Economics 241B
Review of Limit Theorems for Sequences of Random Variables

Convergence in Distribution

The previous definitions of convergence focus on the outcome sequences of a random variable. Convergence in distribution refers to the probability distribution of each element of the sequence directly.

Definition. *A sequence of random variables* $\{\bar{Y}_1, \bar{Y}_2, \ldots\}$ *is said to converge in distribution to a random variable* Y *if*

$$\lim_{n\to\infty} P\left(\bar{Y}_n < c\right) = P\left(Y < c\right)$$

at all c *such that* F_Y *is continuous.*

We express this as $\bar{Y}_n \xrightarrow{D} Y$. An equivalent definition is

Definition. *A sequence of random variables* $\{\bar{Y}_1, \bar{Y}_2, \ldots\}$ *is said to converge in distribution to* F_Y *if*

$$\lim_{n\to\infty} F_{\bar{Y}_n}(c) = F_Y(c)$$

at all c *such that* F_Y *is continuous.*

We express this as $\bar{Y}_n \xrightarrow{D} F_Y$. The distribution F_Y is the asymptotic (or limiting) distribution of $\bar{Y}_n$. Convergence in distribution is simply the pointwise convergence of $F_{\bar{Y}_n}$ to F_Y. (The requirement that F_Y be continuous at all c will hold for all applications in this course, with the exception of binary dependent variables.) In most cases F_Y is a Gaussian distribution, for which we write

$$\bar{Y}_n \xrightarrow{D} N\left(\mu, \sigma^2\right).$$

Convergence in distribution is closely related to convergence in probability, if the limit quantity is changed. In the discussion above, a sequence of random variables was shown to converge (in probability) to a constant. One can also establish that a sequence of random variables converges in probability to a random variable. Recall, for convergence (i.p.) to a constant, the probability that $\bar{Y}_n$ is in an ϵ neighborhood of μ must be high, $P(\left|\bar{Y}_n - \mu\right| < \epsilon) > 1 - \delta$. That is

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

For convergence to a random variable, Y, we need

$$P\left\{\omega : \left|\bar{Y}_n(\omega) - Y(\omega)\right| < \epsilon\right\} > 1 - \delta, \tag{1}$$

that is, for large n we want the probability that the histogram for $\bar{Y}_n$ is close to the histogram for Y. There is no natural measure of distance here, although we think of ϵ as defining the histogram bin width. If the two histograms are arbitrarily close as n increases, then we write

$$\bar{Y}_n \xrightarrow{P} Y.$$

Because the two histograms are arbitrarily close, it should not be surprising that

$$\bar{Y}_n \xrightarrow{P} Y \Rightarrow \bar{Y}_n \xrightarrow{D} F_Y \text{ (sometimes written } \bar{Y}_n \xrightarrow{D} Y\text{)}.$$

Why doesn't the reverse hold? Because of the basic workings of probability spaces, about which we rarely concern ourselves. By the fact that both $\bar{Y}_n$ and Y are indexed by the same ω in (1), both quantities are defined on the same probability space. No such requirement is made in establishing convergence of distribution, we simply discuss a sequence of distribution functions. Not only could $\bar{Y}_n$ and Y be defined on different probability spaces, for each n, $\bar{Y}_n$ could be defined on a different probability space. Yet the two definitions are almost iff. The Skorhorod construction is a proof that establishes that if $\bar{Y}_n \xrightarrow{D} Y$, then it is always possible to construct sequences $\bar{Y}_n'$ and Y' that are defined on the same probability space and are similar to $\bar{Y}_n$ and Y so that $\bar{Y}_n' \xrightarrow{P} Y'$.

The extension to a sequence of random vectors is immediate: $\bar{Y}_n \xrightarrow{D} Y$ if the joint c.d.f. F_n converges to the joint c.d.f. F of Y at every continuity point of F. Note, however, that unlike other concepts of convergence, for convergence in distribution, element by element convergence does not necessarily mean convergence for the vector sequence. That is "each element $\bar{Y}_n \xrightarrow{D}$corresponding element of Y" does not imply "$\bar{Y}_n \xrightarrow{D} Y$". A common way to connect scalar convergence and vector convergence in distribution is

Multivariate Convergence in Distribution Theorem: *Let $\left\{\bar{Y}_n\right\}$ be a sequence of K-dimensional random vectors. Then*

$$\bar{Y}_n \xrightarrow{D} Y \Leftrightarrow \lambda'\bar{Y}_n \xrightarrow{D} \lambda'Y$$

for any K-dimensional vector of real numbers λ.

Convergence in Distribution vs. Convergence in Moments

The moments of the limit distribution of $\bar{Y}_n$ are not necessarily equal to the limit of the moments of $\bar{Y}_n$. Thus $\bar{Y}_n \xrightarrow{D} Y$ does not imply that $\lim_{n\to\infty} E\left(\bar{Y}_n\right) = E\left(Y\right)$. However

Lemma (convergence in distribution and moments): *Let μ_{sn} be the s-th moment of $\bar{Y}_n$ and $\lim_{n\to\infty}\mu_{sn} = \mu_s$ where μ_s is finite (i.e. a real number). Then*

$$\bar{Y}_n \xrightarrow{D} Y \Rightarrow \mu_s \text{ is the } s\text{-th moment of } Y.$$

Thus, for example, if the variance of a sequence of random variables (converging in distribution) converges to some *finite* number, then that number is the variance of the limiting distribution.

Relation among Modes of Convergence

As noted above, some modes of convergence are weaker than others.

Lemma (relation among the four modes of convergence):

(a) $\bar{Y}_n \xrightarrow{m.s.} \mu \Rightarrow \bar{Y}_n \xrightarrow{p} \mu$. *So* $\bar{Y}_n \xrightarrow{m.s.} Y \Rightarrow \bar{Y}_n \xrightarrow{p} Y$.

(b) $\bar{Y}_n \xrightarrow{a.s.} \mu \Rightarrow \bar{Y}_n \xrightarrow{p} \mu$. *So* $\bar{Y}_n \xrightarrow{a.s.} Y \Rightarrow \bar{Y}_n \xrightarrow{p} Y$.

(c) $\bar{Y}_n \xrightarrow{p} \mu \Leftrightarrow \bar{Y}_n \xrightarrow{d} \mu$.

Part *(c)* states that if the limiting random variable is a constant (a trivial random variable) then convergence in distribution is the same as convergence in probability.

Three Useful Results

The great use of asymptotic analysis is that we can easily define the convergence properties of ratios, while we cannot easily define the expectation of ratios (and so cannot easily do finite sample analysis). These results, which highlight such features, are essential for large-sample theory.

Lemma (continuous mapping theorem): *Suppose that $g(\cdot)$ is a* **continuous function that does not depend on** n. *Then:*

$$\begin{aligned} \bar{Y}_n \xrightarrow{P} \mu \quad &\Rightarrow \quad g\left(\bar{Y}_n\right) \xrightarrow{P} g(\mu) \\ \bar{Y}_n \xrightarrow{D} Y \quad &\Rightarrow \quad g\left(\bar{Y}_n\right) \xrightarrow{D} g(Y). \end{aligned}$$

The basic idea is that because $g(\cdot)$ is continuous, $g\left(\bar{Y}_n\right)$ will be close to $g(\mu)$ provided that $\bar{Y}_n$ is close to μ. Note $g\left(\bar{Y}_n\right) = n^{\frac{1}{2}}\bar{Y}_n$ depends on n and is not covered by this proposition. An immediate implication is that the usual arithmetic operations preserve convergence in probability. For example

$$
\begin{aligned}
\bar{Y}_n \xrightarrow{P} \mu,\ \bar{X}_n \xrightarrow{P} \alpha \quad &\Rightarrow \quad \bar{Y}_n + \bar{X}_n \xrightarrow{P} \mu + \alpha \\
\bar{Y}_n \xrightarrow{P} \mu,\ \bar{X}_n \xrightarrow{P} \alpha \quad &\Rightarrow \quad \bar{Y}_n \bar{X}_n \xrightarrow{P} \mu\alpha \\
\bar{Y}_n \xrightarrow{P} \mu,\ \bar{X}_n \xrightarrow{P} \alpha \quad &\Rightarrow \quad \bar{Y}_n / \bar{X}_n \xrightarrow{P} \mu/\alpha \text{ provided that } \alpha \neq 0 \\
Y_n \xrightarrow{P} \Gamma \quad &\Rightarrow \quad Y_n^{-1} \xrightarrow{P} \Gamma^{-1} \text{ provided that } \Gamma \text{ is invertible.}
\end{aligned}
$$

The next result combines convergence in probability and distribution, and is used repeatedly to derive the limit theory for our estimators.

Lemma (parts a and c are called Slutzky's Theorem):

(a) $\bar{Y}_n \xrightarrow{P} \mu$ *and* $X_n \xrightarrow{D} X \Rightarrow \bar{Y}_n + X_n \xrightarrow{D} X + \mu$

(b) $\bar{Y}_n \xrightarrow{P} 0$ *and* $X_n \xrightarrow{D} X \Rightarrow \bar{Y}_n' X_n \xrightarrow{P} 0$

(c) $A_n \xrightarrow{P} A$ *and* $X_n \xrightarrow{D} X \Rightarrow A_n X_n \xrightarrow{D} AX$

(d) $A_n \xrightarrow{P} A$ *and* $X_n \xrightarrow{D} X \Rightarrow X_n' A_n^{-1} X_n \xrightarrow{D} X' A^{-1} X$

Parts (c)-(d) require that X and A be conformable and (for part (d)) that A be invertible.

In particular, if $X \sim N(0, V)$, then $\bar{Y}_n + X_n \xrightarrow{D} N(\mu, V)$ and $A_n X_n \xrightarrow{D} N(0, AVA')$. Also, if we set $\mu = 0$, then part (a) implies

$$\bar{Y}_n \xrightarrow{P} 0 \text{ and } X_n \xrightarrow{D} X \Rightarrow \bar{Y}_n + X_n \xrightarrow{D} X.$$

Let $Z_n = \bar{Y}_n + X_n$, so $\bar{Y}_n \xrightarrow{P} 0$ (i.e. $Z_n - X_n \xrightarrow{p} 0$) implies that the asymptotic distribution of Z_n is the same as that of X_n. When $Z_n - X_n \xrightarrow{p} 0$ we sometimes say that the two sequences are asymptotically equivalent ($Z_n \underset{a}{\sim} X_n$) and write

$$Z_n = X_n + o_P$$

where o_P is some suitable random variable ($\bar{Y}_n$ here) that converges to zero in probability.

We employ the lemma in a standard trick to derive the limit distribution of a sequence of random variables. Suppose we wish to find the limit distribution of $\bar{Y}_n' X_n$, where we know $X_n \xrightarrow{D} X$ and $\bar{Y}_n \xrightarrow{P} \mu$. Then we replace $\bar{Y}_n$ with μ, and determine the limit distribution of $\mu' X_n$. Why does this work? Because $\bar{Y}_n' X_n \underset{a}{\sim} \mu' X_n$, which is more easily seen in the equivalent statement

$$\bar{Y}_n' X_n = \mu' X_n + o_P$$

where o_P here is $\left(\bar{Y}_n - \mu\right)' X_n$.

To test nonlinear hypotheses, given the asymptotic distribution of the estimator, we need

Lemma (delta method): *Let $\{B_n\}_{n\geq 1}$ be a sequence of K-dimensional random variables such that*

$$B_n \xrightarrow{P} \beta$$
$$n^{\frac{1}{2}} \left(B_n - \beta\right) \xrightarrow{D} N\left(0, V\right).$$

Let $g\left(\cdot\right) : R^K \rightarrow R^Q$ have continuous first derivatives, with $G\left(\beta\right)$ denoting the $Q \times K$ matrix of first derivatives evaluated at β:

$$G\left(\beta\right) = \frac{\partial g\left(\beta\right)}{\partial \beta'}.$$

Then

$$n^{\frac{1}{2}} \left(g\left(B_n\right) - g\left(\beta\right)\right) \xrightarrow{D} N\left(0, G\left(\beta\right) V G\left(\beta\right)'\right).$$

Proof: A mean-value expansion yields $n^{\frac{1}{2}} g\left(B_n\right) = n^{\frac{1}{2}} g\left(\beta\right) + n^{\frac{1}{2}} G\left(B_n^*\right)\left(B_n - \beta\right)$, where B_n^* is the mean value between B_n and β. Because $B_n \xrightarrow{P} \beta$, $B_n^* \xrightarrow{P} \beta$ and, by continuity, $G\left(B_n^*\right) \xrightarrow{P} G\left(\beta\right)$. Thus

$$n^{\frac{1}{2}} \left[g\left(B_n\right) - g\left(\beta\right)\right] = n^{\frac{1}{2}} G\left(B_n^*\right)\left(B_n - \beta\right) \xrightarrow{D} G\left(\beta\right) N\left(0, V\right).$$

Viewing Estimators as Sequences of Random Variables

Let B_n be an estimator of β. If $B_n \xrightarrow{P} \beta$, then B_n is consistent for β.

The asymptotic bias of an estimator is not an agreed upon concept. The most common definition is

$$\lim_{n\rightarrow\infty} E\left(B_n\right) - \beta,$$

so consistency and asymptotic unbiasedness are distinct concepts. Hayashi uses $p\lim_{n\rightarrow\infty} B_n - \beta$, so consistency and asymptotic unbiasedness are identical for him.

If , in addition, $n^{\frac{1}{2}} \left(B_n - \beta\right) \xrightarrow{D} N\left(0, V\right)$, then B_n is asymptotically Gaussian (and $n^{\frac{1}{2}}$ consistent, hence CAN). The asymptotic variance is V.

Laws of Large Numbers and Central Limit Theorems

To verify convergence in probability from basic principles, we would need to know the joint distribution of the first n elements in the sequence. To verify convergence almost surely, we would need to know the joint distribution of the entire (infinite dimensional) sequence. To verify convergence in quadratic mean, we would need to know the first two moments of the finite sample distribution of the estimator. In general, we do not know any of these distributions. Instead, we turn to laws of large numbers. Laws of large numbers are one method of establishing convergence in probability (or almost surely) to a constant. They are applied to sequences of random variables, for which each member of the sequence is an (increasing) sum of random variables, such as $\left\{\bar{Y}_n\right\}$. An LLN is strong if the convergence is almost sure and weak if the convergence is in probability.

Chebychev's Weak LLN:

$$\text{"}\lim_{n\to\infty} E\left(\bar{Y}_n\right) = \mu, \ \lim_{n\to\infty} Var\left(\bar{Y}_n\right) = 0\text{"} \Rightarrow \bar{Y}_n \xrightarrow{p} \mu.$$

The following strong LLN assumes that $\{Y_t\}$ is i.i.d. but the variance does not need to be finite

Kolmogorov's Second Strong LLN: *Let $\{Y_t\}$ be i.i.d. with $EY_t = \mu$. Then $\bar{Y}_n \xrightarrow{a.s.} \mu$.*

Central limit theorems are a principal method of establishing convergence in distribution, as they govern the limit behavior of the difference between $\bar{Y}_n$ and $E\bar{Y}_n$ (which equals EY_t if $\{Y_t\}$ is i.i.d.) blown up by $\sqrt{n}$. For i.i.d. sequences, the only CLT we need is

Lindberg-Levy CLT: Let $\{Y_t\}$ be i.i.d. with $EY_t = \mu$ and $Var\left(Y_t\right) = \Omega$. Then

$$\sqrt{n}\left(\bar{Y}_n - \mu\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(Y_t - \mu\right) \xrightarrow{d} N\left(0, \Omega\right).$$

We read the above CLT as: a sequence of random vectors $\sqrt{n}\left(\bar{Y}_n - \mu\right)$ converges in distribution to a random vector whose distribution is $N\left(0, \Omega\right)$. To understand how to relate the above CLT for random vectors, with the CLT for random scalars, let λ be any vector of real numbers with the same dimension as Y_t. Now $\{\lambda' Y_t\}$ is a sequence of random variables with $E\left(\lambda' Y_t\right) = \lambda'\mu$ and $Var\left(\lambda' Y_t\right) = \lambda'\Omega\lambda$. The scalar version of Lindberg-Levy implies

$$\sqrt{n}\left(\lambda'\bar{Y}_n - \lambda'\mu\right) = \lambda'\sqrt{n}\left(\bar{Y}_n - \mu\right) \xrightarrow{d} N\left(0, \lambda'\Omega\lambda\right).$$

Yet this limit distribution is the distribution of $\lambda' Z$, where $Z \sim N(0, \Omega)$, hence the multivariate convergence in distribution theorem establishes the multivariate Lindberg-Levy CLT.