Asymptotic bias

Unbiasedness as a criterion for point estimators is discussed in §2.3.2. In some cases, however, there is no unbiased estimator. Furthermore, having a “slight” bias in some cases may not be a bad idea. Let $T_n(X)$ be a point estimator of $\theta$ for every $n$.

If $ET_n$ exists for every $n$ and $\lim_{n \to \infty} E(T_n - \theta) = 0$ for any $P \in \mathcal{P}$, then $T_n$ is said to be approximately unbiased.

There are many reasonable point estimators whose expectations are not well defined. It is desirable to define a concept of asymptotic bias for point estimators whose expectations are not well defined.

**Definition 2.11.** (i) Let $\xi, \xi_1, \xi_2, \ldots$ be random variables and $\{a_n\}$ be a sequence of positive numbers satisfying $a_n \to \infty$ or $a_n \to a > 0$. If $a_n \xi_n \to_d \xi$ and $E|\xi| < \infty$, then $E\xi/a_n$ is called an asymptotic expectation of $\xi_n$.

(ii) Let $T_n$ be a point estimator of $\theta$ for every $n$. An asymptotic expectation of $T_n - \theta$, if it exists, is called an asymptotic bias of $T_n$ and denoted by $\tilde{b}_{T_n}(P)$ (or $\tilde{b}_{T_n}(\theta)$ if $P$ is in a parametric family). If $\lim_{n \to \infty} \tilde{b}_{T_n}(P) = 0$ for any $P \in \mathcal{P}$, then $T_n$ is said to be asymptotically unbiased.

Like the consistency, the asymptotic expectation (or bias) is a concept relating to sequences $\{\xi_n\}$ and $\{E\xi/a_n\}$ (or $\{T_n\}$ and $\{\tilde{b}_{T_n}(P)\}$).

The exact bias $b_{T_n}(P)$ is not necessarily the same as $\tilde{b}_{T_n}(P)$ when both of them exist.

**Proposition 2.3.** Let $\{\xi_n\}$ be a sequence of random variables. Suppose that both $E\xi/a_n$ and $E\eta/b_n$ are asymptotic expectations of $\xi_n$ defined according to Definition 2.11(i). Then, one of the following three must hold: (a) $E\xi = E\eta = 0$; (b) $E\xi \neq 0$, $E\eta = 0$, and $b_n/a_n \to 0$; or $E\xi = 0$, $E\eta \neq 0$, and $a_n/b_n \to 0$; (c) $E\xi \neq 0$, $E\eta \neq 0$, and $(E\xi/a_n)/(E\eta/b_n) \to 1$.

If $T_n$ is a consistent estimator of $\theta$, then $T_n = \theta + o_p(1)$ and, by Definition 2.11(ii), $T_n$ is asymptotically unbiased, although $T_n$ may not be approximately unbiased.

In Example 2.34, $X_{(n)}$ has the asymptotic bias $\tilde{b}_{X_{(n)}}(P) = h_n(\theta)EY$, which is of order $n^{-(m+1)^{-1}}$.

When $a_n(T_n - \theta) \to_d Y$ with $EY = 0$ (e.g., $T_n = \bar{X}^2$ and $\theta = \mu^2$ in Example 2.33), a more precise order of the asymptotic bias of $T_n$ may be obtained (for comparing different estimators in terms of their asymptotic biases).

Suppose that there is a sequence of random variables $\{\eta_n\}$ such that

$$a_n \eta_n \to_d Y \quad \text{and} \quad a_n^2 (T_n - \theta - \eta_n) \to_d W;$$

where $Y$ and $W$ are random variables with finite means, $EY = 0$ and $EW \neq 0$.

Then we may define $a_n^{-2}$ to be the order of $\tilde{b}_{T_n}(P)$ or define $EW/a_n^2$ to be the $a_n^{-2}$ order.

---

Lecture 27: Asymptotic bias, variance, and mse
asymptotic bias of \( T_n \).

However, \( \eta_n \) in (1) may not be unique.

Some regularity conditions have to be imposed so that the order of asymptotic bias of \( T_n \) can be uniquely defined.

We consider the case where \( X_1, ..., X_n \) are i.i.d. random \( k \)-vectors with finite \( \Sigma = \text{Var}(X_1) \).

Let \( \bar{X} = n^{-1} \sum_{i=1}^{n} X_i \), and \( T_n = g(\bar{X}) \), where \( g \) is a function on \( \mathbb{R}^k \) that is second-order differentiable at \( \mu = E X_1 \in \mathbb{R}^k \).

Consider \( T_n \) as an estimator of \( \vartheta = g(\mu) \).

By Taylor’s expansion,

\[
T_n - \vartheta = [\nabla g(\mu)]^T (\bar{X} - \mu) + \frac{1}{2} (\bar{X} - \mu)^T \nabla^2 g(\mu) (\bar{X} - \mu) + o\left(\frac{1}{n}\right),
\]

where \( \nabla g \) is the \( k \)-vector of partial derivatives of \( g \) and \( \nabla^2 g \) is the \( k \times k \) matrix of second-order partial derivatives of \( g \).

By the CLT and Theorem 1.10(iii),

\[
\frac{n}{2} (\bar{X} - \mu)^T \nabla^2 g(\mu) (\bar{X} - \mu) \to_d \frac{Z_{\Sigma}^T \nabla^2 g(\mu) Z_{\Sigma}}{2},
\]

where \( Z_{\Sigma} = N_k(0, \Sigma) \). Thus,

\[
\frac{E[ Z_{\Sigma}^T \nabla^2 g(\mu) Z_{\Sigma} ]}{2n} = \frac{\text{tr}(\nabla^2 g(\mu) \Sigma)}{2n},
\]

is the \( n^{-1} \) order asymptotic bias of \( T_n = g(\bar{X}) \), where \( \text{tr}(A) \) denotes the trace of the matrix \( A \).

**Example 2.35.** Let \( X_1, ..., X_n \) be i.i.d. binary random variables with \( P(X_i = 1) = p \), where \( p \in (0, 1) \) is unknown.

Consider first the estimation of \( \vartheta = p(1 - p) \).

Since \( \text{Var}(\bar{X}) = p(1 - p)/n \), the \( n^{-1} \) order asymptotic bias of \( T_n = \bar{X}(1 - \bar{X}) \) according to (2) with \( g(x) = x(1 - x) \) is \(-p(1 - p)/n\).

On the other hand, a direct computation shows

\[
E[\bar{X}(1 - \bar{X})] = E \bar{X} - E \bar{X}^2 = p - (E \bar{X})^2 - \text{Var}(\bar{X}) = p(1 - p) - p(1 - p)/n.
\]

Hence, the exact bias of \( T_n \) is the same as the \( n^{-1} \) order asymptotic bias.

Consider next the estimation of \( \vartheta = p^{-1} \).

In this case, there is no unbiased estimator of \( p^{-1} \) (Exercise 84 in §2.6).

Let \( T_n = \bar{X}^{-1} \).

Then, an \( n^{-1} \) order asymptotic bias of \( T_n \) according to (2) with \( g(x) = x^{-1} \) is \((1 - p)/(p^2 n)\).

On the other hand, \( ET_n = \infty \) for every \( n \).

Asymptotic variance and mse

Like the bias, the mse of an estimator \( T_n \) of \( \vartheta \), \( \text{mse}_{T_n}(P) = E(T_n - \vartheta)^2 \), is not well defined if the second moment of \( T_n \) does not exist.

We now define a version of *asymptotic mean squared error* (amse) and a measure of assessing different point estimators of a common parameter.
**Definition 2.12.** Let $T_n$ be an estimator of $\vartheta$ for every $n$ and $\{a_n\}$ be a sequence of positive numbers satisfying $a_n \to \infty$ or $a_n \to a > 0$. Assume that $a_n(T_n - \vartheta) \to_d Y$ with $0 < EY^2 < \infty$.

(i) The asymptotic mean squared error of $T_n$, denoted by $\text{amse}_{T_n}(P)$ or $\text{amse}_{T_n}(\theta)$ if $P$ is in a parametric family indexed by $\theta$, is defined to be the asymptotic expectation of $(T_n - \vartheta)^2$, i.e., $\text{amse}_{T_n}(P) = EY^2/a_n^2$. The asymptotic variance of $T_n$ is defined to be $\sigma_{T_n}^2(P) = \text{Var}(Y)/a_n^2$.

(ii) Let $T'_n$ be another estimator of $\vartheta$. The asymptotic relative efficiency of $T'_n$ w.r.t. $T_n$ is defined to be $e_{T'_n, T_n}(P) = \text{amse}_{T_n}(P)/\text{amse}_{T'_n}(P)$.

(iii) $T_n$ is said to be asymptotically more efficient than $T'_n$ if and only if $\limsup_n e_{T'_n, T_n}(P) \leq 1$ for any $P$ and $<1$ for some $P$.

The amse and asymptotic variance are the same if and only if $EY = 0$.

By Proposition 2.3, the amse or the asymptotic variance of $T_n$ is essentially unique and, therefore, the concept of asymptotic relative efficiency in Definition 2.12(ii)-(iii) is well defined.

In Example 2.33, $\text{amse}_{X^2}(P) = \sigma_{X^2}^2(P) = 4\mu^2\sigma^2/n$.

In Example 2.34, $\sigma_{X^{(\alpha)}}^2(P) = [h_\alpha(\theta)]^2\text{Var}(Y)$ and $\text{amse}_{X^{(\alpha)}}(P) = [h_\alpha(\theta)]^2EY^2$.

When both $\text{mse}_{T_n}(P)$ and $\text{mse}_{T'_n}(P)$ exist, one may compare $T_n$ and $T'_n$ by evaluating the relative efficiency $\text{mse}_{T_n}(P)/\text{mse}_{T'_n}(P)$.

However, this comparison may be different from the one using the asymptotic relative efficiency in Definition 2.12(ii), since the mse and amse of an estimator may be different (Exercise 115 in §2.6).

The following result shows that when the exact mse of $T_n$ exists, it is no smaller than the amse of $T_n$.

It also provides a condition under which the exact mse and the amse are the same.

**Proposition 2.4.** Let $T_n$ be an estimator of $\vartheta$ for every $n$ and $\{a_n\}$ be a sequence of positive numbers satisfying $a_n \to \infty$ or $a_n \to a > 0$. Suppose that $a_n(T_n - \vartheta) \to_d Y$ with $0 < EY^2 < \infty$. Then

(i) $EY^2 \leq \liminf_n E[a_n^2(T_n - \vartheta)^2]$ and

(ii) $EY^2 = \lim_{n \to \infty} E[a_n^2(T_n - \vartheta)^2]$ if and only if $\{a_n^2(T_n - \vartheta)^2\}$ is uniformly integrable.

**Proof.** (i) By Theorem 1.10(iii),

$$\min\{a_n^2(T_n - \vartheta)^2, t\} \to_d \min\{Y^2, t\}$$

for any $t > 0$. Since $\min\{a_n^2(T_n - \vartheta)^2, t\}$ is bounded by $t$,

$$\lim_{n \to \infty} E(\min\{a_n^2(T_n - \vartheta)^2, t\}) = E(\min\{Y^2, t\})$$

(Theorem 1.8(viii)). Then

$$EY^2 = \lim_{t \to \infty} E(\min\{Y^2, t\})$$

$$= \lim_{t \to \infty} \limsup_{n \to \infty} E(\min\{a_n^2(T_n - \vartheta)^2, t\})$$

$$= \lim_{t,n} E(\min\{a_n^2(T_n - \vartheta)^2, t\})$$

$$\leq \limsup_n E[a_n^2(T_n - \vartheta)^2],$$
where the third equality follows from the fact that $E(\min\{a_n^2(T_n - \vartheta)^2, t\})$ is nondecreasing in $t$ for any fixed $n$.

(ii) The result follows from Theorem 1.8(viii).

**Example 2.36.** Let $X_1, ..., X_n$ be i.i.d. from the Poisson distribution $P(\theta)$ with an unknown $\theta > 0$.

Consider the estimation of $\vartheta = P(X_i = 0) = e^{-\theta}$.

Let $T_{1n} = F_n(0)$, where $F_n$ is the empirical c.d.f.

Then $T_{1n}$ is unbiased and has $\text{mse}_{T_{1n}}(\theta) = e^{-\theta}(1 - e^{-\theta})/n$.

Also, $\sqrt{n}(T_{1n} - \vartheta) \rightarrow_d N(0, e^{-\theta}(1 - e^{-\theta}))$ by the CLT.

Thus, in this case $\text{amse}_{T_{1n}}(\theta) = \text{mse}_{T_{1n}}(\theta)$.

Consider $T_{2n} = e^{-\bar{X}}$.

Note that $ET_{2n} = e^{n\theta(e^{-1/n} - 1)}$.

Hence $\text{nbr}_{2n}(\theta) \rightarrow \theta e^{-\theta}/2$.

Using Theorem 1.12 and the CLT, we can show that $\sqrt{n}(T_{2n} - \vartheta) \rightarrow_d N(0, e^{-2\theta})$.

By Definition 2.12(i), $\text{amse}_{T_{2n}}(\theta) = e^{-2\theta}/n$.

Thus, the asymptotic relative efficiency of $T_{1n}$ w.r.t. $T_{2n}$ is

$$e_{T_{1n},T_{2n}}(\theta) = \theta/(e^\theta - 1),$$

which is always less than 1.

This shows that $T_{2n}$ is asymptotically more efficient than $T_{1n}$.

The result for $T_{2n}$ in Example 2.36 is a special case (with $U_n = \bar{X}$) of the following general result.

**Theorem 2.6.** Let $g$ be a function on $\mathcal{R}^k$ that is differentiable at $\theta \in \mathcal{R}^k$ and let $U_n$ be a $k$-vector of statistics satisfying $a_n(U_n - \theta) \rightarrow_d Y$ for a random $k$-vector $Y$ with $0 < E\|Y\|^2 < \infty$ and a sequence of positive numbers $\{a_n\}$ satisfying $a_n \rightarrow \infty$. Let $T_n = g(U_n)$ be an estimator of $\vartheta = g(\theta)$. Then, the amse and asymptotic variance of $T_n$ are, respectively, $E\{(\nabla g(\theta))^\top Y\}^2/a_n^2$ and $[\nabla g(\theta)]^\top \text{Var}(Y)\nabla g(\theta)/a_n^2$. 
