Day 4A
Asymptotic Theory

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

Advanced 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.

July 22-26, 2013
1. Introduction

- Consider sample mean $\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$ as an estimator of the population mean $\mu$
  
  - asymptotic theory gives properties of $\bar{y}$ as $N \to \infty$.

- Make assumptions about the data generating process (dgp)
  
  - $y_i \sim (\mu, \sigma^2)$ i.i.d. (independent and identically distributed)
  
  - $\bar{y} \sim (\mu, \sigma^2/N)$. 
The distribution of $\bar{y}$ has all mass at the mean of $\mu$ as $N \to \infty$

- Intuition: $V[\bar{y}] = E[(\bar{y} - \mu)^2] = \sigma^2 / N \to 0$.
- Formally $\bar{y}$ converges in probability (defined below) to $\mu$
- This is denoted $\bar{y} \xrightarrow{p} \mu$ or $\text{plim } \bar{y} = \mu$.
- Then $\bar{y}$ is consistent for $\mu$.
- The proof uses a law of large numbers (defined below) for an average.
1. Introduction

Statistical inference based on $\bar{y}$ with $N \to \infty$ requires scaling $\bar{y}$ up

- Use standardized statistic $z = \frac{\bar{y} - \mathbb{E}[\bar{y}]}{\sqrt{\mathbb{V}[\bar{y}]}} = \frac{\bar{y} - \mu}{\sigma / \sqrt{N}}$.
- $z \sim (0, 1)$, so $z$ may have a nondegenerate distribution.
- A central limit theorem (defined below) proves $z \sim \mathcal{N}(0, 1)$ as $N \to \infty$.
- Formally $z$ converges in distribution ($\overset{d}{\to}$, defined below) to $\mathcal{N}[0, 1]$.
- Equivalently $\sqrt{N}(\bar{y} - \mu) \overset{d}{\to} \mathcal{N}[0, \sigma^2]$.
- Note: $\bar{y}$ has been scaled up by the multiple $\sqrt{N}$.
For simplicity, the formal result $\sqrt{N}(\bar{y} - \mu) \xrightarrow{d} N[0, \sigma^2]$ is often re-expressed in terms of $\bar{y}$

- $\bar{y} \sim \mathcal{N}[0, \sigma^2 / N]$
- $\sim$ means “is asymptotically distributed as”
- this means $N$ is large enough that the normal is a good approximation
- but $N$ is not so large that $\sigma^2 / N = 0$. 
Outline

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 \( N \to \infty \)?
  - 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. \( a_N = 2 + 3/N \to 2 \) since
    \[
    |a_N - a| = |2 + 3/N - 2| = |3/N| < \varepsilon \text{ for all } N > N^* = 3/\varepsilon.
    \]
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. $b_N = \hat{\beta} = \left( \frac{1}{N} \sum_{i=1}^{N} x_i^2 \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} x_i u_i$

What happens as $N \to \infty$?

- $|b_N - b|$ may exceed $\epsilon$ due to randomness, so $b_N \not\to b$ exactly
- instead use convergence in probability.
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 \( N > N^* \),

\[
\Pr[|b_N - b| < \varepsilon] > 1 - \delta.
\]

- We write \( \text{plim } b_N = b \) or \( b_N \xrightarrow{p} b \)
  - the limit \( b \) may be a constant or a random variable.
Consistency

- Suppose the sequence $b_N$ is an estimator, say $b_N = \hat{\beta}$.
  - If $\hat{\beta} \xrightarrow{p} \beta$, a constant, then we say $\hat{\beta}$ is consistent for $\beta$.
- A simple consistency proof uses convergence in mean square
  - that is \( \lim_{N \to \infty} E[(b_N - b)^2] = 0 \)
  - \( \xrightarrow{ms} \) implies convergence in probability.
- Suppose $\hat{\beta}$ is used to estimate $\beta$
  - \( E[(\hat{\beta} - \beta)^2] = V[\hat{\beta}] + (\text{bias}[\hat{\beta}])^2 \) as MSE = variance + bias$^2$
  - so $\hat{\beta} \xrightarrow{ms} \beta$ if $V[\hat{\beta}] \to 0$ and $\text{bias}[\hat{\beta}] \to 0$ as $N \to \infty$
  - it follows that $\hat{\beta} \xrightarrow{p} \beta$ if the variance and bias go to zero as $N \to \infty$.
- We use the weaker convergence in probability as $\hat{\beta} \xrightarrow{p} \beta$ is possible even if the mean and variance of $\hat{\beta}$ do not exist.
4. Law of large numbers

- Easy way to get probability limit when $b_N$ is an average
  \[ b_N = \bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i. \]

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

- Khinchine’s Theorem (WLLN):
  Let $\{X_i\}$ be i.i.d. (independent and identically distributed).
  If and only if $\mathbb{E}[X_i] = \mu$ exists, then $\bar{X}_N - \mu \xrightarrow{P} 0$. 
Other LLN’s are Kolmogorov and, for i.n.i.d. data, Markov

- these are given later.

If a LLN can be applied then

\[
\text{plim } \bar{X}_N = \lim E[\bar{X}_N] = \lim N^{-1} \sum_{i=1}^{N} E[X_i] = \mu \text{ if } X_i \text{ i.i.d.}
\]
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
  \[
  \lim_{N \to \infty} F_N = F, \text{ where } F \text{ is the c.d.f. of } b
  \]
  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 \( \bar{X}_N \xrightarrow{p} \lim E[\bar{X}_N] \) by a LLN.
- So rescale \( \bar{X}_N \) to standardized variate

\[
\begin{align*}
    b_N &= Z_N = \frac{\bar{X}_N - E[\bar{X}_N]}{\sqrt{V[\bar{X}_N]}} \\
    &\sim [0, 1].
\end{align*}
\]

- **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 \( E[X_i] = \mu \) and \( V[X_i] = \sigma^2 \).
  Then

\[
    Z_N = \sqrt{N}(\bar{X}_N - \mu) / \sigma \xrightarrow{d} \mathcal{N}[0,1].
\]
Note that

\[ Z_N = \frac{\bar{X}_N - E[\bar{X}_N]}{\sqrt{\text{V}[\bar{X}_N]}} \]

in general

\[ = \sum_{i=1}^{N} (X_i - E[X_i]) / \sqrt{\sum_{i=1}^{N} \text{V}[X_i]} \]

if \( X_i \) independent over \( i \)

\[ = \frac{\sqrt{N}(\bar{X}_N - \mu)}{\sigma} \]

if \( X_i \) i.i.d.

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 \( \sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}[0, \text{V}]. \)

- Scale consistent \( \hat{\beta} \) up by \( \sqrt{N} \) to get a limit distribution.
Multivariate central limit theorem

- Consider vector $\overline{X}_N$ with mean $\mu_N$ and variance $V_N$
  $$\overline{X}_N \sim [\mu_N, V_N].$$

- Rescale $\overline{X}_N$ to standardized variate
  $$Z_N = V_N^{-1/2}(\overline{X}_N - \mu_N) \sim [0, I].$$

- **Central Limit Theorem (CLT):**
  A CLT specifies the conditions on the individual terms $X_i$ in $\overline{X}_N$ under which
  $$Z_N \xrightarrow{d} \mathcal{N}[0, I].$$
Often \( \lim N \mathbf{V}_N \) is finite nonzero
- for example if \( X_i \sim (\mu, \Sigma) \) then \( \mathbf{V}_N = \mathbb{V}[\bar{X}_N] = N^{-1} \Sigma \), so \( N \mathbf{V}_N = \Sigma \).

Then \( V_N^{-1/2} (\bar{X}_N - \mu_N) \xrightarrow{d} \mathcal{N}[0, I] \) implies

\[
\sqrt{N}(\bar{X}_N - \mu_N) \xrightarrow{d} \mathcal{N}[0, \lim N^{-1} \mathbf{V}_N]
\]

- scaling the average \( \bar{X}_N \) by a multiple \( \sqrt{N} \) gives a limit distribution with a finite nonzero variance.
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

  $$\mathbf{b}_N \xrightarrow{p} \mathbf{b}, \text{ a constant} \quad \Rightarrow \quad g(\mathbf{b}_N) \xrightarrow{p} g(\mathbf{b}) \quad \text{Probability Continuity}$$

  $$\mathbf{b}_N \xrightarrow{d} \mathbf{b} \quad \Rightarrow \quad g(\mathbf{b}_N) \xrightarrow{d} g(\mathbf{b}) \quad \text{Continuous Mapping}$$

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

  (i) $a_N + b_N \xrightarrow{d} a + b$

  (ii) $a_N b_N \xrightarrow{d} ab$

  (iii) $a_N / b_N \xrightarrow{d} a / b$, provided $\Pr[b = 0] = 0$.

  - We use especially a matrix version of (ii).
• **Product Limit Normal Rule:**
  For vector \(a_N\) and matrix \(H_N\), if

\[
a_N \xrightarrow{d} \mathcal{N} [\mu, A]
\]

\[
H_N \xrightarrow{p} H, \quad \text{where } H \text{ is positive definite}
\]

then

\[
H_N a_N \xrightarrow{d} \mathcal{N} [H\mu, HAH']
\]

• Leading example is OLS:

\[
\sqrt{N} (\hat{\beta} - \beta_0) = \left( \frac{1}{N} X'X \right)^{-1} \times \frac{1}{\sqrt{N}} (X'u)
\]

\[
\xrightarrow{d} \mathcal{N} [A^{-1} \times 0, A^{-1} BA^{-1}']
\]
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

<table>
<thead>
<tr>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0.5459987</td>
<td>0.0511899</td>
</tr>
<tr>
<td></td>
<td>0.4413036</td>
<td>0.6506939</td>
</tr>
</tbody>
</table>

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

For LLN and CLT

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0.4988239</td>
<td>0.0009133</td>
<td>0.4970339 0.5006138</td>
</tr>
</tbody>
</table>

For $N = 100,000$: $\bar{x} = 0.499$ is very close to $\mu = 0.500$. 
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**: $\lim_{N \to \infty} E[(b_N - b)^2] = 0$.
- **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 $b_N$ and $b$. 
**Strong Law of Large Numbers (LLN):**
- The convergence is instead almost surely ($\overset{as}{\to}$).

**Kolmogorov SLLN:**
Let $\{X_i\}$ be i.i.d. If and only if $E[X_i] = \mu$ exists and $E[|X_i|] < \infty$, then $(\bar{X}_N - E[\bar{X}_N]) \overset{as}{\to} 0$.
- Compared to Khinchine's Theorem $\overset{a.s.}{\to}$ requires $E[|X_i|] < \infty$.

**Markov SLLN:**
Let $\{X_i\}$ be i.n.i.d. with $E[X_i] = \mu_i$.
If $\sum_{i=1}^{\infty} \left( E[|X_i - \mu_i|^{1+\delta}] / i^{1+\delta} \right) < \infty$, for some $\delta > 0$, then $(\bar{X}_N - E[\bar{X}_N]) \overset{as}{\to} 0$.
- 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 \( \text{E}[X_i] = \mu_i \) and \( \text{V}[X_i] = \sigma^2_i \).
  If \( \lim \left( \sum_{i=1}^N \text{E}[|X_i - \mu_i|^{2+\delta}] \right) / \left( \sum_{i=1}^N \sigma^2_i \right)^{(2+\delta)/2} = 0 \), for some choice of \( \delta > 0 \), then \( Z_N \xrightarrow{d} N[0, 1] \).
  - The Liapounov CLT relaxes i.i.d. assumption but needs existence of \((2 + \delta)^{th}\) absolute moment.

- **Cramer-Wold Device:**
  If \( \lambda' b_N \xrightarrow{d} N[\ , \ ] \) for all \( \lambda \neq 0 \) then \( b_N \xrightarrow{d} N[\ , \ ] \).
  - So prove a multivariate CLT by proving a scalar CLT for \( \lambda' b_N \).
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 \(x_iu_i\) 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 \(x_iu_i\) 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 \(x_iu_i\) 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 $\mu_x$ & $u_i$ i.i.d. with mean 0.

1. As $x_i u_i$ are i.i.d. apply Khinchine’s Theorem.
   Then $N^{-1} \sum_i x_i u_i \xrightarrow{p} E[xu] = E[x] \times E[u] = 0$.

2. As $x_i^2$ are i.i.d. apply Khinchine’s Theorem.
   Then $N^{-1} \sum_i x_i^2 \xrightarrow{p} E[x^2]$ which we assume exists.

3. The probability limit is obtained by combining

\[
\text{plim} \hat{\beta} = \beta + \text{plim} \left( \frac{\frac{1}{N} \sum_{i=1}^{N} x_i u_i}{\frac{1}{N} \sum_{i=1}^{N} x_i^2} \right)
\]

\[
= \beta + \frac{\text{plim} \left( \frac{1}{N} \sum_{i=1}^{N} x_i u_i \right)}{\text{plim} \left( \frac{1}{N} \sum_{i=1}^{N} x_i^2 \right)}
\]

\[
= \beta + \frac{0}{E[x^2]} = \beta,
\]

using probability limit continuity ($\text{plim} \left[ a_N / b_N \right] = a/b$ if $b \neq 0$).
SRS: assume $x_i$ i.i.d. with mean $\mu_x$ and second moment $E[x^2]$ and assume $u_i$ i.i.d. with mean 0 and variance $\sigma^2$.

Then $x_i u_i$ are i.i.d. with mean $E[xu] = E[x] \times E[u] = 0$, and


1. Lindeberg-Levy CLT for $N^{-1} \sum_{i=1}^N x_i u_i$ yields

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

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

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N x_i u_i = \sqrt{\sigma^2 E[x^2]} \times \frac{1}{\sqrt{\sigma^2 E[x^2]}} \sum_{i=1}^N x_i u_i$$

$$\xrightarrow{d} \sqrt{\sigma^2 E[x^2]} \times \mathcal{N}[0, 1]$$

$$\xrightarrow{d} \mathcal{N}[0, \sigma^2 E[x^2]]$$

using product limit normal rule.
3. The limit distribution is obtained by combining

\[ \sqrt{N} (\hat{\beta} - \beta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i u_i / \frac{1}{N} \sum_{i=1}^{N} x_i^2 \]

\[ \overset{d}{\rightarrow} \mathcal{N} \left[ 0, \sigma^2 \mathbb{E}[x^2] \right] / \text{plim} \frac{1}{N} \sum_{i=1}^{N} x_i^2 \]

\[ \overset{d}{\rightarrow} \mathcal{N} \left[ 0, \sigma^2 \mathbb{E}[x^2] \right] / \mathbb{E}[x^2] \]

\[ \overset{d}{\rightarrow} \mathcal{N} \left[ 0, \sigma^2 \left( \mathbb{E}[x^2] \right)^{-1} \right] \]

using \( \text{plim} N^{-1} \sum_{i=1}^{N} x_i^2 = \mathbb{E}[x^2] \) from consistency proof and the product normal limit rule

(or \( a_N \times b_N \overset{d}{\rightarrow} a \times b \) if \( a_N \overset{d}{\rightarrow} a \) and \( b_N \overset{p}{\rightarrow} b \)).