

Asymptotic Approximations

1 Introduction

When exact sampling distributions for estimators and test statistics are not available, econometricians often rely on approximations obtained from asymptotic arguments. These approximations are sometimes quite accurate and can often be constructed without a complete specification of the population distribution for the data. Suppose $F_n(x)$ is the unknown cumulative distribution function (cdf) for some statistic based on a sample of size n . If it can be shown that the sequence of functions $F_1(x), F_2(x), \dots$ converges rapidly to a known limit $F(x)$ as n tends to infinity, then we might use $F(x)$ as an approximation to $F_n(x)$ even for moderate values of n . The quality of the approximation depends on the speed of convergence, but can be checked by computer simulation.

The simplest example of this approach is the average of independent draws from a distribution possessing a finite variance. Let $\bar{Y}_n = n^{-1} \sum_1^n Y_i$, where the Y 's are i.i.d. with $E[Y_i] = \mu$ and $\text{var}[Y_i] = \sigma^2$. By an easy calculation, we find that \bar{Y}_n has mean μ and variance σ^2/n . Although the exact distribution of \bar{Y}_n depends on the distribution of the Y 's, a simple asymptotic approximation is always available. The cdf F_n for \bar{Y}_n is quite sensitive to the value of n so we would not expect the limit of the sequence F_1, F_2, \dots to yield a good approximation to F_n unless n is very large. But the standardized random variable $S_n = \sqrt{n}(\bar{Y}_n - \mu)/\sigma$ has mean zero and variance one for every n ; its cdf, say F_n^* , is much less sensitive to the value of n . Thus, if we could find the limit F^* of the sequence F_1^*, F_2^*, \dots , we might be willing to use it as an approximation to the distribution of S_n even if n is quite small. As discussed in section 3 below, the sequence F_1^*, F_2^*, \dots necessarily converges to the standard normal cdf. This leads us to approximate F_n by the cdf of a $N(\mu, \sigma^2/n)$ distribution.

Many estimators and test statistics used in econometrics are complicated nonlinear functions of the data and are not covered by the standard theorems which deal with sums of random variables. However, these statistics can often be approximated by sums and are still amenable to asymptotic analysis. A general algorithm for approximating distributions is given in Section 5 below. A brief introduction to the theory underlying these approximations is sketched in Sections 2-4.

2 Convergence in Probability

A sequence of random variables $\{S_n\}$ is said to converge in probability to a constant c if, for any $\varepsilon > 0$, $P[|S_n - c| > \varepsilon]$ tends to zero as n tends to infinity. In that case we write $S_n \xrightarrow{p} c$ or $\text{plim } S_n = c$ (short for "the probability limit of S_n is c ."). If $\{X_n\}$ is a sequence of random variables such that $\text{plim}(n^r X_n) = 0$, we sometimes write $X_n = o_p(n^{-r})$. Thus $X_n = c + o_p(1)$ is another way of saying that $\text{plim } X_n = c$.

The *weak law of large numbers* states that the sample average $\bar{Y} = n^{-1} \sum_i Y_i$ of n i.i.d. random variables having finite population expectation θ converges in probability to that expectation. We say that \bar{Y} is a *consistent* estimator of θ .

Proof of consistency when Y 's have finite variance: If $\text{var}(Y_i) = \sigma^2$ then $\text{var}(\bar{Y}) = \sigma^2/n$ and, by Chebyshev's inequality, $P[|\bar{Y} - \theta| > \varepsilon] \leq \sigma^2/\varepsilon^2 n$. If the variance is not finite, a more delicate argument is needed.

We need not restrict ourselves to the i.i.d. case. Consider again the sample average $\bar{Y} = n^{-1} \sum_{t=1}^n Y_t$, where now the Y 's are a stationary time series with mean θ and covariances of the form $\text{cov}(Y_t, Y_s) = \gamma_{|t-s|}$. By Chebyshev's inequality, $\bar{Y} \xrightarrow{p} \theta$ as long as

$$\text{var}[\bar{Y}] = n^{-2} \sum_{t=1}^n \sum_{s=1}^n \text{cov}(Y_t, Y_s) = n^{-1} \sum_{r=-n+1}^{n-1} (1 - \frac{|r|}{n}) \gamma_r$$

tends to zero. This will necessarily occur if the sequence of autocovariances γ_r tends to zero as r tends to infinity.

Probability limits can be found for sequences other than those constructed from sample sums. For example, the sample median of n i.i.d. draws from a continuous distribution can be shown to converge in probability to the population median. The concept of convergence in probability can also be generalized to vectors and matrices. If $\{S_n\}$ is a sequence of $p \times q$ random matrices where p and q do not depend on n , then we say that $\text{plim } S_n = A$ if each element of S_n converges in probability to the corresponding element of the $p \times q$ matrix A . For example, if $S_n = n^{-1} X'u$, where the $n \times p$ matrix X is nonrandom and the n elements of the vector u are zero-mean uncorrelated random variables with common variance σ^2 , a sufficient condition for $\text{plim } S_n = 0$ is that the elements of X are bounded in absolute value by some number M .

PROOF: For nonrandom c , $E[c'S_n] = 0$ and $\text{var}[c'S_n] = \sigma^2 c'X'Xc/n^2 \leq \sigma^2 M^2 k'c/n \rightarrow 0$.

3 Convergence in Distribution

A sequence of random variables S_n is said to converge in distribution to the random variable X if the cdf of S_n converges to the cdf of X at all continuity points of the later function. The Lindeberg-Levy *central limit theorem* says that, if the random variables Y_1, \dots, Y_n are i.i.d. with mean μ and variance σ^2 , the cdf for $S_n = n^{-1/2} \sum_{j=1}^n (Y_j - \mu)/\sigma$ converges to the cdf of a $N(0, 1)$ variate as n tends to infinity. This is often written $S_n \xrightarrow{d} N(0, 1)$.

SKETCH OF PROOF: For t in a neighborhood of the origin, define the characteristic function for $(Y - \mu)/\sigma$ as $\psi(t) \equiv Ee^{it(Y-\mu)/\sigma}$ where $i = \sqrt{-1}$. Note that ψ is necessarily twice differentiable with $\psi(0) = 1$, $\psi'(0) = 0$ and $\psi''(0) = -1$. By Taylor's Theorem we have $\psi(t) = 1 - t^2/2 + o(t^2)$ as $t \rightarrow 0$. But the characteristic function for S_n is

$$\psi_{S_n}(t) = Ee^{itS} = [\psi(t/\sqrt{n})]^n = [1 - t^2/2n + o(n^{-1})]^n.$$

Hence, as $n \rightarrow \infty$, $\psi_{S_n}(t) \rightarrow e^{-t^2/2}$. It can be shown that the characteristic function uniquely determines a distribution. Since the limiting characteristic function is that of a $N(0, 1)$, the limiting distribution must also be $N(0, 1)$.

Again, it is possible to drop the i.i.d. assumption and extend to the vector case. We state, without proof, a few important limit theorems for sample sums.

1. If the random vectors Y_1, \dots, Y_n are i.i.d. with mean μ and variance matrix Ω , the joint cdf for $S_n = n^{-1/2} \sum_{j=1}^n (Y_j - \mu)$ converges to the cdf of a multivariate normal vector with mean zero and variance matrix Ω .

2. (Lyapunov) Let Y_1, \dots, Y_n be mutually independent k -dimensional random vectors with $EY_i = \mu_i$. Define $A_n = \sum_{i=1}^n \text{var}(Y_i)$. If, as n tends to infinity, A_n/n tends to a positive definite limit matrix A and the absolute third moments $E|Y_{ki} - \mu_{ki}|^3$ are uniformly bounded as $n \rightarrow \infty$, then $A_n^{-1/2} \sum_{i=1}^n (Y_i - \mu_i)$ converges to a $N(0, I_k)$ distribution.
3. Let X be an $n \times k$ nonrandom matrix and let u be a vector of n i.i.d. random variables with zero mean and unit variance. If the elements of X are uniformly bounded and the smallest characteristic root of $X'X$ grows without bound as n tends to infinity, then the limiting distribution of $(X'X)^{-1/2} X'u$ is $N(0, I)$.
4. Suppose Y_1, \dots, Y_n are consecutive observations from a stationary scalar time series process with moving average representation $Y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$, where the ε_t are i.i.d. with mean zero and variance σ^2 . If $\sum_{j=0}^{\infty} |c_j| < \infty$ and $\sum_{j=0}^{\infty} c_j = d \neq 0$, then the limiting distribution of $n^{-1/2} \sum_1^n Y_t$ is $N(0, \sigma^2 d^2)$.
5. Suppose the time series Y_1, \dots, Y_n is a martingale difference sequence; that is, the conditional expectations $E(Y_t | Y_1, Y_2, \dots, Y_{t-1}) = 0$ for all integer $t \leq n$. Suppose the third absolute moments $E|Y_n^3|$ are bounded and $\lim n^{-1} \sum_{t=1}^n E(Y_t^2) = \sigma^2$ as $n \rightarrow \infty$. Then, if $\text{plim } n^{-1} \sum_{t=1}^n Y_t^2 = \sigma^2$, the limiting distribution of $n^{-1/2} \sum_1^n Y_t$ is $N(0, \sigma^2)$.

Asymptotic normality is not restricted to statistics based on sums. Consider again the sample median M of n i.i.d. draws from a continuous density f with population median θ . Then it can be shown that $\sqrt{n}(M - \theta)2f(\theta)$ has a limiting $N(0,1)$ distribution as long as $f(\theta) > 0$. Thus, in large samples, one might approximate the distribution of M by a normal with mean θ and variance $[4nf^2(\theta)]^{-1}$.

4 Some Useful Facts

There are many results that allow us to deduce the limiting behavior of functions of convergent random sequences. Let $\{X_n\}$ and $\{Y_n\}$ be sequences of p -dimension random vectors, let $\{W_n\}$ be a sequence of $p \times q$ random matrices, and let $\{Z_n\}$ be a sequence of q -dimensional random vectors. Let $g : R^p \rightarrow R$ and $h : R^q \rightarrow R$ be functions not depending on n . Then, using standard limit arguments, one can prove the following useful facts:

1. *Slutsky's Theorem*: If $\text{plim } X_n = a$ and g is continuous at a , then $\text{plim } g(X_n) = g(a)$.
2. If $X_n \xrightarrow{d} X$ and $\text{plim}(X_n - Y_n) = 0$, then $Y_n \xrightarrow{d} X$.
3. If $\text{plim } X_n = a$, $\text{plim } W_n = B$, and $Z_n \xrightarrow{d} Z$, then the random vector $X_n + B_n Z_n$ converges in distribution to $a + BZ$.
4. If $Z_n \xrightarrow{d} Z$ and $h(\cdot)$ is everywhere continuous, then $h(Z_n) \xrightarrow{d} h(Z)$.
5. *The delta method*: If $g(\cdot)$ is differentiable at a and $\sqrt{n}(X_n - a) \xrightarrow{d} X$, then

$$\sqrt{n}[g(X_n) - g(a)] \xrightarrow{d} c'X,$$

where c is the gradient vector Dg evaluated at a .

5 Approximating Distributions

Using the concepts of convergence in probability and convergence in distribution, we can now describe more precisely the asymptotic approach to approximating the distribution of a p -dimensional sample statistic T_n :

1. If the center or the dispersion of T_n changes substantially with sample size n , find a new variable $S_n = h(T_n, n)$ whose center and dispersion are not sensitive to n . (We assume the mapping from T to S is one-to-one.) If $\text{plim } T_n = \theta$, the linear mapping $S_n = \sqrt{n}(T_n - \theta)$ often works. S_n is sometimes referred to as a "standardized" or a "normalized" version of the statistic T_n .
2. Find the limiting distribution of S_n . Often this can be accomplished by showing that S_n can be rewritten as $A_n X_n + b_n$, where A_n is a matrix converging in probability to a nonrandom matrix A , b_n is a vector converging in probability to a zero vector, and X_n converges in distribution to a random vector X with cdf F ; the limiting distribution of S_n is then the distribution of AX . If X is $N(0, \Sigma)$, then the limiting distribution of S_n is $N(0, A\Sigma A')$.
3. Approximate the distribution of T by inverting the standardizing transformation. If, for example, the limiting distribution of $\sqrt{n}(T_n - \theta)$ is $N(0, A\Sigma A')$, this means approximating the distribution of T_n by a normal with mean θ and covariance matrix $A\Sigma A'/n$.

As an example, let $y = X\beta + u$ where X is an $n \times K$ matrix and $E[u] = 0$. The OLS estimator $b = \beta + (X'X)^{-1}X'u$ is normal if X is nonrandom and u is normal. But a normal approximation can be justified under much weaker conditions. Suppose we can show that $\text{plim } X'u/n = 0$ and $\text{plim } X'X/n = A$, where A is nonsingular. Then, $\text{plim } b = \beta + A0 = \beta$. If, in addition, $\text{plim } u'u/n = \sigma^2$, then $s^2 \xrightarrow{p} \sigma^2$, where $s^2 = (y - Xb)'(y - Xb)/(n - K)$. If, using one of the central limit theorems given in Section 3 above, we can show that $(X'X)^{-1/2}X'u \xrightarrow{d} N(0, \sigma^2 I)$, then it follows that $(X'X)^{1/2}(b - \beta)/s \xrightarrow{d} N(0, I)$. In large samples, we might act as though b were distributed as $N[\beta, s^2(X'X)^{-1}]$. We could form an approximate confidence set for β using the fact that $(b - \beta)'X'X(b - \beta)/s^2$ is asymptotically chi-squared.

As a second example, consider the approximate distribution of $b_1 b_2$, the product of the first two components of the LS estimator b just considered. By Slutsky's theorem, $\text{plim}(b_1 b_2) = \beta_1 \beta_2$ as long as $\text{plim } b = \beta$. Furthermore, the delta method says that $\sqrt{n}(b_1 b_2 - \beta_1 \beta_2)$ has the same limiting distribution as $\beta_2 \sqrt{n}(b_1 - \beta_1) + \beta_1 \sqrt{n}(b_2 - \beta_2)$. Thus, in large samples, we might behave as though $b_1 b_2$ were normal with mean $\beta_1 \beta_2$ and variance $\beta_2^2 q_{11} + \beta_1^2 q_{22} + 2\beta_1 \beta_2 q_{12}$ where q_{ij} is the ij element of $s^2(X'X)^{-1}$. [Note: if β_1 and β_2 are both zero, the limiting distribution of $\sqrt{n}(b_1 b_2 - \beta_1 \beta_2)$ is degenerate; normalizing by n , instead of \sqrt{n} is needed.]