

Large Sample Theory: Lecture XVI

Charles B. Moss

October 1, 2026

1 Basic Sample Theory

2 Modes of Convergence

Basic Sample Theory

- The problems set up is that we want to discuss sample theory.
 - First assume that we want to make an inference, either estimation or some test, based on a sample.
 - We are interested in how well parameters or statistics based on that sample represent the parameters or statistics of the whole population.
- The complete statistical term is known as convergence.
 - Specifically, we are interested in whether or not the statistics calculated on the sample converge toward the population estimates.
 - Let $\{X_n\}$ be a sequence of samples. We want to demonstrate that statistics based on $\{X_n\}$ converge toward the population statistics for X .

- Taking a slightly different tack: The classical assumptions for ordinary least squares (OLS) as presented in White, Halbert Asymptotic Theory for Econometricians.
 - **Theorem 1.1:** The following are the assumptions of the classical linear model
 - The model is known to be $y = X\beta + \epsilon$, $\beta < \infty$.
 - X is a nonstochastic and finite $n \times k$ matrix.
 - $X'X$ is nonsingular for all $n \leq k$.
 - $E(\epsilon) = 0$.
 - $\epsilon \sim N(0, \sigma_0^2 I)$, $\sigma_0^2 < \infty$.

- Continued

- Given these assumptions, we can conclude that
 - Existence* given (i) - (iii) β_n exists for all $n \geq k$ and is unique.
 - Unbiasedness* given (i) - (v) $E[\beta_n] = \beta_0$.
 - Normality* given (i) - (v) $\beta_n \sim N\left(\beta_0, \sigma^2 (X'X)^{-1}\right)$.
 - Efficiency* given (i) - (v) β_n is the maximum likelihood estimator and the best unbiased estimator in the sense that any other unbiased estimator exceeds that of β_n by a positive semi-definite matrix regardless of the value of β_0 .

- Continued

- Existence, unbiasedness normality and efficiency are small sample analogs of asymptotic theory.
 - Unbiased implies that the distribution of β_n is centered around β_0 .
 - Normality allows us to construct t -distribution or F -distribution tests for restrictions.
 - Efficiency guarantees that the ordinary least squares estimates have the greatest possible precision.
- Asymptotic theory involves the behavior of the estimator under the failure of certain assumptions. Specifically, assumptions (ii) or (v).
- Finally, within the classical linear model the normality of the error term is required to strictly apply t -distributions or F -distributions. However, the central limit theorem can be used if n is large to guarantee that β_n is approximately normal.

Modes of Convergence

- **Definition 6.1.1** A sequence of real numbers $\{ \alpha_n \}$, $n = 1, 2, \dots$ is said to converge to a real number α if for any $\epsilon > 0$ there exists an integer N such that for all $n > N$ we have

$$|\alpha_n - \alpha| < \epsilon \quad (1)$$

- This convergence is expressed as $\alpha_n \rightarrow \alpha$ as $n \rightarrow \infty$ or $\lim_{n \rightarrow \infty} \alpha_n = \alpha$.
- This definition must be changed for random variables because we cannot require a random variable to approach a specific value. Instead, we require the probability of the variable to approach a given value. Specifically, we want the probability of the event to equal 1 or zero as n goes to infinity.

- **Definition 6.1.2** (convergence in probability) A sequence of random variables $\{X_n\}$, $n = 1, 2, \dots$ is said to converge to a random variable X in probability if for any $\epsilon > 0$ and $\delta > 0$ there exists an integer N such that for all $n > N$ we have $P(|X_n - X| < \epsilon) > 1 - \delta$. We write

$$X_n \xrightarrow{P} X \quad (2)$$

$n \rightarrow \infty$ or $\text{plim}_{n \rightarrow \infty} X_n = X$. The last equality reads the probability limit of X_n is X . (Alternatively, the if clause may be paraphrased as follows: if $\lim P(|X_n - X| < \epsilon) = 1$ for any $\epsilon > 0$).

- Since we are dealing with convergence in the probability, we need to introduce two concepts: convergence in mean square and convergence in distribution.

- **Definition 6.1.3.** (*Convergence in Mean Square*) A sequence $\{X_n\}$ is said to *converge* to X in *mean square* if $\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0$. We write

$$X_n \xrightarrow{M} X \quad (3)$$

- **Definition 6.1.4.** (*Convergence in Distribution*) A sequence $\{X_n\}$ is said to *converge to X in distribution* if the distribution function F_n of X_n converges to the distribution function F of X at every continuity point of F . We write

$$X_n \xrightarrow{d} X \quad (4)$$

and call F the limit distribution of $\{X_n\}$. If $\{X_n\}$ and $\{Y_n\}$ have the same limit distribution, we write

$$X_n \stackrel{LD}{=} Y_n \quad (5)$$

Chebyshev

- **Theorem 6.1.1** (*Chebyshev*)

$$X_n \xrightarrow{M} X \Rightarrow X_n \xrightarrow{P} X \quad (6)$$

Amemiya p. 102.

- **Theorem 6.1.2**

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X \quad (7)$$

Chebyshev's Inequality

- Chebyshev's Inequality

$$P[g(X_n) \geq \epsilon^2] \leq \frac{E[g(X_n)]}{\epsilon^2} \quad (8)$$

- A simple proof from Hogg and Craig (p. 58). We start by reformulating Equation 6 slightly

$$P[u(x) \geq c] \leq \frac{E[u(x)]}{c} \quad (9)$$

next, we divide the range of x s into two components

$A = \{x; u(x) \geq c\}$ and A^* . Then

$$E[u(x)] = \int_{-\infty}^{\infty} u(x) f(x) dx \quad (10)$$

$$\int_A u(x) f(x) dx + \int_{A^*} u(x) f(x) dx$$

Chebyshev's Continued

- Given that the integral over entire range is greater than the integral over the range A , we have

$$E[u(x)] \geq \int_A u(x) f(x) dx. \quad (11)$$

Further since $u(x) \geq c$ in this range

$$E[u(x)] \geq \int_A u(x) f(x) dx \geq c \int_A f(x) dx. \quad (12)$$

Therefore

$$E[u(x)] \geq cP[u(x) \geq x] \Rightarrow P[u(x)] \leq \frac{E[u(x)]}{c} \quad (13)$$

Chebyshev's Inequality - Confidence Intervals

- Using a slight modification of the Chebychev formulation

$$P[|x - \mu| \geq k\sigma] \leq \frac{1}{k^2} \quad (14)$$

From our previous discussion take $u(x) = (x - \mu)^2$ and $c = k^2\sigma^2$ which yields

$$P[(x - \mu)^2 \geq k^2\sigma^2] \leq \frac{E[(x - \mu)^2]}{k^2\sigma^2} \quad (15)$$

Given that $E[(x - \mu)^2] = \sigma^2$ the expression becomes

$$P[|x - \mu| \geq k\sigma] \leq \frac{1}{k^2} \quad (16)$$

- **Theorem 6.1.3** Let X_n be a vector of random variables with a fixed finite number of elements. Let g be a function continuous at a constant vector point α . Then

$$X_n \xrightarrow{P} \alpha \Rightarrow g(X_n) \xrightarrow{P} g(\alpha) \quad (17)$$

- **Theorem 6.1.4 (Slutsky)** If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} \alpha$, then

$$X_n + Y_n \xrightarrow{d} X + \alpha$$

$$X_n Y_n \xrightarrow{d} \alpha X \tag{18}$$

$$\left(\frac{X_n}{Y_n} \right) \xrightarrow{d} \frac{X}{\alpha} \text{ if } \alpha \neq 0$$

Readings

- Amemiya, Takeshi (1994) *Introduction to Statistics and Econometrics* Cambridge, Massachusetts: Harvard University Press.
 - Chapter 6 - Large Sample Theory – pp. 100-111.
- Moss, Charles B. (2015) *Mathematical Statistics for Applied Econometrics* Boca Raton: CRC Press.
 - Chapter 6 – Large Sample Theory - pp. 133-151.
- Herman, Russell L. (2017) *An Introduction to Fourier Analysis* Boca Raton: Taylor & Francis.
 - Chapter 1 – Review of Sequences and Infinite Series
- Moss, Charles B. (2022) *Production Economics: An Empirical Approach* New Jersey: World Scientific.
 - 2.7.2.1 Fourier Expansions – pp. 135 - 144.
- Billingsley, Patrick (1995) *Probability and Measure* New York: John Wiley & Sons.
 - Section 26 - Characteristic Functions