Laws of Large Numbers and Central Limit Theorems

Robert A. Miller

Structural Econometrics

November 2021

- One theme of the preceding lectures it that:
  1. unless there are traces of the past in the future, drawing upon data does not enhance prediction.
  2. with only a finite amount of data the predictions of our models are contaminated by sampling error.
  3. our models are often too complex to evaluate the probability distributions of most interest to us.
- *Laws of Large Numbers* (LLN) and *Central Limit Theorems* (CLT) partially compensate for our inability to compute the finite or exact distributional properties of our estimates and test statistics.
- These theorems are vital in:
  1. providing a measure of the central tendency for our predictions.
  2. indicating the precision of the predictions.

Cumulative Distribution Functions

Definition

- To explain LLN and CLT we first:
  - define cumulative distribution functions.
  - introduce a fourth mode of convergence.
- Suppose $X(\omega) : \Omega \rightarrow \mathbb{R}^k$ is measurable with respect to $(\Omega, \mathcal{F}, P)$.
- Define $F(b) : \mathbb{R}^k \rightarrow [0, 1]$, the (cumulative) probability distribution function, for $X(\omega)$ as:

$$F(b) \equiv \Pr\{\omega : X(\omega) \leq b\} = \Pr\{X \leq b\}$$

- For example if $\mathcal{F} = \{A_i\}_{i=1}^{I}$ with $I < \infty$, and $p_i \equiv \Pr\{\omega : \omega \in A_i\}$, we can express $F(b)$ as the expected value of $1\{\omega : X(\omega) \leq b\}$:

$$F(b) \equiv \Pr\{\omega : X(\omega) \leq b\} = \sum_{i=1}^{I} p_i 1\{\omega : X(\omega) \leq b\}$$

- Notice that $F(b) : \mathbb{R} \longrightarrow [0,1]$, is not directly concerned with the domain of $X(\omega)$, that is the event space and its σ-algebra $(\Omega, \mathcal{F})$, but only with its range (outcomes).
- Directly from its definition $F(b)$ is:
   1. increasing.
   2. $\lim_{b \to -\infty} F(b) = 0$ and $\lim_{b \to \infty} F(b) = 1$.
   3. right continuous.
   4. continuous at b iff $F(b - \varepsilon) \to F(b)$ for $\varepsilon > 0$ as $\varepsilon \to 0$.

Lemma

F is continuous at b if and only if $\Pr\{\omega : X(\omega) = b\} = 0$, which is to say there is no mass point at b.

Proof.

$$\begin{aligned} \{\omega : X(\omega) \leq b\} &= \{\omega : X(\omega) < b\} \cup \{\omega : X(\omega) = b\} \\ \Rightarrow \quad \Pr\{\omega : X(\omega) \leq b\} &= \Pr\{\omega : X(\omega) < b\} + \Pr\{\omega : X(\omega) = b\} \\ \Rightarrow \quad \Pr\{\omega : X(\omega) = b\} &= \Pr\{\omega : X(\omega) \leq b\} - \Pr\{\omega : X(\omega) < b\} \\ &= F(b) - F\left(b^{-}\right) \end{aligned}$$

The RHS $= 0$ if and only if F is continuous. Thus $F_x(\cdot)$ is continuous at b if and only if b has measure zero. □

Cumulative Distribution Functions

Expectation and Lebesgue integration defined

- For $(\Omega, \mathcal{F}, P)$ and $\mathcal{F}$-measurable $X(\omega) : \Omega \to [0, \infty]$ define:

$$\int X(\omega)\, dP(\omega) = \sup \left\{ \sum_{i=1}^{I} \left[\inf_{\omega \in A_i} X(\omega) \right] P(A_i) \right\}$$

  where $I < \infty$ and:
  - the supremum is taken over all finite partitions of Ω into sets $A_i \in \mathcal{F}$.
  - $\int X(\omega)\, dP(\omega) = \infty$ if the supremum does not exist.
- Thus simple functions $X(\omega)$, defined on a finite partition $\{A_i\}_{i=1}^{I}$ of Ω with $X(\omega) = b_i \in [0, \infty]$ for $\omega \in A_i$ have expected value:

$$E[X(\omega)] = \sum_{i=1}^{I} P(A_i)\, b_i$$

- More generally for $X(\omega) : \Omega \to [-\infty, \infty]$, define (if possible):

$$\int X(\omega)\, dP(\omega) = \int X^{+}(\omega)\, dP(\omega) - \int X^{-}(\omega)\, dP(\omega)$$

  where $X^{+}(\omega) = \max\{X(\omega), 0\}$ and $X^{-}(\omega) = -\min\{X(\omega), 0\}$.

Cumulative Distribution Functions

An illustration of Lebesgue integration and measurability

Weak Convergence

Convergence in distribution defined

- Let $F_1, F_2, \ldots$ be distribution functions for $X_1, X_2, \ldots$
- Then X_N converges weakly to X, or X_N converges in distribution/law, notated as:

$$X_N \xrightarrow{d} X$$

if and only if:

$$F_N(b) \to F(b) \text{ for all } b \text{ where } F \text{ is continuous.}$$

- For example in an assignment you are asked to show:

$$F_N(b) = \begin{cases} 0 & b < \frac{1}{N} \\ 1 & b \geq \frac{1}{N} \end{cases}$$

$$F(b) = \begin{cases} 0 & b < 0 \\ 1 & b \geq 0 \end{cases}$$

Weak Convergence

Comparing convergence in distribution with convergence in probability

- Convergence in probability implies convergence in distribution:

Lemma

See Dhyrmes (1989 pages 161-162): "Topics in Advance Econometrics: Probability Foundations."

$$\text{If } X_N \xrightarrow{p} X, \text{ then } X_N \xrightarrow{d} X.$$

- However the reverse is not true:

$$X_N \xrightarrow{d} X \text{ does not imply } X_N \xrightarrow{p} X.$$

- For example let $X \sim \mathcal{N}(0,1)$ and $X_1 = X$, $X_2 = -X$, $X_3 = X$, ... :
  - Then (trivially) $X_N \xrightarrow{d} X$ but $X_N \overset{p}{\nrightarrow} X$.
- In contrast to the other three notions of convergence we discussed, weak convergence is not directly concerned with the mapping $X(\omega) : \Omega \to \mathbb{R}$, but only the measure P assigned to $\mathcal{F}$.

Lemma

Let $X_1, X_2, \ldots$ *and* $Y_1, Y_2, \ldots$ *be two sequences of random variables with respect to* $(\Omega, \mathcal{F}, P)$*:*

$$\text{If } X_N - Y_N = o_p(1) \text{ and } X_N \xrightarrow{d} X \text{ then } Y_N \xrightarrow{d} X$$

Proof.

See Fuller (1976, Theorem 5.2.1. pages 193 -194): "Introduction to Statistical Time Series". □

- Consider a sequence of random variables $X_1(\omega), X_2(\omega), \ldots$ and denote the mean of the sample by:

$$\overline{X}_N(\omega) \equiv N^{-1} \sum_{n=1}^{N} X_n(\omega)$$

- *Laws of Large Numbers* (LLN) relate to the convergence of $\overline{X}_N(\omega)$:
  - Weak LLN refer to convergence in probability.
  - Strong LLN refer to almost sure convergence.
- *Central Limit Theorems* (CLT) give conditions for the convergence in distribution of $N^{1/2}\overline{X}_N(\omega)$ when $\overline{X}_N(\omega)$ is *centered*, that is when:
  - $E\left[\overline{X}_N(\omega)\right]$ exists and $E\left[\overline{X}_N(\omega)\right] = 0$.

Asymptotic Measures of Central Tendency

Twin premise

- Both LLN and CLT are based on assumptions that formalize the following intuition:
  1. Outliers play a limited role.
  2. The dependence of current realizations on past realizations is limited too.
- Two sufficient conditions for the theorems and results we review below are that:
  1. $|X_n(\omega)| \leq M$ for some $M < \infty$ (limited variation).
  2. $\Pr[X_n(\omega), X_{n+k}(\omega)] = \Pr[X_n(\omega)] \Pr[X_{n+k}(\omega)]$ for all $n \in \{1, 2, \ldots\}$ and $k \in \{K, K+1, K+2, \ldots\}$ with $K < \infty$ (asymptotic independence).
- More general LLN and CLT amount to further relaxations of these two conditions.

Asymptotic Measures of Central Tendency

What can happen when variation is unbounded?

- To see why some form of bounded variation is necessary consider the following example. For all $n \in \{1, 2, \ldots\}$ let:
  - $X_n(\omega) : \Omega \to \{-n, n\}$.
  - $\Pr\{X_n(\omega) = n\} = \theta$.
  - For all $k \in \{1, 2, \ldots\}$:

$$\Pr\{X_n(\omega), X_{n+k}(\omega)\} = \Pr\{X_n(\omega)\} \Pr\{X_{n+k}(\omega)\}$$

- Hence $E[X_n] = n(2\theta - 1)$ and $E\left[(X_n - E[X_n])^2\right] = 4n^2\theta(1-\theta)$.
- For a sample of length N:

$$\begin{aligned} N^{-1}\sum_{n=1}^{N} E[X_n(\omega)] &= N^{-1}\sum_{n=1}^{N} E[X_n^+(\omega)] - N^{-1}\sum_{n=1}^{N} E[X_n^-(\omega)] \\ &= \theta(N+1)/2 - (1-\theta)(N+1)/2 \end{aligned}$$

- Thus the limit as $N \to \infty$ of $N^{-1}\sum_{n=1}^{N} E[X_n(\omega)]$ is indeterminate.

Asymptotic Measures of Central Tendency

What can happen when there is asymptotic dependence?

- An example violating the second condition is:
  - $X_n(\omega) : \Omega \to \{0, 1\}$ and $\Pr\{X_n(\omega) = 1\} = \theta$.
  - $\Pr\{X_{n+1}(\omega) = x \,|X_n(\omega) = x\} = 1$.
- For future reference we note that $X_n(\omega)$ is stationary because for all $k \in \{1, 2, \ldots\}$:

$$\Pr(x_1, x_2, \ldots) = \Pr(x_k, x_{k+1}, \ldots)$$

- But $\overline{X}_N(\omega) : \Omega \to \{0, 1\}$ and $\Pr\{X_n(\omega) = 1\} = \theta$.
- Therefore $\overline{X}_N(\omega) \overset{p}{\nrightarrow} E[X_n(\omega)] = \theta$.

Theorem

Define:

$$S_N = \sum_{n=1}^{N} X_n$$

for independently distributed random variables $X_1, X_2, \ldots$. Suppose:

$$\mathrm{E}(X_n) = \mu_n \quad \textit{and} \quad \mathrm{var}(X_n) = \sigma_n^2$$

and

$$\mathrm{var}(S_N) \equiv V_N = \sum_{n=1}^{N} \sigma_n^2 = o\left(N^2\right).$$

Then

$$\frac{1}{N}\left(S_N - \sum_{n=1}^{N} \mu_n\right) \xrightarrow{p} 0$$

- For example if $\mathrm{var}(X_n) = \sigma^2$ then $V_N = N\sigma^2$ implying $V_N / N^2 \longrightarrow 0$.

Proof.

First note:

$$E\left[N^{-1}\left(S_N - \sum_{n=1}^{N}\mu_n\right)\right] = 0$$

$$V\left[N^{-1}\left(S_N - \sum_{n=1}^{N}\mu_n\right)\right] = N^{-2}V_N = N^{-2}o\left(N^2\right) = o\left(1\right)$$

Hence for all $\epsilon > 0$ and $\gamma > 0$, there exists $N_0 < \infty$ such that:

$$N^{-2}V_N < \epsilon^2\gamma.$$

for all $N \geq N_0$. Chebychev's inequality implies that for all $\epsilon > 0$:

$$\Pr\left\{|Y| \geq \epsilon\right\} \leq \epsilon^{-2}\mathrm{E}\left[Y^2\right]$$

when the first two moments of the random variable Y exist. Combining both inequalities:

$$\Pr\left\{\left|N^{-1}\left(S_N - \sum_{n=1}^{N}\mu_n\right)\right| \geq \epsilon\right\} \leq \epsilon^{-2}N^{-2}V_N < \gamma$$

$$\implies N^{-1}\left(S_N - \sum_{n=1}^{N}\mu_n\right) \xrightarrow{p} 0$$

Theorem

Suppose $X_n(\omega)$ *is independent with finite variance* σ_n^2. *If* $\sum_{n=1}^{\infty} \sigma_n^2 / n^2 < \infty$ *then* $\overline{X}_N(\omega) - E\left[\overline{X}_N(\omega)\right] \xrightarrow{a.s.} 0$.

Theorem

Suppose $X_n(\omega)$ *is independent and identically distributed (iid). Then a necessary and sufficient condition for* $\overline{X}_N(\omega) \xrightarrow{a.s.} \mu$ *is that* $E[X_n(\omega)]$ *exists and that* $\mu = E[X_n(\omega)]$.

- For proofs see:
  - Rao (1973, pages 114 -116): "Linear Statistical Inference and its Application".
  - Billingsley (1979, pages 250 -251): "Probability and Measure."

- Further extending these results, define the indicator random variable:

$$d_n(\omega) = \begin{cases} 1 \text{ if } X_n(\omega) \leq x \\ 0 \text{ if } X_n(\omega) > x \end{cases}$$

for any $x \in \mathbb{R}$. Then:

$$F(x) \equiv \Pr\{X_n(\omega) \leq x\} = E[d_n(\omega)]$$

- Thus if $\overline{X}_N(\omega) - E[\overline{X}_N(\omega)] \overset{a.s.}{\to} 0$ so does $F_N(x) - F(x)$.
- The Glivenko Cantelli theorem strengthens this pointwise convergence to uniform convergence.
- If the conditions for one of Kolmogorov's LLN are satisfied:

$$\|F_N - F\|_\infty \equiv \sup_{x \in \mathbb{R}} |F_N(x) - F(x)| \overset{a.s.}{\to} 0$$

- The discrete case follows trivially from the strong LLN using the notation above.

- Consider a continuous random variable X and fix:

$$-\infty \equiv x_0 < x_1 < \ldots < x_{r-1-} < x_r = \infty$$

spaced so that $F(x_i) - F(x_{i-1}) = r^{-1}$.

- Then there exists $i \in \{1, \ldots, r\}$ for all $x \in \mathbb{R}$ such that:

$$\begin{aligned} F_N(x) - F(x) &\leq F_N(x_i) - F(x_{i-1}) = F_N(x_i) - F(x_i) + r^{-1} \\ F_N(x) - F(x) &\geq F_N(x_{i-1}) - F(x_i) = F_N(x_{i-1}) - F(x_{i-1}) + r^{-1} \end{aligned}$$

- Therefore:

$$\sup_{x \in \mathbb{R}} |F_N(x) - F(x)| \leq \max_{i \in \{1,\ldots,r\}} |F_N(x_i) - F(x_i)| + r^{-1}$$

- Completing the proof exploits the facts that:
  - we can make r^{-1} infinitesimal.
  - $\max_{i \in \{1,\ldots,r\}} |F_N(x_i) - F(x_i)| \overset{a.s.}{\to} 0$.

- The assumption of independence can be relaxed in several ways.
- Given a probability space $(\Omega, \mathcal{F}, P)$ equipped with an increasing sequence of σ-algebras $\ldots \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$ and suppose $X_n(\omega)$ is measurable with respect to $\mathcal{F}_n$.
- In this case we say that $\{\mathcal{F}_n\}_{n=-\infty}^{\infty}$ is *adapted* to $\{X_n\}_{n=-\infty}^{\infty}$.
- A *martingale difference sequence* (MDS) is an adapted sequence on $\{X_n, \mathcal{F}_n\}_{t=-\infty}^{\infty}$ with the property that $E\left[X_n \,|\mathcal{F}_{n-1}\right] = 0$.

Theorem

Suppose $\{X_n, \mathcal{F}_n\}_{n=0}^{\infty}$ is an MDS with variances $\{\sigma_n^2\}_{n=0}^{\infty}$ and $\{a_n,\}_{n=0}^{\infty}$ is a sequence of constants with $\lim_{n\to\infty} a_n = \infty$. Then $S_N / a_N \xrightarrow{a.s.} 0$ if $\sum_{n=1}^{\infty} \sigma_n^2 / a_n^2 < \infty$.

- $X_1, X_2, \ldots$ is *stationary* if and only if for all $B \in \mathcal{B}_\infty$ and $k = 1, 2, \ldots$:

$$\begin{aligned} & \Pr\left(\{\omega : X_1(\omega), X_2(\omega), \ldots\} \in B\right) \\ = \; & \Pr\left(\{\omega : X_{k+1}(\omega), X_{k+2}(\omega), \ldots\} \in B\right) \end{aligned}$$

- A stationary sequence $X_1, X_2, \ldots$ is *ergodic* if for any two bounded mappings $f : \mathbb{R}^k \rightarrow \mathbb{R}$ and $g : \mathbb{R}^k \rightarrow \mathbb{R}$:

$$\begin{aligned} & \lim_{N\rightarrow\infty} \mathrm{E}\left[f\left(X_1, \ldots, X_k\right) g\left(X_{1+N}, \ldots, X_{l+N}\right)\right] \\ = \; & \mathrm{E}\left[f\left(X_1, \ldots, X_k\right)\right] \mathrm{E}\left[g\left(X_{1+N}, \ldots, X_{l+N}\right)\right] \end{aligned}$$

Theorem

If $\{X_n\}$ is stationary and ergodic, and the expected value of X_n exists with $\mathrm{E}(X_n) < \infty$, then:

$$S_N / N \xrightarrow{a.s.} \mathrm{E}(X_n)$$

Theorem

Let X_n be a sequence of independent and identically distributed random variables with zero men variance σ^2. Then:

$$N^{-1/2}S_N = N^{1/2}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

- Note when X_n is normally distributed $N^{-1/2}S_N \sim \mathcal{N}\left(0, \sigma^2\right)$ for all N not just in the limit as $N \to \infty$.
- Again, there are several generalizations to this result:
  - It can be extended to vectors.
  - The assumption that X_n is identically distributed can be relaxed.
  - We can also relax the independence assumption.

Central Limit Theorems

Intuition for CLT when all the moments exist (and fully characterize the distribution)

- Noting $N^{-1/2}S_N$ has mean 0 and variance σ^2, consider the fourth moment:

$$\mathrm{E}\left(N^{-2}S_N^4\right) = N^{-2}\sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{t=1}^{N}\sum_{u=1}^{N}\mathrm{E}\left(X_r X_s X_t X_u\right)$$

- Since X_n are independently distributed, $\mathrm{E}\left(X_r X_s X_t X_u\right) \neq 0$ only when:

$$r = s = t = u, \quad r = s \neq t = u, \quad r = t \neq s = u, \quad r = u \neq t = s$$

$$\begin{aligned} \Longrightarrow \mathrm{E}\left[N^{-2}S_N^4\right] &= N^{-2}\left\{N\mathrm{E}\left[X_n^4\right] + 3N\left(N-1\right)\sigma^4\right\} \\ &= o\left(N\right) + 3\left(N-1\right)N^{-1}\sigma^4 \end{aligned}$$

- Thus the value of the fourth moment only depends on σ^2.

- A similar argument applies to all the even moments.
- All the odd moments are asymptotically negligible.
- Consequently, given a value for σ^2, the asymptotic distribution of $N^{-1/2}S_N$ does not depend on the distribution of X_n.
- In particular, when X_n is standard normal $\mathcal{N}\left(0, \sigma^2\right)$, so is $N^{-1/2}S_N \sim \mathcal{N}\left(0, \sigma^2\right)$ for all N.

Central Limit Theorems

A general CLT for independent random variables (the Lindeberg-Feller theorem)

- Given independent random variables $\{X_n\}_{n=1}^{\infty}$ with means $\{\mu_n\}_{n=1}^{\infty}$ define the triangular array:

$$X_{nN} \equiv (X_n - \mu_n) / s_N \text{ where } s_N^2 = \sum_{n=1}^{N} E\left[(X_n - \mu_n)^2\right]$$

- Let $S_N = \sum_{n=1}^{N} X_{nN}$. By independence (and construction):

$$E[S_N] = 0 \text{ and } E\left[S_N^2\right] = \sum_{n=1}^{N} \int X_{nN}^2 \mathrm{d}P = 1$$

- Lindeberg proved $S_N \xrightarrow{d} \mathcal{N}(0,1)$ if for all $\varepsilon > 0$:

$$\lim_{N\to\infty} \left\{ \sum_{n=1}^{N} \int_{|X_{nN}|\geq\varepsilon} X_{nN}^2 \mathrm{d}P \right\} = 0 \qquad (1)$$

- Conversely Feller proved (1) holds if $S_N \xrightarrow{d} \mathcal{N}(0,1)$ and for all $\epsilon > 0$:

$$\lim_{N\to\infty} \left[\max_{1\leq n\leq N} \Pr\{|X_{nN}| > \epsilon\} \right] = 0$$

Central Limit Theorems

Liaponov's sufficient condition

- A stronger condition than (1) is that for some $\delta > 0$:

$$\lim_{N\to\infty} \sum_{n=1}^{N} E\left[\left|X_{nN}\right|^{2+\delta}\right] = 0 \tag{2}$$

- To show (2) is stronger than (1), note that for any $\epsilon > 0$:

$$\begin{aligned} E\left[\left|X_{nN}\right|^{2+\delta}\right] &\geq \int_{|X_{nN}|\geq\varepsilon} \left|X_{nN}\right|^{2+\delta} \mathrm{d}P \\ &\geq \int_{|X_{nN}|\geq\varepsilon} X_{nN}^2 \epsilon^{\delta} \mathrm{d}P \\ &= \epsilon^{\delta} \int_{|X_{nN}|\geq\varepsilon} X_{nN}^2 \mathrm{d}P \\ \Longrightarrow \lim_{N\to\infty} \sum_{n=1}^{N} E\left[\left|X_{nN}\right|^{2+\delta}\right] &\geq \epsilon^{\delta} \lim_{N\to\infty} \sum_{n=1}^{N} \int_{|X_{nN}|\geq\varepsilon} X_{nN}^2 \mathrm{d}P \end{aligned}$$

Lindeberg's condition, (1), now follows from (2) because $\epsilon^{\delta} > 0$ and does not depend on N.

Theorem

Let $\{X_{nN}, \mathcal{F}_{nN}\}$ *be a martingale difference array with finite unconditional variances* $\{\sigma^2_{nN}\}$, *and* $\sum_{n=1}^{N} \sigma^2_{nN} = 1$. *Then:*

$$S_N \equiv \sum_{n=1}^{N} X_{nN} \xrightarrow{d} \mathcal{N}(0,1)$$

if:

$$\sum_{n=1}^{N} X^2_{nN} \xrightarrow{p} 1 \text{ and } \max_{1 \leq n \leq N} |X_{nN}| \xrightarrow{p} 0.$$

Proof.

See Davidson (1994, Theorem 24.3, pages 383 - 384): "Stochastic Limit Theory". □