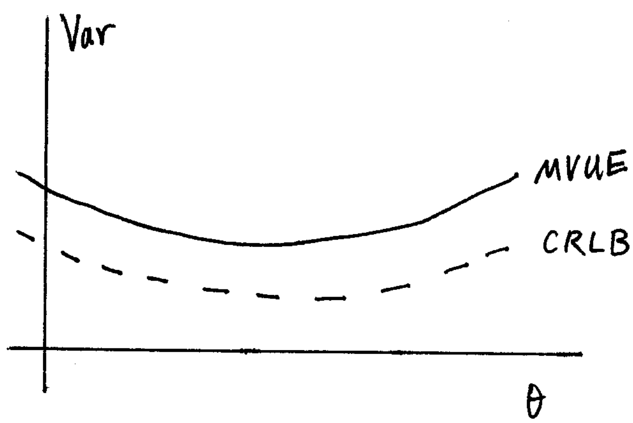
MVUE VIA THE RAO-BLACKWELL THEOREM AND COMPLETE SUFFICIENT STATISTICS

The Cramer-Rao lower bound gives a necessary and sufficient condition for the existence of an efficient estimator.



However, MVUE's are not necessarily efficient. What can we do in such cases?

The Rao-Blackwell theorem, when applied in conjunction with a complete suff. stat., gives another way to find MVUE's that applies even when the CRLB is not defined.

Rao-Blackwell Theorem, Version 2

<u>Theorem</u> Let $\underline{Y}, \underline{Z}$ be random variables and define the function

$$g(\underline{z}) = E[\underline{Y} \mid \underline{Z} = \underline{z}].$$

Then

$$E[g(\underline{Z})] = E[\underline{Y}]$$

and

$$Var\left(g(\underline{Z})\right) \leq Var\left(\underline{Y}\right)$$

with equality iff $\underline{Y} = g(\underline{Z})$ almost surely.

Note that this version of R.B. is quite general and has nothing to do with estimation of parameters. However, we can apply it to parameter estimation as follows.

Consider $\underline{X} \sim f_{\underline{\theta}}(\underline{x})$. Let $\hat{\underline{\theta}}_1$ be an unbiased estimator of $\underline{\theta}$, and let $\underline{T} = \tau(\underline{X})$ be a sufficient statistic for $\underline{\theta}$. Apply Rao-Blackwell with

$$\underline{Y} = \hat{\underline{\theta}}_1(\underline{X})$$

$$\underline{Z} = \underline{T} = \tau(\underline{X}).$$

Consider the new estimator

$$\underline{\hat{\theta}}_2(\underline{x}) = g(\tau(\underline{x}))$$

$$= E\left[\underline{\hat{\theta}}_1(\underline{X}) \mid \underline{T} = \tau(\underline{x})\right].$$

Then we may conclude

① $\underline{\hat{\theta}}_2$ is unbiased

② $Var_{\underline{\theta}}(\underline{\hat{\theta}}_2) \le Var_{\underline{\theta}}(\underline{\hat{\theta}}_1)$

In words, if $\underline{\hat{\theta}}_1$ is any unbiased estimator, then smoothing $\underline{\hat{\theta}}_1$ w.r.t. a sufficient stat decreases the variance while preserving unbiasedness.

Therefore, we can restrict our search for the MVUE to functions of a sufficient statistic.

Proof of Rao-Blackwell, Version 2

First we must show $E[g(\underline{Z})] = E[\underline{Y}]$.

This follows by the law of total expectation:

$$E[g(\underline{Z})] = E[E[\underline{Y} \mid \underline{Z}]] = E[\underline{Y}]$$

Second we must show

$$E[(g(\underline{Z})-\underline{\theta})^T (g(\underline{Z})-\underline{\theta})] \leq E[(\underline{Y}-\underline{\theta})^T(\underline{Y}-\underline{\theta})],$$

where $\underline{\theta} = E[\underline{Y}] = E[g(\underline{Z})]$. To see this, write

$$\begin{aligned}
\mathrm{Var}(\underline{Y}) &= E[(\underline{Y}-\underline{\theta})^T(\underline{Y}-\underline{\theta})] \\
&= E[(\underline{Y}-g(\underline{Z}) + g(\underline{Z})-\underline{\theta})^T(\underline{Y}-g(\underline{Z})+g(\underline{Z})-\underline{\theta})] \\
&= E[(\underline{Y}-g(\underline{Z}))^T(\underline{Y}-g(\underline{Z}))] + \mathrm{Var}(g(\underline{Z})) \\
&\quad + 2E[(\underline{Y}-g(\underline{Z}))^T(g(\underline{Z})-\underline{\theta})]
\end{aligned}$$

Consider the third term:

$$E[(\underline{Y}-g(\underline{Z}))^T(g(\underline{Z})-\underline{\theta})]$$

$$=$$

(a)

Thus, we have shown

$$\begin{aligned}
Var(\underline{Y}) &= Var(g(\underline{Z})) + E\left[(\underline{Y}-g(\underline{Z}))^T(\underline{Y}-g(\underline{Z}))\right] \\
&= Var(g(\underline{Z})) + E\left[\|\underline{Y}-g(\underline{Z})\|^2\right] \\
&\geq Var(g(\underline{Z}))
\end{aligned}$$

with equality iff $\underline{Y} = g(\underline{Z})$ with probability one. ∎

<u>Example 1</u> Consider $\underline{X} \sim N(\theta \cdot \underline{1}, \sigma^2 I_{N\times N})$, σ^2 known. Let $\hat{\theta}_1(\underline{x}) = x_1$. Then

$$E[\hat{\theta}_1] = \theta$$

$$Var(\hat{\theta}_1) = \sigma^2.$$

Consider the sufficient statistic $T = \sum_{i=1}^{N} X_i$ and define

$$\hat{\theta}_2(\underline{x}) = E\left[\hat{\theta}_1(\underline{X}) \mid T = \Sigma x_i\right]$$

How can we find a formula for $\hat{\theta}_2$?

Observe that X_1, T are jointly Gaussian:

$$\begin{bmatrix} X_1 \\ T \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{A} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

Then

$$\begin{bmatrix} X_1 \\ T \end{bmatrix} \sim N\left(A \cdot \theta \underline{1} \, , \, A \cdot \sigma^2 I_{N \times N} \cdot A^T \right)$$

$$\sim N\left(\begin{bmatrix} \theta \\ N\theta \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 1 \\ 1 & N \end{bmatrix} \right)$$

Recall the following property of the MVG: If

$$\underline{W} = \begin{bmatrix} \underline{U} \\ \underline{V} \end{bmatrix} \sim N\left(\begin{bmatrix} \underline{\mu}_u \\ \underline{\mu}_v \end{bmatrix}, \begin{bmatrix} R_{uu} & R_{uv} \\ R_{vu} & R_{vv} \end{bmatrix} \right)$$

Then

$$\underline{U} \mid \underline{V} = \underline{v} \sim N\Big(\underline{\mu}_u + R_{uv} R_{vv}^{-1} (\underline{v} - \underline{\mu}_v), \; R_{uu} - R_{uv} R_{vv}^{-1} R_{vu} \Big)$$

Applying this to $\begin{bmatrix} X_1 \\ T \end{bmatrix}$ we obtain

$$X_1 | T = t \sim N\left(\theta + 1 \cdot \frac{1}{N}(t - N\theta),\ \sigma^2\left(1 - 1 \cdot \frac{1}{N} \cdot 1\right)\right)$$

$$\sim N\left(\frac{t}{N},\ \sigma^2\left(1 - \frac{1}{N}\right)\right)$$

Therefore

$$\hat{\theta}_2(\underline{x}) = E\left[X_1 \mid T = \Sigma x_i\right]$$

$$= \frac{1}{N} \sum_{i=1}^{N} x_i$$

Notice the reduction in variance:

$$Var(\hat{\theta}_2) = \frac{\sigma^2}{N} < \sigma^2 = Var(\hat{\theta}_1).$$

The Rao-Blackwell Theorem tells us how to decrease the variance of an unbiased estimator. But when can we know that we get a MVUE?

The answer: When $\underline{T}$ is a <u>complete</u> suff. stat.

<u>Theorem</u> (Lehmann-Scheffe)

If $\underline{T}$ is complete, there is at most one unbiased estimator that is a function of $\underline{T}$.

<u>Proof</u> Suppose $E[\hat{\underline{\theta}}_1] = E[\hat{\underline{\theta}}_2] = \underline{\theta}$ and

$\hat{\underline{\theta}}_1(\underline{x}) = g_1(\tau(\underline{x}))$, $\hat{\underline{\theta}}_2(\underline{x}) = g_2(\tau(\underline{x}))$.

Define

$$\phi(\underline{t}) = g_1(\underline{t}) - g_2(\underline{t}).$$

Then

$$E\{\phi(\underline{T})\} = E\{g_1(\underline{T}) - g_2(\underline{T})\}$$

$$= E\{\hat{\underline{\theta}}_1(\underline{X}) - \hat{\underline{\theta}}_2(\underline{X})\}$$

$$= \underline{0}.$$

By definition of completeness,

$$P(\phi(\underline{I}) = 0) = 1 \qquad \forall \underline{\theta}.$$

In other words,

$$\hat{\underline{\theta}}_1 = \hat{\underline{\theta}}_2 \quad \text{with prob. } 1$$

$\blacksquare$

This result suggests the following recipe for finding an MVUE:

1.) Find a complete sufficient statistic $\underline{I} = \tau(\underline{X})$

2.a) Find any unbiased estimator $\hat{\underline{\theta}}'$ and set

$$\hat{\underline{\theta}}(x) = E\{\hat{\underline{\theta}}'(\underline{X}) \mid \underline{I} = \tau(\underline{x})\}$$

2.b) Find a function g such that

$$\hat{\underline{\theta}}(\underline{x}) = g(\tau(\underline{x}))$$

is unbiased.

<u>Theorem</u> If $\hat{\underline{\theta}}$ is constructed by the recipe above, then $\hat{\underline{\theta}}$ is the <u>unique</u> MVUE.

<u>Proof</u> Note that in either construction, $\underline{\hat{\theta}}$ is a function of $\underline{T}$.

Let $\underline{\hat{\theta}}_1$ be any unbiased estimator. We must show

$$\mathrm{Var}(\underline{\hat{\theta}}) \le \mathrm{Var}(\underline{\hat{\theta}}_1).$$

Define

$$\underline{\hat{\theta}}_2(\underline{x}) = E\{\underline{\hat{\theta}}_1(\underline{X}) \mid \underline{T} = \tau(\underline{x})\}.$$

By Rao-Blackwell, it suffices to show

$$\mathrm{Var}(\underline{\hat{\theta}}) \le \mathrm{Var}(\underline{\hat{\theta}}_2).$$

But $\underline{\hat{\theta}}$ and $\underline{\hat{\theta}}_2$ are both unbiased and functions of a complete s.s.

$$\Longrightarrow \quad \underline{\hat{\theta}} = \underline{\hat{\theta}}_2 \qquad \text{w.p. } 1.$$

To show uniqueness, in the above argument suppose $\mathrm{Var}(\underline{\hat{\theta}}_1) = \mathrm{Var}(\underline{\hat{\theta}})$. Then the Rao-Blackwell bound holds with equality

$$\Longrightarrow \quad \underline{\hat{\theta}}_1 = \underline{\hat{\theta}}_2 \qquad \text{w.p. } 1$$

$$\Longrightarrow \quad \underline{\hat{\theta}}_1 = \underline{\hat{\theta}} \qquad \text{w.p. } 1$$

because $\underline{\hat{\theta}}_2 = \underline{\hat{\theta}}$ w.p. 1. $\blacksquare$

We have seen previously that sufficient statistics arising from the exponential family of distributions are complete. Typically, however, MVUE's for the exponential family can be found using the CRLB.

A strength of the Rao-Blackwell approach is that it can produce MVUE's even when CRLB can't.

<u>Example</u> Suppose $\underline{x} = [x_1 \cdots x_N]^T$ where

$$x_i \overset{iid}{\sim} \text{unif}[0, \theta], \qquad i = 1, \ldots, N.$$

Note that the CRLB cannot be applied because

$$\log f(\underline{x}; \theta)$$

is not differentiable w.r.t θ.

What is an unbiased estimator of θ?

$$\hat{\theta}_1 = \frac{2}{N}\sum_{i=1}^{N} X_i$$

is unbiased. However, it is not MVUE.

From

$$f_\theta(\underline{x}) = \prod_{i=1}^{N} \frac{1}{\theta} I_{[0,\theta]}(x_i)$$

$$= \underbrace{\frac{1}{\theta^N} I_{[\max x_i, \infty)}(\theta)}_{g_\theta(\underline{t})} \cdot \underbrace{I_{(-\infty, \min x_i]}(0)}_{h(\underline{x})}$$

we see that

$$T = \max_i X_i$$

is a sufficient statistic. It is left as an exercise to show that T is in fact complete.

Since $\hat{\theta}_1$ is not a function of T, it is not MVUE.

However

$$\hat{\theta}_2(\underline{x}) = E\{\hat{\theta}_1(\underline{X}) \mid T = \tau(\underline{x})\}$$

is the MVUE.

It is also left as an exercise to find the precise form of $\hat{\theta}_2$.

$\boxed{\text{Summary}}$

- Rao - Blackwell Thm
  - decreases estimator variance by conditioning on a sufficient statistic
  - filters out the randomness (noise) in the data not captured by suff. stat.

- Complete suff. stat. $\Longrightarrow$ Rao- Blackwellization results in ~~the~~ unique MVUE

Key

a. $$\mathbf{E}\left[\mathbf{E}\left[(\underline{Y} - g(\underline{Z}))^T (g(\underline{Z}) - \underline{\theta}) \mid \underline{Z}\right]\right]$$

$$= E\left[\mathbf{E}\left[(\underline{Y} - g(\underline{Z}))^T \mid \underline{Z}\right] \cdot (g(\underline{Z}) - \underline{\theta})\right]$$

$$= E\left[(g(\underline{Z}) - g(\underline{Z}))^T (g(\underline{Z}) - \underline{\theta})\right]$$

$$= 0$$