Convergence in Distribution

A random variable, \( Y_n \), with distribution function \( F_n(y) \), converges in distribution to a random variable, \( Y \), with distribution function \( F(y) \), if and only if

\[
\lim_{n \to \infty} F_n(y) = F(y)
\]

at any point of continuity in \( F(y) \). The distribution function \( F(y) \) is said to be the "limiting distribution" of the random variable \( Y_n \). This is often denoted as \( Y_n \overset{D}{\to} Y \).

Given the unique relationship between the distribution function and MGF of a random variable, this definition may be restated as follows. A random variable, \( Y_n \), with MGF, \( M_n(t) \), converge in distribution to a random variable, \( Y \), with MGF, \( M(t) \), if and only if

\[
\lim_{n \to \infty} M_n(t) = M(t)
\]

for \( t \) in an open region about the origin.

The Central Limit Theorem

We have seen that the limiting distribution of \( \bar{X} \) is degenerate for random variables with finite variances. The form of the limiting distribution is dominated by the collapsing sample variance. By standardizing the sample mean, however, a remarkable invariance in the form of the limiting distribution emerges. This result is known as the Central Limit Theorem (CLT).
If \( \{X_n\} \) denotes a sequence of iid random variables with common mean \( \mu \) and common variance \( \sigma^2 \), then the standardized variable 
\[
Z_n = \sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow{d} N(0,1).
\]

Note that 
\[
Z_n = \sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right),
\]
where \( Z_i = (X_i - \mu)/\sigma \). Clearly, the \( Z_i \) are iid with mean 0 and variance 1. The common MGF of the \( Z_i \) is \( M(t) = E[\exp(tZ)] \). (The observation index may be dropped from since identically-distributed observations have a common MGF.) A first-order McClaurin series with remainder gives
\[
M(t) = M(0) + M_1(0)t + \frac{1}{2}M_2(\delta)t^2 \quad \text{for } 0 < \delta < t
\]
where the subscript indicates the order of the partial derivative of \( M(t) \) with respect to \( t \). This may be written
\[
M(t) = 1 + E(Z)t + \frac{1}{2}M_2(\delta)t^2
\]
\[
= 1 + \frac{1}{2}t^2 + \frac{1}{2}[M_2(\delta)-1]t^2 \quad \text{for } 0 < \delta < t
\]
The MGF of \( Z_n \) is
\[
M_Z(t) = E[\exp(tZ_n)]
\]
\[
= E\{\exp[(tZ_1/\sqrt{n}) + \ldots + (tZ_n/\sqrt{n})]\}
\]
\[
= E[\exp(tZ_1/\sqrt{n})]\cdots E[\exp(tZ_n/\sqrt{n})]
\]
\[
= [M(t/\sqrt{n})]^n
\]
\[
= \{ 1 + (\frac{1}{2}t^2)/n + [M_2(\delta)-1](\frac{1}{2}t^2)/n \}^n
\]
where the second equality follows from the definition of \( Z_n \), the third from statistical independence of the \( Z \), and the fourth from the fact that the \( Z \) are identically distributed, and where \( \delta \) must now satisfy \( 0 < \delta < t/\sqrt{n} \).

Note that \( \delta \to 0 \) as \( n \to \infty \). Since the \( Z \) have unit variance, this implies that
\[
\lim_{n \to \infty} [M_2(\delta)-1] = 0
\]
In addition, since
\[
\lim_{n \to \infty} \left(1 + \frac{a}{n}\right)^n = \exp(a)
\]
the limiting MGF of \( Z_n \) is
\[
\lim_{n \to \infty} M_Z(t) = \exp(\frac{1}{2}t^2).
\]
This is just the MGF of a standard normal! That is, \( Z_n = \sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right) \xrightarrow{D} N(0,1) \).

The CLT states that, subject to the conditions of the theorem, the limiting distribution of the standardized sample mean is invariant to the form of the small-sample distribution generating the data. For sufficiently large samples, this invariance may be exploited to obtain a useful approximation to the small-sample distribution.

**Application**

Consider a SI sequence of Bernoulli trials. Each trial \( X_i \) has mean \( p \) and variance \( p(1-p) \).

The sample proportion \( \hat{p} = \bar{X} \) has mean \( p \) and variance \( p(1-p)/n \). Since this example satisfies the conditions of the CLT, we know that
\[
\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}} \xrightarrow{D} N(0,1)
\]

**Application**

Assume that \( X_i \sim \text{iidU}(0, \alpha) \). Then \( E(\bar{X}) = \frac{\alpha}{2} \) and \( \text{Var}(\bar{X}) = \frac{\alpha^2}{12n} \). Since the conditions of the CLT are satisfied, we know that
\[
\frac{\sqrt{n} \left( \bar{X} - \frac{\alpha}{2} \right)}{\frac{\alpha^2}{\sqrt{12}}} \xrightarrow{D} N(0,1)
\]
or expressed in terms of the method of moments estimator $\hat{\alpha} = 2\bar{X}$, we have

$$\frac{\sqrt{n}(\hat{\alpha} - \alpha)}{\sqrt{\frac{\alpha^2}{3}}} \overset{D}{\rightarrow} N(0,1)$$

Asymptotic Distributions

The Central Limit Theorem states that $Z_n \overset{D}{\rightarrow} N(0,1)$ under fairly general conditions. Clearly, $Z_n \overset{D}{\rightarrow} N(0,1)$ does not imply $Z_n \sim N(0,1)$ for finite $n$. For sufficiently large samples, however, the assumption $Z_n \sim N(0,1)$ may provide a useful approximation to the exact (small sample) distribution of $Z_n$. Is anything changed if we assume that $\bar{X} \sim N(\mu, \sigma^2/n)$ instead? The answer is no, since for given values of $n$, $Z_n \sim N(0,1)$ if and only if $\bar{X} \sim N(\mu, \sigma^2/n)$. The same approximation is obtained with either assumption. The term “asymptotic distribution” is used to refer to this process of approximation. That is, if $Z_n \overset{D}{\rightarrow} N(0,1)$, we say that $\bar{X} \overset{A}{\sim} N(\mu, \sigma^2/n)$.

Note that it is incorrect to say that $\bar{X} \overset{D}{\rightarrow} N(\mu, \sigma^2/n)$.

The concept of an asymptotically normal estimator can be generalized to other estimation problems. Consider an estimator $t(X)$ for a parameter vector $\Theta$. If $\sqrt{n}[t(X)-\tau] \overset{D}{\rightarrow} N(0,\Sigma)$, then we say that $t(X) \overset{A}{\sim} N(\tau, n^{-1}\Sigma)$. If the asymptotic distribution exists, then $\tau$ is the asymptotic mean of $t(X)$ and $n^{-1}\Sigma$ is the asymptotic covariance matrix of $t(X)$. The estimator $t(X)$ is an asymptotically unbiased estimator of $\Theta$, if $\tau = \Theta$ for any valid $\Theta$. The estimator $t(X)$ is an asymptotically efficient estimator of $\Theta$ if $t(X)$ is asymptotically unbiased and has no larger asymptotic variance than any other asymptotically unbiased estimator.
The Asymptotic CR Bound

Assume that the estimator \( t(X) \) is asymptotically normal and asymptotically unbiased. That is, \( t(X) \converges \text{N}(\theta, n^{-1}\Sigma) \). Then a sufficient condition for \( t(X) \) to be an asymptotically efficient estimator of \( \theta \) is that its covariance matrix \( n^{-1}\Sigma \) equals the lower bound \( n^{-1}[\text{lim}_{n \to \infty} n^{-1}I(\theta)]^{-1} \) where \( I(\theta) \) is the small-sample information matrix.

Application

Consider a SI sequence of Bernoulli trials. We have seen that the sample proportion \( \hat{p} = \bar{X} \) has mean \( p \) and variance \( p(1-p)/n \), and that by the CLT,

\[
\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}} \overset{d}{\to} \text{N}(0,1)
\]

which implies that \( \sqrt{n}(\hat{p} - p) \overset{d}{\to} \text{N}[0,p(1-p)] \). Thus, \( \hat{p} \converges \text{N}[p, n^{-1}p(1-p)] \).

The estimator \( \hat{p} \) is asymptotically unbiased since the mean of the asymptotic distribution is \( p \). The estimator \( \hat{p} \) is also asymptotically efficient. Recall that the small sample information matrix for this problem is \( I(p) = n/[p(1-p)] \). The asymptotic CR bound is \( n^{-1}[\text{lim}_{n \to \infty} n^{-1}I(p)]^{-1} \). This is just \( n^{-1}p(1-p) \). Since \( \hat{p} \) is asymptotically unbiased and has asymptotic variance that equals the asymptotic CR bound, it is asymptotically efficient.
**Application**

Assume that $X \sim N(\mu, \sigma^2 I_n)$. Since this case satisfies the conditions of the CLT, we know that $\bar{X} \overset{d}{\sim} N(\mu, \sigma^2/n)$. (This is also the small sample distribution of $\bar{X}$ for normal populations, so there is no error in approximation here.) The estimator $\bar{X}$ is asymptotically unbiased since its asymptotic mean is $\mu$. We have seen that the small-sample information matrix in this case is

$$
\begin{bmatrix}
(n/\sigma^2) & 0 \\
0 & (n/2\sigma^4)
\end{bmatrix}
$$

Thus, the asymptotic CR bound, $n^{-1}[\nu_{n \to \infty}^{-1} n^{-1} I(\Theta)]^{-1}$, is

$$
\begin{bmatrix}
\sigma^2/n & 0 \\
0 & (2\sigma^4/n)
\end{bmatrix}
$$

Since $\bar{X}$ is asymptotically unbiased and has asymptotic variance that equals the asymptotic CR bound, it is asymptotically efficient.