

557: MATHEMATICAL STATISTICS II

METHODS OF EVALUATING ESTIMATORS

An estimator, $T(\underline{X})$, 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 $n \rightarrow \infty$. 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 $\underline{X}$) 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

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

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 $\text{Var}_{f_{T|\theta}}[T]$, defined

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

For an unbiased estimator,

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

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

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

By elementary calculation, it follows that

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

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^* , is the estimator with the **smallest variance** of all unbiased estimators of $\tau(\theta)$, that is, if T is any other unbiased estimator of $\tau(\theta)$,

$$\text{Var}_{f_{T|\theta}}[T] \geq \text{Var}_{f_{T^*|\theta}}[T^*]$$

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

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

then T is the Best Unbiased Estimator.

Theorem (The Cramér-Rao Inequality)

Suppose that $X_1, \dots, X_n$ is a sample of random variables from probability model described by pmf/pdf $f_{X|\theta}$, and let $T(\underline{X})$ be an estimator of $\tau(\theta)$. Suppose that

$$\frac{d}{d\theta} \left\{ E_{f_{T|\theta}}[T] \right\} = \int_{\mathcal{X}} \frac{\partial}{\partial \theta} \left\{ T(\underline{x}) f_{\underline{X}|\theta}(\underline{x}|\theta) \right\} d\underline{x}$$

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

$$\text{Var}_{f_{T|\theta}}[T] \geq \frac{\left(\frac{d}{d\theta} \left\{ E_{f_{T|\theta}}[T] \right\} \right)^2}{E_{f_{\underline{X}|\theta}}[S(\underline{X}; \theta)^2]}$$

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

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

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

$$\left\{ \text{Cov}_{f_{U,V}}[U, V] \right\}^2 \leq \text{Var}_{f_U}[U] \text{Var}_{f_V}[V] \quad \therefore \quad \text{Var}_{f_U}[U] \geq \frac{\left\{ \text{Cov}_{f_{U,V}}[U, V] \right\}^2}{\text{Var}_{f_V}[V]}. \quad (1)$$

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

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

as $E_{f_{S|\theta}}[S] \equiv E_{f_{\underline{X}|\theta}}[S(\underline{X}; \theta)] = 0$, by standard results. Similarly

$$\text{Var}_{f_{\underline{X}|\theta}}[S(\underline{X}; \theta)] = E_{f_{\underline{X}|\theta}}[S(\underline{X}; \theta)^2].$$

Therefore, using the covariance inequality

$$\text{Var}_{f_{T|\theta}}[T] \geq \frac{\left\{ \frac{d}{d\theta} \left\{ E_{f_{T|\theta}}[T] \right\} \right\}^2}{E_{f_{\underline{X}|\theta}}[S(\underline{X}; \theta)^2]}$$

as required. ■

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

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

where $\mathcal{I}(\theta)$ is the **Fisher Information**

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

Recall that, if second derivatives exist

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

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, $E_{f_{T|\theta}}[T] = b_\theta(T) + \tau(\theta)$, so

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

Vector Parameter Case

A similar result can be derived in the vector parameter case. Suppose that $\underline{\theta} = (\theta_1, \dots, \theta_k)^\top$. If $\underline{T}(\underline{X})$ is a d -dimensional estimator of a vector function of $\underline{\theta}$, 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

$$\underline{x}^\top (A - B) \underline{x} \geq 0 \quad \underline{x} \in \mathbb{R}^k.$$

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

$$\text{Var}_{f_{\underline{T}|\underline{\theta}}}[\underline{T}] \geq \dot{\ell}(\underline{\theta}) \mathcal{I}(\underline{\theta})^{-1} \dot{\ell}(\underline{\theta})^\top \quad (2)$$

where

$$\mathcal{I}(\underline{\theta}) = E_{f_{X|\underline{\theta}}}[\underline{S}(X; \underline{\theta}) \underline{S}(X; \underline{\theta})^\top]$$

and $\underline{S}(X; \underline{\theta})$ is the $k \times 1$ vector score function with j th component

$$\frac{\partial}{\partial \theta_j} \log f_{X|\underline{\theta}}(x|\underline{\theta}) \quad j = 1, \dots, k.$$

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

$$\frac{\partial}{\partial \theta_j} \left\{ E_{f_{T_l|\underline{\theta}}} [T_l] \right\} \quad l = 1, \dots, d, \quad j = 1, \dots, k$$

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

$$\mathcal{I}(\underline{\theta}) = -E_{f_{X|\underline{\theta}}}[\underline{\Psi}(X; \underline{\theta})]$$

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

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

Attaining the Lower Bound.

Theorem

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

$$\text{Var}_{f_{T|\theta}}[T] = B(\theta) = \frac{\left(\frac{d}{d\theta} \{E_{f_{T|\theta}}[T]\} \right)^2}{E_{f_{\underline{X}|\theta}}[S(\underline{X}; \theta)^2]}$$

if and only if

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

for some function $a(\theta)$.

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

$$U \equiv T(\underline{X}) \quad V \equiv \frac{\partial}{\partial \theta} \log L(\theta|\underline{X})$$

so that

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

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

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

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 $f_{\underline{X}|\theta}$ on both sides of equation (3), and noting that the expectation on the right-hand side is zero, we must have

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

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 $\underline{X}$

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

so that

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

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

$$\frac{T(\underline{X})}{n}$$

is an unbiased estimator of $\tau(\theta)$ 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

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

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

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

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,

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

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

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

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

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

Then T^* is clearly unbiased, and by elementary results

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

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

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

But, by the covariance equality above,

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

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 $E_{f_{T|\theta}}[T] = \tau(\theta)$ and T is uncorrelated with all estimators U such that

$$E_{f_{U|\theta}}[U] = 0.$$

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

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

Thus choosing a so that

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

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

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

that is, that T and U are uncorrelated.

Now suppose that $E_{f_{T|\theta}}[T] = \tau(\theta)$, 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

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

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

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

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

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

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

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

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