Lecture 17: Characteristic of UMVUE and Fisher information bound

When a complete and sufficient statistic is not available, it is usually very difficult to derive a UMVUE.
In some cases, the following result can be applied, if we have enough knowledge about unbiased estimators of 0.

Theorem 3.2

Let $\mathscr{U}$ be the set of all unbiased estimators of 0 with finite variances and T be an unbiased estimator of ϑ with $E(T^2) < \infty$.

(i) A necessary and sufficient condition for $T(X)$ to be a UMVUE of ϑ is that $E[T(X)U(X)] = 0$ for any $U \in \mathscr{U}$ and any $P \in \mathscr{P}$.

(ii) Suppose that $T = h(\tilde{T})$, where $\tilde{T}$ is a sufficient statistic for $P \in \mathscr{P}$ and h is a Borel function.
Let $\mathscr{U}_{\tilde{T}}$ be the subset of $\mathscr{U}$ consisting of Borel functions of $\tilde{T}$.
Then a necessary and sufficient condition for T to be a UMVUE of ϑ is that $E[T(X)U(X)] = 0$ for any $U \in \mathscr{U}_{\tilde{T}}$ and any $P \in \mathscr{P}$.

Proof of Theorem 3.2(i)

Suppose that T is a UMVUE of ϑ.
Then $T_c = T + cU$, where $U \in \mathscr{U}$ and c is a fixed constant, is also unbiased for ϑ and, thus,

$$\mathrm{Var}(T_c) \geq \mathrm{Var}(T) \qquad c \in \mathscr{R},\ P \in \mathscr{P},$$

which is the same as

$$c^2\,\mathrm{Var}(U) + 2c\,\mathrm{Cov}(T,U) \geq 0 \qquad c \in \mathscr{R},\ P \in \mathscr{P}.$$

This is impossible unless $\mathrm{Cov}(T,U) = E(TU) = 0$ for any $P \in \mathscr{P}$.
Suppose now $E(TU) = 0$ for any $U \in \mathscr{U}$ and $P \in \mathscr{P}$.
Let T_0 be another unbiased estimator of ϑ with $\mathrm{Var}(T_0) < \infty$.
Then $T - T_0 \in \mathscr{U}$ and, hence,

$$E[T(T - T_0)] = 0 \qquad P \in \mathscr{P},$$

which with the fact that $ET = ET_0$ implies that

$$\mathrm{Var}(T) = \mathrm{Cov}(T, T_0) \qquad P \in \mathscr{P}.$$

Note that $[\mathrm{Cov}(T,T_0)]^2 \leq \mathrm{Var}(T)\,\mathrm{Var}(T_0)$.
Hence $\mathrm{Var}(T) \leq \mathrm{Var}(T_0)$ for any $P \in \mathscr{P}$.

Proof of Theorem 3.2(ii)

It suffices to show that $E(TU) = 0$ for any $U \in \mathscr{U}_{\tilde{T}}$ and $P \in \mathscr{P}$ implies that $E(TU) = 0$ for any $U \in \mathscr{U}$ and $P \in \mathscr{P}$.
If $U \in \mathscr{U}$, then $E(U|\tilde{T}) \in \mathscr{U}_{\tilde{T}}$.
The result follows from the fact that $T = h(\tilde{T})$ and

$$E(TU) = E[E(TU|\tilde{T})] = E[E(h(\tilde{T})U|\tilde{T})] = E[h(\tilde{T})E(U|\tilde{T})].$$

Theorem 3.2 can be used to

- find a UMVUE,
- check whether a particular estimator is a UMVUE, and
- show the nonexistence of any UMVUE.

Theorem 3.2(ii) is more convenient to use.

Corollary 3.1

(i) If T_j is a UMVUE of ϑ_j, $j = 1,...,k$, then $\sum_{j=1}^k c_j T_j$ is a UMVUE of $\vartheta = \sum_{j=1}^k c_j \vartheta_j$ for any constants $c_1,...,c_k$.

(ii) If T_1 and T_2 are two UMVUE's of ϑ, then $T_1 = T_2$ a.s. P for any $P \in \mathscr{P}$.

Proof

(i) Obviously, $\sum_{j=1}^{k} c_j T_j$ is a unbiased for $\vartheta = \sum_{j=1}^{k} c_j \vartheta_j$
For each j,

$$E(T_j U) = 0, \qquad U \in \mathscr{U}$$

Then

$$E\left[\left(\sum_{j=1}^{k} c_j T_j\right) U\right] = \sum_{j=1}^{k} c_j E(T_j U) = 0, \qquad U \in \mathscr{U}$$

(ii) Let T_1 and T_2 be two UMVUE's of ϑ.

Then $T_1 - T_2 \in \mathscr{U}$ and

$$E[T_j(T_1 - T_2)] = 0 \qquad j = 1, 2.$$

Then

$$E(T_1 - T_2)^2 = E[T_1(T_1 - T_2)] - E[T_2(T_1 - T_2)] = 0$$

Hence, $T_1 = T_2$ a.s. P for any $P \in \mathscr{P}$.

Example 3.7

Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on the interval $(0, \theta)$.
In Example 3.1, $(1 + n^{-1})X_{(n)}$ is shown to be the UMVUE for θ when the parameter space is $\Theta = (0, \infty)$.
Suppose now that $\Theta = [1, \infty)$.
Then $X_{(n)}$ is not complete, although it is still sufficient for θ.
Thus, Theorem 3.1 does not apply to $X_{(n)}$.
We now illustrate how to use Theorem 3.2(ii) to find a UMVUE of θ.

Let $U(X_{(n)})$ be an unbiased estimator of 0.
Since $X_{(n)}$ has the Lebesgue p.d.f. $n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$,

$$0 = \int_0^1 U(x)x^{n-1}dx + \int_1^\theta U(x)x^{n-1}dx \quad \text{for all } \theta \geq 1.$$

This implies that $U(x) = 0$ a.e. Lebesgue measure on $[1, \infty)$ and

$$\int_0^1 U(x)x^{n-1}dx = 0.$$

Consider $T = h(X_{(n)})$.
To have $E(TU) = 0$, we must have

$$\int_0^1 h(x)U(x)x^{n-1}dx = 0.$$

Thus, we may consider the following function:

$$h(x) = \begin{cases} c & 0 \le x \le 1 \\ bx & x > 1, \end{cases}$$

where c and b are some constants.
From the previous discussion,

$$E[h(X_{(n)})U(X_{(n)})] = 0, \qquad \theta \ge 1.$$

Since $E[h(X_{(n)})] = \theta$, we obtain that

$$\begin{aligned} \theta &= cP(X_{(n)} \le 1) + bE[X_{(n)} I_{(1,\infty)}(X_{(n)})] \\ &= c\theta^{-n} + [bn/(n+1)](\theta - \theta^{-n}). \end{aligned}$$

Thus, $c = 1$ and $b = (n+1)/n$.
The UMVUE of θ is then

$$h(X_{(n)}) = \begin{cases} 1 & 0 \le X_{(n)} \le 1 \\ (1+n^{-1})X_{(n)} & X_{(n)} > 1. \end{cases}$$

- This estimator is better than $(1+n^{-1})X_{(n)}$, which is the UMVUE when $\Theta = (0,\infty)$ and does not make use of the information about $\theta \geq 1$.
- When $\Theta = (0,\infty)$, this estimator is not unbiased.
- In fact, $h(X_{(n)})$ is complete and sufficient for $\theta \in [1,\infty)$.

Example 3.8

Let X be a sample (of size 1) from the uniform distribution $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathscr{R}$.
We now apply Theorem 3.2 to show that there is no UMVUE of $\vartheta = g(\theta)$ for any nonconstant function g.

Note that an unbiased estimator $U(X)$ of 0 must satisfy

$$\int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} U(x)dx = 0 \qquad \text{for all } \theta \in \mathscr{R}.$$

Differentiating both sides of the previous equation and applying the result of differentiation of an integral lead to

$$U(x) = U(x+1) \quad \text{a.e. } m,$$

where m is the Lebesgue measure on $\mathscr{R}$.
If T is a UMVUE of $g(\theta)$, then $T(X)U(X)$ is unbiased for 0 and, hence,

$$T(x)U(x) = T(x+1)U(x+1) \quad \text{a.e. } m,$$

where $U(X)$ is any unbiased estimator of 0.
Since this is true for all U,

$$T(x) = T(x+1) \quad \text{a.e. } m.$$

Since T is unbiased for $g(\theta)$,

$$g(\theta) = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} T(x)dx \qquad \text{for all } \theta \in \mathscr{R}.$$

Differentiating both sides of the previous equation and applying the result of differentiation of an integral, we obtain that

$$g'(\theta) = T\left(\theta + \tfrac{1}{2}\right) - T\left(\theta - \tfrac{1}{2}\right) = 0 \quad \text{a.e. } m.$$

Hence g is a constant a.e.

Information inequality

Theorem 3.3 (Cramér-Rao lower bound)

Let $X = (X_1, ..., X_n)$ be a sample from $P \in \mathscr{P} = \{P_\theta : \theta \in \Theta\}$, where Θ is an open set in $\mathscr{R}^k$.
Suppose that $T(X)$ is an estimator with $E[T(X)] = g(\theta)$ being a differentiable function of θ; P_θ has a p.d.f. f_θ w.r.t. a measure ν for all $\theta \in \Theta$; and f_θ is differentiable as a function of θ and satisfies

$$\frac{\partial}{\partial \theta} \int h(x) f_\theta(x) d\nu = \int h(x) \frac{\partial}{\partial \theta} f_\theta(x) d\nu, \qquad \theta \in \Theta, \tag{1}$$

for $h(x) \equiv 1$ and $h(x) = T(x)$.
Then

$$\mathrm{Var}(T(X)) \geq \left[\tfrac{\partial}{\partial \theta} g(\theta)\right]^\tau [I(\theta)]^{-1} \tfrac{\partial}{\partial \theta} g(\theta), \tag{2}$$

where

$$I(\theta) = E\left\{ \frac{\partial}{\partial \theta} \log f_\theta(X) \left[\frac{\partial}{\partial \theta} \log f_\theta(X) \right]^\tau \right\}$$

is assumed to be positive definite for any $\theta \in \Theta$.

Discussion

Suppose that we have a lower bound for the variances of all unbiased estimators of ϑ.
If there is an unbiased estimator T of ϑ whose variance is always the same as the lower bound, then T is a UMVUE of ϑ.
Although this is not an effective way to find UMVUE's, it provides a way of assessing the performance of UMVUE's.

Proof of Theorem 3.3

We prove the univariate case ($k = 1$) only.
When $k = 1$, (2) reduces to

$$\mathrm{Var}(T(X)) \geq \frac{[g'(\theta)]^2}{E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2}.$$

From the Cauchy-Schwartz inequality, we only need to show that

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 = \mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)$$

Proof of Theorem 3.3 (continued)

and

$$g'(\theta) = \text{Cov}\left(T(X), \frac{\partial}{\partial\theta} \log f_\theta(X) \right).$$

From condition (1) with $h(x) = 1$,

$$E\left[\frac{\partial}{\partial\theta} \log f_\theta(X) \right] = \int \frac{\partial}{\partial\theta} f_\theta(X) d\nu = \frac{\partial}{\partial\theta} \int f_\theta(X) d\nu = 0.$$

From condition (1) with $h(x) = T(x)$,

$$E\left[T(X) \frac{\partial}{\partial\theta} \log f_\theta(X) \right] = \int T(x) \frac{\partial}{\partial\theta} f_\theta(X) d\nu = \frac{\partial}{\partial\theta} \int T(x) f_\theta(X) d\nu,$$

which $= g'(\theta)$.

The $k \times k$ matrix

$$I(\theta) = E\left\{ \frac{\partial}{\partial\theta} \log f_\theta(X) \left[\frac{\partial}{\partial\theta} \log f_\theta(X) \right]^\tau \right\}$$

is called the *Fisher information matrix*.

The greater $I(\theta)$ is, the easier it is to distinguish θ from neighboring values and, therefore, the more accurately θ can be estimated.

Thus, $I(\theta)$ is a measure of the information that X contains about θ.

The inequality in (2) is called *information inequalities*.

The following result is helpful in finding the Fisher information matrix.

Proposition 3.1

(i) If X and Y are independent with the Fisher information matrices $I_X(\theta)$ and $I_Y(\theta)$, respectively, then the Fisher information about θ contained in (X, Y) is $I_X(\theta) + I_Y(\theta)$.
In particular, if $X_1, ..., X_n$ are i.i.d. and $I_1(\theta)$ is the Fisher information about θ contained in a single X_i, then the Fisher information about θ contained in $X_1, ..., X_n$ is $nI_1(\theta)$.

(ii) Suppose that X has the p.d.f. f_θ that is twice differentiable in θ and that (1) holds with $h(x) \equiv 1$ and f_θ replaced by $\partial f_\theta / \partial \theta$.
Then

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta^\tau}\log f_\theta(X)\right].$$

Proof

Result (i) follows from the independence of X and Y and the definition of the Fisher information.

Result (ii) follows from the equality

$$\frac{\partial^2}{\partial\theta\partial\theta^\tau}\log f_\theta(X) = \frac{\frac{\partial^2}{\partial\theta\partial\theta^\tau} f_\theta(X)}{f_\theta(X)} - \frac{\partial}{\partial\theta}\log f_\theta(X)\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^\tau.$$

Example 3.9

Let $X_1,...,X_n$ be i.i.d. with the Lebesgue p.d.f. $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$, where $f(x) > 0$ and $f'(x)$ exists for all $x \in \mathscr{R}$, $\mu \in \mathscr{R}$, and $\sigma > 0$ (a location-scale family).

Let $\theta = (\mu,\sigma)$. Then, the Fisher information about θ contained in $X_1,...,X_n$ is (exercise)

$$I(\theta) = \frac{n}{\sigma^2}\begin{pmatrix} cc\int \frac{[f'(x)]^2}{f(x)}dx & \int \frac{f'(x)[xf'(x)+f(x)]}{f(x)}dx \\ \int \frac{f'(x)[xf'(x)+f(x)]}{f(x)}dx & \int \frac{[xf'(x)+f(x)]^2}{f(x)}dx \end{pmatrix}.$$

Remarks

- Note that $I(\theta)$ depends on the particular parameterization.
- If $\theta = \psi(\eta)$ and ψ is differentiable, then the Fisher information that X contains about η is

$$\frac{\partial}{\partial\eta}\psi(\eta) I(\psi(\eta)) \left[\frac{\partial}{\partial\eta}\psi(\eta)\right]^{\tau}.$$

- However, the Cramér-Rao lower bound in (2) is not affected by any one-to-one reparameterization.
- If we use inequality (2) to find a UMVUE $T(X)$, then we obtain a formula for $\mathrm{Var}(T(X))$ at the same time.
- On the other hand, the Cramér-Rao lower bound in (2) is typically not sharp.
- Under some regularity conditions, the Cramér-Rao lower bound is attained iff f_θ is in an exponential family; see Propositions 3.2 and 3.3 and the discussion in Lehmann (1983, p. 123).
- Some improved information inequalities are available (see, e.g., Lehmann (1983, Sections 2.6 and 2.7)).

Proposition 3.2.

Suppose that the distribution of X is from an exponential family $\{f_\theta : \theta \in \Theta\}$, i.e., the p.d.f. of X w.r.t. a σ-finite measure is

$$f_\theta(x) = \exp\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\}c(x), \quad (3)$$

where Θ is an open subset of $\mathscr{R}^k$.

(i) The regularity condition (1) is satisfied for any h with $E|h(X)| < \infty$ and

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta^\tau}\log f_\theta(X)\right].$$

(ii) If $\underline{I}(\eta)$ is the Fisher information matrix for the natural parameter η, then the variance-covariance matrix $\mathrm{Var}(T) = \underline{I}(\eta)$.

(iii) If $\bar{I}(\vartheta)$ is the Fisher information matrix for the parameter $\vartheta = E[T(X)]$, then $\mathrm{Var}(T) = [\bar{I}(\vartheta)]^{-1}$.

A direct consequence of Proposition 3.2(ii) is that the variance of any linear function of T in (3) attains the Cramér-Rao lower bound.

Proof

(i) This is a direct consequence of Theorem 2.1.

(ii) The p.d.f. under the natural parameter η is

$$f_\eta(x) = \exp\{\eta^\tau T(x) - \zeta(\eta)\}\, c(x).$$

From Theorem 2.1, $E[T(X)] = \frac{\partial}{\partial \eta}\zeta(\eta)$.
The result follows from

$$\frac{\partial}{\partial \eta}\log f_\eta(x) = T(x) - \frac{\partial}{\partial \eta}\zeta(\eta).$$

(iii) Since $\vartheta = E[T(X)] = \frac{\partial}{\partial \eta}\zeta(\eta)$,

$$\underline{I}(\eta) = \frac{\partial \vartheta}{\partial \eta}\bar{I}(\vartheta)\left(\frac{\partial \vartheta}{\partial \eta}\right)^\tau = \frac{\partial^2}{\partial \eta \partial \eta^\tau}\zeta(\eta)\bar{I}(\vartheta)\left[\frac{\partial^2}{\partial \eta \partial \eta^\tau}\zeta(\eta)\right]^\tau.$$

By Theorem 2.1 and the result in (ii),

$$\frac{\partial^2}{\partial \eta \partial \eta^\tau}\zeta(\eta) = \mathrm{Var}(T) = \underline{I}(\eta).$$

Hence

$$\bar{I}(\vartheta) = [\underline{I}(\eta)]^{-1}\underline{I}(\eta)[\underline{I}(\eta)]^{-1} = [\underline{I}(\eta)]^{-1} = [\mathrm{Var}(T)]^{-1}.$$

Example 3.10

Let $X_1, ..., X_n$ be i.i.d. from the $N(\mu, \sigma^2)$ distribution with an unknown $\mu \in \mathscr{R}$ and a known σ^2.

Let f_μ be the joint distribution of $X = (X_1, ..., X_n)$.

Then

$$\frac{\partial}{\partial\mu} \log f_\mu(X) = \sum_{i=1}^n (X_i - \mu)/\sigma^2.$$

Thus, $I(\mu) = n/\sigma^2$.

Consider the estimation of μ.

It is obvious that $\mathrm{Var}(\bar{X})$ attains the Cramér-Rao lower bound in (2).

Consider now the estimation of $\vartheta = \mu^2$.

Since $E\bar{X}^2 = \mu^2 + \sigma^2/n$, the UMVUE of ϑ is $h(\bar{X}) = \bar{X}^2 - \sigma^2/n$.

A straightforward calculation shows that

$$\mathrm{Var}(h(\bar{X})) = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}.$$

On the other hand, the Cramér-Rao lower bound in this case is $4\mu^2\sigma^2/n$: $\mathrm{Var}(h(\bar{X}))$ does not attain the Cramér-Rao lower bound.

The difference is $2\sigma^4/n^2$.