Chapter 6
Asymptotic Distribution Theory

© R. Susmel, 2022 (for private use, not to be posted/shared online).

Asymptotic Distribution Theory

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

• Do not confuse with a*symptotic 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

  $f(n) = ((n(n+1)+3)/(2n + 3n^2 + 5) \quad n = 1, 2, 3, ...$

  Ordinary limit:

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

  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,

$$S_n(X; \theta) = \Sigma_i\, S(x_n; \theta)/\, n$$

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 θ 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 θ.

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

Notation: $x_n \xrightarrow{p} \theta$
$\text{plim } x_n = \theta$

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 θ.

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 (for example, *i.i.d.* with *finite mean*), 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:

$$(1/n)\sum_i^n f(z_i) \xrightarrow{p} \text{E}[f(z_i)]$$

• The usual version in Greene assumes *i.i.d.* with finite mean. This is the Khinchin's (1929) (weak) LLN. (Khinchin is also spelled *Khintchine)*

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 $x_n \xrightarrow{p} \theta$ and $g(x)$ is continuous at $x = \theta$, then

$$g(x_n) \xrightarrow{p} g(\theta) \qquad \text{(provided } g(\theta) \text{ exists.)}$$

Slutsky's Theorem

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

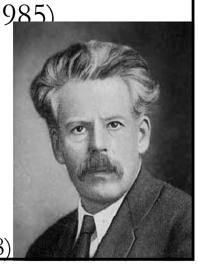
$$\text{plim}[g(x_n)] = g[\text{plim}(x_n)] = g(\theta) \qquad \text{(provided } g[\text{plim}(x_n)] \text{ exists)}$$

When $g(.)$ is continuous, this result is sometimes referred as the *continuity theorem*.

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 plim x_n?

– E[x_n] reflects an average

– plim x_n reflects a (probabilistic) limit of a sequence.

Slutsky's Theorem works for plims, but not for expectations. That is,

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

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

Properties of plims

Let x_n have plim $x_n = \theta$ and y_n have plim $y_n = \psi$. Let c be a constant. Then,

1) plim $c = c$.

2) plim $(x_n + y_n) = \theta + \psi$.

3) plim $(x_n * y_n) = \theta * \psi$. (plim $(c\, x_n) = c\,\theta$.)

4) plim $(x_n / y_n) = \theta/\psi$. (provided $\psi \neq 0$)

5) plim[$g(x_n, y_n)$] = $g(\theta,\psi)$. (assuming it exists and $g(.)$ is cont. diff.)

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 plim $\mathbf{A}_n = \mathbf{A}$ and plim $\mathbf{B}_n = \mathbf{B}$ (element by element). Then
1) $\text{plim}(\mathbf{A}_n^{-1}) = [\text{plim } \mathbf{A}_n]^{-1} = \mathbf{A}^{-1}$
2) $\text{plim}(\mathbf{A}_n\mathbf{B}_n) = \text{plim}(\mathbf{A}_n)\ \text{plim}(\mathbf{B}_n) = \mathbf{AB}$

Convergence in Mean (*r*)

Definition: Convergence in mean r
Let θ be a constant, and n be the index of the sequence of RV x_n. If

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

we say that x_n *converges in mean r* to θ.

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

Notation: $x_n \xrightarrow{p} \theta$
$x_n \xrightarrow{m.s.} \theta$ (when $r = 2$)

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

Theorem: $x_n \xrightarrow{m.s.} \theta \quad \Rightarrow x_n \xrightarrow{p} \theta$

Consistency: Brief Remarks

• *Consistency*

A consistent estimator of a population characteristic satisfies two conditions:

(1) It possesses a probability limit –its distribution collapses to a spike as the sample size becomes large, and

(2) The spike is located at the true value of the population characteristic.

• The sample mean in our example satisfies both conditions and so it is a consistent estimator of μ_X. Most estimators, in practice, satisfy the first condition, because their variances tend to zero as the sample size becomes large.

• Then, the only issue is whether the distribution collapses to a spike at the true value of the population characteristic.

20

Consistency: Brief Remarks

- A *sufficient* condition for consistency is that the estimator should be unbiased and that its variance should tend to zero as *n* becomes large.

- However the condition is only sufficient, not *necessary*. It is possible that an estimator may be biased in a finite sample, but the bias disappears as the sample size tends to infinity.

⇒ Such an estimator is biased (in finite samples), but consistent because its distribution collapses to a spike at the true value.

20

Almost Sure Convergence

• Definition: Almost sure convergence

Let θ be a constant, and n be the index of the sequence of RV x_n. If

$$P[\lim_{n\to\infty} x_n = \theta] = 1,$$

we say that x_n *converges almost surely* to θ.

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

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

This is a stronger convergence than convergence in probability.

Theorem: $x_n \xrightarrow{a.s.} \theta \quad \Rightarrow \quad x_n \xrightarrow{p} \theta$

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 plim $x_n = \theta$ (a constant), then $F_n(x_n)$ becomes a point.

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

Convergence to a Random Variable

Theorem: If $x_n \xrightarrow{d} x$ & plim $y_n = c$. Then, $x_n\, y_n \xrightarrow{d} c\, x$.

That is the limiting distribution of $x_n\, y_n$ is the distribution of $c\, x$.

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

• Extension

Let $x_n \xrightarrow{d} x$ & $g(x_n, \theta) \xrightarrow{d} g(x)$ (θ: parameter).
Let plim $y_n = \theta$ (y_n is a consistent estimator of θ)

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

That is, replacing θ by a consistent estimator leads to the same limiting distribution.

Extension of Slutsky's Theorem: Examples

Example 1: t_n statistic

$$z = \sqrt{n}\,(\bar{x} - \mu)/\sigma \xrightarrow{d} \text{N}(0, 1)$$

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

Example 2: F-statistic for testing restricted regressions.

$$F = [(\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e})/J]/[\mathbf{e}'\mathbf{e}/(T-k)]$$
$$= [(\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e})/(\sigma^2 J)]/[\mathbf{e}'\mathbf{e}/(\sigma^2(T-k))]$$

The denominator: $\mathbf{e}'\mathbf{e}/[\sigma^2(T-k)] \xrightarrow{p} 1$.

Then, 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, $z = \sqrt{n}\,(\bar{x} - \mu)/\sigma \xrightarrow{d}$ 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 $\rho = E(|X|) < \infty$ 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 μ and finite σ^2.

Lindeberg-Feller: $\{x_n\}$ are independent, with finite $\mu_i, \sigma_i^2 < \infty$,
$S_n = \sum_i^n x_i$, $s_n^2 = \sum_i^n \sigma_i^2$ and for ε > 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^δ) (*order less than* n^δ) 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^δ) (*at most of order* n^δ) if $|n^{-\delta}\, x_n| \to \psi$, as $n \to \infty$ $(0 < \psi < \infty$, constant).
Example: f(z) = $(6z^4 - 2z^3 + 5)$ is $O(z^4)$ and o($n^{4+\delta}$) for every δ>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 $\{\bar{x}\}$?
Var[$\bar{x}$] = σ^2/n, which is O($1/n$) -or O(n^{-1}).

Asymptotic Distribution

• An asymptotic distribution is a hypothetical distribution that is the *limiting* distribution of a sequence of distributions.

We will use the asymptotic distribution as a finite sample *approximation* to the true distribution of a RV when n -i.e., the sample size- is *large*.

Practical question: When is n large?

Asymptotic Distribution

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

Transform x_n 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 Var[$\bar{x}$] = σ^2/n

1) Multiply $\bar{x}$ by $\sqrt{n}$. Then, Var[$\sqrt{n}\,\bar{x}$] = σ^2.

2) Check mean of transformed variable: E[$\sqrt{n}\,\bar{x}$] = $\sqrt{n}\,\mu$.

The stabilizing transformation is: $\sqrt{n}\,(\bar{x} - \mu)$

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 get asymptotic distribution

Example: $\sqrt{n}\,(\bar{x}-\mu)/\sigma \xrightarrow{d}$ N(0,1)

Assume this limiting distribution works well for finite samples. Then,

$$\sqrt{n}\,(\bar{x}-\mu)/\sigma \xrightarrow{d} \text{N}(0,1)$$

$$\sqrt{n}\,(\bar{x}-\mu) \xrightarrow{a} \text{N}(0,\sigma^2) \qquad \text{(Note we have replaced } d \text{ for } a\text{)}$$

$$(\bar{x}-\mu) \xrightarrow{a} \text{N}(0,\sigma^2/n)$$

$$\bar{x} \xrightarrow{a} \text{N}(\mu,\sigma^2/n) \qquad \text{(Asymptotic distribution of } \bar{x}\text{)}$$

The Delta Method

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

- **Univariate case**

Let x_n be a RV, with plim $x_n = \theta$ and Var$(x_n) = \sigma^2 < \infty$.

We can apply the CLT to obtain $\sqrt{n}\,(x_n - \mu)/\sigma \xrightarrow{d}$ N(0, 1).

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

Steps:

(1) Taylor series approximation around θ:

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

We will assume the higher order terms are o(n).

The Delta Method

Remark: o(n): as n grows the higher order terms vanish.

(2) Use Slutsky theorem: plim $g(x_n) = g(\theta)$
plim $g'(x_n) = g'(\theta)$

Then, as n grows, $g(x_n) \approx g(\theta) + g'(\theta)\,(x_n - \theta)$

$$\Rightarrow \quad \sqrt{n}\,[g(x_n) - g(\theta)]) \approx g'(\theta)\,[\sqrt{n}(x_n - \theta)]$$

$$\Rightarrow \quad \sqrt{n}\,([g(x_n) - g(\theta)]\,/\sigma) \approx g'(\theta)\,[\sqrt{n}(x_n - \theta)/\sigma]$$

If $g(.)$ does not behave badly, the asymptotic distribution of $(g(x_n) - g(\theta))$ is given by that of $[\sqrt{n}(x_n - \theta)/\sigma]$, which is a standard normal. For the approximation to work well, we want σ to be "small."

Then,

$$\sqrt{n}\,[g(x_n) - g(\theta)] \xrightarrow{a} N(0, [g'(\theta)]^2\,\sigma^2).$$

Delta Method: Example

Then,

$$\sqrt{n}\,[g(x_n) - g(\theta)] \xrightarrow{a} N(0, [g'(\theta)]^2\,\sigma^2).$$

After some work ("inversion"), we obtain:

$$g(x_n) \xrightarrow{a} N(g(\theta), [g'(\theta)]^2\,\sigma^2).$$

• If we want to test H_0: $g(\theta) = g_0$, we can do a Wald test:

$$W = [g(x_n) - g_0]^2/[[g(x_n)]^2\,s^2/n] \xrightarrow{a} \chi^2_1$$