{\partial \lambda} = -(a'\Sigma_{22}\underline{a} - 1) = 0$$

$$\Rightarrow \underline{a}'\Sigma_{11}\underline{a} = 1$$

From (i) and (ii) we get

$$4\lambda^2 = \underline{b}'\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11}\Sigma_{11}^{-1}\Sigma_{12}\underline{b} \qquad (\Sigma_{21} = \Sigma_{12}')$$

$$= \underline{b}'\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\underline{b}$$

$$2\lambda = \pm\sqrt{\underline{b}'\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\underline{b}}$$

$$\Rightarrow a = \pm\frac{\Sigma_{11}^{-1}\Sigma_{12}\underline{b}}{\sqrt{\underline{b}'\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\underline{b}}}$$

Since $\rho^* \geq 0$ $(\psi(\underline{a}) = -\psi(-\underline{a}))$

$$\rho^* = \frac{\underline{b}' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \underline{b}}{\sqrt{\underline{b}' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \underline{b}} \ \sqrt{\underline{b}' \Sigma_{22} \underline{b}}}$$

$$= \frac{\sqrt{\underline{b}' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \underline{b}}}{\sqrt{\underline{b}' \Sigma_{22} \underline{b}}}$$

(b) Since $\rho^{*2} \leq 1$ we have

$$\underline{b}' ( \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} ) \underline{b} \geq 0 \qquad \forall \ \underline{b} \in \mathbb{R}^{\nu} - \{0\}$$

$$\Rightarrow \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \text{ is n.n.d.}$$

$\boxed{///}$

**Note:** $\rho^* \geq \mathrm{Cov}( X_i, \sum_{j=1}^{\nu} b_j Y_j ) \quad \forall \ \underline{b} \in \mathbb{R}^{\nu}.$

Now let $\underline{X}$ be a random vector with p.d.f/p.m.f. $f_{\underline{\theta}}(\cdot)$, where $\underline{\theta} = (\theta_1, \dots, \theta_p)' \in \Theta \subseteq \mathbb{R}^p$.

Regularity Conditions $R_2$:

(a) $\Theta$ is an open rectangle (finite, infinite or semi-infinite).

(b) $f_{\underline{\theta}}(\cdot)$, $\underline{\theta} \in \Theta$, have the common support so that, WLOG, the set $A = \{ \underline{x} : f_{\underline{\theta}}(\underline{x}) > 0 \}$ is independent of $\underline{\theta} \in \Theta$;

(c) $\forall \ \underline{x} \in A$, $\underline{\theta} \in \Theta$ and $i = 1, \dots, p$, the derivatives $\frac{\partial f_{\underline{\theta}}(\underline{x})}{\partial \theta_i}$ exist and are finite. Also for $i = 1, \dots, p$, $\int f_{\underline{\theta}}(\underline{x})\, d\underline{x} = 1$ can be differentiated w.r.t. $\theta_i$ by differentiating under the integral sign so that

$$E_{\underline{\theta}} \left( \frac{\partial}{\partial \theta_i} \ln f_{\underline{\theta}}(\underline{X}) \right) = \int \frac{\partial}{\partial \theta_i} \ln f_{\underline{\theta}}(\underline{x}) = 0, \quad \forall \ \underline{\theta} \in \Theta, \ i = 1, \dots, p$$

(d) the derivative w.r.t. each $\theta_j$ $(j = 1, \dots, p)$ of $\int \frac{\partial}{\partial \theta_i} \ln f_{\underline{\theta}}(\underline{x})\, d\underline{x} = 0$ can be obtained by differentiating under the integral sign.

For $i, j = 1, \dots, p$, define

$$I_{ij}(\underline{\theta}) = E_\theta\left( \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) \frac{\partial}{\partial\theta_j} \ln f_{\underline{\theta}}(\underline{x}) \right)$$

$$= Cov_{\underline{\theta}}\left( \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}), \frac{\partial}{\partial\theta_j} \ln f_{\underline{\theta}}(\underline{x}) \right).$$

Let

$$\boxed{I(\underline{\theta}) = ((I_{ij}(\underline{\theta}))) \longrightarrow \text{Fisher's Information matrix}}$$

Note that $I(\underline{\theta})$ (being a Variance-Covariance matrix) is n.n.d.

In fact $I(\underline{\theta})$ is p.d. unless $\frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x})$, $i=1 \dots p$, are affinely dependent, i.e., $\exists$ constants $c_1, \dots, c_p$ not all $c_i$s zero such that

$$\sum_{i=1}^{p} c_i \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) = 0.$$

Note that

$$\int \frac{\partial}{\partial\theta_i} f_{\underline{\theta}}(\underline{x})\, dx = 0, \quad \forall\, i = 1 \dots p$$

$$\Rightarrow \frac{\partial}{\partial\theta_j} \int \frac{\partial}{\partial\theta_i} f_{\underline{\theta}}(\underline{x})\, d\underline{x} = 0 \quad \forall\, i = 1 \dots p,\ j = 1 \dots p$$

$$\Rightarrow \frac{\partial}{\partial\theta_j} \int \left( \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) \right) f_{\underline{\theta}}(\underline{x})\, d\underline{x} = 0 \quad \forall\, i, j = 1 \dots p$$

$$\Rightarrow \int \left( \frac{\partial^2}{\partial\theta_j \partial\theta_i} \ln f_{\underline{\theta}}(\underline{x})\, f_{\underline{\theta}}(\underline{x}) + \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) \frac{\partial}{\partial\theta_j} f_{\underline{\theta}}(\underline{x}) \right) d\underline{x} = 0$$

$$\Rightarrow E_{\underline{\theta}}\left[ \frac{\partial^2}{\partial\theta_j \partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) \right] = -E_{\underline{\theta}}\left[ \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) \frac{\partial}{\partial\theta_j} \ln f_{\underline{\theta}}(\underline{x}) \right]$$

$$\Rightarrow \boxed{\begin{aligned} I_{ij}(\underline{\theta}) &= E_\theta\left[ \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) \frac{\partial}{\partial\theta_j} \ln f_{\underline{\theta}}(\underline{x}) \right] \\ &= -E_\theta\left[ \frac{\partial^2}{\partial\theta_j \partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}) \right] \\ &= Cov_{\underline{\theta}}\left( \frac{\partial}{\partial\theta_i} \ln f_{\underline{\theta}}(\underline{x}), \frac{\partial}{\partial\theta_j} \ln f_{\underline{\theta}}(\underline{x}) \right), \\ &\qquad i, j = 1 \dots p. \end{aligned}}$$

**Theorem** Let

$$f_{\underline{\theta}}(\underline{x}) = \exp\left[\sum_{i=1}^{p} \eta_i(\underline{\theta}) T_i(\underline{x}) - B(\underline{\theta})\right] h(\underline{x}), \quad \underline{x} \in A,$$

where $\theta_i = E_{\underline{\theta}}(T_i(\underline{X}))$, $i=1 \ldots p$. here $\eta_i$'s, $T_i$'s and $B$ are real-valued functions. Then

$$I(\underline{\theta}) = [Var(\underline{T})]^{-1}.$$

**Proof.** On writing $f_{\underline{\theta}}(\cdot)$ in Cannonical form, we have the p.d.f/p.m.f of $\underline{X}$ as

$$f^*_{\eta}(\underline{x}) = \exp\left[\sum_{i=1}^{p} \eta_i T_i(\underline{x}) - A(\underline{\eta})\right] h(\underline{x})$$

$$I^*_{ij}(\underline{\eta}) = E_{\eta}\left(\frac{\partial}{\partial \eta_i} \ln f^*_{\eta}(\underline{X}) \frac{\partial}{\partial \eta_j} \ln f^*_{\eta}(\underline{X})\right), \quad i,j = 1 \ldots p$$

$$= E_{\underline{\eta}}\left(\left(T_i(\underline{X}) - \frac{\partial}{\partial \eta_i} A(\underline{\eta})\right)\left(T_j(\underline{X}) - \frac{\partial}{\partial \eta_j} A(\underline{\eta})\right)\right)$$

Also

$$\int \exp\left[\sum_{i=1}^{p} \eta_i T_i(\underline{x}) - A(\underline{\eta})\right] h(\underline{x})\, d\underline{x} = 1 \quad \forall\ \underline{\eta}$$

On differentiating both sides w.r.t. $\eta_j$ we get

$$E_{\underline{\eta}}(T_j(\underline{X})) = \frac{\partial}{\partial \eta_j} A(\underline{\eta}), \quad j = 1 \ldots p$$

$$\Rightarrow \quad I^*_{ij}(\underline{\eta}) = \cancel{Var}\ Cov\,(T_i, T_j)$$

$$I^*(\underline{\eta}) = ((I^*_{ij}(\underline{\eta}))) = Var(\underline{T}).$$

$$I^*_{ij}(\underline{\eta}) = E_{\theta}\left(\frac{\partial}{\partial \eta_i} \ln f_{\underline{\theta}}(\underline{X}) \frac{\partial}{\partial \eta_j} \ln f_{\underline{\theta}}(\underline{X})\right)$$

$$= E_{\underline{\theta}}\left(\sum_{k=1}^{p} \frac{\partial}{\partial \theta_k} \ln f_{\underline{\theta}}(\underline{X}) \frac{\partial \theta_k}{\partial \eta_i}, \sum_{\ell=1}^{p} \frac{\partial}{\partial \theta_\ell} \ln f_{\underline{\theta}}(\underline{X}) \frac{\partial \theta_\ell}{\partial \eta_j}\right)$$

$$= E_{\theta}\left(\sum_{k=1}^{p} \sum_{\ell=1}^{p} \frac{\partial}{\partial \theta_k} \ln f_{\underline{\theta}}(\underline{X}) \frac{\partial}{\partial \theta_\ell} \ln f_{\underline{\theta}}(\underline{X}) \frac{\partial \theta_k}{\partial \eta_i} \frac{\partial \theta_\ell}{\partial \eta_j}\right),$$

$i, j = 1 \ldots p$

Also

$$\theta_i = E_{\underline{\theta}}(T_i(\underline{X})) = \int T_i(\underline{x}) \exp\left[\sum_{l=1}^{k} \eta_l T_l(\underline{x}) - A(\eta)\right] h(\underline{x})\, d\underline{x}$$

$$\frac{\partial \theta_i}{\partial \eta_j} = \int T_i(\underline{x}) \left( T_j(\underline{x}) - \frac{\partial}{\partial \eta_j} A(\eta) \right) \exp\left[\sum_{l=1}^{k} \eta_l T_l(\underline{x}) - A(\eta)\right] h(\underline{x})\, d\underline{x}$$

$$= Cov(T_i(\underline{X}), T_j(\underline{X})) = Cov(T_j(\underline{X}), T_i(\underline{X})) = \frac{\partial \theta_j}{\partial \eta_i}.$$

$$\Rightarrow \quad I^*_{ij}(\eta|\underline{\theta}) = E_{\theta}\left( \sum_{k=1}^{p} \sum_{l=1}^{p} \frac{\partial \theta_i}{\partial \eta_k} \frac{\partial}{\partial \theta_k} \ln f_{\underline{\theta}}(\underline{X}) \frac{\partial}{\partial \theta_l} \ln f_{\underline{\theta}}(\underline{X}) \frac{\partial \theta_l}{\partial \eta_j} \right)$$

$$\Rightarrow \quad I^*(\eta|\underline{\theta}) = Var(T)\, I(\underline{\theta})\, Var(T)$$

$$\Rightarrow \quad Var(T) = Var(T)\, I(\underline{\theta})\, Var(T)$$

$$\Rightarrow \quad I(\underline{\theta}) = [Var(T)]^{-1}.$$

__Theorem__ (Rao-Cramer Lower Bound). Let $X$ be a random vector with p.d.f./p.m.f. $f_{\underline{\theta}}(\cdot)$, where $\underline{\theta} \in \Theta$. Under the regularity conditions $R_2$, let $\delta(\underline{X}) = (\delta_1(\underline{X}), \ldots, \delta_m(\underline{X}))'$ be an estimator of estimand $\psi(\underline{\theta}) = (\psi_1(\underline{\theta}), \ldots, \psi_m(\underline{\theta}))'$. Define

$$J = ((J_{ij})),$$

where $J_{ij} = \frac{\partial}{\partial \theta_j} E_{\underline{\theta}}(\delta_i(\underline{X}))$, $j = 1, \ldots, p$, $i = 1, \ldots, m$

Then

$$Var_{\theta}(\delta) - J\, I^{-1}(\underline{\theta})\, J'$$

is n.n.d.

__Proof.__ Define

$$\underline{Y} = \left( \frac{\partial}{\partial \theta_1} \ln f_{\underline{\theta}}(\underline{X}), \ldots, \frac{\partial}{\partial \theta_m} \ln f_{\underline{\theta}}(\underline{X}) \right).$$

so that

$$\mathrm{Var}_\theta(\underline{Y}) = \Sigma_{11} = I(\underline{\theta})$$

$$\Sigma_{22} = V_\theta(\underline{\delta}(\underline{X}))\ , \text{ say}$$

$$\Sigma_{21} = \mathrm{Cov}_{\underline{\theta}}(\delta(\underline{X}), \underline{Y})$$

Then

$$\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \text{ is n.n.d.}$$

$$(\Sigma_{21})_{ij} = E_{\underline{\theta}}\left(\delta_i(\underline{X}) \frac{\partial}{\partial\theta_j} \ln f_{\underline{\theta}}(\underline{X})\right)$$

$$= \int \delta_i(\underline{x}) \frac{\partial}{\partial\theta_j} \ln f_{\underline{\theta}}(\underline{x})\, f_{\underline{\theta}}(\underline{x})\, d\underline{x}$$

$$= \int \delta_i(\underline{x}) \frac{\partial}{\partial\theta_j} f_{\underline{\theta}}(\underline{x})\, d\underline{x}$$

$$= \frac{d}{d\theta_j} \int \delta_i(\underline{x}) f_{\underline{\theta}}(\underline{x})\, d\underline{x}$$

$$= \frac{d}{d\theta_j} E_\theta(\delta_i(\underline{X})) = J_{ij}\ , \quad i=1\ldots m,\ j=1\ldots p$$

$$\Rightarrow \quad \Sigma_{21} = J$$

$$\Rightarrow \quad V_\theta(\delta(\underline{X})) - J\, I^{-1}(\theta)\, J \text{ is n.n.d.}$$

Example Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, where $\underline{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$. Let $\psi(\underline{\theta}) = (\mu, \sigma^2)$. Then for any for any unbiased estimator $\underline{\delta}(\underline{x}) = (\delta_1(\underline{x}), \delta_2(\underline{x}))$ of $\underline{\theta} = (\mu, \sigma^2)$

$$J = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

It can be seen that

$$I(\underline{\theta}) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

$\Rightarrow \quad Var_\theta(\underline{\delta}) - \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$

is n.n.d.

In particular

$V_\theta(\delta_1) \geq \frac{\sigma^2}{n} \qquad V_\theta(\delta_2) \geq \frac{2\sigma^4}{n}$

We know that $\bar{X}$ is the UMVUE for $\mu$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the UMVUE for $\sigma^2$ with

$$Var_\theta(\bar{X}) = \frac{\sigma^2}{n}$$

$$Var_\theta(S^2) = \frac{2\sigma^4}{n-1} \geq \frac{2\sigma^4}{n}$$

**Example** Let $X_1 \ldots X_n$ be i.i.d. $Bin(1, \theta)$, $\theta \in \Theta = (0, 1)$. Let $\underline{\delta}(\underline{X}) = (\delta_1(\underline{X}), \delta_2(\underline{X}), \delta_3(\underline{X}))$ be an unbiased estimator of $\underline{\psi}(\theta) = (\theta, 1-\theta, \theta(1-\theta))$. Then

$$J = \begin{pmatrix} 1 \\ -1 \\ 1-2\theta \end{pmatrix}$$

It is easy to see that

$$I(\theta) = \frac{n}{\theta(1-\theta)}$$

$V_\theta(\underline{\delta}) - \begin{pmatrix} 1 \\ -1 \\ 1-2\theta \end{pmatrix} \frac{\theta(1-\theta)}{n} \begin{pmatrix} 1 & -1 & 1-2\theta \end{pmatrix}$ is n.n.d

$$\Leftrightarrow \quad V_\theta(\underline{\delta}) - \begin{bmatrix} \frac{\theta(1-\theta)}{n} & -\frac{\theta(1-\theta)}{n} & \frac{\theta(1-\theta)(1-2\theta)}{n} \\ -\frac{\theta(1-\theta)}{n} & \frac{\theta(1-\theta)}{n} & -\frac{\theta(1-\theta)(1-2\theta)}{n} \\ \frac{\theta(1-\theta)(1-2\theta)}{n} & -\frac{\theta(1-\theta)(1-2\theta)}{n} & \frac{\theta(1-\theta)(1-2\theta)^2}{n} \end{bmatrix}$$

is n.n.d.

In particular if $\delta_1$, $\delta_2$ and $\delta_3$ are unbiased estimators of $\theta$, $(1-\theta)$ and ~~for~~ $\theta(1-\theta)$, respectively, then

$$Var_\theta(\delta_1) \geq \frac{\theta(1-\theta)}{n} = Var(\bar{X})$$

$$Var_\theta(\delta_2) \geq \frac{\theta(1-\theta)}{n} = Var(1-\bar{X})$$

$$Var(\delta_3) \geq \frac{\theta(1-\theta)(1-2\theta)^2}{n}$$

It follows that $\bar{X}$ and $1-\bar{X}$ are the UMVUEs of $\theta$ and $1-\theta$, respectively.

Also

$$\delta_3^* = \frac{T(n-T)}{n(n-1)} \text{ is the UMVUE of } \theta(1-\theta)$$

$$\Rightarrow \quad Var\left(\frac{T(n-T)}{n(n-1)}\right) \geq \frac{\theta(1-\theta)(1-2\theta)^2}{n}$$

## Assignment Problems

(1) Let $\{T_n\}_{n\geq 1}$ be a sequence of UMVUEs and $T$ be a statistic with $E_\theta(T^2) < \infty$ such that $E_\theta(T_n - T)^2 \to 0$ as $n \to \infty$ $\forall\, \theta \in \Theta$. Show that $T$ is also the UMVUE of $\psi(\theta) = E_\theta(T)$.

(2) If a sample consists of $n$ i.i.d. observations show that the UMVUE, if it exists, is a permutation symmetric function of $X_i$s.

(3) Let $X_1, \dots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, where $\underline{\theta} = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$ is unknown. Find the UMVUE of:
(i) $\mu + \sigma$ ; (ii) $\frac{\mu}{\sigma}$

(4) Let $X_1, \dots, X_n$ be a random sample from the p.m.f. $f_\theta(x) = \frac{1}{\theta}$, if $x \in \{1, 2, \dots, \theta\}$, $= 0$ otherwise, where $\theta$ is an unknown parameter.
(a) Find the UMVUE of $\theta$ when $\theta \in \Theta_1 = \{1, 2, \dots\}$
(b) Find the UMVUE of $\theta$, where $\theta \in \Theta_2 = \Theta - \{\theta_0\}$, for some fixed positive integer $\theta_0$

(5) Let $X_1, \ldots, X_n$ be i.i.d. from a distribution in the family $\mathcal{P}$. Find an unbiased estimator of $\psi(\theta)$, if it exists, in each of the following cases: (a) $\mathcal{P} = \{\text{Bin}(1, \theta),\ 0<\theta<1\}$ $\psi(\theta) = \theta^2$. (b) $\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma>0\}$, $\psi(\underline{\theta}) = \mu^p$ for a positive integer $p$, $\underline{\theta} = (\mu, \sigma)$; (c) $\mathcal{P} = \{\text{Bin}(1, \theta),\ 0<\theta<1\}$, $\psi(\theta) = \sqrt{\theta}$

(d) $\mathcal{P} = \{\text{Poisson}(\theta) : \theta > 0\}$, $\psi(\theta) = \frac{1}{\theta}$

(6) Let $X_1, \ldots, X_n$ be i.i.d. from a distribution in the family $\mathcal{P}$. Find the UMVUE in each of the following cases:

(a) $\mathcal{P} = \{\text{Bin}(1, \theta),\ 0<\theta<1\}$, $\psi_1(\theta) = \theta^\wedge$, $\psi_2(\theta) = \theta^\wedge + (1-\theta)^{n-\wedge}$,
where $\wedge \in (0, n)$ is an integer,

(b) $\mathcal{P} = \{NB(1, \theta),\ 0<\theta<1\}$ $\psi(\theta) = P_\theta(X=0)$

(c) $f_\theta(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x>0$, $\theta>0$, $\psi(\theta) = P_\theta(X_1 \le t_0)$,
where $t_0$ is a fixed real number.

(d) $\mathcal{P} = \{\text{Poisson}(\theta) : \theta>0\}$, $\psi_1(\theta) = \sum_{k=0}^{a} c_k \theta^k$, for real constants $c_1, \ldots, c_k$, $\psi_2(\theta) = (1-\theta)^{-1}$. $\psi_3(\theta) = \theta^\wedge$, for some fixed integer $\wedge>0$; $\psi_4(\theta) = P_\theta(X=0 \text{ or } X=1)$.
$\psi_5(\theta) = P_\theta(X=k)$, where $k$ is a fixed positive integer.

(e) $f_\theta(x) = \frac{1}{\theta}$, if $x \in \{1, \ldots, \theta\}$, where $\theta \in \Theta = \{1, 2, \ldots\}$. Find the UMVUE of $\psi(\theta)$ for arbitrary real valued function $\psi(\cdot)$

(7) In each of the following cases find the lower R-C bound for unbiased ~~UMVUE~~ estimators of $\psi(\theta) = \theta$ based on a random sample $X_1 \ldots X_n$ from a pdf/pmf $f_\theta(\cdot)$, $\theta \in \Theta$.

(a) $X_1 \sim N(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$; (b) $X_1 \sim \text{Bin}(1, \theta)$, $\theta \in \Theta = (0, 1)$

(c) $X_1 \sim \text{Poisson}(\theta)$, $\theta \in \Theta = \mathbb{R}$ (d) $N(\mu_0, \theta)$, $\theta \in \Theta = \mathbb{R}$,

where $\mu_0$ is known

(8) Let $X_1, \ldots, X_n$ be a random sample from $U(0, \theta)$ distribution, where $\theta \in \Theta = [1, \infty)$. Find the UMVUE of $\psi(\theta) = \theta$.

(9) Let $X$ be a sample of size 1 from $U(\theta - 1, \theta + 1)$ distribution where $\theta \in \Theta = \mathbb{R}$. Let $\psi(\cdot)$ be an arbitrary non-constant function. Show that there is no UMVUE of $\psi(\theta)$.

(10) Let $X_1, \ldots, X_n$ be iid $Bin(1, \theta)$, where $\theta \in \Theta = (0, 1)$. Find the UMVUE of

(a) $\theta^m$, where $m \in \{1, 2, \ldots, n\}$ is a fixed integer;

(b) $P_\theta(X_1 + \cdots + X_m = k)$, where $m$ and $k$ are fixed positive integers $\leq n$;

(c) $P_\theta(X_1 + \cdots + X_{n-1} > X_n)$

(11) Let $X_1, \ldots, X_m$ be i.i.d having the p.d.f $\frac{1}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right)$ and let $Y_1, \ldots, Y_n$ be iid having the pdf $\frac{1}{\sigma_2} f\left(\frac{x - \mu_2}{\sigma_2}\right)$, where $\mu_1, \mu_2, \sigma_1$ and $\sigma_2$ are unknown parameters.

(a) If $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty < x < \infty$, find the UMVUE of

(i) $\psi_1(\underline{\theta}) = \mu_1 - \mu_2$ and $\psi_2(\underline{\theta}) = \left(\frac{\sigma_1}{\sigma_2}\right)^r$, where $\underline{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2) \in \mathbb{R}^2 \times (0, \infty)^2$ is unknown and $r$ is a fixed positive constant;

(ii) $\psi_1(\underline{\theta}) = \sigma_1^2$ and $\psi_2(\theta) = \frac{\mu_1 - \mu_2}{\sigma_1}$, by assuming that $\mu_1 \in \mathbb{R}$, $\mu_2 \in \mathbb{R}$ and $\sigma_1^2 = \sigma_2^2 \in (0, \infty)$ are unknown;

(iii) $\psi_1(\underline{\theta}) = \mu_1$, by assuming that $\mu_1 = \mu_2 \in \mathbb{R}$, $\sigma_1 > 0$, and $\sigma_2 > 0$ are unknown and $\frac{\sigma_1^2}{\sigma_2^2} = \vartheta$ is unknown.

(iv) $\psi_1(\underline{\theta}) = P_\theta(X_1 \leq Y_1)$ by assuming that $\mu_1 = \mu_2 \in \mathbb{R}$, $\sigma_1 > 0$ and $\sigma_2 > 0$ are unknown

(v) Repeat (iv) under the assumption that $\sigma_1 = \sigma_2$

(b) Repeat (a) under $f(x) = e^{-x}$, $x > 0$.

(12) Let $X_1, \ldots, X_n$ be iid $E(\mu, \sigma)$, where $\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ is unknown.

(a) Find the UMVUE of $\mu$ when $\sigma$ is known;

(b) Find the UMVUE of $\sigma$, when $\mu$ is known;

(c) Find the UMVUEs of $\mu$ and $\sigma$ when both are unknown.

(d) Assume that $\sigma$ is known. Find the UMVUE of $P(X_1 \geq t_0)$ and $\frac{d}{dt} P(X_1 \geq t)$, for a fixed $t > 0$

(13) Suppose that $X = (X_1, \ldots, X_k) \sim \text{Mult}(n, p_1, \ldots, p_k)$, where $p_i \in (0, 1)$, $i = 1, \ldots, k$ and $\sum_{i=1}^{k} p_i = 1$. For non-negative integers $r_1, \ldots, r_k$ with $\sum_{i=1}^{k} r_i \leq n$, find the UMVUE of $p_1^{r_1} \cdots p_k^{r_k}$.

(14) Let $X_1, \ldots, X_n$ be iid from $P \in \mathcal{P}$, where $\mathcal{P}$ is the family of all symmetric cdfs with finite mean and Lebesgue pdfs on $\mathbb{R}$. Show that there is no UMVUE of $\mu = E(X_1)$

(15) Let $X_1, \ldots, X_n$ be iid $\text{Bin}(1, \theta)$, where $\theta \in \Theta = (0, 1)$. Show that the UMVUE of $\psi(\theta) = \theta(1-\theta)$ does not attain the Rao-Cramer lower bound.

(16) Let $X_1, \ldots, X_n$ be iid $\text{Exp}(\theta)$ r.v.s where $\theta = E_\theta(X_1)$. Let $\psi_1(\theta) = \theta$, $\psi_2(\theta) = e^{-\frac{1}{\theta}}$ and $\psi_3(\theta) = \theta^2$. Find a R-C lower bound for unbiased estimator $\delta = (\delta_1, \delta_2, \delta_3)'$ of $\psi(\theta) = (\psi_1(\theta), \psi_2(\theta), \psi_3(\theta))'$.

(17) If $\delta_i$ is UMVUE of $\psi_i(\theta)$, $i = 1, \ldots, p$. Show that $\delta = \sum_{i=1}^{p} \alpha_i \delta_i$ is the UMVUE of $\psi(\theta) = \sum_{i=1}^{p} \alpha_i \psi_i(\theta)$ here $\alpha_1, \ldots, \alpha_p$ are fixed real constants.

(18) Let $X_1, \ldots, X_m$ be iid $N(\mu, \sigma_1^2)$ and $Y_1, \ldots, Y_n$ be iid $N(\mu, \sigma_2^2)$ where $\mu \in \mathbb{R}$, $\sigma_1 > 0$ and $\sigma_2 > 0$ are unknown. Further assume that $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_n)$ are independent. Show that a UMVUE of $\mu$ does not exist.