# Economics 583: Econometric Theory I
# A Primer on Asymptotics

Eric Zivot

January 14, 2013

The two main concepts in asymptotic theory that we will use are

- Consistency
- Asymptotic Normality

Intuition

- consistency: as we get more and more data, we eventually know the truth
- asymptotic normality: as we get more and more data, averages of random variables behave like normally distributed random variables

Motivating Example:

Let $X_1, \ldots, X_n$ denote an independent and identically distributed (iid ) random sample with $E[X_i] = \mu$ and var$(X_i) = \sigma^2$.

We don't know the probability density function (pdf) $f(X_i, \boldsymbol{\theta})$, but we know the value of $\sigma^2$.

Goal: Estimate the mean value $\mu$ from the random sample of data, compute 95% confidence interval for $\mu$, and perform 5% test of $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

Remarks

1. A natural question is: how large does $n$ have to be in order for the asymptotic distribution to be accurate?

2. The CLT is a first order asymptotic approximation. Sometimes higher order approximations are possible.

3. We can use *Monte Carlo simulation* experiments to evaluate the asymptotic approximations for particular cases.

4. We can often use *bootstrap* techniques to provide numerical estimates for $\mathsf{avar}(\hat{\mu})$ and confidence intervals. These are alternatives to the analytic formulas derived from asymptotic theory.

5. If we don't know $\sigma^2$, we have to estimate $\text{avar}(\hat{\mu})$. This gives the estimated asymptotic variance

$$\widehat{\text{avar}}(\hat{\mu}) = \frac{\hat{\sigma}^2}{n},$$

where $\hat{\sigma}^2$ is a consistent estimate of $\sigma^2$.

**Probability Theory Tools**

The probability theory tools (theorems) for establishing consistency of estimators are

- *Laws of Large Numbers* (LLNs).

The tools (theorems) for establishing asymptotic normality are

- *Central Limit Theorems* (CLTs).

A comprehensive reference is White (1994), *Asymptotic Theory for Econometricians*, Academic Press.

## Laws of Large Numbers

Let $X_1, \ldots, X_n$ be a iid random variables with pdf $f(X, \boldsymbol{\theta})$. For a given function $g$, define the sequence of random variables based on the sample

$$\begin{aligned}
Y_1 &= g(X_1) \\
Y_2 &= g(X_1, X_2) \\
&\vdots \\
Y_n &= g(X_1, \ldots, X_n)
\end{aligned}$$

Example: Let $X \sim N(\mu, \sigma^2)$ so that $\boldsymbol{\theta} = (\mu, \sigma^2)$ and define

$$Y_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Hence, sample statistics may be treated as a sequence of random variables.

**Definition 1** *Convergence in Probability*

Let $Y_1, \ldots, Y_n$ be a sequence of random variables. We say that $Y_n$ converges in probability to $c$, which may be a constant or a random variable, and write

$$Y_n \xrightarrow{p} c$$

or

$$p \lim_{n \to \infty} Y_n = c$$

if $\forall\ \varepsilon > 0$,

$$\lim_{n \to \infty} \Pr(|Y_n - c| > \varepsilon) = 0$$

Remarks

1. $Y_n \xrightarrow{p} c$ is the same as $Y_n - c \xrightarrow{p} 0$

2. For a vector process, $\mathbf{Y}_n = (Y_{n1}, \ldots, Y_{nk})'$, $\mathbf{Y}_n \xrightarrow{p} \mathbf{c}$ if $Y_{ni} \xrightarrow{p} c_i$ for $i = i, \ldots, n$.

**Definition 2** *Consistent Estimator*

If $\hat{\theta}$ is an estimator of the scalar parameter $\theta$, then $\hat{\theta}$ is consistent for $\theta$ if

$$\hat{\theta} \xrightarrow{p} \theta$$

If $\hat{\boldsymbol{\theta}}$ is an estimator of the $n \times 1$ vector $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}$ if $\hat{\theta}_i \xrightarrow{p} \theta_i$ for $i = 1, \ldots, n$.

Remark:

All consistency proofs are based on a particular LLN. A LLN is a result that states the conditions under which a sample average of random variables converges to a population expectation. There are many LLN results. The most straightforward is the LLN due to Chebychev.

## Chebychev's LLN

**Theorem 1** *Chebychev's LLN*

Let $X_1, \ldots, X_n$ be iid random variables with $E[X_i] = \mu < \infty$ and $\text{var}(X_i) = \sigma^2 < \infty$. Then

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} E[X_i] = \mu$$

The proof is based on the famous Chebychev's inequality.

**Lemma 2** *Chebychev's Inequality*

Let $X$ by any random variable with $E[X] = \mu < \infty$ and $\text{var}(X) = \sigma^2 < \infty$. Then for every $\varepsilon > 0$

$$\Pr(|X - \mu| \geq \varepsilon) \leq \frac{\text{var}(X)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

Proof of Chebychev's LLN

Applying Chebychev's inequality to $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ gives

$$\Pr(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\text{var}(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

so that

$$\lim_{n\to\infty} \Pr(|\bar{X} - \mu| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0$$

Remark

The proof of Chebychev's LLN relies on the concept of convergence in mean square. That is,

$$\text{MSE}(\bar{X}, \mu) = E[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n} \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

In general, if MSE$(\bar{X}, \mu) \longrightarrow 0$ then $\bar{X} \xrightarrow{p} \mu$. However, it may be the case that $\bar{X} \xrightarrow{p} \mu$ but MSE$(\bar{X}, \mu) \nrightarrow 0$. This would occur, for example, if var$(X)$ does not exist.

Recall,

$$\text{MSE}(\bar{X}, \mu) = E[(\bar{X} - \mu)^2] = \text{bias}(\bar{X}, \mu)^2 + \text{var}(\bar{X})$$

So convergence in MSE implies that as $n \to \infty$

$$\begin{aligned} \text{bias}(\bar{X}, \mu)^2 &= \left(E[\bar{X}] - \mu\right)^2 \to 0 \\ \text{var}(\bar{X}) &= E\left(\bar{X} - E[\bar{X}]\right)^2 \to 0 \end{aligned}$$

## Kolmogorov's and Khinchine's LLN

The LLN with the weakest set of conditions on the random sample $X_1, \ldots, X_n$ is due to Kolmogorov and Khinchine.

**Theorem 3** *Kolmongorov's LLN*

Let $X_1, \ldots, X_n$ be iid random variables with $E[|X_i|] < \infty$ and $E[X_i] = \mu$. Then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E[X_i] = \mu$$

Remark

Kolmogorov's LLN does not require var$(X_i)$ to exist. Only the mean needs to exist. That is, this LLN covers random variables with fat-tailed distributions (e.g., Student's $t$ with 2 degrees of freedom).

Markov's LLN

If we relax the iid assumption then we can still get convergence of the sample mean. However, further assumptions are required on the sequence of random variables $X_1, \ldots, X_n$. For example, a LLN for non iid random variables that is particularly useful is due to Markov.

**Theorem 4** *Markov's LLN*

Let $X_1, \ldots, X_n$ be a sample of uncorrelated random variables with finite means $E[X_i] = \mu_i < \infty$ and uniformly bounded variances var$(X_i) = \sigma_i^2 \leq M < \infty$ for $i = 1, \ldots, n$. Then

$$\bar{X} - \bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mu_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \xrightarrow{p} 0$$

Remarks:

1. The proof follows directly from Chebychev's inequality.

2. Sometimes the uniformly bounded variance assumption, var$(X_i) = \sigma_i^2 \leq M < \infty$ for $i = 1, \ldots, n$, is stated as

$$\sup_i \ \sigma_i^2 < \infty$$

   where sup denotes the supremum or least upper bound.

3. Notice that when the iid assumption is relaxed, stronger restrictions need to be placed on the variances of the random variables. This is a general principle with LLNs. If some assumptions are weakened then other assumptions must be strengthened. Think of this as an instance of the no free lunch principle applied to probability theory.

**LLNs for Serially Correlated Random Variables**

If we go further and relax the uncorrelated assumption, then we can still get a LLN result. However, we must control the dependence among the random variables. In particular, if $\sigma_{ij} =\text{cov}(X_i, X_j)$ exists for all $i, j$ and are close to zero for $|i-j|$ large; e.g., if

$$\sigma_{ij} \leq M \cdot \rho^{|i-j|},\ 0 < \rho < 1 \text{ and } M < \infty$$

then it can be shown that

$$\Pr(|\bar{X} - \bar{\mu}| > \varepsilon) \leq \frac{M}{n\varepsilon^2} \to 0$$

as $n \to \infty$. Further details on LLNs for serially correlated random variables will be discussed in the section on time series concepts.

**Theorem 5** *Slutsky's Theorem 1*

Let $\{Y_n\}$ and $\{Z_n\}$ be a sequences of random variables and let $b$, $c$ and $d$ be constants.

1. If $Y_n \xrightarrow{p} c$ then $bY_n \xrightarrow{p} bc$
2. If $Y_n \xrightarrow{p} c$ and $Z_n \xrightarrow{p} d$ then $Y_n + Z_n \xrightarrow{p} c + d$
3. If $Y_n \xrightarrow{p} c$ and $Z_n \xrightarrow{p} d$ then $\frac{Y_n}{Z_n} \xrightarrow{p} \frac{c}{d}$, provided $d \neq 0$; $Y_n Z_n \xrightarrow{p} cd$
4. If $Y_n \xrightarrow{p} c$ and $h(\cdot)$ is a continuous function at $c$ then $h(Y_n) \xrightarrow{p} h(c)$

**Example 1** *Convergence of the sample variance and standard deviation*

Let $X_1, \ldots, X_n$ be iid random variables with $E[X_1] = \mu$ and $\text{var}(X_1) = \sigma^2 < \infty$. Then the sample variance, given by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

is a consistent estimator for $\sigma^2$; i.e. $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

Remarks:

1. Notice that in order to prove the consistency of the sample variance we used the fact that the sample mean is consistent for the population mean. If $\bar{X}$ was not consistent for $\mu$, then $\hat{\sigma}^2$ would not be consistent for $\sigma^2$. The consisitency of $\hat{\sigma}^2$ implies that the finite sample bias of $\hat{\sigma}^2 \longrightarrow 0$ as $n \longrightarrow \infty$.

2. We may write

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\mu - \bar{X})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 + o_p(1)
\end{aligned}
$$

Here, $Y_n = -(\mu - \bar{X})^2 \xrightarrow{p} 0$ is a sequence of random variables that converge in probability to zero. In this, case we often use the notation

$$Y_n = o_p(1).$$

This short-hand notation often simplifies the exposition of certain derivations involving probability limits.

3. Given that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and the square root function is continuous it follows from Slutsky's Theorem that the sample standard deviation, $\hat{\sigma}$, is consistent for $\sigma$.

## Convergence in Distribution and the Central Limit Theorem

Let $Y_1, \ldots, Y_n$ be a sequence of random variables. For example, let $X_1, \ldots, X_n$ be an iid sample with $E[X_i] = \mu$ and $\mathrm{var}(X_i) = \sigma^2$ and define $Y_n = \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)$. We say that $Y_n$ *converges in distribution* to a random variable $W$ and write

$$Y_n \xrightarrow{d} W$$

if

$$F_{Y_n}(y) = \Pr(Y_n \leq y) \to F_W(y) = \Pr(W \leq y) \text{ as } n \to \infty$$

for every continuity point of the *cumulative distribution function* (CDF) of $W$.

Remarks

1. In most applications, $W$ is either a normal or chi-square distributed random variable.

2. Convergence in distribution is usually established through *Central Limit Theorems* (CLTs).

3. If $n$ is large, we can use the convergence in distribution results to justify using the distribution of $W$ as an approximating distribution for $Y_n$. That is, for $n$ large enough we use the approximation

$$\Pr(Y_n \in A) \approx \Pr(W \in A)$$

for any set $A \subset \mathbb{R}$.

4. Let $\mathbf{Y}_n = (Y_{n1}, \ldots, Y_{nk})'$ be a multivariate sequence of random variables. Then

$$\mathbf{Y}_n \xrightarrow{d} \mathbf{W}$$

if and only if

$$\boldsymbol{\lambda}'\mathbf{Y}_n \xrightarrow{d} \boldsymbol{\lambda}'\mathbf{W}$$

for any $\boldsymbol{\lambda} \in \mathbb{R}^k$.

**Relationship between Convergence in Distribution and Convergence in Probability**

Result 1: If $\mathbf{Y}_n$ converges in probability to the random vector $\mathbf{Y}$, then $\mathbf{Y}_n$ converges in distribution to the probability distribution of $\mathbf{Y}$

Result 2: If $\mathbf{Y}_n$ converges in distribution to a non-random vector $\mathbf{c}$, i.e., if the probability distribution of $\mathbf{Y}_n$ "converges" to the point mass distribution at $\mathbf{c}$, then $\mathbf{Y}_n$ converges in probability to $\mathbf{c}$.

**Central Limit Theorems**

The most famous CLT is due to Lindeberg and Levy.

**Theorem 6** *Lindeberg-Levy CLT (Greene, 2003 p. 909)*

Let $X_1, \ldots, X_n$ be an iid sample with $E[X_i] = \mu$ and $\text{var}(X_i) = \sigma^2 < \infty$. Then

$$Y_n = \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right) \xrightarrow{d} Z \sim N(0,1) \text{ as } n \to \infty$$

That is, for all $y \in \mathbb{R}$,

$$\Pr(Y_n \leq y) \to \Phi(y) \text{ as } n \to \infty$$

where

$$\Phi(y) = \int_{-\infty}^{y} \phi(z)dz$$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

Remarks

1. The CLT suggests that we may approximate the distribution of $Y_n = \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)$ by a standard normal distribution. This, in turn, suggests approximating the distribution of the sample average $\bar{X}$ by a normal distribution with mean $\mu$ and variance $\frac{\sigma^2}{n}$.

2. A common short-hand notation is

$$Y_n \overset{A}{\sim} N(0,1)$$

which implies (by Slutsky's theorem) that

$$\bar{X} \overset{A}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\text{avar}(\bar{X}) = \frac{\sigma^2}{n}$$

3. A consistent estimate of $\text{avar}(\bar{X})$ is

$$\widehat{\text{avar}}(\bar{X}) = \frac{\hat{\sigma}^2}{n} \text{ s.t. } \hat{\sigma}^2 \xrightarrow{p} \sigma^2$$

**Theorem 7** *Multivariate Lindeberg-Levy CLT (Greene, 2003 p. 912)*

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $k-$dimensional iid random vectors with $E[\mathbf{X}_i] = \boldsymbol{\mu}$ and var$(\mathbf{X}_i) = E[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a $k \times k$ nonsingular matrix. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2\prime}$. Then

$$\sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \xrightarrow{d} \mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_k)$$

Equivalently, we may write

$$\begin{aligned}
\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) &\overset{A}{\sim} N(0, \boldsymbol{\Sigma}) \\
\sqrt{n}\bar{\mathbf{X}} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{X}_i \overset{A}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\bar{\mathbf{X}} &\overset{A}{\sim} N(\boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})
\end{aligned}$$

This result implies that

$$\text{avar}(\bar{\mathbf{X}}) = n^{-1}\boldsymbol{\Sigma}$$

Remark

If $\mathbf{\Sigma}$ is unknown and if $\hat{\mathbf{\Sigma}} \xrightarrow{p} \mathbf{\Sigma}$ then we can consistently estimate avar($\bar{\mathbf{X}}$) using

$$\widehat{\text{avar}}(\bar{\mathbf{X}}) = n^{-1}\hat{\mathbf{\Sigma}}$$

For example, we could estimate $\mathbf{\Sigma}$ using the sample covariance

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

The Lindeberg-Levy CLT is restricted to iid random variables, which limits its usefulness. In particular, it is not applicable to the least squares estimator in the linear regression model with fixed regressors. To see this, consider the simple linear model with a single fixed regressor

$$y_i = x_i\beta + \varepsilon_i$$

where $x_i$ is fixed and $\varepsilon_i$ is iid $(0, \sigma^2)$. The least squares estimator is

$$\begin{aligned} \hat{\beta} &= \left(\sum_{i=1}^{n} x_i^2\right)^{-1} \sum_{i=1}^{n} x_i y_i \\ &= \beta + \left(\sum_{i=1}^{n} x_i^2\right)^{-1} \sum_{i=1}^{n} x_i \varepsilon_i \end{aligned}$$

Re-arranging and multiplying both sides by $\sqrt{n}$ gives

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \varepsilon_i$$

The CLT needs to be applied to the random variable

$$w_i = x_i \varepsilon_i.$$

However, even though $\varepsilon_i$ is iid, $w_i$ is not iid since $\text{var}(w_i) = x_i^2 \sigma^2$ and, thus, varies with $x_i$. Hence, the Lindeberg-Levy CLT is not appropriate.

The following Lindeberg-Feller CLT is applicable for the linear regression model with fixed regressors.

**Theorem 8** *Lindeberg-Feller CLT (Greene, 2003 p. 901)*

Let $X_1, \ldots, X_n$ be independent (but not necessarily identically distributed) random variables with $E[X_i] = \mu_i$ and $\text{var}(X_i) = \sigma_i^2 < \infty$. Define $\bar{\mu}_n = n^{-1}\sum_{i=1}^n \mu_i$ and $\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \sigma_i^2$. Suppose

$$\lim_{n\to\infty} \max_i \frac{\sigma_i^2}{n\bar{\sigma}_n^2} = 0, \quad \lim_{n\to\infty} \bar{\sigma}_n^2 = \bar{\sigma}^2 < \infty$$

Then

$$\sqrt{n}\left(\frac{\bar{X} - \bar{\mu}_n}{\bar{\sigma}_n}\right) \xrightarrow{d} Z \sim N(0,1)$$

$$\sqrt{n}\left(\bar{X} - \bar{\mu}_n\right) \xrightarrow{d} \bar{\sigma} \cdot Z \sim N(0, \bar{\sigma}^2)$$

A CLT result that is equivalent to the Lindeberg-Feller CLT but with conditions that are easier to understand and verify is due to Liapounov.

**Theorem 9** *Liapounov's CLT (Greene, 2003 p. 912)*

Let $X_1, \ldots, X_n$ be independent (but not necessarily identically distributed) random variables with $E[X_i] = \mu_i$ and var$(X_i) = \sigma_i^2 < \infty$. Suppose further that

$$E[|X_i - \mu_i|^{2+\delta}] \leq M < \infty$$

for some $\delta > 0$. If $\bar{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \sigma_i^2$ is positive and finite for all $n$ sufficiently large, then

$$\sqrt{n}\left(\frac{\bar{X} - \bar{\mu}_n}{\bar{\sigma}_n}\right) \xrightarrow{d} Z \sim N(0,1)$$

Equivalently,

$$\sqrt{n}\left(\bar{X} - \bar{\mu}_n\right) \xrightarrow{d} N(0, \bar{\sigma}^2)$$
$$\lim_{n\to\infty} \bar{\sigma}_n^2 = \bar{\sigma}^2 < \infty.$$

Remark

There is a multivariate version of the Lindeberg-Feller CLT (See Greene, 2003 p. 913) that can be used to prove that the OLS estimator in the multiple regression model with fixed regressors converges to a normal random variable. For our purposes, we will use a different multivariate CLT that is applicable for random regressors in a cross section or time series context. Details will be given in the section of time series concepts.

## Asymptotic Normality

A consistent estimator $\hat{\boldsymbol{\theta}}$ is *asymptotically normally distributed* (asymptotically normal) if

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma})$$

Equivalently, $\hat{\boldsymbol{\theta}}$ is asymptotically normal if

$$\hat{\boldsymbol{\theta}} \overset{A}{\sim} N(\boldsymbol{\theta}, n^{-1}\boldsymbol{\Sigma}),\ \text{avar}(\hat{\boldsymbol{\theta}}) = n^{-1}\boldsymbol{\Sigma}$$

Remark

If $\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}$ then

$$\hat{\boldsymbol{\theta}} \overset{A}{\sim} N(\boldsymbol{\theta}, n^{-1}\hat{\boldsymbol{\Sigma}}),\ \widehat{\text{avar}}(\hat{\boldsymbol{\theta}}) = n^{-1}\hat{\boldsymbol{\Sigma}}$$

This result is justified by an extension of Slutsky's theorem.

**Theorem 10** *Slutsky's Theorem 2 (Extension of Slutsky's Theorem to convergence in distribution)*

Let $Y_n$ and $Z_n$ be sequences of random variables such that

$$Y_n \xrightarrow{d} W, Z_n \xrightarrow{p} c$$

where $W$ is a random variable and $c$ is a constant. Then the following results hold:

1. $Z_n Y_n \xrightarrow{d} cW$

2. $Y_n / Z_n \xrightarrow{d} W/c$ provided $c \neq 0$

3. $Y_n + Z_n \xrightarrow{d} W + c$

Remark

Suppose $Y_n$ and $Z_n$ are sequences of random variables such that

$$Y_n \xrightarrow{d} W$$
$$Z_n \xrightarrow{d} Z$$

where $W$ and $Z$ are (possibly correlated) random variables. Then it is not necessarily true that

$$Y_n + Z_n \xrightarrow{d} W + Z$$

We have to worry about the dependence between $Y_n$ and $Z_n$.

**Theorem 11** *Continuous Mapping Theorem (CMT)*

Suppose $h(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is continuous everywhere and $Y_n \xrightarrow{d} W$ as $n \rightarrow \infty$. Then

$$h(Y_n) \xrightarrow{d} h(W) \text{ as } n \rightarrow \infty$$

Remark

This is one of the most useful results in statistics. In particular, we will use it to deduce the asymptotic distribution of test statistics (e.g. Wald, LM, LR, etc.)

## The Delta Method

Suppose we have an asymptotically normal estimator $\hat{\theta}$ for the scalar parameter $\theta$; i.e.,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} W \sim N(0, \sigma^2)$$

Often we are interested in some function of $\theta$, say $\eta = g(\theta)$. Suppose $g(\cdot) : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous and differentiable at $\theta$ and that $g' = \frac{dg}{d\theta}$ is continuous. Then the *delta method* result is

$$\sqrt{n}(\hat{\eta} - \eta) = \sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} W^* \sim N(0, g'(\theta)^2\sigma^2)$$

Equivalently,

$$g(\hat{\theta}) \overset{A}{\sim} N\left(g(\theta), \frac{g'(\theta)^2\sigma^2}{n}\right)$$

Remark

The asymptotic variance of $\hat{\eta} = g(\hat{\theta})$ is

$$\text{avar}(\hat{\eta}) = \text{avar}\left(g(\hat{\theta})\right) = \frac{g'(\theta)^2\sigma^2}{n}$$

which depends on $\theta$ and $\sigma^2$. Typically, $\theta$ and/or $\sigma^2$ are unknown and so $\text{avar}\left(g(\hat{\theta})\right)$ must be estimated. A consistent estimate of $\text{avar}\left(g(\hat{\theta})\right)$ has the form

$$\widehat{\text{avar}}(\hat{\eta}) = \widehat{\text{avar}}\left(g(\hat{\theta})\right) = \frac{g'(\hat{\theta})^2\hat{\sigma}^2}{n}$$

where

$$\hat{\theta} \xrightarrow{p} \theta$$
$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2$$

Proof of delta method result.

The delta method gets its name from the use of a first order Taylor series expansion. Consider an exact first order Taylor series expansion of $g(\hat{\theta})$ at $\hat{\theta} = \theta$ (i.e., apply the mean value theorem)

$$\begin{aligned} g(\hat{\theta}) &= g(\theta) + g'(\tilde{\theta})(\hat{\theta} - \theta) \\ \tilde{\theta} &= \lambda\hat{\theta} + (1 - \lambda)\theta,\ 0 \le \lambda \le 1 \end{aligned}$$

Multiplying both sides by $\sqrt{n}$ and re-arranging gives

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) = g'(\tilde{\theta})\sqrt{n}(\hat{\theta} - \theta)$$

Since $\tilde{\theta}$ is between $\hat{\theta}$ and $\theta$ and since $\hat{\theta} \xrightarrow{p} \theta$ we have that $\tilde{\theta} \xrightarrow{p} \theta$. Further since $g'$ is continuous, by Slutsky's Theorem $g'(\tilde{\theta}) \xrightarrow{p} g'(\theta)$. It follows from the convergence in distribution results that

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} g'(\theta) \cdot N(0, \sigma^2) \sim N(0, g'(\theta)^2\sigma^2)$$

Now suppose $\boldsymbol{\theta} \in \mathbb{R}^k$ and we have an asymptotically normal estimator

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$
$$\hat{\boldsymbol{\theta}} \overset{A}{\sim} N(\boldsymbol{\theta}, n^{-1}\boldsymbol{\Sigma})$$

Let $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\theta}) : \mathbb{R}^k \longrightarrow \mathbb{R}^j$; i.e.,

$$\underset{(j \times 1)}{\boldsymbol{\eta}} = \mathbf{g}(\boldsymbol{\theta}) = \begin{pmatrix} g_1(\boldsymbol{\theta}) \\ g_2(\boldsymbol{\theta}) \\ \vdots \\ g_j(\boldsymbol{\theta}) \end{pmatrix}$$

denote the parameter of interest where $\boldsymbol{\eta} \in \mathbb{R}^j$ and $j \leq k$.

Assume that $\mathrm{g}(\boldsymbol{\theta})$ is continuous with continuous first derivatives. Define the Jacobian matrix

$$\underset{(j\times k)}{\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}} = \begin{pmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_k} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_k} \end{pmatrix}$$

Then

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) &= \sqrt{n}(g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \\ &\xrightarrow{d} N\left(\mathbf{0}, \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right) \boldsymbol{\Sigma} \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)'\right)\end{aligned}$$

Remark

If $\hat{\boldsymbol{\Sigma}} \rightarrow \boldsymbol{\Sigma}$ then a practically useful result is

$$g(\hat{\boldsymbol{\theta}}) \overset{A}{\sim} N\left(g(\boldsymbol{\theta}), n^{-1} \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right) \hat{\boldsymbol{\Sigma}} \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right)'\right)$$

$$\widehat{\mathsf{avar}}(g(\hat{\boldsymbol{\theta}})) = n^{-1} \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right) \hat{\boldsymbol{\Sigma}} \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right)'$$

**Example 2** *Estimation of Generalized Learning Curve*

Consider the generalized learning curve (see Berndt, 1992, chapter 3)

$$C_t = C_1 N_t^{\alpha_c/R} Y_t^{(1-R)/R} \exp(u_t)$$

where

$$\begin{aligned}
C_t &= \text{real unit cost at time } t \\
N_t &= \text{cumulative production up to time } t \\
Y_t &= \text{production in time } t \\
&\quad u_t \sim iid\ (0, \sigma^2) \\
\alpha_c &= \text{learning curve parameter} \\
R &= \text{returns to scale parameter}
\end{aligned}$$

Intuition: Learning is proxied by cumulative production.

- If the learning curve effect is present, then as cumulative production (learning) increases real unit costs should fall.

- If production technology exhibits constant returns to scale, then real unit costs should not vary with the level of production.

- If returns to scale are increasing, then real unit costs should decline as the level of production increases.

The generalized learning curve may be converted to a linear regression model by taking logs:

$$\begin{aligned}\ln C_t &= \ln C_1 + \left(\frac{\alpha_c}{R}\right)\ln N_t + \left(\frac{1-R}{R}\right)\ln Y_t + u_t \\ &= \beta_0 + \beta_1 \ln N_t + \beta_2 \ln Y_t + u_t \\ &= \mathbf{x}_t'\boldsymbol{\beta} + u_t\end{aligned}$$

where

$$\begin{aligned}\beta_0 &= \ln C_1 \\ \beta_1 &= \alpha_c/R \\ \beta_2 &= (1-R)/R \\ \mathbf{x}_t &= (1, \ln N_t, \ln Y_t)'.\end{aligned}$$

The learning curve parameters may be recovered using

$$\begin{aligned}\alpha_c &= \frac{\beta_1}{1+\beta_2} = g_1(\boldsymbol{\beta}) \\ R &= \frac{1}{1+\beta_2} = g_2(\boldsymbol{\beta})\end{aligned}$$

Least squares gives consistent and asymptotically normal estimates

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \xrightarrow{p} \boldsymbol{\beta} \\
\hat{\sigma}^2 &= n^{-1}\sum_{t=1}^{n}(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}})^2 \xrightarrow{p} \sigma^2
\end{aligned}$$

$$\hat{\boldsymbol{\beta}} \overset{A}{\sim} N(\boldsymbol{\beta}, \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1})$$

Then from Slutsky's Theorem

$$\begin{aligned}
\hat{\alpha}_c &= \frac{\hat{\beta}_1}{1+\hat{\beta}_2} \xrightarrow{p} \frac{\beta_1}{1+\beta_2} = \alpha_c \\
\hat{R} &= \frac{1}{1+\hat{\beta}_2} \xrightarrow{p} \frac{1}{1+\beta_2} = R
\end{aligned}$$

provided $\beta_2 \neq -1$.

We can use the delta method to get the asymptotic distribution of $\hat{\eta} = (\hat{\alpha}_c, \hat{R})'$ :

$$\begin{pmatrix} \hat{\alpha}_c \\ \hat{R} \end{pmatrix} = \begin{pmatrix} g_1(\hat{\boldsymbol{\beta}}) \\ g_2(\hat{\boldsymbol{\beta}}) \end{pmatrix} \overset{A}{\sim} N\left(\mathbf{g}(\boldsymbol{\beta}), \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}'}\right) \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}'}\right)'\right)$$

where

$$\frac{\partial \mathbf{g}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \begin{pmatrix} \frac{\partial g_1(\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial g_1(\boldsymbol{\beta})}{\partial \beta_2} & \frac{\partial g_1(\boldsymbol{\beta})}{\partial \beta_3} \\ \frac{\partial g_2(\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial g_2(\boldsymbol{\beta})}{\partial \beta_2} & \frac{\partial g_2(\boldsymbol{\beta})}{\partial \beta_2} \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{1+\beta_2} & \frac{-\beta_1}{(1+\beta_2)^2} \\ 0 & 0 & \frac{-1}{(1+\beta_2)^2} \end{pmatrix}$$

Remark

Asymptotic standard errors for $\hat{\alpha}_c$ and $\hat{R}$ are given by the square root of the diagonal elements of

$$\left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}'}\right) \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1} \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}'}\right)'$$

where

$$\frac{\partial \mathbf{g}(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}'} = \begin{pmatrix} 0 & \frac{1}{1+\hat{\beta}_2} & \frac{-\hat{\beta}_1}{(1+\hat{\beta}_2)^2} \\ 0 & 0 & \frac{-1}{(1+\hat{\beta}_2)^2} \end{pmatrix}$$