14.383 Econometrics II
Handout[1] 1.2: Asymptotic Theory

TA: Matthew C. Harding

mharding@mit.edu

In this handout I present a very brief summary of asymptotic theory, including the main theorems and results used when deriving asymptotic distributions in most of the topics we have covered so far[2].

1   Modes of Convergence

When we think about convergence, we usually have in mind a sequence that converges to a limit X, i.e. a sequence X_n that after some $n > N$, stays in some neighborhood of X. When thinking of convergence of random variables we talk about convergence of a sequence of functions. However, the usual notions of convergence for a sequence of functions are not very useful in this case. In probability theory there are four different ways to measure convergence:

- *Almost-Sure Convergence*: Probabilistic version of pointwise convergence. We only require that the set on which $X_n(\omega)$ converges has probability 1. The notation is the following

$$P(\omega \in \Omega : X_n(\omega) \to X(\omega)) = 1 \tag{1}$$

  or also written as

$$P(\lim_n X_n = X) = 1 \tag{2}$$

  or $X_n \to^{a.s.} X$.

- *Convergence in Probability*: a sequence X_n converges in probability to X if $\forall \epsilon > 0$ and $\eta > 0$ $\exists$ an $N(\epsilon, \eta)$ such that $P(|X_n - X| \geq \epsilon) < \eta \ \forall \ n > N(\epsilon, \eta)$. Equivalently one can write

$$\lim_n P(|X_n - X| > \epsilon) = 0 \qquad \forall \epsilon > 0 \tag{3}$$

  which is also written as $X_n \to^p X$.

- *Convergence in r^{th} Mean*: If $E|X_n|^r < \infty$ for all n and

$$E(|X_n - X|^r) \to 0 \text{ as } n \to \infty \tag{4}$$

  then $X_n \to^r X$.

- *Convergence in distribution*: I think the easiest way to define this concept is using the following condition. X_n converges in distribution to X if

$$F_n(x) = P(X_n \leq x) \to P(X \leq x) = F(x) \tag{5}$$

  for all points at which $F(x) = P(X \leq x)$ is continuous. The usual notation is $X_n \to^d X$.

Lemma 1 *If $X_n \to^p X$ then $X_n \to^d X$ but the converse does not hold in general. If $X_n \to^d c$ where c is a constant then implies $X_n \to^p c$.*

Lemma 2 *If $r > s \geq 1$ and $X_n \to^r X$ then $X_n \to^s X$. In addition, if $X_n \to^r X$ then $X_n \to^p X$ but the converse is false in general.*

---

[1] Thanks to Jose Tessada.

[2] I will not mention anything about the stochastic orders of magnitude, if you feel like reading something about them, check the handout on asymptotic theory we covered in 14.381.

2   Some Asymptotic Theory

- *Continuous Mapping Theorem*: (CMT) If $P(X \in \mathcal{C}) = 1$, $g(x)$ is continuous on $\mathcal{C}$, and $X_n \to^d X$ then $g(X_n) \to^d g(X)$.

  There also exists a convergence in probability version of the CMT.

- *Slutzky Theorem*: If $X_n \to^d X$ and $Y_n \to^p c$, where c is a constant, then $Y_n X_n \to^d cX$ and $Y_n + X_n \to^d c + X$.

  There is also a version of this for the case when both sequences converge in probability. This is *Slutzky Theorem* for convergence in probability, and comes from the fact that you can apply the CMT to any function of X_n and Y_n. This is true because of the following property holds for convergence in probability:

  $$X_n \to^p X \text{ and } Y_n \to^p Y \Leftrightarrow \left[\begin{array}{c} X_n \\ Yn \end{array} \right] \to^p \left[\begin{array}{c} X \\ Y \end{array} \right]$$

  Be careful because this property doesn't hold for convergence in distributions, i.e. it is not enough to look at the marginals to prove joint convergence.

  This result follows from the CMT and the fact that $X_n \to^d X$ and $Y_n \to^p c \Rightarrow (X_n, Y_n) \to^d (X, c)$.

2.1   Laws of Large Numbers

Basically, a Law of Large Numbers (LLN) states the conditions for a sample and population averages to be close to each other, i.e. when the sample average plims to the population average we are trying to approximate[3].

- *Khintchine's LLN*: If Y_i are iid and $E[\|Y_i\|] < \infty$ then $\overline{Y} \to^p E[Y_i]$.

- *Chebyshev's LLN*: If $Var(\overline{Y}) \to 0$ then $\overline{Y} - E\left[\overline{Y}\right] \to^p 0$.

Notice that the main difference between both LLN is that Chebyshev's does not require the data to be iid, but has less primitive conditions. In general, Khintchine's is enough for many econometric problems unless you really think you might have some kind of dependence in the observations[4].

2.2   Central Limit Theorems

The Central Limit Theorems (CLT) gives the conditions for sample averages to have an asymptotic normal distribution. In general, if you have a sequence of random vectors $Y_1, Y_2, \ldots$ a CLT gives you the conditions for

$$\sqrt{n}(\overline{Y} - E\left[\overline{Y}\right]) \to^d N(0, \lim_{n \to \infty} \left[n Var(\overline{Y})\right])$$

Beware of the fact that the previous statement requires the existence of the limit for this statement to be valid. This can be relaxed so to be able to apply CLT to certain cases when this is not true.

- *Lindberg-Levy CLT*: if Y_i are iid and $E\left[\|Y_i\|^2\right], \infty$ then

$$\sqrt{n}(\overline{Y} - E\left[\overline{Y}\right]) \to^d N(0, Var(Y_i))$$

This is the basic CLT for iid data, and should be sufficient for many cross-section or panel data applications. However, when the regressors have non-stochastic trends and we want to show that the t- and F-statistics are valid asymptotically we need to weaken the iid assumption and we also need to allow for $n Var(\overline{Y})$ not to converge. Let $1(\mathcal{A})$ be the indicator function for the event $\mathcal{A}$. Let $Y_{1n}, \ldots, Y_{nn}$ be scalar random variables where at least one is not constant, and $W_{in} = Var(\overline{Y})^{-\frac{1}{2}} \dfrac{(Y_{in} - E\left[Y_{in}\right])}{n}$.

---

[3] In this section I follow a handout prepared by W. Newey that Victor handed out in 14.382.

[4] I am not sure of this, but probably you might worry about what LLN you use when you are doing some asymptotic analysis of a panel estimator. Of course, in time series analysis it is a suicide to use Khintchine's LLN.

- *Lindberg-Feller CLT*: If for every n, $Y_{1n}, \ldots, Y_{nn}$ are independent, and for every $\delta > 0$,
$$\sum_{i=1}^{n} E\left[1(|W_{in}| > \delta)W_{in}^2\right] \to 0, \text{ then}$$

$$Var(\overline{Y})^{-\frac{1}{2}}(\overline{Y} - E\left[\overline{Y}\right]) \to^d N(0,1).$$

If we want to use the Lindberg-Feller CLT with random vectors we can make use of the following:

- *Cramer-Wold Device*: If $c'Y_n \to^d c'Y$ for all c with $\|c\| = 1$ then $Y_n \to^d Y$.

So according to this, to prove joint convergence we just need to prove that every linear combination of the random vector Y_n converges. In this case, to apply the Lindberg-Feller CLT to a vector, need to prove first that in fact jointly converges, and as it is explained before proving marginal convergence of each element is not sufficient.

Finally, notice that independence of the observations is present in all of the primitive conditions for the CLT and LLN stated here, that means that we cannot apply them to any example where we know this does not hold. It can be relaxed but applying some limits to the degree of dependence; a very "informal" way to explain it is to say that "too much correlation (=dependence) makes a LLN fail".