557: MATHEMATICAL STATISTICS II
METHODS OF EVALUATING ESTIMATORS

An estimator, $T(\underset{\sim}{X})$, of θ can be evaluated via its statistical properties. Typically, two aspects are considered:

- Expectation
- Variance

either in terms of **finite** n behaviour, or the hypothetical **limiting case** as $n \longrightarrow \infty$. In a frequentist setting, these assessments are made **conditional** on a given value of θ, by examining the distribution of T given θ, $f_{T|\theta}$.

Bias, Variance And Mean Square Error

For estimator T (in general a function of sample $\underset{\sim}{X}$) of parameter $\tau(\theta)$, the following quantities will be used to evaluate T.

- **Bias:** The **bias** of T is denoted $b_T(\theta)$, and is defined by

  $$b_T(\theta) = \mathrm{E}_{f_{T|\theta}}[T] - \tau(\theta).$$

  If $b_T(\theta) = 0$ for all θ, then T is termed **unbiased** for $\tau(\theta)$.
- **Variance:** The **variance** of T is denoted in the usual way by $\mathrm{Var}_{f_{T|\theta}}[T]$, defined

  $$\mathrm{Var}_{f_{T|\theta}}[T] = \mathrm{E}_{f_{T|\theta}}[(T - \mathrm{E}_{f_{T|\theta}}[T])^2]$$

  For an unbiased estimator,

  $$\mathrm{Var}_{f_{T|\theta}}[T] = \mathrm{E}_{f_{T|\theta}}[(T - \tau(\theta))^2].$$

- **Mean Square Error:** The Mean Square Error (MSE) of T is denoted $\mathrm{MSE}_\theta(T)$ and defined by

  $$\mathrm{MSE}_\theta(T) = \mathrm{E}_{f_{T|\theta}}[(T - \tau(\theta))^2]$$

  By elementary calculation, it follows that

  $$\mathrm{MSE}_\theta(T) = \mathrm{Var}_{f_{T|\theta}}[T] + (\mathrm{E}_{f_{T|\theta}}[T] - \tau(\theta))^2$$

  so that

  $$\text{Mean Square Error} = \text{Variance} + (\text{Bias})^2$$

The **Best Unbiased Estimator**, or **Uniform Minimum Variance Unbiased Estimator** (UMVUE), of $\tau(\theta)$, denoted $T^\star$, is the estimator with the **smallest variance** of all unbiased estimators of $\tau(\theta)$, that is, if T is any other unbiased estimator of $\tau(\theta)$,

$$\mathrm{Var}_{f_{T|\theta}}[T] \geq \mathrm{Var}_{f_{T^\star|\theta}}[T^\star]$$

It transpires that there is a lower bound, $B(\theta)$, on the variance of unbiased estimators of $\tau(\theta)$, given by the following result. The result does not in general guarantee that an estimator with variance $B(\theta)$ exists, and does not give a method of constructing such an estimator, but it does confirm that if T is such an unbiased estimator, and

$$\mathrm{Var}_{f_{T|\theta}}[T] = B(\theta)$$

then T is the Best Unbiased Estimator.

Theorem (The Cramér-Rao Inequality)
Suppose that $X_1, \ldots, X_n$ is a sample of random variables from probability model described by pmf/pdf $f_{X|\theta}$, and let $T(\underset{\sim}{X})$ be an estimator of $\tau(\theta)$. Suppose that

$$\frac{d}{d\theta}\left\{\mathrm{E}_{f_{T|\theta}}[T]\right\} = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\left\{T(\underset{\sim}{x}) f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\right\}\, d\underset{\sim}{x}$$

that is, exchanging the order of integration and differentiation is legitimate, and that $\mathrm{Var}_{f_{T|\theta}}[T] < \infty$. Then

$$\mathrm{Var}_{f_{T|\theta}}[T] \geq \frac{\left(\dfrac{d}{d\theta}\left\{\mathrm{E}_{f_{T|\theta}}[T]\right\}\right)^2}{\mathrm{E}_{f_{\underset{\sim}{X}|\theta}}\left[S(\underset{\sim}{X};\theta)^2\right]}$$

where $S(\underset{\sim}{x};\theta)$ is the score function

$$S(\underset{\sim}{x};\theta) = \frac{\partial}{\partial\theta}\left\{\log f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\right\}$$

Proof For any two random variables U and V, by a previous result (the Cauchy-Schwarz inequality)

$$\left\{\mathrm{Cov}_{f_{U,V}}[U,V]\right\}^2 \leq \mathrm{Var}_{f_U}[U]\,\mathrm{Var}_{f_V}[V] \qquad \therefore \qquad \mathrm{Var}_{f_U}[U] \geq \frac{\left\{\mathrm{Cov}_{f_{U,V}}[U,V]\right\}^2}{\mathrm{Var}_{f_V}[V]}. \tag{1}$$

with equality if and only if U and V are linearly related. Now, note that, under the assumptions of the theorem,

$$\begin{aligned}
\frac{d}{d\theta}\left\{\mathrm{E}_{f_{T|\theta}}[T]\right\} &= \int_{\mathcal{X}} T(\underset{\sim}{x})\frac{\partial}{\partial\theta}\left\{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\right\}\, d\underset{\sim}{x} \\
&= \int_{\mathcal{X}} T(\underset{\sim}{x})\frac{\dfrac{\partial}{\partial\theta}\left\{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\right\}}{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\, d\underset{\sim}{x} \\
&= \int_{\mathcal{X}} T(\underset{\sim}{x})\frac{\partial}{\partial\theta}\left\{\log f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\right\} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\, d\underset{\sim}{x} \\
&= \mathrm{E}_{f_{\underset{\sim}{X}|\theta}}\left[T(\underset{\sim}{X})S(\underset{\sim}{X};\theta)\right] \\
&\equiv \mathrm{Cov}_{f_{T,S|\theta}}[T,S]
\end{aligned}$$

as $\mathrm{E}_{f_{S|\theta}}[S] \equiv \mathrm{E}_{f_{\underset{\sim}{X}|\theta}}[S(\underset{\sim}{X};\theta)] = 0$, by results from MATH 556. Similarly

$$\mathrm{Var}_{f_{\underset{\sim}{X}|\theta}}[S(\underset{\sim}{X};\theta)] = \mathrm{E}_{f_{\underset{\sim}{X}|\theta}}[S(\underset{\sim}{X};\theta)^2].$$

Therefore, using the covariance inequality

$$\mathrm{Var}_{f_{T|\theta}}[T] \geq \frac{\left\{\dfrac{d}{d\theta}\left\{\mathrm{E}_{f_{T|\theta}}[T]\right\}\right\}^2}{\mathrm{E}_{f_{\underset{\sim}{X}|\theta}}[S(\underset{\sim}{X};\theta)^2]}$$

as required. ■

Corollary : If $X_1, \ldots, X_n$ are a random sample, then

$$\text{Var}_{f_{T|\theta}}[T] \geq \frac{\left\{ \frac{d}{d\theta} \left\{ \text{E}_{f_{T|\theta}}[T] \right\} \right\}^2}{n\mathcal{I}(\theta)}$$

where $\mathcal{I}(\theta)$ is the **Fisher Information** as defined in MATH 556 as

$$\mathcal{I}(\theta) = \text{E}_{f_{X|\theta}}[S(X;\theta)^2]$$

Recall that, if second derivatives exist

$$\mathcal{I}(\theta) = -\text{E}_{f_{X|\theta}}[\Psi(X;\theta)]$$

where

$$\Psi(X;\theta) = \frac{\partial^2}{\partial\theta^2} \left\{ \log f_{X|\theta}(x|\theta) \right\}$$

is the second derivative function.

Corollary : By definition, $\text{E}_{f_{T|\theta}}[T] = b_\theta(T) + \tau(\theta)$, so

$$\text{Var}_{f_{T|\theta}}[T] \geq \frac{\left\{ \dot{b}_T(\theta) + \dot{\tau}(\theta) \right\}^2}{\text{E}_{f_{\underset{\sim}{X}|\theta}}[S(\underset{\sim}{X};\theta)^2]}$$

Vector Parameter Case

A similar result can be derived in the vector parameter case. Suppose that $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)^\mathsf{T}$. If $\underset{\sim}{T}(\underset{\sim}{X})$ is a d-dimensional estimator of a vector function of $\underset{\sim}{\theta}$, then we have a similar bound for the variance-covariance matrix of the estimator. Recall first that for two $(k \times k)$ matrices A and B, we write $A \geq B$ if $A - B$ is **non-negative definite**, that is

$$\underset{\sim}{x}^\mathsf{T}(A - B)\underset{\sim}{x} \geq 0 \qquad \underset{\sim}{x} \in \mathbb{R}^k.$$

Under the same assumptions as in the single parameter case, that differentiation and integration orders may exchanged, and the required expectations and variances are finite, it follows that

$$\text{Var}_{f_{\underset{\sim}{T}|\theta}}[\underset{\sim}{T}] \geq \dot{\ell}(\underset{\sim}{\theta})\, \mathcal{I}(\underset{\sim}{\theta})^{-1}\, \dot{\ell}(\underset{\sim}{\theta})^\mathsf{T} \tag{2}$$

where

$$\mathcal{I}(\underset{\sim}{\theta}) = \text{E}_{f_{X|\theta}}[\underset{\sim}{S}(X;\underset{\sim}{\theta})\underset{\sim}{S}(X;\underset{\sim}{\theta})^\mathsf{T}]$$

and $\underset{\sim}{S}(X;\underset{\sim}{\theta})$ is the $k \times 1$ vector score function with jth component

$$\frac{\partial}{\partial\theta_j} \log f_{X|\underset{\sim}{\theta}}(x|\underset{\sim}{\theta}) \qquad j = 1, \ldots, k.$$

and $\dot{\ell}(\underset{\sim}{\theta})$ is the $d \times k$ matrix with (l, j)th element

$$\frac{\partial}{\partial\theta_j} \left\{ E_{f_{T_l|\underset{\sim}{\theta}}}[T_l] \right\} \qquad l = 1, \ldots, d,\ j = 1 \ldots, k$$

Note that in equation (2), the left-hand and right-hand side are $d \times d$ matrices. Note also that if the second-derivative matrix can be defined, then

$$\mathcal{I}(\underset{\sim}{\theta}) = -\text{E}_{f_{X|\theta}}[\boldsymbol{\Psi}(X;\underset{\sim}{\theta})]$$

where the (l, j)th element of the $k \times k$ matrix $\boldsymbol{\Psi}$ is

$$\frac{\partial^2}{\partial\theta_j \partial\theta_l} \left\{ \log f_{X|\underset{\sim}{\theta}}(x|\underset{\sim}{\theta}) \right\}$$

Attaining the Lower Bound.

Theorem
Suppose that $X_1, \ldots, X_n$ is a sample of random variables from probability model described by pmf/pdf $f_{X|\theta}$, with likelihood $L(\theta|\underset{\sim}{x})$. Let $T(\underset{\sim}{X})$ be an unbiased estimator of $\tau(\theta)$. Then $T(\underset{\sim}{X})$ attains the Cramér-Rao lower bound, that is

$$\mathrm{Var}_{f_{T|\theta}}[T] = B(\theta) = \frac{\left(\dfrac{d}{d\theta}\left\{\mathrm{E}_{f_{T|\theta}}[T]\right\}\right)^2}{\mathrm{E}_{f_{\underset{\sim}{X}|\theta}}\left[S(\underset{\sim}{X};\theta)^2\right]}$$

if and only if

$$a(\theta)(T(\underset{\sim}{X}) - \tau(\theta)) = \frac{\partial}{\partial\theta}\log L(\theta|\underset{\sim}{X})$$

for some function $a(\theta)$.

Proof In the variance inequality in equation (1), set

$$U \equiv T(\underset{\sim}{X}) \qquad\qquad V \equiv \frac{\partial}{\partial\theta}\log L(\theta|\underset{\sim}{X})$$

so that

$$\left\{\mathrm{Cov}_{f_{\underset{\sim}{X}|\theta}}\left[T(\underset{\sim}{X}), \frac{\partial}{\partial\theta}\log L(\theta|\underset{\sim}{X})\right]\right\}^2 \leq \mathrm{Var}_{f_{T|\theta}}[T]\,\mathrm{Var}_{f_{\underset{\sim}{X}|\theta}}\left[\frac{\partial}{\partial\theta}\log L(\theta|\underset{\sim}{X})\right]$$

with equality if and only if T and $\dfrac{\partial}{\partial\theta}\log L(\theta|\underset{\sim}{X})$ are linearly related, that is

$$m(\theta)T + c(\theta) = \frac{\partial}{\partial\theta}\log L(\theta|\underset{\sim}{X}) = \frac{\partial}{\partial\theta}\left\{\sum_{i=1}^{n}\log f_{X_i|\theta}(X_i|\theta)\right\} \qquad (3)$$

for some functions $m(\theta)$ and $c(\theta)$ that do not depend on X, but may in general depend on θ. Taking expectations with respect to $f_{\underset{\sim}{X}|\theta}$ on both sides of equation (3), and noting that the expectation on the right-hand side is zero, we must have

$$c(\theta) = -\mathrm{E}_{f_{T|\theta}}[T] = \tau(\theta)$$

and the result follows. ■

If an estimator can be found such that the bound is met, then that estimator is the best unbiased estimator. Note that, in the one-parameter Exponential Family, for a random sample $\underset{\sim}{X}$

$$L(\theta|\underset{\sim}{x}) = f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta) = h(\underset{\sim}{x})\{c(\theta)\}^n\exp\{w(\theta)T(\underset{\sim}{x})\}$$

so that

$$\frac{\partial}{\partial\theta}\log L(\theta|\underset{\sim}{X}) = n\frac{\dot{c}(\theta)}{c(\theta)} + \dot{w}(\theta)T(\underset{\sim}{x}) = \dot{w}(\theta)\left(T(\underset{\sim}{x}) - \frac{n\dot{c}(\theta)}{c(\theta)\dot{w}(\theta)}\right) = a(\theta)\left(T(\underset{\sim}{x}) - n\tau(\theta)\right)$$

say, where $\dot{c}(\theta)$ is the partial derivative of $c(\theta)$ with respect to θ. Hence, taking expectations on left and right hand sides, we note that

$$E_{f_{T|\theta}}[T] = n\tau(\theta)$$

so that

$$\frac{T(\underset{\sim}{X})}{n}$$

is an unbiased estimator of $\tau(\theta)$ that that has minimum variance.

Sufficiency and Unbiasedness.

Theorem The Rao-Blackwell Theorem
Let T be an unbiased estimator of $\tau(\theta)$, and S be a sufficient statistic for θ. Define statistic U by

$$U \equiv g(S) = \mathrm{E}_{f_{T|S,\theta}}[T|S]$$

Then U is an unbiased estimator of $\tau(\theta)$, and for all θ

$$\mathrm{Var}_{f_{U|\theta}}[U] \leq \mathrm{Var}_{f_{T|\theta}}[T].$$

Proof Clearly $U = g(S)$ is a valid estimator, as it does not depend on the θ; the conditional distribution of T given S does not depend on θ by sufficiency. By iterated expectation,

$$\mathrm{E}_{f_{U|\theta}}[U] = \mathrm{E}_{f_{S|\theta}}[g(S)] = \mathrm{E}_{f_{S|\theta}}[\mathrm{E}_{f_{T|S,\theta}}[T|S]] = \mathrm{E}_{f_{T|\theta}}[T] = \tau(\theta)$$

so U is unbiased for $\tau(\theta)$, and similarly

$$\begin{aligned} \mathrm{Var}_{f_{T|\theta}}[T] &= \mathrm{E}_{f_{S|\theta}}[\mathrm{Var}_{f_{T|S,\theta}}[T|S]] + \mathrm{Var}_{f_{S|\theta}}[\mathrm{E}_{f_{T|S,\theta}}[T|S]] \\ &\geq \mathrm{Var}_{f_{S|\theta}}[\mathrm{E}_{f_{T|S,\theta}}[T|S]] \\ &= \mathrm{Var}_{f_{S|\theta}}[g(S)] = \mathrm{Var}_{f_{U|\theta}}[U] \end{aligned}$$

and thus U is a better estimator of $\tau(\theta)$ than T, as it has lower variance. ■

Uniqueness.

Theorem
If T is a best unbiased estimator of $\tau(\theta)$, that is, it achieves the lower bound on variance $B(\theta)$, then T is unique.

Proof Let T' be another best unbiased estimator. Let

$$T^\star = \frac{1}{2}(T + T').$$

Then $T^\star$ is clearly unbiased, and by elementary results

$$\begin{aligned} \mathrm{Var}_{f_{T^\star|\theta}}[T^\star] &= \frac{1}{4}\mathrm{Var}_{f_{T|\theta}}[T] + \frac{1}{4}\mathrm{Var}_{f_{T'|\theta}}[T'] + \frac{1}{2}\mathrm{Cov}_{f_{T,T'|\theta}}[T, T'] \\ &\leq \frac{1}{4}\mathrm{Var}_{f_{T|\theta}}[T] + \frac{1}{4}\mathrm{Var}_{f_{T'|\theta}}[T'] + \frac{1}{2}\left(\mathrm{Var}_{f_{T|\theta}}[T]\,\mathrm{Var}_{f_{T'|\theta}}[T']\right)^{1/2} \\ &= \mathrm{Var}_{f_{T|\theta}}[T] \end{aligned}$$

with equality if and only if T and T' are linearly related, as the variances of T and T' are equal. Thus, to avoid contradiction, we must have a linear relationship, that is

$$T' = m(\theta)T + c(\theta)$$

say. But, in this case

$$\mathrm{Cov}_{f_{T,T'|\theta}}[T, T'] = \mathrm{Cov}_{f_{T|\theta}}[T, m(\theta)T + c(\theta)] = \mathrm{Cov}_{f_{T|\theta}}[T, m(\theta)T] = m(\theta)\mathrm{Var}_{f_{T|\theta}}[T]$$

But, by the covariance equality above,

$$\mathrm{Cov}_{f_{T,T'|\theta}}[T, T'] = \mathrm{Var}_{f_{T|\theta}}[T]$$

implying that $m(\theta) \equiv 1$. Hence, as T and T' both have expectation $\tau(\theta)$, we must also have $c(\theta) = 0$, so that T and T' are identical. ■

Characterizing Best Unbiased Estimators.

Theorem
An estimator T of $\tau(\theta)$ is the best unbiased estimator of $\tau(\theta)$ if and only if $\mathrm{E}_{f_{T|\theta}}[T] = \tau(\theta)$ and T is uncorrelated with all estimators U such that

$$\mathrm{E}_{f_{U|\theta}}[U] = 0.$$

U is termed an **unbiased estimator of zero**.

Proof Suppose first that T is the best unbiased estimator of $\tau(\theta)$, and U is an unbiased estimator of zero. Then estimator

$$S = T + aU$$

for constant a is also unbiased for $\tau(\theta)$, and

$$\mathrm{Var}_{f_{S|\theta}}[S] = \mathrm{Var}_{f_{T|\theta}}[T] + a^2\mathrm{Var}_{f_{U|\theta}}[U] + 2a\mathrm{Cov}_{f_{T,U|\theta}}[T, U].$$

Thus choosing a so that

$$a^2 < -\frac{2a\,\mathrm{Cov}_{f_{T,U|\theta}}[T, U]}{\mathrm{Var}_{f_{U|\theta}}[U]}$$

renders $\mathrm{Var}_{f_{S|\theta}}[S] < \mathrm{Var}_{f_{T|\theta}}[T]$ and a contradiction. Such a choice can always be made if $\mathrm{Cov}_{f_{T,U|\theta}}[T, U]$ is non-zero. Hence we must have

$$\mathrm{Cov}_{f_{T,U|\theta}}[T, U] = 0,$$

that is, that T and U are uncorrelated.

Now suppose that $\mathrm{E}_{f_{T|\theta}}[T] = \tau(\theta)$, and that T is uncorrelated with all unbiased estimators of zero. Let T' be any other unbiased estimator of $\tau(\theta)$. Now, writing

$$T' = T + (T' - T) = T + Z$$

say, yields

$$\begin{aligned}\mathrm{Var}_{f_{T'|\theta}}[T'] &= \mathrm{Var}_{f_{T|\theta}}[T] + \mathrm{Var}_{f_{Z|\theta}}[Z] + 2\mathrm{Cov}_{f_{T,Z|\theta}}[T, Z] \\ &\geq \mathrm{Var}_{f_{T|\theta}}[T]\end{aligned}$$

as Z is an unbiased estimator of zero, and is thus uncorrelated with T by assumption, and also $\mathrm{Var}_{f_{Z|\theta}}[Z] \geq 0$. ■

Corollary : If T is a complete sufficient statistic for parameter θ, and $h(T)$ is an estimator which is a function of T only, then $h(T)$ is the unique best unbiased estimator of $\tau(\theta) = \mathrm{E}_{f_{T|\theta}}[h(T)]$.

Proof If T is complete, then the only function g with

$$\mathrm{E}_{f_{T|\theta}}[g(T)] = 0.$$

is $g(t) = 0$ for all t, that is, the only unbiased estimator of zero is zero itself. But the previous result states that an estimator is a best unbiased estimator if it is uncorrelated with all unbiased estimators of zero. As

$$\mathrm{Cov}_{f_{T|\theta}}[h(T), 0] = 0$$

for any $h(T)$, it follows that $h(T)$ is the unique best unbiased estimator of its expectation. ■