Lecture 16

Minimum Variance Unbiased Estimator (II)

1. Proof of Factorization Theorem

($\Rightarrow$) Suppose $T(Y)$ is a suff. stat. Then, by definition $f_{Y|T(Y)}(y|t;\theta)$ must be indep of $\theta$. Therefore, by conditional probability definition,

$$f_Y(y;\theta) = \underbrace{f_{Y|T(Y)}(y|t;\theta)}_{\text{def } h(y)} \underbrace{f_T(Y)(t;\theta)}_{\text{def } g_\theta(t)}.$$  

($\Leftarrow$) Suppose that

$$f_Y(y;\theta) = g_\theta(T(y))h(y).$$

Since

$$f_{Y|T(Y)}(y|t;\theta) = \frac{f_{Y,T(Y)}(y,t;\theta)}{f_T(Y)(t;\theta)},$$

and since $T(Y)$ is a function of $Y$,

$$f_{Y|T(Y)}(y|t;\theta) = \begin{cases} \frac{f_Y(y;\theta)}{f_T(Y)(t;\theta)}, & \text{if } T(y) = t \\ 0, & \text{if } T(y) \neq t. \end{cases}$$

Also, since

$$f_T(Y)(t;\theta) = \int_{T(y)=t} f_Y(y;\theta) \, dy,$$
So \( f_y(y; \theta) = g_\theta(T(y))h(y) \) implies that

\[
\frac{f_{Y|T(Y)}(y|t; \theta)}{f_{T(Y)}(t; \theta)} = \begin{cases} 
\frac{g_\theta(T(y))h(y)}{\int_{T(y)=t} g_\theta(T(y))h(y)\,dy}, & T(y)=t \\
0, & T(y) \neq t
\end{cases}
\]

\[
= \frac{g_\theta(T(y))h(y) \mathbb{1}\{T(y)=t\}}{\int_{T(y)=t} h(y)\,dy} = \frac{h(y) \mathbb{1}\{T(y)=t\}}{\int_{T(y)=t} h(y)\,dy},
\]

which is independent of \( \theta \).

\( \mathbb{X} \)

Rao-Blackwell Theorem

(A procedure to "improve" any unbiased estimator)

Theorem

Suppose that \( \hat{g}(y) \) is an unbiased estimate of \( g(\theta) \) and that \( T(y) \) is a sufficient statistic for \( \theta \). Then

1. \( \hat{g}(T(y)) \) defined as \( \mathbb{E}_{Y|T(Y)}[\hat{g}(Y) \mid T(Y) = T(y)] \) is also an unbiased estimator of \( g(\theta) \)

2. \( \text{Var}_{T(Y)}[\hat{g}(T(y))] \leq \text{Var}_{Y}[\hat{g}(Y)] \)

with equality holds iff

\[
P\left( \hat{g}(T(Y)) \neq \hat{g}(Y) \right) = 1.
\]
Proof:

(1) To show $\tilde{g}(T(Y))$ is an unbiased estimator, note that

$$E_{T(Y)} \left[ \tilde{g}(T(Y)) \right]$$

$$= E_{T(Y)} \left[ E_{Y|T(Y)} \left( \tilde{g}(Y) | T(Y) = T(y) \right) \right]$$

$$= E_Y \left[ \tilde{g}(Y) \right] = g(\theta).$$

(2) $Var_{T(Y)} \left[ \tilde{g}(T(Y)) \right]$

$$= E_{T(Y)} \left[ \tilde{g}(T(Y))^2 \right] - g(\theta)^2$$

$$Var_Y \left[ \tilde{g}(Y) \right]$$

$$= E_Y \left[ \tilde{g}(Y)^2 \right] - g(\theta)^2.$$

Note that

$$E_{T(Y)} \left[ \tilde{g}(T(Y))^2 \right] = E_{T(Y)} \left[ E_{Y|T(Y)} \left[ \tilde{g}(Y)^2 \right] \right]$$

where (a) holds because of Jensen's inequality, and (b) is the result of iterated expectation. Equality holds when Jensen's inequality holds, i.e.,

$$P(\tilde{g}(T(Y)) = \tilde{g}(Y)) = 1.$$
Implication of Rao-Blackwell:

(1) With a sufficient statistic, we can improve any unbiased estimator that is not already a function of \( T \) by conditioning it on \( T(Y) \).

(2) If \( T \) is sufficient for \( \Theta \), and if there is only one function of \( T \) that is an unbiased estimator of \( g(\Theta) \) (i.e. \( \hat{g}(T(Y)) \)) then that function must be MVUE.

To see this:

Suppose \( \hat{g}^*(T(Y)) \) is the only function of \( T(Y) \) such that \( E_{T(Y)}[\hat{g}^*(T(Y))] = g(\Theta) \).

Let \( \hat{g}(Y) \) be any unbiased estimator. Then, RB says \( \hat{g}(T(Y)) = E_Y[\hat{g}(Y)|T(Y) = T(Y)] \) is also unbiased and has a variance

\[
\text{Var}_{T(Y)}[\hat{g}(T(Y))] \leq \text{Var}_T[\hat{g}(Y)].
\]

But since \( \hat{g}^*(T(Y)) \) is arbitrary, we have

\[
\text{Var}_{T(Y)}[\hat{g}^*(T(Y))] \leq \text{Var}_{T(Y)}[\hat{g}(T(Y))],
\]

So \( \hat{g}^*(T(Y)) \) is an MVUE.
Jensen's inequality

For any convex function $f$,

$$ f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]. $$

Think of: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. Therefore,

$$ \mathbb{E}[X]^2 \leq \mathbb{E}[X^2] , \text{ and so } $$

$$ \mathbb{E}_{Y \mid T(Y)} \left[ \hat{g}(Y) \right]^2 \leq \mathbb{E}_{Y \mid T(Y)} \left[ \hat{g}(Y)^2 \right]. $$