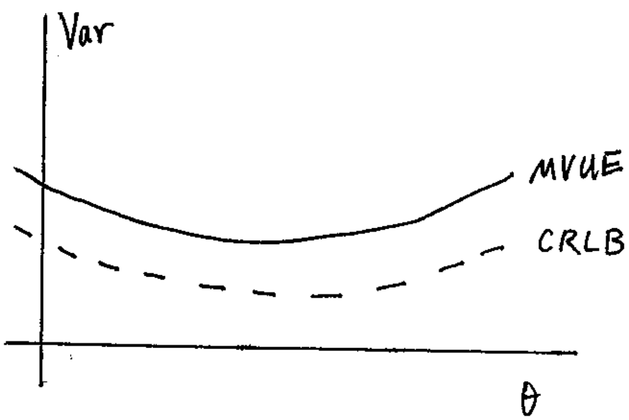
MVUE VIA THE RAO-BLACKWELL THEOREM AND COMPLETE SUFFICIENT STATISTICS

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.

$\boxed{\text{Rao-Blackwell Theorem}}$

Theorem | 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

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

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 $\hat{\underline{\theta}}$ be an unbiased estimator of $\underline{\theta}$, and let $\underline{I} = \tau(\underline{X})$ be a sufficient statistic for $\underline{\theta}$. Apply Rao-Blackwell with

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

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

Consider the new estimator

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

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

Then we may conclude

① $\underline{\hat{\theta}}_{RB}$ is unbiased

② $Var_{\underline{\theta}}(\underline{\hat{\theta}}_{RB}) \leq Var_{\underline{\theta}}(\underline{\hat{\theta}})$

In words, if $\underline{\hat{\theta}}$ is any unbiased estimator, then smoothing $\underline{\hat{\theta}}$ 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.

Why did we assume $\underline{Z}$ is a sufficient statistic

ⓐ as opposed to some more general function of $\underline{X}$?

Proof

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

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

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

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

Consider the third term:

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

$$=$$

(b)

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. ∎

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

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

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

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

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

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

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

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

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

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

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

Therefore

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

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

Notice the reduction in variance:

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

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 $\mathcal{I}$ is a <u>complete</u> suff. stat.

Complete Sufficient Statistics

<u>Definition</u> A sufficient statistic $\mathcal{I} = \tau(\underline{X})$ is <u>complete</u> iff for all real-valued functions ϕ

$$\left(E_{\underline{\theta}}\left[\phi(\mathcal{I}) \right] = 0 \quad \forall \underline{\theta} \right) \Longrightarrow \left(P_{\underline{\theta}}\left[\phi(\mathcal{I}) = 0 \right] = 1 \ \forall \underline{\theta} \right)$$

<u>Example</u> Bernoulli trials, part IV

Consider N independent Bernoulli trials

$$X_i \overset{iid}{\sim} \text{Bernoulli}(\theta) \qquad \theta \in [0,1]$$

Recall $K = \sum_{i=1}^{N} X_i$ is sufficient for θ.

Suppose $E_\theta[\phi(K)] = 0 \quad \forall \theta$. But

$$E[\phi(K)] = \sum_{k=0}^{N} \phi(k) \binom{N}{k} \theta^k (1-\theta)^{N-k}$$

$$= \text{poly}(\theta)$$

where $\text{poly}(\theta)$ is an N^{th} degree polynomial.

Then $\quad poly(\theta) = 0 \quad \forall \theta \in [0,1]$

$\Longrightarrow poly(\theta)$ is the zero polynomial

$\Longrightarrow \phi(k) = 0 \quad \forall k$

$\Longrightarrow K$ is complete

Note/ The definition of completeness depends on the parameter space Θ. In the last example we had $\Theta = [0,1]$. What would happen if $\Theta = \{\frac{1}{3}, \frac{2}{3}\}$?

The Lehmann–Scheffé Theorem

Theorem/ (Lehmann–Scheffé)

If T is complete, there is at most one unbiased estimator that is a function of T.

Proof/ 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,

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

In other words,

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

This result suggests the following recipe for finding an MVUE:

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

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

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

OR 2.b) Find a function g such that

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

is unbiased.

Corollary If $\hat{\underline{\theta}}_{RB}$ is constructed by the preceeding recipe, then $\hat{\underline{\theta}}_{RB}$ is the <u>unique</u> MVUE.

Proof By the Lehmann-Scheffé Theorem, there is at most one unbiased estimator that is a function of a complete suff. stat. $\hat{\underline{\theta}}_{RB}$ is such an estimator.

Let $\hat{\underline{\theta}}'$ be an unbiased estimator such that

$$Var(\hat{\underline{\theta}}') \leq Var(\hat{\underline{\theta}}_{RB})$$

Set

$$\hat{\underline{\theta}}'_{RB}(\underline{x}) = E[\hat{\underline{\theta}}'(\underline{X}) \mid \underline{T} = \tau(\underline{x})].$$

Then $\hat{\underline{\theta}}'_{RB}$ is unbiased, $Var(\hat{\underline{\theta}}'_{RB}) \leq Var(\hat{\underline{\theta}}')$, and $\hat{\underline{\theta}}'_{RB}$ is a function of $\underline{T}$. It follows that $\hat{\underline{\theta}}'_{RB} = \hat{\underline{\theta}}_{RB}$ w.p. 1 and so

$$Var(\hat{\underline{\theta}}'_{RB}) = Var(\hat{\underline{\theta}}') = Var(\hat{\underline{\theta}}_{RB}).$$

This shows $\hat{\underline{\theta}}_{RB}$ has minimum variance. To establish unique, the condition for equality in the Rao-Blackwell Theorem tells us $\hat{\underline{\theta}}' = \hat{\underline{\theta}}'_{RB}$ w.p. 1, which implies $\hat{\underline{\theta}}' = \hat{\underline{\theta}}_{RB}$ w.p. 1 ∎

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?

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

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

However

$$\hat{\theta}_{RB}(\underline{x}) = E\{\hat{\theta}_1(\underline{X}) \mid T = t(\underline{x})\}$$

is the MVUE.

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

The Exponential Family

In general, sufficient statistics, especially ones that are minimal and complete, can be difficult to find (if they even exist).

For a special family of distributions, however, we can immediately identify a complete and minimal suff. stat.

<u>Definition</u> We say the distribution of $\underline{X}$ belongs to the <u>exponential family</u> of distributions if its pdf/pmf can be written

$$f_{\underline{\theta}}(\underline{x}) = a(\underline{\theta})\, b(\underline{x}) \exp\left\{ c(\underline{\theta})^T \tau(\underline{x}) \right\}$$

for some a, b, c and τ, where the dimension p of $\underline{\theta}$ is also the dimension of $c(\underline{\theta})$ and $\tau(\underline{x})$.

Example | Bernoulli trials, part V

$$p_\theta(\underline{x}) = \theta^k (1-\theta)^{N-k} \qquad \left[k = \sum_{i=1}^{N} x_i\right]$$

$$= \exp\{\log(\theta^k (1-\theta)^{N-k})\}$$

$$= \exp\{k \log\theta + (N-k)\log(1-\theta)\}$$

$$= \underbrace{\exp\{N\log(1-\theta)\}}_{a(\theta)} \cdot \exp\{\underbrace{[\log\theta - \log(1-\theta)]}_{c(\theta)} \cdot \underbrace{k}_{\tau(x)}\}$$

$$[b(\underline{x}) = 1]$$

Many common distributions belong to the exponential family, including Gaussian w/ unknown mean and/or variance, Poisson, exponential, gamma, binomial, and multinomial.

Exercise | Suppose $\underline{X} = [X_1, \ldots, X_N]^T$ where

$$X_i \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$$

Recall that if $X \sim \text{Gamma}(\alpha, \beta)$, then

$$f_{\underline{\theta}}(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot x^{\alpha - 1} e^{-\beta x}, \quad x \geq 0$$

where $\underline{\theta} = [\alpha \;\; \beta]^T$. Show that the distribution of $\underline{X}$ belongs to the exponential family.

Solution

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

$$= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^N \cdot \prod_{i=1}^{N} x_i^{\alpha-1} \exp\left\{-\beta \sum_{i=1}^{N} x_i\right\}$$

$$= \left(\frac{\beta}{\Gamma(\alpha)}\right)^N \exp\left\{\log\left[\left(\prod x_i\right)^{\alpha-1}\right] - \beta \sum x_i\right\}$$

$$= \left(\frac{\beta}{\Gamma(\alpha)}\right)^N \exp\left\{(\alpha-1)\sum_{i=1}^{N} \log x_i - \beta \sum x_i\right\}$$

$$\Longrightarrow = \underbrace{\left(\frac{\beta}{\Gamma(\alpha)}\right)^N}_{a(\underline{\theta})} \underbrace{\left(\prod_{i=1}^{N} x_i\right)^{-1}}_{b(\underline{x})} \exp\left\{\underbrace{\begin{bmatrix}\alpha & -\beta\end{bmatrix}}_{c(\underline{\theta})^T} \underbrace{\begin{bmatrix}\sum \log x_i \\ \sum x_i\end{bmatrix}}_{\tau(\underline{x})}\right\}$$

[Note: representation is not unique.]

Proposition | If the distribution of $\underline{X}$ belongs to the exponential family, then $\underline{T} = \tau(\underline{X})$ is a sufficient statistic for $\underline{\theta}$.

Proof | $\underline{T}$ is sufficient for $\underline{\theta}$ by the FNFT:

$$f_{\underline{\theta}}(\underline{x}) = a(\underline{\theta})\, b(\underline{x}) \exp\left\{ c(\underline{\theta})^T \tau(\underline{x}) \right\}$$

$$= \underbrace{a(\underline{\theta}) \exp\left\{ c(\underline{\theta})^T \tau(\underline{x}) \right\}}_{g_{\theta}(\tau(\underline{x}))} \cdot \underbrace{b(\underline{x})}_{h(\underline{x})}$$

Proposition | Under certain "reasonable" conditions, $\underline{T} = \tau(\underline{X})$ is a complete and minimal sufficient statistic for $\underline{\theta}$.

Sketch of proof | Before we argued that the pdf/pmf of $\underline{X}$ depends on $\underline{\theta}$ only through $f_{\underline{\theta}}(\underline{t})$.

Thus

$$f_{\underline{\theta}}(\underline{t}) \propto \exp\left\{ c(\underline{\theta})^T \underline{t} \right\}$$

Suppose ϕ is a real-valued function such that

$$E_{\underline{\theta}}\left\{ \phi(\underline{T}) \right\} = 0 \qquad \forall \theta.$$

We must show $P_{\theta}\left\{ \phi(\underline{T}) = 0 \right\} = 1 \quad \forall \theta.$

For each $\underline{\theta}$ we can write

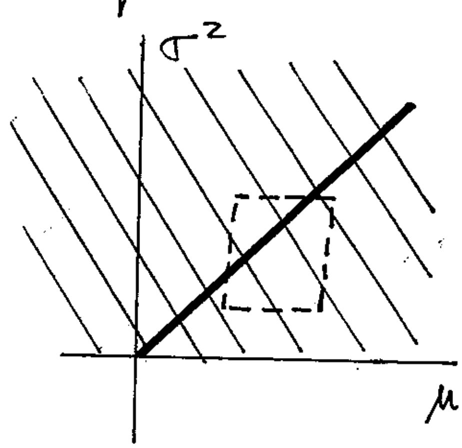
$$
\begin{aligned}
0 &= E\{\phi(\underline{T})\} \\
&= \int \phi(\underline{t})\, f_{\underline{\theta}}(\underline{t})\, d\underline{t} \\
&\propto \int \phi(\underline{t}) \exp\{c(\underline{\theta})^T \underline{t}\}\, d\underline{t}
\end{aligned}
$$

which is the Laplace transform of ϕ at $c(\underline{\theta})$. Inverting the Laplace transform we find $\phi \equiv 0$. ∎

The "reasonable" conditions under which the above arguments hold are needed to ensure the uniqueness (invertibility) of the Laplace transform:

- The parameter space Θ must contain an open rectangle: think back to Bernoulli example. As another example, in Gaussian case with $\mu = \sigma^2$, $\underline{T}$ is not complete.

- The image of $c(\underline{\theta})$, $\underline{\theta} \in \Theta$, should have full dimensionality.



Note that minimality of $\underline{T} = \underline{\tau}(\underline{X})$ for the exponential family follows from completeness – see Appendix.

Efficiency

It can be shown that an efficient estimator exists iff $f(\underline{x};\underline{\theta})$ is in the exponential family and has a <u>mean</u> <u>parameterization</u>, i.e. $E[\underline{\tau}(\underline{X})] = \underline{\theta}$.

See Hero or Poor for details.

Summary

Rao-Blackwell Theorem

- decreases estimator variance by conditioning on a sufficient statistic
- filter out the noise in the data not captured by the suff. stat.
- results in the MVUE when applied with a complete suff. stat.

Key

(a) So that the distribution of $\hat{\underline{\theta}}$ given $\underline{T}$ is independent of $\underline{\theta}$, and therefore $\hat{\underline{\theta}}_{RB}$ can be computed without knowledge of $\underline{\theta}$.

(b)
$$E\left[E\left[(\underline{Y} - g(\underline{Z}))^T (g(\underline{Z}) - \underline{\theta}) \mid \underline{Z}\right]\right.$$
$$= E\left[E\left[(\underline{Y} - g(\underline{Z}))^T \mid \underline{Z}\right] \cdot (g(\underline{Z}) - \underline{\theta})\right)$$
$$= E\left[(g(\underline{Z}) - g(\underline{Z}))^T \cdot (g(\underline{Z}) - \underline{\theta})\right]$$
$$= 0$$

Appendix: Minimal Sufficient Statistics

If we observe $\underline{x}$, then $\underline{x}$ itself is a sufficient statistic, albeit not a very interesting one. When is a suff. stat. as compressed as it can possibly be?

<u>Definition</u> A sufficient statistic is <u>minimal</u> if it is a function of every other sufficient statistic.

<u>Example</u>

For iid Gaussian observations with unknown mean, the following statistics are sufficient:

- $\underline{x} = [x_1, \cdots, x_n]^T$
- $[x_1 + x_3 + \cdots, \; x_2 + x_4 + \cdots]^T$
- $[\bar{x}, \; s^2]^T$
- $\bar{x}$

However, the first 3 are not minimal because they are not functions of the 4th.

Since $\bar{x}$ is 1-dimensional, it is minimal.

Remark | When we say "is a function of every other sufficient statistic," we exclude functions that increase dimensionality. Otherwise

$$[\bar{x}, \bar{x}]^T$$

would be sufficient in the previous example.

The dimension of a minimal suff. stat. cannot be less than the dimension of $\underline{\theta}$. If we are lucky the dimensions will be equal. Sometimes, the dim. of a minimal suff. stat. is as large as N. See, for example, the Cauchy distribution.

Proving Minimality

Proposition | $T = \tau(\underline{X})$ is a minimal suff. stat. if

$$\frac{f_{\underline{\theta}}(\underline{x})}{f_{\underline{\theta}}(\underline{y})} \text{ is independent of } \underline{\theta} \iff \tau(\underline{x}) = \tau(\underline{y})$$

Proof | First let's show T is a sufficient stat. under the given assumption.

For each $\underline{t}$ in the range of τ, assign a vector $\underline{y}(\underline{t})$ such that $\tau(\underline{y}(\underline{t})) = \underline{t}$.

Then

$$f_{\underline{\theta}}(\underline{x}) = \frac{f_{\underline{\theta}}(\underline{x})}{f_{\underline{\theta}}(\underline{y}(\tau(\underline{x})))} \cdot f_{\underline{\theta}}(\underline{y}(\tau(\underline{x})))$$

$$= h(\underline{x}) \cdot g_{\theta}(\tau(\underline{x}))$$

independent of θ since $\tau(\underline{x}) = \tau(\underline{y}(\tau(\underline{x})))$.

To show T is minimal, we must show that for any other suff. stat. $T' = \tau'(\underline{X})$, T is a function of T'.

So suppose T' takes on the value $\underline{t}'$. We must show that T is uniquely determined by $\underline{t}'$.

That is, if $\underline{x}$ and $\underline{y}$ are such that $\tau'(\underline{x}) = \tau'(\underline{y}) = \underline{t}'$, then $\tau(\underline{x}) = \tau(\underline{y})$.

So suppose $\underline{x}$ any $\underline{y}$ are such that $\tau'(\underline{x}) = \tau'(\underline{y}) = \underline{t}'$.

Then

$$\frac{f_{\underline{\theta}}(\underline{x})}{f_{\underline{\theta}}(\underline{y})} = \frac{g'_{\underline{\theta}}(\tau'(\underline{x})) \cdot h'(\underline{x})}{g'_{\underline{\theta}}(\tau'(\underline{y})) \cdot h'(\underline{y})} = \frac{h'(\underline{x})}{h'(\underline{y})}$$

which is independent of $\underline{\theta}$. Therefore $\tau(\underline{x}) = \tau(\underline{y})$ ▨

Exercise | Show that $[\Sigma X_i^2, \Sigma X_i]^T$ is a minimal suff. stat. for $[\mu, \sigma^2]^T$, where $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$.

Solution $\underline{\theta} = [\mu, \sigma^2]^T$.

$$\frac{f_{\underline{\theta}}(\underline{x})}{f_{\underline{\theta}}(\underline{y})} = \frac{\exp\left\{-\frac{1}{2\sigma^2}\left[\sum x_i^2 - 2\mu \sum x_i + n\mu^2\right]\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left[\sum y_i^2 - 2\mu \sum y_i + n\mu^2\right]\right\}}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left[\sum x_i^2 - \sum y_i^2\right] + \frac{\mu}{\sigma^2}\left[\sum x_i - \sum y_i\right]\right\}$$

which is independent of $\underline{\theta} \iff \left[\sum x_i^2 \;\; \sum x_i\right]^T = \left[\sum y_i^2 \;\; \sum y_i\right]^T$

$$\iff \tau(\underline{x}) = \tau(\underline{y}).$$

Since $[\bar{X} \;\; \bar{S}^2]^T$ is a function of $[\sum X_i^2 \;\; \sum X_i]^T$, it is also minimal. ∎

A second way to show that a suff. stat. is minimal is to show that it is <u>complete</u>.

<u>Proposition</u> Under very general conditions, if $\underline{T}$ is a complete s.s., then $\underline{T}$ is minimal.

<u>Proof</u> Let $\underline{T}' = \tau'(\underline{X})$ be any s.s.

We need to show $\underline{T}$ is determined completely by $\underline{T}'$.

Define

$$\psi(\underline{t}') := E_{\underline{\theta}}[\underline{T} \mid \underline{T}' = \underline{t}'].$$

We will show in particular that $\underline{T} = \psi(\underline{T}')$.

Introduce

$$\rho(\underline{t}) := E_{\underline{\theta}}[\psi(\underline{T}') \mid \underline{T} = \underline{t}]$$

Note that

$$\begin{aligned} E_{\underline{\theta}}[\underline{T}] &= E[E[\underline{T} \mid \underline{T}']] \\ &= E[\psi(\underline{T}')] \\ &= E[E[\psi(\underline{T}') \mid \underline{T}]] \\ &= E[\rho(\underline{T})] \end{aligned}$$

(law of total expectation)

By completeness, we deduce

$$P_{\underline{\theta}}\left[\,\underline{T} = \rho(\underline{T})\,\right] = 1 \quad \forall\,\underline{\theta} \tag{1}$$

This implies

$$\psi(\underline{t}') = E\left[\,\underline{T} \mid \underline{T}' = \underline{t}'\,\right] = E\left[\rho(\underline{T}) \mid \underline{T}' = \underline{t}'\right] \tag{2}$$

Now recall the "law of total variance":

$$\mathrm{Var}[Y] = E\left[\mathrm{Var}[Y \mid Z]\right] + \mathrm{Var}\left[E[Y \mid Z]\right]$$

for scalar random variables Y and Z. Using the subscript "i" to denote the i^{th} component, we have

$$\begin{aligned}
\mathrm{Var}[\rho_i(\underline{T})] &= E\left[\mathrm{Var}[\rho_i(\underline{T}) \mid \underline{T}']\right] + \mathrm{Var}\left[E[\rho_i(\underline{T}) \mid \underline{T}']\right] \\
&= \qquad '' \qquad + \mathrm{Var}[\psi_i(\underline{T}')] \qquad (\text{by } (2)) \\
&= \qquad '' \qquad + E\left[\mathrm{Var}[\psi_i(\underline{T}') \mid \underline{T}]\right] + \mathrm{Var}\left[E[\psi_i(\underline{T}') \mid \underline{T}]\right] \\
&= \qquad '' \qquad + \qquad '' \qquad + \mathrm{Var}[\rho_i(\underline{T})]
\end{aligned}$$

Since $\mathrm{Var}[\text{anything}] \geq 0$ we deduce

$$\mathrm{Var}\left[\rho_i(\underline{T}) \mid \underline{T}'\right] = 0 \qquad \text{with prob. } 1$$

$$\mathrm{Var}\left[\psi_i(\underline{T}') \mid \underline{T}\right] = 0 \qquad \text{with prob. } 1$$

How does this imply the desired result?

$\rho_i(\underline{I})$ is a deterministic function of $\underline{I}'$

In particular (returning to vector notation)

$$\rho(\underline{I}) = E[\rho(\underline{I}) \,/\, \underline{I}']$$

$$= \psi(\underline{I}')$$

Since $\underline{I} = \rho(\underline{I})$ with prob. 1, we conclude

$$\underline{I} = \rho(\underline{I}) = \psi(\underline{I}') \qquad \text{w.p. } 1$$

as was to be shown. ∎

The "general conditions" under which the result is valid are that the various means and variances used throughout the proof are well-defined and finite.