Chapter 6
Asymptotic Distribution Theory

Asymptotic Distribution Theory

• *Asymptotic distribution theory* studies the hypothetical distribution -the limiting distribution- of a sequence of distributions.

• Do not confuse with *asymptotic theory* (or *large sample theory*), which studies the properties of asymptotic expansions.

• Definition Asymptotic expansion

An *asymptotic expansion* (*asymptotic series* or *Poincaré expansion*) is a formal series of functions, which has the property that truncating the series after a finite number of terms provides an approximation to a given function as the argument of the function tends towards a particular, often infinite, point.

(In asymptotic distribution theory, we do use asymptotic expansions.)
Asymptotic Distribution Theory

- In Chapter 5, we derive exact distributions of several sample statistics based on a random sample of observations.

- In many situations an exact statistical result is difficult to get. In these situations, we rely on approximate results that are based on what we know about the behavior of certain statistics in large samples.

- Example from basic statistics: What can we say about $1/\bar{X}$? We know a lot about $\bar{X}$. What do we know about its reciprocal? Maybe we can get an approximate distribution of $1/\bar{X}$ when $n$ is large.

Convergence

- Convergence of a non-random sequence.

Suppose we have a sequence of constants, indexed by $n$

$$f(n) = \frac{(n(n+1)+3)}{(2n + 3n^2 + 5)} \quad n=1, 2, 3, \ldots.$$

Ordinary limit: $\lim_{n \to \infty} \frac{(n(n+1)+3)}{(2n + 3n^2 + 5)} = 1/3$

There is nothing stochastic about the limit above. The limit will always be $1/3$.

- In econometrics, we are interested in the behavior of sequences of real-valued random scalars or vectors. In general, these sequences are averages or functions of averages. For example,

$$S_n(X; \theta) = \Sigma_i S(x_i; \theta)/n$$
Convergence

• For example,
  \[ S_n(X; \theta) = \frac{\sum_i S(x_i; \theta)}{n} \]

Since the \(X_i\)'s are RV, then different realizations of \(\{X_n\}\) can produce a different limit for \(S_n(X; \theta)\).

Now, convergence to a particular value is a random event.

• We are interested in cases where non convergence is rare (in some defined sense).

Convergence

• Classes of convergence for random sequences as \(n\) grows large:
  1. To a constant.
     Example: the sample mean converges to the population mean. (LLN is applied)
  2. To a random variable.
     Example: a \(t\) statistic with \(n-1\) degrees of freedom converges to a standard normal distribution. (CLT is applied)
**Probability Limit (plim)**

- **Definition**: Convergence in probability

Let \( \theta \) be a constant, \( \varepsilon > 0 \), and \( n \) be the index of the sequence of RV \( x_n \).

If \( \lim_{n \to \infty} \text{Prob}[|x_n - \theta| > \varepsilon] = 0 \) for any \( \varepsilon > 0 \), we say that \( x_n \) **converges in probability** to \( \theta \).

That is, the probability that the difference between \( x_n \) and \( \theta \) is larger than any \( \varepsilon > 0 \) goes to zero as \( n \) becomes bigger.

**Notation:**

\[
\begin{align*}
&\quad \ x_n \xrightarrow{p} \theta \\
&\quad \text{plim} \ x_n = \theta
\end{align*}
\]

- If \( x_n \) is an estimator (for example, the sample mean) and if \( \text{plim} \ x_n = \theta \), we say that \( x_n \) is a **consistent** estimator of \( \theta \).

Estimators can be **inconsistent**. For example, when they are consistent for something other than our parameter of interest.

---

**Probability Limit (plim)**

- **Theorem**: Convergence for sample moments.

Under certain assumptions, sample moments converge in probability to their population counterparts.

We saw this theorem before. It’s the (Weak) Law of Large Numbers (LLN). Different assumptions create different versions of the LLN.

**Note**: The LLN is very general:

\[
(1/n)\sum f(z_i) \xrightarrow{p} \text{E}[f(z)].
\]
Slutsky’s Theorem

• We would like to extend the limit theorems for sample averages to statistics, which are functions of sample averages.

• Asymptotic theory uses smoothness properties of those functions -i.e., continuity and differentiability- to approximate those functions by polynomials, usually constant or linear functions.

• The simplest of these approximation results is the continuity theorem, which states that plims share an important property of ordinary limits: the plim of a continuous function is the value of that function evaluated at the plim. That is,

If $x_n \rightarrow \theta$ and $g(x)$ is continuous at $x = \theta$, then

$$g(x_n) \rightarrow g(\theta)$$

(provided $g(\theta)$ exists.)

Let $x_n$ be a RV such that plim $x_n = \theta$. (We assume $\theta$ is a constant.) Let $g(\cdot)$ be a continuous function with continuous derivatives. $g(\cdot)$ is not a function of $n$. Then

$$\text{plim}[g(x_n)] = g[\text{plim}(x_n)] = g(\theta)$$

(provided $g[\text{plim}(x_n)]$ exists.)

This theorem is also attributed to Harald Cramer (1893–1985).

This is a very important and useful result. Many results for estimators will be derived from this theorem.

Somehow, there are many “Slutsky’s Theorems.”

Eugen E. Slutsky, Russia (1880 – 1948)
Plims and Expectations

Q: What is the difference between $E[x_n]$ and $\text{plim } x_n$?
- $E[x_n]$ reflects an average
- $\text{plim } x_n$ reflects a (probabilistic) limit of a sequence.

Slutsky’s Theorem works for plim, but not for expectations. That is,

$$\text{plim}[\bar{x}] = \mu \Rightarrow \text{plim}[1/\bar{x}] = 1/\mu$$

$$E[\bar{x}] = \mu \Rightarrow E[1/\bar{x}] = ?$$

Properties of plims

These properties are derived from Slutsky’s Theorem.
Let $x_n$ have plim $\Delta_n = \theta$ and $y_n$ have plim $\Delta_n = \psi$. Let $\epsilon$ be a constant. Then,

1) $\text{plim } \epsilon = \epsilon$.
2) $\text{plim } (x_n + y_n) = \theta + \psi$.
3) $\text{plim } (x_n \times y_n) = \theta \times \psi$. (plim $(\epsilon x_n) = \epsilon \theta$)
4) $\text{plim } (x_n / y_n) = \theta / \psi$. (provided $\psi \neq 0$)
5) $\text{plim}[g(x_n, y_n)] = g(\theta, \psi)$. (assuming it exists and $g(.)$ is continuous differentiable)
Properties of plims for Matrices

Functions of matrices are continuous functions of the elements of the matrices. Thus, we can generalize Slutsky's Theorem to matrices.

Let \( \text{plim} A_n = A \) and \( \text{plim} B_n = B \) (element by element). Then

1) \( \text{plim}(A_n^{-1}) = [\text{plim} A_n]^{-1} = A^{-1} \)

2) \( \text{plim}(A_n B_n) = \text{plim}(A_n) \text{plim}(B_n) = AB \)

Convergence in Mean (\( r \))

**Definition:** Convergence in mean \( r \)

Let \( \theta \) be a constant, and \( n \) be the index of the sequence of RV \( x_n \). If

\[
\lim_{n \to \infty} E[(x_n - \theta)^r] = 0 \text{ for any } r \geq 1,
\]

we say that \( x_n \) **converges in mean** \( r \) to \( \theta \).

The most used version is mean-squared convergence, which sets \( r = 2 \).

**Notation:**

\[
x_n \overset{r}{\longrightarrow} \theta \quad x_n \overset{\infty}{\longrightarrow} \theta \quad (\text{when } r = 2)
\]

For the case \( r = 2 \), the sample mean converges to a constant, since its variance converges to zero.

**Theorem:**

\[
x_n \overset{r}{\longrightarrow} \theta \quad \Rightarrow \quad x_n \overset{\infty}{\longrightarrow} \theta
\]
Almost Sure Convergence

- **Definition:** Almost sure convergence

Let \( \theta \) be a constant, and \( n \) be the index of the sequence of RV \( X_n \). If

\[
P\left[ \lim_{n \to \infty} X_n = \theta \right] = 1,
\]

we say that \( X_n \) converges almost surely to \( \theta \).

The probability of observing a realization of \( \{X_n\} \) that does not converge to \( \theta \) is zero. \( \{X_n\} \) may not converge everywhere to \( \theta \), but the points where it does not converge form a zero measure set (probability sense).

**Notation:** \( X_n \xrightarrow{sa} \theta \)

This is a stronger convergence than convergence in probability.

**Theorem:** \( X_n \xrightarrow{sa} \theta \Rightarrow X_n \xrightarrow{p} \theta \)

Almost Sure Convergence

- In almost sure convergence, the probability measure takes into account the joint distribution of \( \{X_n\} \). With convergence in probability we only look at the joint distribution of the elements of \( \{X_n\} \) that actually appear in \( X_n \).

- **Strong Law of Large Numbers**

We can state the LLN in terms of almost sure convergence:

Under certain assumptions, sample moments converge almost surely to their population counterparts.

This is the Strong Law of Large Numbers.

- From the previous theorem, the Strong LLN implies the (Weak) LLN.
Convergence to a Random Variable

• Definition: Limiting Distribution
Let \( x_n \) be a random sequence with cdf \( F_n(x_n) \). Let \( x \) be a random variable with cdf \( F(x) \).

When \( F_n \) converges to \( F \) as \( n \to \infty \), for all points \( x \) at which \( F(x) \) is continuous, we say that \( x_n \) converges in distribution to \( x \). The distribution of that random variable is the limiting distribution of \( x_n \).

Notation: \( x_n \xrightarrow{d} x \)

Remark: If \( \text{plim} \ x_n = \theta \) (a constant), then \( F_n(x_n) \) becomes a point.

Example: The \( t_n \) statistic converges to a standard normal: \( t_n \xrightarrow{d} \text{N}(0,1) \)

Convergence to a Random Variable

Theorem: If \( x_n \xrightarrow{d} x \) and \( \text{plim} \ y_n = c \). Then, \( x_n y_n \xrightarrow{d} cx \). That is the limiting distribution of \( x_n y_n \) is the distribution of \( cx \).

Also, \( x_n + y_n \xrightarrow{d} x + c \)
\( x_n/y_n \xrightarrow{d} x/c \) (provided \( c \neq 0 \)).

Note: This theorem may be also referred as Slutsky’s theorem.
Slutsky’s Theorem for RVs

Let $x_n$ converge in distribution to $x$ and let $g(.)$ be a continuous function with continuous derivatives. $g(.)$ is not a function of $n$. Then,

$$g(x_n) \xrightarrow{d} g(x).$$

**Example:** $t_n \xrightarrow{d} N(0,1) \Rightarrow g(t_n) = (t_n)^2 \xrightarrow{d} [N(0,1)]^2$.

• **Extension**

Let $x_n \xrightarrow{d} x$ and $g(x_n, \theta) \xrightarrow{d} g(x)$ ($\theta$ parameter).

Let $\text{plim } y_n = \theta$ ($y_n$ is a consistent estimator of $\theta$)

Then, $g(x_n, y_n) \xrightarrow{d} g(x)$.

That is, replacing $\theta$ by a consistent estimator leads to the same limiting distribution.

---

Extension of Slutsky’s Theorem: Examples

• **Example 1:** $t_n$ statistic

$$z = n^{1/2} \left( \bar{x} - \mu \right) / \sigma \xrightarrow{d} N(0,1)$$

$$t_n = n^{1/2} \left( \bar{x} - \mu \right) / s_n \xrightarrow{d} N(0,1) \text{ (where plim } s_n = \sigma)$$

• **Example 2:** F-statistic for testing restricted regressions.

$$F = \left[ (e'*e* - e'e) / J \right] / \left[ e'e / (n-k) \right] = \left[ (e'*e* - e'e) / \sigma^2 \right] / \left[ e'e / \sigma^2 (n-k) \right]$$

The denominator: $[e'e / \sigma^2 (n-k)] \xrightarrow{d} 1$.

The limiting distribution of the F statistic will be given by the limiting distribution of the numerator.
Revisiting the CLT

• The CLT states conditions for the sequence of RV \( \{x_n\} \) under which the mean or a sum of a sufficiently large number of \( x_i \)'s will be approximately normally distributed.

CLT: Under some conditions, \( z = n^{1/2} \left( \bar{X} - \mu \right) / \sigma \xrightarrow{d} N(0,1) \)

• It is a general result. When sums of random variables are involved, eventually (sometimes after transformations) the CLT can be applied.

• The Berry–Esseen theorem (Berry–Esseen inequality) attempts to quantify the rate at which the convergence to normality takes place.

\[ |F_n(x) - \Phi(x)| \leq \frac{C\rho}{\sigma^3 n^{1/2}} \]

where \( \rho = \text{E}(|X|) < \infty \) and \( C \) is a constant (best current \( C=0.7056 \)).

Revisiting the CLT

• Two popular versions used in economics and finance:

  * **Lindeberg-Levy**: \( \{x_n\} \) are i.i.d., with finite \( \mu \) and finite \( \sigma^2 \).

  * **Lindeberg-Feller**: \( \{x_n\} \) are independent, with finite \( \mu_i, \sigma_i^2 \), and for \( \varepsilon > 0 \),

    \[ \lim_{n \to \infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|x_i - \mu_i| > \varepsilon s_n} (x_i - \mu_i)^2 f(x_i) \, dx = 0 \]

  **Note:**

  * Lindeberg-Levy assumes random sampling – observations are i.i.d., with the same mean and same variance.
  * Lindeberg-Feller allows for heterogeneity in the drawing of the observations --through different variances. The cost of this more general case: More assumptions about how the \( \{x_n\} \) vary.
Order of a Sequence: Big O and Little o

• “Little o” o().
A sequence \( \{x_n\} \) is \( o(n^\delta) \) (order less than \( n^\delta \)) if \( |n^\delta x_n| \to 0 \), as \( n \to \infty \).

Example: \( x_n = n^3 \) is \( o(n^4) \) since \( |n^4 x_n| = 1/n \to 0 \), as \( n \to \infty \).

• “Big O” O().
A sequence \( \{x_n\} \) is \( O(n^\delta) \) (at most of order \( n^\delta \)) if \( |n^\delta x_n| \to \psi \), as \( n \to \infty \) \((0<\psi<\infty, \text{ constant})\).

Example: \( f(\zeta) = (6\zeta^4 - 2\zeta^3 + 5) \) is \( O(\zeta^4) \) and \( o(n^4+\delta) \) for every \( \delta > 0 \).

Special case: \( O(1) \): constant

• Order of a sequence of RV
The order of the variance gives the order of the sequence.

Example: What is the order of the sequence \( \{\bar{x}\} \)?
\[
\text{Var}[\bar{x}] = \sigma^2/n, \text{ which is } O(1/n) \text{ -or } O(n^1).
\]

Asymptotic Distribution

• An asymptotic distribution is a hypothetical distribution that is the \textit{limiting} distribution of a sequence of distributions.

We will use the asymptotic distribution as a finite sample \textit{approximation} to the true distribution of a RV when \( n \)-i.e., the sample size- is \textit{large}.

Practical question: When is \( n \) large?
**Asymptotic Distribution**

- Trick to obtain a limiting distribution: *Stabilizing transformation*

Transform $x_a$ to make sure the moments do not depend on $n$.

Steps:
Multiply the sample statistic $x_n$ by $n^a$ such that the limiting distribution of $n^a x_n$ has a finite, non-zero variance.
Then, transform $x_n$ to make sure the mean does not depend on $n$ either.

**Example:** $\bar{x}$ has a limiting variance equal to zero, since $\text{Var}(\bar{x})=\sigma^2/n$.
1) Multiply $\bar{x}$ by $n^{\alpha}$. Then, $\text{Var}(n^{\alpha} \bar{x}) = \sigma^2$.
2) Check mean of transformed variable: $E[n^{\alpha} \bar{x}] = n^{\alpha} \mu$.
The stabilizing transformation is: $n^{\alpha}(\bar{x} - \mu)$

---

**Asymptotic Distribution**

- Obtaining an asymptotic distribution from a limiting distribution

Steps:
1) Obtain the limiting distribution via a stabilizing transformation
2) Assume the limiting distribution can be used in finite samples
3) Invert the stabilizing transformation to obtain the asymptotic distribution

**Example:** $n^{\alpha}(\bar{x} - \mu)/\sigma \xrightarrow{d} N(0,1)$
Assume this limiting distribution works well for finite samples. Then,

\[
\begin{align*}
  n^{\alpha}(\bar{x} - \mu)/\sigma &\xrightarrow{d} N(0,1) \quad \text{(Note we have replaced } d \text{ for } a) \\
  n^{\alpha}(\bar{x} - \mu) &\xrightarrow{a} N(0, \sigma^2) \\
  (\bar{x} - \mu) &\xrightarrow{a} N(0, \sigma^2/n) \\
  \bar{x} &\xrightarrow{a} N(\mu, \sigma^2/n) \quad \text{(Asymptotic distribution of } \bar{x})
\end{align*}
\]
The Delta Method

- The delta method is used to obtain the asymptotic distribution of a non-linear function of a random variable (usually, an estimator). It uses a first-order Taylor series expansion and Slutsky’s theorem.

- Let $x_n$ be a RV, with $\lim_{n \to \infty} x_n = \theta$ and $\text{Var}(x_n) = \sigma < \infty$.
  We can apply the CLT to obtain $n^{1/2}(x_n - \theta)/\sigma \xrightarrow{d} \mathcal{N}(0,1)$

- Goal: $g(x_n) \xrightarrow{} ?$ ($g(x_n)$ is a continuous differentiable function, independent of $n$.)

Steps:
(1) Taylor series approximation around $\theta$:

$$g(x_n) = g(\theta) + g'( \theta ) (x_n - \theta) + \text{higher order terms}$$

We will assume the higher order terms are $o(n)$. That is, as $n$ grows the higher order terms vanish.

The Delta Method

(2) Use Slutsky theorem:

$$\lim g(x_n) = g(\theta)$$
$$\lim g'(x_n) = g'(\theta)$$

Then, as $n$ grows, using the Taylor series expansion

$$\Rightarrow \quad n^{1/2}(g(x_n) - g(\theta)) \approx g'(\theta) [n^{1/2}(x_n - \theta)].$$

If $g(.)$ does not behave badly, the asymptotic distribution of $(g(x_n) - g(\theta))$ is given by that of $[n^{1/2}(x_n - \theta)]$, which is $\mathcal{N}(0, \sigma^2)$. Then,

$$n^{1/2}((g(x_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, [g'(\theta)]^2 \sigma^2).$$

After some work (“inversion”), we obtain:

$$g(x_n) \xrightarrow{} \mathcal{N}(g(\theta), [g'(\theta)]^2 \sigma^2/n).$$
Delta Method: Example

Let \( x_n \xrightarrow{d} N(\theta, \sigma^2 / n) \)

Q: \( g(x_n) = \delta / x_n \xrightarrow{} ? \) \hspace{1cm} (\delta \text{ is a constant})

First, calculate \( g(x_n) \) & \( g'(x_n) \) and evaluate their plims:
\[
\begin{align*}
g(x_n) &= \delta / x_n \\
g'(x_n) &= -\left(\delta / x_n^2\right)
\end{align*}
\]
\[
\begin{align*}
\text{plim } g(x_n) &= \left(\delta / \theta\right) \\
\text{plim } g'(x_n) &= -\left(\delta / \theta^2\right)
\end{align*}
\]

Recall delta method formula: \( g(x_n) \xrightarrow{d} N(g(\theta), [g'(\theta)]^2 \sigma^2 / n) \).
Then,
\[
\begin{align*}
g(x_n) &\xrightarrow{d} N(\delta / \theta, (\delta^2 / \theta^4) \sigma^2 / n)
\end{align*}
\]