Large Sample Properties

To this point, we have been concerned with small sample properties of estimators. A small sample is any finite sample. This, of course, includes all cases of practical interest. We turn now to a discussion of large sample properties. These properties are based upon an examination of the behavior of an estimator as the sample size is increased without bound. It is often the case that an estimator has a particularly troublesome small sample distribution, the form of which simplifies greatly as the sample size increases without bound. In addition, small sample differences between estimators often disappear in the limit. Thus, an examination of the large sample behavior of an estimator may provide useful insight into the behavior of an estimator for sufficiently large, but finite, samples.

The Basic Inequality

Let $b$ denote a positive constant, and $h(X)$ a non-negative function of the random vector $X$, then $P[h(X) \geq b] \leq E[h(X)]/b$, provided the expectation exists.

Let $A = \{X| h(X) \geq b\}$, then by properties of conditional expectation,

$$E[h(X)] = E[h(X)|A]P(A) + E[h(X)|A^c]P(A^c)$$
The basic inequality follows immediately.

The Chebyshev Inequality

Let $h(X) = (X-\mu)^2$ and $b = K^2\sigma^2$ in the Basic inequality above. Then,

$$P[(X-\mu)^2 \geq K^2\sigma^2] \leq \frac{E[(X-\mu)^2]}{(K^2\sigma^2)}$$

or

$$P[|X-\mu| \geq K\sigma] \leq \frac{\sigma^2}{(K^2\sigma^2)}$$

or

$$P[|X-\mu| \geq K\sigma] \leq \frac{1}{K^2}$$

for any $K>0$.

The Chebyshev Inequality bounds the "tail" probabilities of distributions with finite variances. For example, letting $K=3$, we find that the probability of a random variable falling outside an interval $\pm 3$ standard deviations about its mean cannot exceed $1/9$. This inequality is often a "crude" approximation. When specific information about the form of the distribution is available, a more precise figure can often be obtained. For example, the probability of a normal random variable falling outside an interval $\pm 3$ standard deviations about its mean is 0.0026.

Convergence in Probability

A random variable, $\omega$, converges in probability to a constant, $c$, if

$$\lim_{n \to \infty} P(|\omega - c| \geq \epsilon) = 0$$

for any $\epsilon>0$. Convergence in probability is often called stochastic convergence or weak convergence. If $\omega$ converges in probability to $c$, then $c$ is called the "probability limit" of $\omega$. 
This is denoted Plim $\omega = c$. If $\omega$ converges in probability to $c$, then, in the limit, the distribution of $\omega$ collapses about the single point $c$ (or becomes degenerate at $c$).

**Theorem (“Slutsky”)**

If $g(X)$ is a continuous function of $X$ and Plim $X = c$, then Plim $g(X) = g(c)$, provided $g(c)$ exists.

**Consistent Estimation**

An estimator $t(X)$ is consistent for the parameter vector $\theta$, if $t(X)$ converges in probability to $\theta$. That is, $t(X)$ is consistent for $\theta$ if

$$\lim_{n \to \infty} P(|t(X) - \theta| \geq \epsilon) = 0$$

for any $\epsilon > 0$. This is often denoted Plim $t(X) = \theta$ or $t(X) \rightarrow \theta$. If $t(X)$ is a consistent estimator of $\theta$, then the limiting distribution of $t(X)$ is degenerate at $\theta$.

**Law of Large Numbers**

If $X_i \sim iid(\mu, \sigma^2)$, with $\sigma^2$ finite, then $\bar{X}$ is a consistent estimator of $\mu$.

Under these conditions, $\bar{X} \sim (\mu, \sigma^2/n)$. Applying the Chebychev Inequality,

$$P(|\bar{X} - \mu| \geq K\sigma/\sqrt{n}) \leq 1/K^2$$

for any $K > 0$. Letting $\epsilon = K\sigma/\sqrt{n}$, or $K = \epsilon \sqrt{n}/\sigma$, we have

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \sigma^2/n\epsilon^2$$

for any $\epsilon > 0$.

Taking limits gives

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \epsilon) \leq \lim_{n \to \infty} \sigma^2/n\epsilon^2$$
or
\[
\lim_{n \to \infty} P(|\overline{X} - \mu| \geq \epsilon) \leq 0
\]
for any \( \epsilon > 0 \). Thus, \( \overline{X} \) is consistent for \( \mu \).

**Convergence in Mean Square**

A random variable, \( \omega \), converges in mean square to a constant, \( c \), if
\[
\lim_{n \to \infty} E[(\omega - c)^2] = 0.
\]
This is often called convergence in quadratic mean. Convergence in mean square is a sufficient condition for convergence in probability. Setting \( h(X) = (\omega - c)^2 \) and \( b = \epsilon^2 \) in the Basic inequality gives
\[
P[(\omega - c)^2 \geq \epsilon^2] \leq \frac{E[(\omega - c)^2]}{\epsilon^2}
\]
for any \( \epsilon > 0 \), or
\[
P[|\omega - c| \geq \epsilon] \leq \frac{E[(\omega - c)^2]}{\epsilon^2}
\]
for any \( \epsilon > 0 \). This implies that
\[
\lim_{n \to \infty} P[|\omega - c| \geq \epsilon] \leq \lim_{n \to \infty} \frac{E[(\omega - c)^2]}{\epsilon^2}
\]
or
\[
\lim_{n \to \infty} P[|\omega - c| \geq \epsilon] \leq 0 \quad \text{if} \quad \lim_{n \to \infty} E[(\omega - c)^2] = 0.
\]
Note that convergence in mean square is a sufficient condition for convergence in probability, but is not a necessary condition. The value of this result is that if we can show that
\[
\lim_{n \to \infty} \text{MSE}[t(X), \theta] = 0,
\]
then we know that \( t(X) \) is a consistent estimator of \( \theta \).
Application

We have seen that if $X_i \sim \text{iid} (\mu, \sigma^2)$, then $\overline{X} \sim (\mu, \sigma^2/n)$. Since the mean square error of an estimator may be written as the variance plus the square of the bias, we have

$$\text{MSE}(\overline{X}, \mu) = \sigma^2/n + (\mu - \mu)^2 = \sigma^2/n.$$  

Clearly, $\underset{n \to \infty}{\text{i.m.}} \text{MSE}(\overline{X}, \mu) \equiv \sigma^2/n = 0$. Thus, $\overline{X}$ is consistent for $\mu$.

Limitation

To illustrate that convergence in mean square is not a necessary condition for convergence in probability, consider the following contrived example. Assume that the random variable $\omega$ takes the value $\theta$ with probability $1-(1/n)$ and takes the value $n$ with probability $(1/n)$.

The random variable $\omega$ converges to $\theta$ since

$$\underset{n \to \infty}{\text{i.m.}} P(|\omega - \theta| \geq \epsilon) = \underset{n \to \infty}{\text{i.m.}} (1/n) = 0.$$  

Note that $|\omega - \theta| \neq 0$ only when $\omega = n$, which occurs with probability $(1/n)$.

The $\text{MSE}(\omega, \theta)$ is

$$E[(\omega - \theta)^2] = (\theta - \theta)^2[1-(1/n)] + (n - \theta)^2(1/n)$$

$$= (n - \theta)^2(1/n)$$

$$= n - 2\theta + (\theta^2/n)$$

Clearly, $\underset{n \to \infty}{\text{i.m.}} \text{MSE}(\omega, \theta) = \infty$. In summary, $\omega$ converges in probability to $\theta$ even though $\omega$ does not converge in mean square to $\theta$. To understand this result, note that while the probability that $\omega$ differs from $\theta$ is essentially zero for large $n$, that difference is large (of order $n$) when it does occur. Consequently, the variance of this random variable is not finite.
**Theorem** ("Sample moments converge to population moments.")

Let $m_h = \frac{1}{n} \sum_{i=1}^{n} X_i^h$ denote the sample moment of order $h$ and $\mu_h = E(X^h)$ the corresponding population moment. If the $X_i$ are independent and identically distributed, then $m_h \to \mu_h$ if $\mu_{2h}$ exists.

Let $Y_i = X_i^h$, then $m_h = Y$. Clearly,

$$E(m_h) = E(Y) = E(Y_i) = E(X_i^h) = \mu_h$$

In addition,

$$V(m_h) = V(Y) = \frac{1}{n} V(Y_i) = \frac{1}{n} \left[ E(Y_i^2) - E(Y_i)^2 \right]$$

$$= \frac{1}{n} \left[ E(X_i^{2h}) - E(X_i^h)^2 \right]$$

$$= \frac{1}{n} \left[ \mu_{2h} - \mu_h^2 \right]$$

Consequently,

$$MSE(m_h, \mu_h) = \frac{1}{n} \left[ \mu_{2h} - \mu_h^2 \right]$$

If $\mu_{2h}$ is finite, then $\lim_{n \to \infty} MSE(m_h, \mu_h) = 0$, which is a sufficient condition to insure that $m_h \to \mu_h$. 