557: MATHEMATICAL STATISTICS II
BIAS AND VARIANCE

An estimator, $T(\mathbf{X})$, 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 **limiting case** as $n \longrightarrow \infty$. These assessments are made for a given value of θ, by examining the distribution of T given θ, $f_T(t;\theta)$.

Bias, Variance And Mean Square Error

For estimator T of estimand $\tau(\theta)$, the following quantities will be used to evaluate T.

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

$$b_T(\theta) = \mathbb{E}_T[T;\theta] - \tau(\theta).$$

  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 $\text{Var}_T[T;\theta]$, defined

$$\text{Var}_T[T;\theta] = \mathbb{E}_T[(T - \mathbb{E}_T[T;\theta])^2].$$

  For an unbiased estimator,

$$\text{Var}_T[T;\theta] = \mathbb{E}_T[(T - \tau(\theta))^2].$$

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

$$\text{MSE}_T(\theta) = \mathbb{E}_T[(T - \tau(\theta))^2;\theta].$$

  By elementary calculation, it follows that

$$\text{MSE}_T(\theta) = \text{Var}_T[T;\theta] + (\mathbb{E}_T[T;\theta] - \tau(\theta))^2$$

  so that

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

Minimum Variance Unbiased Estimation

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 $\tau(\theta)$,

$$\text{Var}_T[T;\theta] \geq \text{Var}_{T^\star}[T^\star;\theta]$$

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

$$\text{Var}_T[T;\theta] = B(\theta)$$

then T is the Best Unbiased Estimator.

RESULT 1: The Cramér-Rao Variance Bound.

Theorem (The Cramér-Rao Bound in the 1-d Case)
Suppose that $X_1, \ldots, X_n$ is a sample of random variables from probability model $f_X(x;\theta)$, and let $T = T(\mathbf{X}) = T(X_{1:n})$ be an estimator of $\tau(\theta)$. Denote $\mathbb{E}_T[T;\theta] = \mu_T(\theta)$, and suppose that

$$\dot{\mu}_T(\theta) = \frac{d}{d\theta}\{\mu_T(\theta)\} = \int \frac{d}{d\theta}\{T(\mathbf{x})f_{\mathbf{X}}(\mathbf{x};\theta)\}\ d\mathbf{x}$$

that is, we may differentiate under the integral, and that $\text{Var}_T[T;\theta] < \infty$. Then

$$\text{Var}_T[T;\theta] \geq \frac{(\dot{\mu}_T(\theta))^2}{n\mathbb{E}_X\left[U(X;\theta)^2\right]}$$

where, in the usual notation, $U(x;\theta)$ is the first derivative of the log density.

Proof For any two random variables V and W, by the Cauchy-Schwarz inequality,

$$\{\text{Cov}_{V,W}[V,W]\}^2 \leq \text{Var}_V[V]\,\text{Var}_W[W] \qquad \therefore \qquad \text{Var}_V[V] \geq \frac{\{\text{Cov}_{V,W}[V,W]\}^2}{\text{Var}_W[W]} \tag{1}$$

with equality if and only if V and W are linearly related. Under the assumptions of the theorem,

$$\begin{aligned}\dot{\mu}_T(\theta) &= \int T(\mathbf{x})\frac{d}{d\theta}\{f_{\mathbf{X}}(\mathbf{x};\theta)\}\ d\mathbf{x} = \int T(\mathbf{x})\frac{\frac{d}{d\theta}\{f_{\mathbf{X}}(\mathbf{x};\theta)\}}{f_{\mathbf{X}}(\mathbf{x};\theta)}f_{\mathbf{X}}(\mathbf{x};\theta)\,d\mathbf{x} \\ &= \int T(\mathbf{x})\frac{d}{d\theta}\{\log f_{\mathbf{X}}(\mathbf{x};\theta)\}\,f_{\mathbf{X}}(\mathbf{x};\theta)\,d\mathbf{x} = \mathbb{E}_{\mathbf{X}}\left[T(\mathbf{X})U(\mathbf{X};\theta);\theta\right] \equiv \text{Cov}_{\mathbf{X}}[T(\mathbf{X})U(\mathbf{X};\theta);\theta]\end{aligned}$$

as $\mathbb{E}_{\mathbf{X}}[U(\mathbf{X};\theta)] = 0$, by standard results. Similarly

$$\text{Var}_{\mathbf{X}}[U(\mathbf{X};\theta)] = \text{Var}_{\mathbf{X}}\left[\sum_{i=1}^{n}U(X_i;\theta);\theta\right] = n\text{Var}_X[U(X;\theta);\theta] = n\mathbb{E}_X[U(X;\theta)^2;\theta].$$

Therefore, using the covariance inequality

$$\text{Var}_T[T;\theta] \geq \frac{\{\dot{\mu}_T(\theta)\}^2}{n\mathbb{E}_X\left[U(X;\theta)^2\right]}$$

as required. ■

Corollary : If $X_1, \ldots, X_n$ are a random sample, then in terms of the bias $b_T(\theta)$

$$\text{Var}_T[T;\theta] \geq \frac{\{\dot{\mu}_T(\theta)\}^2}{n\mathcal{J}_\theta(\theta)} = \frac{\left\{\dot{b}_T(\theta) + \dot{\tau}(\theta)\right\}^2}{n\mathbb{E}_X[U(X;\theta)^2]}$$

where $\mathcal{J}_\theta(\theta) \equiv \mathcal{I}_\theta(\theta)$ equals the **Fisher Information**.

Vector Parameter Case: A similar result can be derived in the vector parameter case. Suppose that $\theta = (\theta_1, \ldots, \theta_k)^\top$. If $\mathbf{T}(\mathbf{X})$ is a d-dimensional estimator of a vector function of θ, 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 $\mathbf{x}^\top(A-B)\mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^k$.

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

$$\text{Var}_{\mathbf{T}}[\mathbf{T};\theta] \geq \frac{1}{n}\{\dot{\mu}_{\mathbf{T}}(\theta)\}\,\{\mathcal{J}_\theta(\theta)\}^{-1}\,\{\dot{\mu}_{\mathbf{T}}(\theta)\}^\top \tag{2}$$

where

$$\mathcal{J}_\theta(\theta) = \mathbb{E}_X[U(X;\theta)U(X;\theta)^\top;\theta]$$

where in this setting $\mathcal{J}_\theta(\theta) \equiv \mathcal{I}_\theta(\theta)$ is the $(k \times k)$ Fisher information matrix.

RESULT 2: Attaining the Lower Bound.

Theorem
Suppose that $X_1, \ldots, X_n$ is a sample of random variables from probability distribution $f_X(x;\theta)$, with log-likelihood $\ell(\mathbf{x};\theta)$. Let $T(\mathbf{X})$ be an unbiased estimator of $\tau(\theta)$. Then $T(\mathbf{X})$ attains the Cramér-Rao lower bound, that is

$$\mathrm{Var}_T[T;\theta] = B(\theta) = \frac{(\dot{\mu}_T(\theta))^2}{n\mathbb{E}_X\left[U(X;\theta)^2\right]}$$

if and only if, for some function $a(\theta)$,

$$a(\theta)(T(\mathbf{x}) - \tau(\theta)) = \frac{d}{d\theta}\ell(\mathbf{x};\theta) = U(\mathbf{x};\theta) = \sum_{i=1}^{n} U(x_i;\theta).$$

Proof In the variance inequality in equation (1), set $V \equiv T(\mathbf{X})$, $W = U(\mathbf{X};\theta)$ so that

$$\{\mathrm{Cov}_{\mathbf{X}}\left[T(\mathbf{X}), U(\mathbf{X};\theta);\theta\right]\}^2 \leq \mathrm{Var}_T[T;\theta]\ \mathrm{Var}_{\mathbf{X}}\left[U(\mathbf{X};\theta);\theta\right]$$

with equality if and only if T and $U(\mathbf{X};\theta)$ are linearly related, that is

$$m(\theta)(T + c(\theta)) = U(\mathbf{X};\theta) = \sum_{i=1}^{n} U(X_i;\theta) \tag{3}$$

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

$$c(\theta) = -\mathbb{E}_T[T;\theta] = \tau(\theta)$$

and the result follows. ■

If an estimator can be found such that the bound is met, then it follows that this estimator is the best unbiased estimator. Note that, in the one-parameter Exponential Family, for a random sample $\mathbf{x} = x_{1:n}$

$$\mathscr{L}(\mathbf{x};\theta) = f_{\mathbf{X}}(\mathbf{x};\theta) = h(\mathbf{x})\{c(\theta)\}^n \exp\{w(\theta)T(\mathbf{x})\}$$

where as previously $T(\mathbf{x}) = \sum\limits_{i=1}^{n} t(x_i)$. Therefore

$$\frac{\partial}{\partial\theta}\ell(\mathbf{x};\theta) = n\frac{\dot{c}(\theta)}{c(\theta)} + \dot{w}(\theta)T(\mathbf{x}) = \dot{w}(\theta)\left(T(\mathbf{x}) + \frac{n\dot{c}(\theta)}{c(\theta)\dot{w}(\theta)}\right) = a(\theta)\left(T(\mathbf{x}) - n\tau(\theta)\right)$$

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 as

$$\mathbb{E}_X[U(X;\theta);\theta] = 0$$

we must have that

$$\mathbb{E}_T[T;\theta] = n\tau(\theta).$$

Thus

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

is the unbiased estimator of its expectation $\tau(\theta)$ that has minimum variance.

RESULT 3: 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

$$U \equiv g(S) = \mathbb{E}_{T|S}[T|S]$$

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

$$\text{Var}_U[U;\theta] \leq \text{Var}_T[T;\theta].$$

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,

$$\mathbb{E}_U[U;\theta] = \mathbb{E}_S[g(S);\theta] = \mathbb{E}_S[\mathbb{E}_{T|S}[T|S];\theta] = \mathbb{E}_T[T;\theta] = \tau(\theta)$$

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

$$\text{Var}_T[T;\theta] = \mathbb{E}_S[\text{Var}_{T|S}[T|S];\theta] + \text{Var}_S[\mathbb{E}_{T|S}[T|S];\theta] \quad \geq \quad \text{Var}_S[\mathbb{E}_{T|S}[T|S];\theta] \equiv \text{Var}_U[U;\theta]$$

and thus U has variance no greater than that of T. ■

RESULT 4: Uniqueness.

Theorem
If T is a best unbiased estimator of $\tau(\theta)$, that is, $\text{Var}_T[T;\theta] = B(\theta)$, then T is unique.

Proof Let T' be another best unbiased estimator. Let $T^\star = (T + T')/2$. Then $T^\star$ is clearly unbiased, and by elementary results

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

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

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

almost surely. But, in this case,

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

But, by the covariance equality above,

$$\text{Cov}_{T,T'}[T,T'] = \text{Var}_T[T;\theta]$$

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. ■

RESULT 5: Characterizing Best Unbiased Estimators.

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

$$\mathbb{E}_U[U;\theta] = 0.$$

Such a statistic 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 the estimator

$$S = T + aU$$

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

$$\mathrm{Var}_S[S;\theta] = \mathrm{Var}_T[T;\theta] + a^2\mathrm{Var}_U[U;\theta] + 2a\mathrm{Cov}_{T,U}[T,U;\theta].$$

Thus choosing a so that

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

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

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

that is, that T and U are uncorrelated.

Conversely suppose that $\mathbb{E}_T[T;\theta] = \tau(\theta)$, and that T is uncorrelated with all unbiased estimators of zero. Let T' be any other unbiased estimator of $\tau(\theta)$. Write

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

say, where $\mathbb{E}_Z[Z;\theta] = 0$ as both T and T' have expectation $\tau(\theta)$. Thus

$$\mathrm{Var}_{T'}[T';\theta] \quad = \quad \mathrm{Var}_T[T;\theta] + \mathrm{Var}_Z[Z;\theta] + 2\mathrm{Cov}_{T,Z}[T,Z;\theta] \geq \mathrm{Var}_T[T;\theta]$$

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

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 $\tau(\theta) = \mathbb{E}_T[h(T);\theta]$.

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

$$\mathbb{E}_T[g(T);\theta] = 0$$

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

$$\mathrm{Cov}_T[h(T), 0;\theta] = 0$$

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