Day 1B
Asymptotic Theory

© A. Colin Cameron
Univ. of Calif.- Davis

Frontiers in Econometrics
Bavarian Graduate Program in Economics

.

Based on A. Colin Cameron and Pravin K. Trivedi (2009, 2010), Microeconometrics using Stata (MUS), Stata Press. and A. Colin Cameron and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications (MMA), C.U.P.

March 21-25, 2011

1. Introduction

OLS for the linear model is the building block for nonlinear regression.

1. Introduction
2. Sequences of random variables
3. Convergence in probability
4. Laws of large numbers (for averages)
5. Convergence in distribution
6. Central limit theorems (for averages)
7. Some Key Results
8. Simulations for LLN and CLT
9. Appendix: Some Further Asymptotic Results
   Appendix: Sampling Schemes
   Appendix: OLS under Simple Random Sampling

2. Sequences of random variables

- Recall a sequence of real numbers
  - e.g. $a_N = 2 + \frac{3}{N}$
- What happens as $N \rightarrow \infty$?
  - mathematical convergence (or divergence)
- A sequence of nonstochastic real numbers $\{a_N\}$ *converges* to a if for any $\varepsilon > 0$, there exists $N^* = N^*(\varepsilon)$ such that for all $N > N^*$,

$$|a_N - a| < \varepsilon.$$

  - e.g. $a_N = 2 + 3/N \rightarrow 2$ since $|a_N - a| = |2 + 3/N - 2| = |3/N| < \varepsilon$ for all $N > N^* = 3/\varepsilon$.

- We instead consider a sequence of random variables b_N.
  - e.g. $b_N = \frac{1}{N}\sum_{i=1}^N x_i^2$
  - e.g. $b_N = \frac{1}{N}\sum_{i=1}^N x_i u_i$
  - e.g. $b_N = \widehat{\beta} = \left(\frac{1}{N}\sum_{i=1}^N x_i^2\right)^{-1} \frac{1}{N}\sum_{i=1}^N x_i u_i$
- What happens as $N \longrightarrow \infty$?
  - $|b_N - b|$ may exceed ε due to randomness, so $b_N \nrightarrow b$ exactly
  - instead use convergence in probability and in distribution.

3. Convergence in probability and consistency

- Informal definition: The sequence $\{b_N\}$ *converges in probability* to b if for any $\varepsilon > 0$
$$\lim_{N\to\infty} \Pr[|b_N - b| < \varepsilon] = 1.$$
- Formal definition: A sequence of random variables $\{b_N\}$ *converges in probability* to b if for any $\varepsilon > 0$ and $\delta > 0$, there exists $N^* = N^*(\varepsilon, \delta)$ such that for all $N > N^*$,
$$\Pr[|b_N - b| < \varepsilon] > 1 - \delta.$$
- Intuition: Convergence in mean square (i.e. $\lim_{N\to\infty} \mathrm{E}[(b_N - b)^2] = 0$) implies convergence in probability.
  - But convergence in probability can happen even if $\mathrm{E}[(b_N - b)^2]$ does not exist.

Consistency

- We write $\text{plim } b_N = b$ or $b_N \xrightarrow{p} b$.
- The limit b may be a constant or a random variable.
  - Often $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$, a constant.
  - Then we say $\widehat{\boldsymbol{\beta}}$ is *consistent* for $\boldsymbol{\beta}$.

4. Law of large numbers

- Easy way to get probability limit when b_N is an average

$$b_N = \bar{X}_N = \frac{1}{N} \sum\nolimits_{i=1}^{N} X_i.$$

  - X_i here is general notation for a random variable. e.g. $X_i = x_i u_i$.

- *Weak Law of Large Numbers (WLLN)*:
  Specifies conditions on the individual terms X_i in $\bar{X}_N$ under which

$$(\bar{X}_N - \mathrm{E}[\bar{X}_N]) \xrightarrow{p} 0.$$

- *Khinchine's Theorem (WLLN)*:
  Let $\{X_i\}$ be i.i.d. (independent and identically distributed).
  If and only if $\mathrm{E}[X_i] = \mu$ exists, then $(\bar{X}_N - \mu) \xrightarrow{p} 0$.

- Other LLN's are Kolmogorov and, for i.n.i.d. data, Markov.
- If a LLN can be applied then

$$\begin{aligned} \text{plim}\,\bar{X}_N &= \lim \mathrm{E}[\bar{X}_N] && \text{in general} \\ &= \lim N^{-1} \textstyle\sum_{i=1}^{N} \mathrm{E}[X_i] && \text{if } X_i \text{ independent over } i \\ &= \mu && \text{if } X_i \text{ i.i.d.} \end{aligned}$$

5. Convergence in distribution

- b_N has (unknown) cumulative distribution function (cdf) F_N.
  Like any other function, F_N may have a limit function.
- *Convergence in Distribution*:
  A sequence of random variables $\{b_N\}$ *converges in distribution* to a random variable b if

$$\lim_{N\rightarrow\infty} F_N = F, \text{ where } F \text{ is the c.d.f. of } b$$

  at every continuity point of F, where convergence is in the usual mathematical sense.
- We write $b_N \xrightarrow{d} b$, and call F the *limit distribution* of $\{b_N\}$.
- Basically F_N is very complicated and F is simple like $\mathcal{N}[0, 1]$.

6. Central limit theorems

- Easy way to get limit distribution when b_N is an average $\bar{X}_N$.
- $\bar{X}_N$ has a degenerate limit distribution with all mass at one point since $\bar{X}_N \xrightarrow{p} \lim \mathrm{E}[\bar{X}_N]$ by a LLN.
- So rescale $\bar{X}_N$ to standardized variate

$$b_N = Z_N = \frac{\bar{X}_N - \mathrm{E}[\bar{X}_N]}{\sqrt{\mathrm{V}[\bar{X}_N]}} \sim [0,\ 1].$$

- *Central Limit Theorem (CLT)*:
  A CLT specifies the conditions on the individual terms X_i in $\bar{X}_N$ under which

$$Z_N \xrightarrow{d} \mathcal{N}[0,1].$$

- *Lindeberg-Levy CLT:*
  Let $\{X_i\}$ be i.i.d. with $\mathrm{E}[X_i] = \mu$ and $\mathrm{V}[X_i] = \sigma^2$.
  Then $Z_N = \sqrt{N}(\bar{X}_N - \mu)/\sigma \xrightarrow{d} \mathcal{N}[0,1]$.

- Note that

$$\begin{aligned} Z_N &= (\bar{X}_N - \mathrm{E}[\bar{X}_N])/\sqrt{\mathrm{V}[\bar{X}_N]} && \text{in general} \\ &= \textstyle\sum_{i=1}^N (X_i - \mathrm{E}[X_i])/\sqrt{\sum_{i=1}^N \mathrm{V}[X_i]} && \text{if } X_i \text{ independent over } i \\ &= \sqrt{N}(\bar{X}_N - \mu)/\sigma && \text{if } X_i \text{ i.i.d.} \end{aligned}$$

- The last expression can be rewritten as

$$\frac{\bar{X}_N - \mu}{\sigma/\sqrt{N}} \xrightarrow{d} \mathcal{N}[0, 1].$$

- It follows that $\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} \mathcal{N}[0, \sigma^2]$.
- More generally we often find $\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}[0, V]$.
  - Scale consistent $\widehat{\beta}$ up by $\sqrt{N}$ to get a limit distribution.

Multivariate central limit theorem

- Consider vector $\overline{\mathbf{X}}_N$ with mean $\boldsymbol{\mu}_N$ and variance $\mathbf{V}_N$

$$\overline{\mathbf{X}}_N \sim [\boldsymbol{\mu}_N,\ \mathbf{V}_N].$$

- Rescale $\overline{\mathbf{X}}_N$ to standardized variate

$$\mathbf{Z}_N = \mathbf{V}_N^{-1/2}(\mathbf{b}_N - \boldsymbol{\mu}_N) \sim [\mathbf{0},\ \mathbf{I}].$$

- *Central Limit Theorem (CLT)*:
  A CLT specifies the conditions on the individual terms $\mathbf{X}_i$ in $\overline{\mathbf{X}}_N$ under which

$$\mathbf{Z}_N \xrightarrow{d} \mathcal{N}[\mathbf{0},\ \mathbf{I}].$$

7. Some Key Results

- *Probability Continuity and Continuous Mapping Theorems*
  Let $\mathbf{b}_N$ be a vector of random variables, and $g(\cdot)$ be a continuous real-valued function. Then

$$\begin{array}{rlr} \mathbf{b}_N \xrightarrow{p} \mathbf{b}, \text{ a constant} & \Rightarrow g(\mathbf{b}_N) \xrightarrow{p} g(\mathbf{b}) & \text{Probability Continuity} \\ \mathbf{b}_N \xrightarrow{d} \mathbf{b} & \Rightarrow g(\mathbf{b}_N) \xrightarrow{d} g(\mathbf{b}) & \text{Continuous Mapping} \end{array}$$

- *Transformation Theorem:*
  If $a_N \xrightarrow{d} a$ (a random variable) and $b_N \xrightarrow{p} b$ (a constant), then

$$\begin{array}{ll} (i) & a_N + b_N \xrightarrow{d} a + b \\ (ii) & a_N b_N \xrightarrow{d} ab \\ (iii) & a_N / b_N \xrightarrow{d} a/b, \quad \text{provided } \Pr[b=0]=0. \end{array}$$

  - We use especially a matrix version of (ii).

- *Product Limit Normal Rule:*
  For vector $\mathbf{a}_N$ and matrix $\mathbf{H}_N$, if

$$\begin{aligned} \mathbf{a}_N &\xrightarrow{d} \mathcal{N}[\boldsymbol{\mu}, \mathbf{A}] \\ \mathbf{H}_N &\xrightarrow{p} \mathbf{H}, \quad \text{where } \mathbf{H} \text{ is positive definite} \end{aligned}$$

  then

$$\mathbf{H}_N \mathbf{a}_N \xrightarrow{d} \mathcal{N}[\mathbf{H}\boldsymbol{\mu}, \mathbf{HAH}'].$$

- Leading example is OLS:

$$\begin{aligned} \sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= (\frac{1}{N}\mathbf{X}'\mathbf{X})^{-1} \times \frac{1}{\sqrt{N}}(\mathbf{X}'\mathbf{u}) \\ &\xrightarrow{d} \mathcal{N}[\mathbf{A}^{-1} \times \mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1\prime}]. \end{aligned}$$

8. Simulations for LLN and CLT

Uniform on $(0,1)$ has mean 0 and variance $1/12$.

Sample average of N uniforms has mean 0 and variance $(1/12)/N$.

```
. * Draw from uniform with population mean 0.5
. * Demonstrate LLN by finding average for a very large sample
. * Demonstrate CLT by simulating to obtain many averages
.
. * Small sample: sample mean differs from population mean
. set obs 30
obs was 0, now 30

. set seed 10101

. quietly generate x = runiform()

. mean x

Mean estimation                    Number of obs    =      30

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
           x |   .5459987   .0511899      .4413036    .6506939
--------------------------------------------------------------
```

For $N = 30$: $\bar{x} = 0.546$ differs appreciably from $\mu = 0.500$.

```
. * Consistency: Large sample: sample mean is very close to population mean
. clear all

. set obs 100000
obs was 0, now 100000

. set seed 10101

. quietly generate x = runiform()

. mean x

Mean estimation                     Number of obs    =   100000

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
           x |   .4988239   .0009133      .4970339    .5006138
--------------------------------------------------------------
```

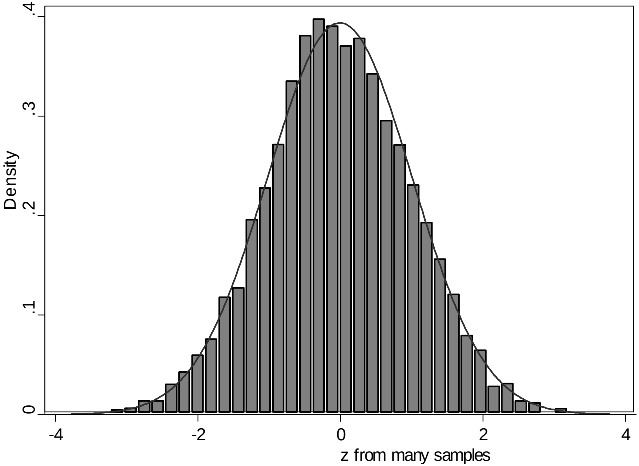
For $N = 100,000$: $\bar{x} = 0.499$ is very close to $\mu = 0.500$.

```
. * Central limit theorem
. * Write program to obtain sample mean for one sample of size numobs (= 30)
. program onesample, rclass
  1.      args numobs
  2.      drop _all
  3.      quietly set obs `numobs'
  4.      generate x = runiform()
  5.      summarize x
  6.      return scalar meanforonesample = r(mean)
  7. end

. * Run this program 10,000 times to get 10,000 sample means
. quietly simulate xbar = r(meanforonesample), seed(10101) reps(10000) nodots: ///
>    onesample 30

. * Summarize the 10,000 sample means and draw histogram
. summarize xbar

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        xbar |     10000    .4995835    .0533809   .3008736   .6990562

. histogram xbar, normal xtitle("xbar from many samples")
(bin=40, start=.30087364, width=.00995456)
```

For $S = 10,000$ simulations each with sample size $N = 30$

$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{10000}$ has distribution with mean 0.4996 close to $\mu = 0.500$ and standard deviation 0.0534, close to $\sigma/\sqrt{N} = \sqrt{1/12}/\sqrt{30} = 0.0527$.

$z = (\bar{x} - \mu)/(\sigma/\sqrt{N}) = (\bar{x} - 0.5)/(\sqrt{1/12}/\sqrt{30}) = (\bar{x} - 0.5)/0.0527.$
Histogram and kernel density estimate for $z_1, z_2, \ldots, z_{10000}$.



This is standard normal, as predicted by the CLT.

9. Appendix: Some Further Asymptotic Results

- Alternative modes of convergence of b_N to b
  - *Mean square*: $\lim_{N\to\infty} \mathrm{E}[(b_N - b)^2] = 0$.
  - *Chebychev's inequality:* $\Pr[(Z-\mu)^2 > k] \leq \sigma^2/k$, for any $k > 0$.
  - *Almost sure:* $\Pr\{\omega | \lim b_N(\omega) = b(\omega)\} = 1$.
  - These imply convergence in probability.
- Consequences:
  - $b_N \xrightarrow{p} b$ implies $b_N \xrightarrow{d} b$.
  - The reverse is generally not true, unless b is a constant.
  - For vector r.v.'s define F_N and F to be cdf's of vectors $\mathbf{b}_N$ and $\mathbf{b}$.
- *Strong Law of Large Numbers (LLN):*
  - The convergence is instead almost surely $(\xrightarrow{as})$.

- *Markov SLLN*:
  Let $\{X_i\}$ be i.n.i.d. with $\mathrm{E}[X_i] = \mu_i$.
  If $\sum_{i=1}^{\infty} \left(\mathrm{E}[|X_i - \mu_i|^{1+\delta}] / i^{1+\delta} \right) < \infty$, for some $\delta > 0$, then $(\bar{X}_N - \mathrm{E}[\bar{X}_N]) \xrightarrow{as} 0$.
  - Relaxes i.i.d. assumption at expense of requiring existence of $(1+\delta)^{th}$ absolute moment.
  - Easiest to set $\delta = 1$, so need variance.
- *Liapounov CLT:*
  Let $\{X_i\}$ be independent with $\mathrm{E}[X_i] = \mu_i$ and $\mathrm{V}[X_i] = \sigma_i^2$.
  If $\lim \left(\sum_{i=1}^{N} \mathrm{E}[|X_i - \mu_i|^{2+\delta}] \right) / \left(\sum_{i=1}^{N} \sigma_i^2 \right)^{(2+\delta)/2} = 0$, for some choice of $\delta > 0$, then $Z_N \xrightarrow{d} \mathcal{N}[0,1]$.
  - The Liapounov CLT relaxes i.i.d. assumption but needs existence of $(2+\delta)^{th}$ absolute moment.
- *Cramer-Wold Device:*
  If $\boldsymbol{\lambda}'\mathbf{b}_N \xrightarrow{d} \mathcal{N}[\ ,\ ]$ for all $\boldsymbol{\lambda} \neq \mathbf{0}$ then $\mathbf{b}_N \xrightarrow{d} \mathcal{N}[\ ,\ ]$.
  - So prove a multivariate CLT by proving a scalar CLT for $\boldsymbol{\lambda}'\mathbf{b}_N$.

9. Appendix: Sampling schemes

- Simple Random Sampling (SRS)
  - Randomly draw (y_i, x_i) from the population with equal probabilities.
  - Then x_i i.i.d. So $x_i u_i$ i.i.d. (if errors u_i are i.i.d.), and x_i^2 i.i.d.
  - Can use Khinchine's LLN and Lindeberg-Levy CLT.
- Fixed regressors
  - Experiment where x_i are fixed and we observe the resulting random y_i.
  - Then x_i fixed, u_i i.i.d. implies $x_i u_i$ i.n.i.d. and x_i^2 nonstochastic.
  - Need to use Markov LLN and Liapounov CLT.
- Exogenous Stratified Sampling
  - Oversample some values of x and undersample others.
  - Then x_i i.n.i.d., so $x_i u_i$ i.n.i.d. and x_i^2 i.n.i.d.
  - Need to use Markov LLN and Liapounov CLT.

- These three different sampling schemes ultimately lead to similar asymptotic results.
- Microeconometrics often use survey data obtained by stratified sampling.
- The simplest results assume simple random sampling.
- Big problems arise if the stratified sampling is Instead Endogenous Stratified Sampling
  - Oversample some values of y and undersample others.
  - Estimators can be inconsistent.
  - Leading examples are Tobit and selection models.

9. Appendix: OLS under simple random sampling

- Scalar regressor: $y_i = \beta x_i + u_i$.
- SRS: (x_i, y_i) i.i.d. with x_i i.i.d. with mean μ_x & u_i i.i.d. with mean 0.
  - **1.** As $x_i u_i$ are i.i.d. apply Khinchine's Theorem.
    Then $N^{-1}\sum_i x_i u_i \xrightarrow{p} \mathrm{E}[xu] = \mathrm{E}[x] \times \mathrm{E}[u] = 0$.
  - **2.** As x_i^2 are i.i.d. apply Khinchine's Theorem.
    Then $N^{-1}\sum_i x_i^2 \xrightarrow{p} \mathrm{E}[x^2]$ which we assume exists.
  - **3.** The probability limit is obtained by combining

$$\begin{aligned}
\operatorname{plim} \widehat{\beta} &= \beta + \operatorname{plim}\left(\frac{\frac{1}{N}\sum_{i=1}^N x_i u_i}{\frac{1}{N}\sum_{i=1}^N x_i^2}\right) \\
&= \beta + \frac{\operatorname{plim} \frac{1}{N}\sum_{i=1}^N x_i u_i}{\operatorname{plim} \frac{1}{N}\sum_{i=1}^N x_i^2} \\
&= \beta + \frac{0}{\mathrm{E}[x^2]} = \beta,
\end{aligned}$$

using probability limit continuity $(\operatorname{plim}[a_N / b_N] = a/b \text{ if } b \neq 0)$.

- SRS: assume x_i i.i.d. with mean μ_x and second moment $\mathsf{E}[x^2]$ and assume u_i i.i.d. with mean 0 and variance σ^2.
- Then $x_i u_i$ are i.i.d. with mean $\mathsf{E}[xu] = \mathsf{E}[x] \times \mathsf{E}[u] = 0$, and $\mathsf{V}[xu] = \mathsf{E}[(xu-0)^2] = \mathsf{E}[x^2 u^2] = \mathsf{E}[x^2]\mathsf{E}[u^2] = \sigma^2 \mathsf{E}[x^2]$.
  - **1.** Lindeberg-Levy CLT for $N^{-1}\sum_{i=1}^N x_i u_i$ yields

$$\sqrt{N}\left(\frac{N^{-1}\sum_{i=1}^N x_i u_i - 0}{\sqrt{\sigma^2 \mathsf{E}[x^2]}}\right) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N x_i u_i}{\sqrt{\sigma^2 \mathsf{E}[x^2]}} \xrightarrow{d} \mathcal{N}[0,1].$$

  - **2.** Convert to $\frac{1}{\sqrt{N}}\sum_{i=1}^N x_i u_i$

$$\begin{aligned} \frac{1}{\sqrt{N}}\sum_{i=1}^N x_i u_i &= \sqrt{\sigma^2 \mathsf{E}[x^2]} \times \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N x_i u_i}{\sqrt{\sigma^2 \mathsf{E}[x^2]}} \\ &\xrightarrow{d} \sqrt{\sigma^2 \mathsf{E}[x^2]} \times \mathcal{N}[0,1] \\ &\xrightarrow{d} \mathcal{N}[0, \sigma^2 \mathsf{E}[x^2]] \end{aligned}$$

using product limit normal rule.

- 3. **3.** The limit distribution is obtained by combining

$$
\begin{aligned}
\sqrt{N}(\widehat{\beta}-\beta) &= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i u_i}{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \\
&\xrightarrow{d} \frac{\mathcal{N}\left[0, \sigma^2 \mathrm{E}[x^2]\right]}{\mathrm{plim}\, \frac{1}{N}\sum_{i=1}^{N} x_i^2} \\
&\xrightarrow{d} \frac{\mathcal{N}\left[0, \sigma^2 \mathrm{E}[x^2]\right]}{\mathrm{E}[x^2]} \\
&\xrightarrow{d} \mathcal{N}\left[0, \sigma^2 \left(\mathrm{E}[x^2]\right)^{-1}\right]
\end{aligned}
$$

using $\mathrm{plim}\, N^{-1}\sum_{i=1}^{N} x_i^2 = \mathrm{E}[x^2]$ from consistency proof and the product normal limit rule (or $a_N \times b_N \xrightarrow{d} a \times b$ if $a_N \xrightarrow{d} a$ and $b_N \xrightarrow{p} b$).