

Limits and the Law of Large Numbers: Lecture XVII

Charles B. Moss

October 1, 2026

Complex Analysis

- Taylor, Michael. *Introduction to Complex Analysis*
<https://mtaylor.web.unc.edu/wp-content/uploads/sites/16915/2018/04/complex.pdf>
- Chapters 1 – 4.

- 1 Almost Sure Convergence
 - Convergence in Probability
- 2 Laws of Large Numbers
- 3 Asymptotic Normality

Linear Estimators

- Next week we start a detailed discussion of estimators, but I wanted to introduce the concept of a linear estimator.
- Let us start with the least squares estimator

$$\hat{\beta} = (x'x)^{-1} (x'y) \quad (1)$$

This estimator can be rewritten as a linear function of y

$$\hat{\beta} = (x'x)^{-1} \sum_{i=1}^N x_i y_i \quad (2)$$

This concept of a linear estimator is magnified if we consider the x s to be non-stochastic. $(x'x)^{-1} x_i$ is simply a weight on the observation y_i .

Special Case – the Mean

- If we let x simply be a vector of ones, the estimator of β becomes

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^N y_i \quad (3)$$

Hence, the mean is also defined as a linear estimator in this way.

Almost Sure Convergence

- Let ω represent the entire random sequence $\{Z_t\}$. As previously, our interest typically centers around the averages of this sequence

$$b_n(\omega) = \frac{1}{n} \sum_{t=1}^n Z_t \quad (4)$$

- Definition 2.9** Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables. We say that $b_n(\omega)$ converges *almost surely* to b , written

$$b_n(\omega) \xrightarrow{a.s.} b \quad (5)$$

if and only if there exists a real number b such that

$$P[\omega : b_n(\omega) \rightarrow b] = 1. \quad (6)$$

- The probability measure P describes the distribution of ω and determines the joint distribution function for the entire sequence $\{Z_t\}$.
- Other common terminology is that $b_n(\omega)$ converges to b with probability 1 (w.p.1) or that $b_n(\omega)$ is strongly consistent for b .
- **Example 2.10:** Let

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_t \quad (7)$$

where $\{Z_t\}$ is a sequence of independently and identically distributed (i.i.d.) random variables with $E[Z_t] = \mu < \infty$.
Then

$$\bar{Z}_n \xrightarrow{a.s.} \mu \quad (8)$$

by the Komolgorov strong law of large numbers (Theorem 3.1).

- **Proposition 2.11:** Given $g : R^k \rightarrow R^l$ ($k, l < \infty$) and any sequence $\{b_n\}$ such that

$$b_n \xrightarrow{a.s.} b \quad (9)$$

where b_n and b are $k \times 1$ vectors, if g is continuous at b , then

$$g(b_n) \xrightarrow{a.s.} g(b). \quad (10)$$

- **Theorem 2.12:** Suppose

- $y = X\beta_0 + \epsilon$;
- $\frac{X'\epsilon}{n} \xrightarrow{a.s.} 0$;
- $\frac{X'X}{n} \xrightarrow{a.s.} M$, is finite and positive definite.

Then β_n exists a.s. for all n sufficiently large, and $\beta_n \xrightarrow{a.s.} \beta_0$.

- Proof: Since $X'X/n \xrightarrow{a.s.} M$, it follows from Proposition 2.11 that $\det(X'X/n) \xrightarrow{a.s.} \det(M)$. Because M is positive definite by (c), $\det(M) > 0$. It follows that $\det(X'X/n) > 0$ a.s. for all n sufficiently large, so $(X'X)^{-1}$ exists for all n sufficiently large. Hence

$$\hat{\beta}_n \equiv \left(\frac{X'X}{n} \right)^{-1} \frac{X'y}{n} \quad (11)$$

exists for all n sufficiently large. In addition,

$$\hat{\beta}_n = \beta_0 + \left(\frac{X'X}{n} \right)^{-1} \frac{X'\epsilon}{n} \quad (12)$$

It follows from Proposition 2.11 that

$$\hat{\beta}_n \xrightarrow{a.s.} \beta_0 + M^{-1}0 = \beta_0 \quad (13)$$

Convergence in Probability

- A weaker stochastic convergence concept is that of convergence in probability.
- **Definition 2.23:** Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables. If there exists a real number b such that for every $\epsilon > 0$,

$$P[\omega : |b_n(\omega) - b| < \epsilon] \rightarrow 1 \quad (14)$$

as $n \rightarrow \infty$, then $b_n(\omega)$ converges in probability to b .

- The almost sure measure of probability takes into account the joint distribution of the entire sequence $\{Z_t\}$, but with convergence in probability, we only need to be concerned with the joint distribution of those elements that appear in $b_n(\omega)$.
- Convergence in probability is also referred to as *weak consistency*.

- **Theorem 2.24:** Let $\{b_n(\omega)\}$ be a sequence of random variables. If

$$b_n \xrightarrow{a.s.} b, \text{ then } b_n \xrightarrow{P} b. \quad (15)$$

If b_n converges in probability to b , then there exists a subsequence $\{b_{n_j}\}$ such that

$$b_{n_j} \xrightarrow{a.s.} b. \quad (16)$$

- Convergence in the r^{th} mean.
 - Definition 2.37:** Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables. If there exists a real number b such that

$$E[|b_n(\omega) - b|^r] \rightarrow 0 \quad (17)$$

As $n \rightarrow \infty$ for some $r > 0$, then $b_n(\omega)$ *converges in the r^{th} mean* to b , written as

$$b_n(\omega) \xrightarrow{r.m.} b. \quad (18)$$

- **Proposition 2.38:** (*Jensen's inequality*) Let $g : R^1 \rightarrow R^1$ be a convex function on the interval $B \subset R^1$ and Z be a random variable such that $P[Z \in B] = 1$. Then $g[E(Z)] \leq E[g(Z)]$. If g is concave on B , then $g[E(Z)] \geq E[g(Z)]$.
- **Proposition 2.41:** (*Generalized Chebyshev Inequality*) Let Z be a random variable such that $E|Z|^r < \infty$, $r > 0$. Then for every $\epsilon > 0$

$$P[|Z| \geq \epsilon] \leq \frac{E(|Z|^r)}{\epsilon^r} \quad (19)$$

- **Theorem 2.42:** If $b_n(\omega) \xrightarrow{r.m.} b$ for some $r > 0$, then $b_n(\omega) \xrightarrow{p} b$.

Law of Large Numbers

- **Proposition 3.0:** Given restrictions on the dependence, heterogeneity, and moments of a sequence of random variables $\{ Z_t \}$,

$$\bar{Z}_n - \bar{\mu}_n \xrightarrow{a.s.} 0 \quad (20)$$

where

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_t \text{ and } \bar{\mu}_n = E(\bar{Z}_n) \quad (21)$$

- Independent and identically distributed observations
 - Theorem 3.1:** (*Komolgorov*) Let $\{Z_t\}$ be a sequence of i.i.d. random variables. Then

$$\bar{Z}_n \xrightarrow{a.s.} \mu \quad (22)$$

if and only if $E|Z_t| < \infty$ and $E(Z_t) = \mu$.

- This result is consistent with Theorem 6.2.1 (*Khinchine*). Let $\{X_i\}$ be independent and identically distributed with $E[X_i] = \mu$. Then

$$\bar{X}_n \xrightarrow{P} \mu. \quad (23)$$

- Continued

- Proposition 3.4:** (*Holder's Inequality*) If $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, and if $E|Y|^p < \infty$ and $E|Z|^q < \infty$, then

$$E|YX| \leq [E|y|^p]^{1/p} [E|X|^q]^{1/q}. \quad (24)$$

- If $p = q = 2$, we have the Cauchy-Schwartz inequality

$$E[YZ] \leq E[Y^2]^{1/2} E[Z^2]^{1/2} \quad (25)$$

Asymptotic Normality

- Under the traditional assumptions of the linear model (fixed regressors and normally distributed error terms) β_n is distributed multivariate normal with

$$\begin{aligned} E \left[\hat{\beta}_n \right] &= \beta_0 \\ V \left[\hat{\beta}_n \right] &= \sigma_0^2 (X'X)^{-1} \end{aligned} \tag{26}$$

for any sample size n .

- However, when the sample size becomes large the distribution of β_n is approximately normal under some general conditions.

- Definition 4.1:** Let $\{b_n\}$ be a sequence of random finite-dimensional vectors with joint distribution functions $\{F_n\}$. If $F_n(z) \rightarrow F(z)$ as $n \rightarrow \infty$ for every continuity point z , where F is the distribution function of a random variable Z , then b_n *converges in distribution* to the random variable Z , denoted

$$b_n \xrightarrow{d} Z. \quad (27)$$

- Intuitively, the distribution of b_n becomes closer and closer to the distribution function of the random variable Z . Hence, the distribution F can be used as an approximation of the distribution function of b_n .

- Continued

- Other ways of stating this concept are that b_n *converges in law* to Z

$$b_n \xrightarrow{L} Z. \quad (28)$$

Or, b_n is *asymptotically distributed* as F

$$b_n \overset{A}{\sim} F \quad (29)$$

In this case, F is called the *limiting distribution* of b_n .

- **Example 4.3:** Let $\{Z_t\}$ be a i.i.d. sequence of random variables with mean μ and variance $\sigma^2 < \infty$. Define

$$b_n = \frac{\bar{Z}_n - E[\bar{Z}_n]}{(V[\bar{Z}_n])^{1/2}} = \left(\frac{1}{n}\right)^{1/2} \sum_{i=1}^n \frac{(Z_t - \mu)}{\sigma} \quad (30)$$

Then by the Lindeberg-Levy central limit theorem (Theorem 6.2.2)

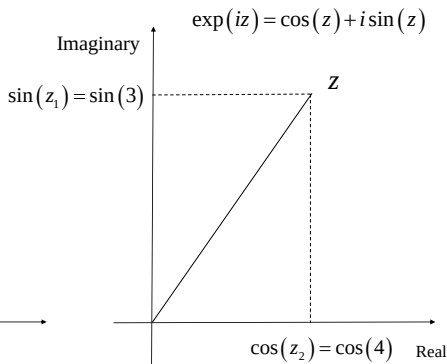
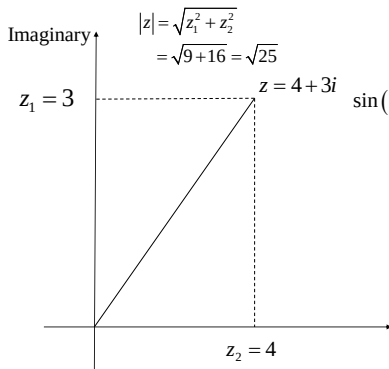
$$b_n \overset{A}{\approx} N(0, 1) \quad (31)$$

- **Theorem 6.2.2:** (Lindeberg-Levy) Let $\{X_i\}$ be i.i.d. with $E[X_i] = \mu$ and $V(X_i) = \sigma^2$. Then $Z_n \rightarrow N(0, 1)$.
- Characteristic Functions
 - **Definition 4.8:** Let Z be a $k \times 1$ random vector with distribution function F . The characteristic function of F is defined as

$$f(\lambda) = E[\exp(i\lambda'Z)] \quad (32)$$

where $i^2 = -1$ and λ is a $k \times 1$ real vector.

Graphic Depiction of Complex Number Space



Moment Generating Function vs Characteristic Function

- As we have defined the moment generating function

$$\mu_x(t) = \int_{-\infty}^{\infty} f(x) \exp(tx) dx \quad (33)$$

where the numbers x are defined in real number space.

- Now consider a slight modification of Euler's equation

$$\exp(i\lambda x) = \cos(\lambda x) + i \sin(\lambda x) \quad (34)$$

The definition of the characteristic function follows the conceptual formulation of the moment generating function

$$f(\lambda) = \int_{-\infty}^{\infty} f(x) \exp(i\lambda x) dx \quad (35)$$

This integral considers complex number space.

- **Example 4.10:** Let $Z \sim N(\mu, \sigma^2)$. Then

$$f(\lambda) = \exp\left(i\lambda\mu - \frac{\lambda^2\sigma^2}{2}\right) \quad (36)$$

- To demonstrate this result

$$\begin{aligned} f(x) \exp(i\lambda) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} + i\lambda x\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2 - 2x\mu + \mu^2 - 2\sigma^2 i\lambda x}{2\sigma^2}\right) \end{aligned} \quad (37)$$

- Focusing on the numerator of the exponent from Equation 37

$$x^2 - 2x\mu + \mu^2 - 2\sigma^2 i\lambda x = x^2 - 2x(\mu + \sigma^2 i\lambda) + \mu^2 \quad (38)$$

- Like our derivation of the moment generating function we are interested in completing the square

$$\begin{aligned} (x - (\mu - \sigma^2 i\lambda))^2 &= x^2 + 2x(\mu - \sigma^2 i\lambda) + (\mu - \sigma^2 i\lambda)^2 \\ (\mu + \sigma^2 i\lambda)^2 &= \mu^2 + 2\sigma^2 i\lambda - \sigma^4 \lambda^2 \end{aligned} \quad (39)$$

- To complete the square we add

$$2\sigma^2 i\lambda - \sigma^4 \lambda^2 - 2\sigma^2 i\lambda + \sigma^4 \lambda^2 = 0. \quad (40)$$

- $f(x) \exp(i\lambda)$ then becomes

$$f(x) \exp(i\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - (\mu + \sigma i\lambda))^2}{2\sigma^2} + i\lambda\mu - \frac{\sigma^2\lambda^2}{2}\right) \quad (41)$$

- Following our approach for the moment generating function, we move the constant outside the integral

$$f(\lambda) = \exp\left(i\lambda\mu - \frac{\sigma^2\lambda^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - (\mu + \sigma i\lambda))^2}{2\sigma^2}\right) dx \quad (42)$$

- **Theorem 4.11** (Uniqueness Theorem) Two distribution functions are identical if and only if their characteristic functions are identical.
- Note that we have a similar theorem for moment generating functions.

- Proof of Lindeberg-Levy:

- First define $f(\lambda)$ as the characteristic function for $Z_t - \mu$ and let $f_n(\lambda)$ be the characteristic function of

$$\frac{\sqrt{n}(\bar{Z}_n - \bar{\mu}_n)}{\bar{\sigma}_n} = \left(\frac{1}{n}\right)^{1/2} \sum_{i=1}^n \left(\frac{Z_t - \mu}{\sigma}\right). \quad (43)$$

- By the structure of the characteristic function we have

$$f_n(\lambda) = \left[f\left(\frac{\lambda}{\sigma\sqrt{n}}\right) \right]^n \quad (44)$$

$$\ln(f_n(\lambda)) = n \ln\left(f\left(\frac{\lambda}{\sigma\sqrt{n}}\right)\right)$$

- Taking a second order Taylor series expansion of $f(\lambda)$ around $\lambda = 0$ gives

$$\begin{aligned}\tilde{f}(\lambda) = & \exp\left(i\mu\lambda_0 - \frac{\sigma^2\lambda_0^2}{2}\right) + \\ & (i\mu - \lambda_0\sigma^2) \exp\left(i\mu\lambda_0 - \frac{\sigma^2\lambda_0^2}{2}\right) (\lambda - \lambda_0) + \\ & \frac{1}{2} (\sigma^2 + i\mu - \lambda_0\sigma^2) \exp\left(i\mu\lambda_0 - \frac{\sigma^2\lambda_0^2}{2}\right) (\lambda - \lambda_0)^2 + \\ & O(\lambda^3)\end{aligned}\tag{45}$$

- Taking $\lambda_0 \rightarrow 0$

$$\tilde{f}(\lambda) = 1 + i\mu\lambda + \frac{1}{2} (-\mu^2 - \sigma^2) \lambda^2 + O(\lambda^3)\tag{46}$$

- Taking $\mu \rightarrow 0$ the result is

$$f(\lambda) = 1 - \frac{\sigma^2 \lambda^2}{2} + o(\lambda^2). \quad (47)$$

- Thus,

$$\ln(f_n(\lambda)) = n \ln \left[1 - \frac{\lambda^2}{2n} + o\left(\frac{\lambda^2}{n}\right) \right] \rightarrow \frac{\lambda^2}{2} \text{ as } n \rightarrow \infty \quad (48)$$

- Thus, by the Uniqueness Theorem the characteristic function of the sample approaches the characteristic function of the standard normal.

Readings

- Moss 6.4 – 6.5
- Popper Chapter 10 – Corroboration, or How a Theory Stands up to Tests