Large Sample Theory (Review)

Ercan Karadas

New York University
Department of Economics

Spring, 2018

Notation

x

- **A**: matrix
- **a**: column vector
  - To denote column k of **A** use $\mathbf{a_k}$
  - To denote a column of ones use **i**
  - To denote row k of **A** use $\mathbf{a_k'}$
- **i**: a vector that contains .

Examples

- For a column vector $\mathbf{x} = (x_1, \ldots, x_n)$
  - $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{i}'\mathbf{x}$
  - $\sum_{i=1}^{n} x_i^2 = \mathbf{x}'\mathbf{x}$
  - $\sum_{i=1}^{n} x_i y_i = \mathbf{x}'\mathbf{y}$
- For $n \times K$ matrix $\mathbf{X}$
  - The inner product of the i^{th} and j^{th} columns of matrix $\mathbf{X}$:

$$[\mathbf{X}'\mathbf{X}]_{ij} = [\mathbf{x}_i'\mathbf{x}_j]$$

  - The $K \times K$ matrix $\mathbf{X}'\mathbf{X}$ is the sum of n $K \times K$ matrices formed from a single row of $\mathbf{X}$:

$$\mathbf{X}'\mathbf{X} = \sum_{i}^{n} \mathbf{x}_i'\mathbf{x}_i$$

M^0: A Useful Idempotent Matrix

$\mathbf{M}^0$ is matrix with all diagonal elements $(1 - 1/n)$, and its off-diagonal elements $-1/n$:

$$\mathbf{M}^0 = \mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}'$$

- Deviations of $\mathbf{x}$ from its mean $\bar{x}$:

$$\mathbf{M}^0\mathbf{x} = \left(\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}'\right)\mathbf{x} = \mathbf{x} - \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x} = \mathbf{x} - \mathbf{i}\bar{x} = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$

- For any constant vector $\mathbf{x} = (x, \ldots, x)$

$$\mathbf{M}^0\mathbf{x} = \mathbf{0}$$

M^0: A Useful Idempotent Matrix (cont'd)

- $\mathbf{M}^0$ is idempotent:

$$\mathbf{M}^{0'} = \mathbf{M}^0 \text{ and } \mathbf{M}^0\mathbf{M}^0 = \mathbf{M}^0$$

- For a vector $\mathbf{x} = (x_1, \ldots, x_n)$ the sum of deviations about the mean:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \mathbf{i}'[\mathbf{M}^0\mathbf{x}] = [\mathbf{M}^0\mathbf{i}]'\mathbf{x} = \mathbf{0}'\mathbf{x} = \mathbf{0}$$

- The sum of squared deviations about the mean:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \mathbf{x}'\mathbf{M}^0\mathbf{x}$$

- The sum of cross products in deviations from the column means:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \mathbf{x}'\mathbf{M}^0\mathbf{y}$$

M^0: A Useful Idempotent Matrix (cont'd)

- 2×2 VC matrix:

$$\begin{bmatrix} \sum_{i=1}^{n}(x_i - \bar{x})^2 & \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ \sum_{i=1}^{n}(y_i - \bar{x})(x_i - \bar{y}) & \sum_{i=1}^{n}(y_i - \bar{y})^2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{M}^0\mathbf{x} & \mathbf{x}'\mathbf{M}^0\mathbf{y} \\ \mathbf{y}'\mathbf{M}^0\mathbf{x} & \mathbf{y}'\mathbf{M}^0\mathbf{y} \end{bmatrix}$$

- Define $n \times 2$ matrix $\mathbf{Z} = [\mathbf{x}\ \mathbf{y}]$, then VC matrix can be written as:

$$\begin{bmatrix} \mathbf{x}'\mathbf{M}^0\mathbf{x} & \mathbf{x}'\mathbf{M}^0\mathbf{y} \\ \mathbf{y}'\mathbf{M}^0\mathbf{x} & \mathbf{y}'\mathbf{M}^0\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix} \mathbf{M}^0 \begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix} = \mathbf{Z}'\mathbf{M}^0\mathbf{Z}$$

Introduction

- We looked at finite-sample properties of the OLS estimator and its associated test statistics (first term)
- These are based on assumptions that are violated very often
- The finite-sample theory breaks down if one of the following three assumptions is violated:
  - the exogeneity of regressors
  - the normality of the error term, and
  - the linearity of the regression equation
- **Asymptotic** or **large-sample theory** provides an alternative approach retaining only the third assumption
- It derives an approximation to the distribution of the estimator and its associated statistics assuming that 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

Two Main Concepts of Asymptotic Theory

- The two main concepts in asymptotic theory: **consistency** and **asymptotic normality**
- Some intuition
  - Consistency: the more data we get, the closer we get to knowing the truth (or we eventually know the truth)
  - Asymptotic normality: as we get more and more data, averages of random variables behave like normally distributed random variables.
- The main probability theory tools for establishing
  - consistency $\rightarrow$ Laws of Large Numbers (LLNs)
  - asymptotic normality $\rightarrow$ Central Limit Theorems (CLTs)

Probability Tools for Asymptotic Theory: LLN, CLT

- Laws of Large Numbers (LLNs)
  - LLN is a result that states the conditions under which a sample average of random variables converges to a population expectation.
  - LLNs concern conditions under which the sequence of sample mean converges either in probability or almost surely
  - There are many LLN results (eg. Chebychev's LLN, Kolmongorov's/Khinchine's LLN, Markov's LLN)
- Central Limit Theorems (CLTs)
  - CLTs are about the limiting behaviour of the difference between a sample mean and its expected value
  - There are many CLTs (eg. Lindeberg-Levy CLT, Lindeberg-Feller CLT, Liapounov's CLT)

Modes of Convergence - Convergence in Probability

- A sequence of random variables $\{x_n\}$ converges in probability to a constant c iff

$$\lim_{n\to\infty} \text{Prob}\left(|x_n - c| > \varepsilon\right) = 0 \quad \text{for any } \varepsilon > 0$$

  Notation: $\text{plim } x_n = c \quad$ or $\quad x_n \xrightarrow{p} c$

- A sequence of $K \times 1$ random vectors $\{\mathbf{x}_n\}$ converges in probability to a constant vector $\mathbf{c}$ iff

$$\text{plim } x_{kn} = c_k \quad \text{for all } k = 1, \ldots, K$$

  where x_{kn} is the k-th element of $\mathbf{x}_n$, c_k is the k-th element of $\mathbf{c}$

- A sequence of random variables $\{x_n\}$ converges in probability to a random variable x iff

$$\lim_{n\to\infty} \text{Prob}\left(|x_n - x| > \varepsilon\right) = 0 \quad \text{for any } \varepsilon > 0$$

Modes of Convergence - Almost Sure Convergence

- A sequence of random variables $\{x_n\}$ converges almost surely to a constant c iff

$$\text{Prob}\left(\lim_{n\to\infty} x_n = c\right) = 1$$

  Notation: $x_n \xrightarrow{a.s.} c$

- A sequence of $K \times 1$ random vectors $\{\mathbf{x}_n\}$ converges almost surely to a constant vector $\mathbf{c}$ iff

$$x_{kn} \xrightarrow{a.s.} c_k \quad \text{for all } k = 1, \ldots, K$$

- A sequence of random variables $\{x_n\}$ converges almost surely to a random variable x iff

$$\lim_{n\to\infty} \text{Prob}\left(|x_i - x| > \varepsilon \text{ for all } i \geq n\right) = 0 \quad \text{for any } \varepsilon > 0$$

Modes of Convergence - Convergence in r-th Mean

- A sequence of random variables x_n converges in r-th mean to a constant c iff

$$E\left[|x_n|^r\right] < \infty \quad \text{and} \quad \lim_{n\to\infty} E\left[|x_n - c|^r\right] = 0$$

  Notation: $x_n \xrightarrow{r.m.} c$

- For $r = 2$ it is called Converges in Mean Square and denoted by $x_n \xrightarrow{m.s.} c$
- A sequence of $K \times 1$ random vectors $\{\mathbf{x}_n\}$ converges in r-th mean to a constant vector $\mathbf{c}$ iff

$$x_{kn} \xrightarrow{r.m.} c_k \quad \text{for all } k = 1, \ldots, K$$

- A sequence of random variables $\{x_n\}$ converges in r-th mean to a random variable x iff

$$E\left[|x_n|^r\right] < \infty \quad \text{and} \quad \lim_{n\to\infty} E\left[|x_n - x|^r\right] = 0$$

Modes of Convergence - Convergence in Distribution

- A sequence of random variables x_n converges in distribution to a random variable x with CDF $F(x)$ iff

$$\lim_{n\to\infty} \mid F_n(x_n) - F(x) \mid = 0 \quad \text{at all continuity points of } F(x)$$

  where $F_n(x_n)$ is the CDF of x_n.
  Notation: $x_n \xrightarrow{d} x$
- If $x_n \xrightarrow{d} x$, then $F(x)$ is called the limiting distribution of x_n.
- A sequence of random vectors $\mathbf{x}_n$ converges in distribution to a random vector $\mathbf{x}$ with (joint) CDF $F(\mathbf{x})$ iff

$$\lim_{n\to\infty} \mid F_n(\mathbf{x}_n) - F(\mathbf{x}) \mid = 0 \quad \text{at all continuity points of } F(\mathbf{x})$$

  where $F_n(\mathbf{x}_n)$ is the CDF of $\mathbf{x}_n$.
- Note that for convergence in distribution, unlike the other concepts of convergence, element-by-element convergence does not necessarily mean convergence for the vector sequence.

Relation among Modes of Convergence

i) $\mathbf{x}_n \xrightarrow{m.s.} \mathbf{c} \implies \mathbf{x}_n \xrightarrow{p} \mathbf{c}$ (so $\mathbf{x}_n \xrightarrow{m.s.} \mathbf{x} \implies \mathbf{x}_n \xrightarrow{p} \mathbf{x}$)

ii) $\mathbf{x}_n \xrightarrow{a.s.} \mathbf{c} \implies \mathbf{x}_n \xrightarrow{p} \mathbf{c}$ (so $\mathbf{x}_n \xrightarrow{a.s.} \mathbf{x} \implies \mathbf{x}_n \xrightarrow{p} \mathbf{x}$)

iii) $\mathbf{x}_n \xrightarrow{p} \mathbf{c} \iff \mathbf{x}_n \xrightarrow{d} \mathbf{c}$

That is, if the limiting random variable is a constant (a trivial random variable), convergence in distribution is the same as convergence in probability.

Preservation of Convergence for Continuous Transformation

Suppose $\mathbf{a}(\cdot)$ is a vector-valued continuous function that does not depend on n, then

i) $\mathbf{x}_n \xrightarrow{p} \mathbf{c} \implies \mathbf{a}(\mathbf{x}_n) \xrightarrow{p} \mathbf{a}(\mathbf{c})$. Alternatively stated

$$\text{plim } \mathbf{a}(\mathbf{x}_n) = \mathbf{a}(\text{plim } \mathbf{x}_n)$$

ii) $\mathbf{x}_n \xrightarrow{d} \mathbf{x} \implies \mathbf{a}(\mathbf{x}_n) \xrightarrow{d} \mathbf{a}(\mathbf{x})$.

Combinations of Modes of Convergences

i) $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$, $\mathbf{y}_n \xrightarrow{p} \mathbf{c} \implies \mathbf{x}_n + \mathbf{y}_n \xrightarrow{d} \mathbf{x} + \mathbf{c}$.

ii) $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$, $\mathbf{y}_n \xrightarrow{p} \mathbf{0} \implies \mathbf{y}_n'\mathbf{x}_n \xrightarrow{p} \mathbf{0}$.

iii) $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$, $\mathbf{A}_n \xrightarrow{p} \mathbf{A} \implies \mathbf{A}_n\mathbf{x}_n \xrightarrow{d} \mathbf{A}\mathbf{x}$,
provided that $\mathbf{A}_n$ and $\mathbf{x}_n$ are conformable.
In particular, if $\mathbf{x} \sim N(\mathbf{0}, \mathbf{\Sigma})$, then $\mathbf{A}_n\mathbf{x}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$.

iv) $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$, $\mathbf{A}_n \xrightarrow{p} \mathbf{A} \implies \mathbf{x}_n'\mathbf{A}_n^{-1}\mathbf{x}_n \xrightarrow{d} \mathbf{x}'\mathbf{A}^{-1}\mathbf{x}$,
provided that $\mathbf{A}_n$ and $\mathbf{x}_n$ are conformable.

Parts (i) and (iii) are sometimes called Slutzky's Theorem.

The Delta Method

Suppose $\mathbf{x}_i$ is a sequence of K-dimensional random vectors such that

$$\mathbf{x}_n \xrightarrow{p} \mathbf{c} \quad \text{and} \quad \sqrt{n}\,(\mathbf{x}_n - \mathbf{c}) \xrightarrow{d} \mathbf{z}$$

and suppose that $\mathbf{a}(\cdot) : \mathbb{R}^K \to \mathbb{R}^r$ has continuous first derivatives with $\mathbf{A}(\mathbf{c})$ denoting the $r \times K$ matrix of first derivatives evaluated at $\mathbf{c}$:

$$\mathbf{A}(\mathbf{c}) \equiv \frac{\partial \mathbf{a}(\mathbf{c})}{\partial \mathbf{c}'}$$

Then

$$\sqrt{n}\,[\mathbf{a}(\mathbf{x}_n) - \mathbf{a}(\mathbf{c})] \xrightarrow{d} \mathbf{A}(\mathbf{c})\mathbf{z}$$

In particular,

$$\sqrt{n}\,(\mathbf{x}_n - \mathbf{c}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}) \implies \sqrt{n}\,[\mathbf{a}(\mathbf{x}_n) - \mathbf{a}(\mathbf{c})] \xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}(\mathbf{c})\boldsymbol{\Sigma}\mathbf{A}(\mathbf{c})'\right)$$

Khinchine Weak Law of Large Numbers (WLLN)

- If x_i, $i = 1, \ldots, n$ is a random (i.i.d.) sample from a distribution with finite mean $E[x_i] = \mu$, then

$$\text{plim } \bar{x}_n = \mu$$

- Extensions:
  - Multivariate Extension (sequence of random vectors $\{\mathbf{x}_i\}$)
  - Relaxation of i.i.d. assumption
  - Functions of random variables $f(x_i)$
  - Vector valued functions $f(\mathbf{x}_i)$

Lindeberg-Levy Central Limit Theorem

If x_i, $i = 1, \ldots, n$ is a random (i.i.d.) sample from a distribution with finite mean $E[x_i] = \mu$ and $Var[x_i] = \sigma^2$, then

$$\sqrt{n}\,(\bar{x}_n - \mu) \xrightarrow{d} N[0, \sigma^2]$$

or

$$\bar{x}_n \overset{a}{\sim} N[\mu, \frac{\sigma^2}{n}]$$

Read $\overset{a}{\sim}$ 'approximately distributed as'

CLT also holds for multivariate extension: sequence of random vectors