Economics 241B
Modes of Convergence

We have concentrated on the (exact) finite-sample distribution for the OLS estimator. The finite-sample theory breaks down if one of the following three assumptions is violated: 1) the exogeneity of the regressors, 2) the normality of the error term, 3) the linearity of the regression. We now develop an alternative approach, which requires only the third assumption. The approach, termed asymptotic or large-sample theory, derives an approximation to the distribution of the estimator and its associated test statistics, assuming the sample size is sufficiently large.

Rather than making assumptions on the sample of a given size, large-sample theory makes assumptions on the stochastic process that generates the sample. We study the limiting behavior of a sequence of random variables $\left(\bar{Y}_1, \bar{Y}_2, \ldots\right)$, which we denote by $\left\{\bar{Y}_n\right\}$.[1]

Modes of Convergence

Proof that an estimator is consistent requires that we construct a limit argument. One might naturally ask, what is the limit of $\bar{Y}_n$ as $n \to \infty$? Because $\bar{Y}_n$ is a random variable, the (deterministic) limit does not exist. To understand the point, recall the following

Definition. *A sequence of real numbers $\{\alpha_1, \alpha_2, \ldots\}$ is said to converge to a real number α if, for any $\epsilon > 0$ there exists an integer N such that for all $n > N$:*

$$|\alpha_n - \alpha| < \epsilon.$$

We express the convergence as $\alpha_n \to \alpha$ as $n \to \infty$ or (more precisely) as $\lim_{n\to\infty} \alpha_n = \alpha$. The definition applies to vectors. Let $\alpha_n \in \mathbb{R}^k$. If $\lim_{n\to\infty} \alpha_{n,i} = \alpha_i$ for each $i = 1, \ldots, k$, then $\lim_{n\to\infty} \alpha_n = \alpha$.

Example. Let $\alpha_n = 1 - \frac{1}{n}$. For any $\epsilon > 0$ there exists an integer N such that for all $n > N$: $|\alpha_n - 1| < \epsilon$, so $\lim_{n\to\infty} \alpha_n = 1$. (For example, $N = [\frac{1}{\epsilon}]$, where $[\cdot]$ yields the nearest integer that is at least as large as $\frac{1}{\epsilon}$.)

[1] A sequence is the most common mathematical construct that may possess a limit, but certainly not the only one. A sequence is a countably infinite collection of numbers (vectors, matrices). A sequence must embody a rule that defines it and is ordered: The rule generates the n^{th} member of the sequence, for any positive integer n.

Example. Let $\alpha_n = (-1)^n$. For values of $\epsilon > 1$ there exists an integer N (for example $N = 1$) such that for all $n > N$: $|\alpha_n - 0| < \epsilon$. For values of $|\epsilon| \leq 1$ and for every value of N, $|\alpha_n - 0| = 1 > \epsilon$ for all $n > N$. Because a similar argument holds for any proposed constant limit value, the limit does not exist.

The definition of a deterministic limit requires that $|\alpha_n - \alpha| < \epsilon$ for *all* values of $n > N$, not merely for "most" or "several" values of $n > N$. Because $\bar{Y}_n$ is a random variable, we cannot be sure that $\left|\bar{Y}_n - \mu\right| < \epsilon$ for *all* values of $n > N$. (Note the problem does not depend on the value of ϵ!)

Convergence in Probability

To construct a limit for a sequence of random variables, we must weaken the definition of a deterministic limit. Because $\left|\bar{Y}_n - \mu\right| < \epsilon$ never holds with certainty, we modify the definition in such a way that the conclusion states that $P(\left|\bar{Y}_n - \mu\right| < \epsilon)$ approaches 1 as $n \to \infty$.

Definition. *A sequence of random variables* $\left\{\bar{Y}_1, \bar{Y}_2, \ldots\right\}$ *is said to converge in probability (or weakly) to a real number* μ *if, for any* $\epsilon > 0$ *and* $\delta > 0$ *there exists an integer* N *such that for all* $n > N$*:*

$$P(\left|\bar{Y}_n - \mu\right| < \epsilon) > 1 - \delta.$$

For this definition, think of selecting (ε, δ) for which there exists an $N(\varepsilon, \delta)$ that satisfies the relation. We express the convergence as $\bar{Y}_n \xrightarrow{P} \mu$ (equivalently, $\bar{Y}_n - \mu \xrightarrow{P} 0$) as $n \to \infty$ or (less precisely because there are different types of probabilistic convergence) as $p\lim_{n\to\infty} \bar{Y}_n = \mu$. The definition applies to vectors. Let $\bar{Y}_n \in \mathbb{R}^k$. If $p\lim_{n\to\infty} \bar{Y}_{n,i} = \mu_i$ for each $i = 1, \ldots, k$, then $p\lim_{n\to\infty} \bar{Y}_n = \mu$.

An alternative way to express the same

Definition. *A sequence of random variables* $\left\{\bar{Y}_1, \bar{Y}_2, \ldots\right\}$ *is said to converge in probability (or weakly) to a real number* μ *if, for any* $\epsilon > 0$*:*

$$\lim_{n\to\infty} P(|\bar{Y}_n - \mu| < \epsilon) = 1.$$

To link the two definitions, fix ε and select N, for which there exists a $\delta(N)$ that satisfies the relation in the first definition. Because $\delta(N) \to 0$ as N grows,

$$\lim_{n\to\infty} P(|\bar{Y}_n - \mu| < \epsilon) = \lim_{n\to\infty} (1 - \delta(N)) = 1.$$

Example. Let $\bar{Y}_n(\omega)$ be the estimator where ω indexes the underlying event over which the probability measure is defined. We begin by drawing a value of ω and

obtain a corresponding value for $\bar{Y}_n(\omega)$. (Note that if ω is fixed, the sequence $\{\bar{Y}_t(\omega)\}_{t=1}^n$ is not random.) Let $n_1 = 100$, let $n_2 = 500$, and assume that μ is known. From the first draw of ω we obtain

$$|\bar{Y}_{n_1} - \mu| = .04 \text{ and } |\bar{Y}_{n_2} - \mu| = .001.$$

For the second draw of ω we obtain

$$|\bar{Y}_{n_1} - \mu| = .01 \text{ and } |\bar{Y}_{n_2} - \mu| = .02.$$

From the two draws of ω we see two important points. First, the accuracy of an estimator with a fixed sample size varies across draws of ω (note the values of $|\bar{Y}_{n_1} - \mu|$ vary over draws as do the values of $|\bar{Y}_{n_2} - \mu|$). Second, the accuracy of an estimator may not increase monotonically within a sequence. (The estimator $\bar{Y}_n$ is more accurate for a sample size of 500 than for a sample size of 100 for the first draw of ω, while $\bar{Y}_n$ is less accurate for a sample size of 500 than for a sample size of 100 for the second draw of ω.)

If we perform the same calculations for every value of $\omega \in \Omega$, then we construct the distribution for each of the random variables $|\bar{Y}_{n_1} - \mu|$ and $|\bar{Y}_{n_2} - \mu|$. Because we could, in principle, construct the distribution for any value of n, let us assume that we have also constructed the distribution for a third value $n_3 = 1000$. From the distributions we can calculate $P(|\bar{Y}_{n_1} - \mu| < \epsilon)$, $P(|\bar{Y}_{n_2} - \mu| < \epsilon)$, and $P(|\bar{Y}_{n_3} - \mu| < \epsilon)$ for any value of ϵ. Let $\epsilon = .01$. Suppose that

$$P(|\bar{Y}_{n_1} - \mu| < .01) = .87,\ P(|\bar{Y}_{n_2} - \mu| < .01) = .85,\ P(|\bar{Y}_{n_3} - \mu| < .01) = .90,$$

that $P(|\bar{Y}_n - \mu| < .01) < .85$ for all $n < n_1$, and that $P(|\bar{Y}_n - \mu| < .01) > .90$ for all $n > n_3$. Two important facts are again revealed. First, that $P(|\bar{Y}_n - \mu| < \epsilon)$ gets large as n increases (this is true for any value of ϵ). Second, that $P(|\bar{Y}_n - \mu| < \epsilon)$ need not be monotonic in n (note that $P(|\bar{Y}_{n_1} - \mu| < .01) > P(|\bar{Y}_{n_2} - \mu| < .01)$). All we need is that $\min_{n>N} P(|\bar{Y}_n - \mu| < \epsilon)$ be monotonic in N.

To motivate the first definition we must specify the value of δ. If $\delta = .15$ and $\epsilon = .01$, then for all $n > n_2$ we have that $P(|\bar{Y}_n - \mu| < .01) > .85$. If $\delta = .10$ and $\epsilon = .01$, then for all $n > n_3$ we have that $P(|\bar{Y}_n - \mu| < .01) > .90$. The value of N that appears in the first definition depends on both ϵ and δ.

To motivate the second definition, we do not need to specify a value for δ. Rather, we must study the value of $P(|\bar{Y}_n - \mu| < \epsilon)$ for all values of n. If we fix $\epsilon = .01$ and let n increase, then, even though the probabilities need not be monotonic as in the above example, $P(|\bar{Y}_n - \mu| < .01)$ must tend to one. To

understand why the two definitions are equivalent, because one can choose a value of δ that is arbitrarily close to 0, the first definition is equivalent to $\lim_{n\to\infty} P(|\bar{Y}_n - \mu| < \epsilon) = 1$.

Almost Sure Convergence

Observe that the definitions of convergence in probability contained reference to $P(|\bar{Y}_n - \mu| < \epsilon)$. Because $\bar{Y}_n$ is constructed from a finite (n) number of random variables, the probability is defined with respect to a finite dimensional sequence of random variables $\{Y_t(\omega)\}_{t=1}^{n}$. To strengthen the definition of convergence, we will need to work with an infinite dimensional sequence of random variables $\{Y_t(\omega)\}_{t=1}^{\infty}$, which requires more complicated mathematics. Fortunately, the complications are hidden beneath the level at which we use the concept.

Definition. *A sequence of random variables* $\left\{\bar{Y}_1, \bar{Y}_2, \ldots\right\}$ *is said to converge almost surely (or strongly) to a real number* μ *if, for any* $\epsilon > 0$*:*

$$\lim_{N\to\infty} P(\sup_{n>N} |\bar{Y}_n - \mu| < \epsilon) = 1.$$

We express the convergence as $\bar{Y}_n \overset{AS}{\to} \mu$ as $n \to \infty$. The definition applies to vectors. Let $\bar{Y}_n \in \mathbb{R}^k$. If $\bar{Y}_{n,i} \overset{AS}{\to} \mu_i$ as $n \to \infty$ for each $i = 1, \ldots, k$, then $\bar{Y}_n \overset{AS}{\to} \mu$.

An alternative way to express the same

Definition. *A sequence of random variables* $\left\{\bar{Y}_1, \bar{Y}_2, \ldots\right\}$ *is said to converge almost surely (or strongly) to a real number* μ *if*

$$P(\lim_{n\to\infty} \bar{Y}_n = \mu) = 1,$$

which is less succinctly written as $P(\omega | \lim_{n\to\infty} \bar{Y}_n(\omega) = \mu) = 1$.

From the above definitions, we begin to learn about the strength of almost sure convergence. Compare the first definition of almost sure convergence with the second definition of convergence in probability. The difference is that in almost sure convergence, the probability is taken with regards to the largest value of $|\bar{Y}_n - \mu|$. For weak convergence it is possible that there are some values of $n > N$ for which $|\bar{Y}_n - \mu| > \epsilon$. Suppose there are m such values. It must simply be the case that $\frac{m}{N} \to 0$ as $N \to \infty$. For strong convergence it must be the case that there are *no* values of $n > N$ for which $|\bar{Y}_n - \mu| > \epsilon$.

To learn more about almost sure convergence, we show that the two definitions are equivalent. We begin by fixing the value of ω to a value for which almost sure convergence holds. From definition two we have that $\lim_{n\to\infty} \bar{Y}_n(\omega) = \mu$. By the

definition of a deterministic limit given earlier, it must be the case that for any $\epsilon > 0$ there exists an integer N such that for all $n > N$ we have $|\bar{Y}_n(\omega) - \mu| < \epsilon$. Yet the statement that for all $n > N$ we have $|\bar{Y}_n(\omega) - \mu| < \epsilon$ implies that $\sup_{n>N} |\bar{Y}_n(\omega) - \mu| < \epsilon$, which is simply definition one. Under the first definition the $\sup_{n\geq N} |\bar{Y}_n(\omega) - \mu| < \epsilon$, so $\lim_{n\to\infty} \bar{Y}_n(\omega) = \mu$.

At first glance it may seem that almost sure convergence is the same as a deterministic limit. But there are two crucial differences. First, the value of N may depend on ω; that is, the value beyond which all elements of the sequence are in an ϵ-neighborhood of the true parameter varies across sequences. Second, there may exist some values of ω, that is, some sequences, for which the limit is not μ. It must simply be the case that the probability of obtaining such a sequence is zero.

Example. Let $\Omega = (\omega_1, \omega_2)$ where $P(\omega_1) = 0$ and $P(\omega_2) = 1$. (Note that ω_1 still occurs.) Let

$$\bar{Y}_n(\omega_1) = \begin{bmatrix} \frac{1}{n} \text{ if } n \text{ is even} \\ \frac{1}{2} \text{ if } n \text{ is odd} \end{bmatrix}, \; \bar{Y}_n(\omega_2) = 1 - \frac{1}{n}.$$

Then with probability 1 we obtain the sequence that converges to 1 and with probability 0 we obtain the sequence that does not converge. Hence $P(\lim_{n\to\infty} \bar{Y}_n = 1) = 1$, so $\bar{Y}_n \overset{AS}{\to} 1$ as $n \to \infty$. The example contains one of the differences mentioned in the previous paragraph. Because there is only one convergent sequence, the value of N for which all elements of the sequence beyond N are in an ϵ-neighborhood does not depend on ω. There is, however, a value of ω for which the limit of the sequence is not 1.

To exhibit both of the differences, we need to enrich our

Example. Let $\Omega = (\omega_1, \omega_2, \omega_3)$ where $P(\omega_1) = 0$ and $P(\omega_2) + P(\omega_3) = 1$. Let $\bar{Y}_n(\omega_1)$ and $\bar{Y}_n(\omega_2)$ be as in the previous example and let

$$\bar{Y}_n(\omega_3) = 1 - \frac{1}{n^{\frac{1}{2}}}.$$

Again with probability 1 we obtain the sequence that converges to 1 and $\bar{Y}_n \overset{AS}{\to} 1$ as $n \to \infty$. For this example, the appropriate value of N does depend on ω because there is more than one convergent sequence. If $\epsilon = .1$, then the value of N such that for all $n > N$ we have that $|\bar{Y}_n(\omega) - 1| < .1$ is: $N = 10$ if $\omega = \omega_2$ and $N = 100$ if $\omega = \omega_3$. It is precisely this uncertainty over the correct value of N that gives rise to the limit that appears in front of the probability statement in the first definition. We have $P(\sup_{n\geq N} |\bar{Y}_n(\omega) - 1| < .1)$ equals: 0 if $N < 10$, $P(\omega_2)$

if $10 \leq N < 100$, and 1 if $100 \leq N$. It is the uncertainty over the appropriate value of N (as well as the possibility of non-convergent sequences) that prevents us from specifying the sample size we need to be sure that we are within a certain distance of the parameter.

What is the relation between convergence in probability and convergence almost surely? As you may have guessed from the alternative names weak and strong convergence, strong convergence implies weak convergence but not vice versa

$$\bar{Y}_n \xrightarrow{AS} \mu \;\Rightarrow\; \bar{Y}_n \xrightarrow{P} \mu.$$

In words, if there exists an $N(\epsilon)$ such that for all $n > N(\epsilon)$ the elements of the sequence are contained within an ϵ-neighborhood of μ, then certainly the probability of obtaining an element in the sequence that is contained within an ϵ-neighborhood is one. The converse is clearly not true. Even though the probability of obtaining an element in the sequence that is contained with an ϵ-neighborhood is one, there can still be elements of the sequence that are not contained within the ϵ-neighborhood. If we excluded the elements that are not contained with the ϵ-neighborhood, then all remaining elements are contained in the ϵ-neighborhood, which is the intuition behind the result that if $\bar{Y}_n \xrightarrow{P} \mu$, then there exists a subsequence $\{\bar{Y}_{n_s}\}$ such that $\bar{Y}_{n_s} \xrightarrow{AS} \mu$.

Why discuss almost sure convergence when consistency requires only convergence in probability? At times, it is easier to prove almost sure convergence, which because convergence in probability is implied, ensures that our estimator is consistent.

Convergence in Mean Square (Quadratic Mean)

Both probabilistic convergence and almost sure convergence are based on probability calculations for the underlying sample space Ω. Convergence in quadratic mean is based on limit calculations for the moments of the random variable. In general, probabilistic and almost sure convergence are used but in some cases, limit calculations for moments are easier to construct.

Definition. A sequence of random variables $\left\{\bar{Y}_1, \bar{Y}_2, \ldots\right\}$ is said to converge in mean square to a real number μ if

$$\lim_{n\to\infty} E\left[\left(\bar{Y}_n - \mu\right)^2\right] = 0.$$

We express the convergence as $\bar{Y}_n \xrightarrow{m.s.} \mu$ as $n \to \infty$. The definition applies to vectors. Let $\bar{Y}_n \in \mathbb{R}^k$. If $\bar{Y}_{n,i} \xrightarrow{m.s.} \mu_i$ as $n \to \infty$ for each $i = 1, \ldots, k$, then $\bar{Y}_n \xrightarrow{m.s.} \mu$.

From the above definition, the intuition of convergence in quadratic mean becomes clear. Recall that $E\left[\left(\bar{Y}_n - \mu\right)^2\right]$ is the mean-square error of $\bar{Y}_n$. Convergence in quadratic mean requires that the mean-square error of $\bar{Y}_n$ equal zero in the limit. In other words, both the variance and the bias must be asymptotically zero. If $\bar{Y}_n$ converges in quadratic mean to μ, then $\bar{Y}_n$ is asymptotically unbiased.

Although there is no direct link between convergence in quadratic mean and convergence almost surely, there is a link between convergence in quadratic mean and probabilistic convergence. Convergence in quadratic mean is similar to convergence almost surely in that

$$\bar{Y}_n \overset{m.s.}{\rightarrow} \mu \;\Rightarrow\; \bar{Y}_n \overset{P}{\rightarrow} \mu,$$

and the converse does not hold. It is a simple matter to verify that convergence in quadratic mean implies probabilistic convergence. By Chebyshev's inequality

$$P\left(\left|\bar{Y}_n - \mu\right| < \varepsilon\right) \geq 1 - \frac{1}{\varepsilon^2} E\left[\left(\bar{Y}_n - \mu\right)^2\right].$$

From the definition of convergence in mean square; if $\bar{Y}_n \overset{m.s.}{\rightarrow} \mu$ then

$$\lim_{n\to\infty} \left\{1 - \frac{1}{\varepsilon^2} E\left[\left(\bar{Y}_n - \mu\right)^2\right]\right\} = 1.$$

If we take the limit, as $n \to \infty$, of both sides of the inequality

$$\lim_{n\to\infty} P\left(\left|\bar{Y}_n - \mu\right| < \varepsilon\right) = 1,$$

which is the definition of probabilistic convergence.

Convergence to a Random Variable

In each of the preceding definitions, the limit could be a random variable. For example, we say that a sequence of K-dimensional random variables $\left\{\bar{Y}_n\right\}$ converges weakly to a K-dimensional random variable Y and write

$$\bar{Y}_n \overset{P}{\rightarrow} Y \;\text{ if }\; \bar{Y}_n - Y \overset{P}{\rightarrow} 0.$$