Handout # 6
14.381 Statistics Fall, 2004

TA: José Tessada (tessada@mit.edu)

This is a brief summary of asymptotic theory, including the main theorems and results that will be used later in this class and in the following classes in the econometrics sequence. There are other results and formulas, but I will only present the more relevant ones for us.[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:

Definition 1 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

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 \tag{2}$$

or $X_n \xrightarrow{a.s.} X$.

Definition 2 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*

$$\lim_{n\to\infty} (P|X_n - X| > \epsilon) = 0 \qquad \forall \epsilon > 0 \tag{3}$$

which is also written as $X_n \xrightarrow{p} X$.

Definition 3 Convergence in rth 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 $X_n \xrightarrow{r} X$.

Definition 4 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 $X_n \xrightarrow{d} X$.

Lemma 5 *If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{d} X$ but the converse does not hold in general. If $X_n \xrightarrow{d} c$ where c is a constant then implies $X_n \xrightarrow{p} c$.*

Lemma 6 *If $r > s \geq 1$ and $X_n \xrightarrow{r} X$ then $X_n \xrightarrow{s} X$. In addition, if $X_n \xrightarrow{r} X$ then $X_n \xrightarrow{p} X$ but the converse is false in general.*

[1] Based on notes by Victor Chernozhukov, Guido Kuersteiner, and Whitney Newey.

Some Asymptotic Theory

Theorem 7 Continuous Mapping Theorem (CMT) *If* $P(X \in \mathcal{C}) = 1$, $g(x)$ *is continuous on* $\mathcal{C}$, *and* $X_n \xrightarrow{d} X$ *then* $g(X_n) \xrightarrow{d} g(X)$.

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

Theorem 8 Slutzky's Theorem *If* $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} c$, *where* c *is a constant, then* $Y_n X_n \xrightarrow{d} cX$ *and* $Y_n + X_n \xrightarrow{d} c + X$.

The previous theorem follows from the CMT and the fact that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c \Rightarrow (X_n, Y_n) \xrightarrow{d} (X, c)$.

There is also a version of this for the case when both sequences converge in probability. This is *Slutzky's 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:

$$X_n \xrightarrow{p} X \text{ and } Y_n \xrightarrow{p} Y \Leftrightarrow \begin{bmatrix} X_n \\ Yn \end{bmatrix} \xrightarrow{p} \begin{bmatrix} X \\ Y \end{bmatrix}$$

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

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[2].

Theorem 9 Khintchine's LLN *If* Y_i *are iid and* $E[\|Y_i\|] < \infty$ *then* $\overline{Y} \xrightarrow{p} E[Y_i]$.

Theorem 10 Chebyshev's LLN *If* $Var(\overline{Y}) \to 0$ *then* $\overline{Y} - E\left[\overline{Y}\right] \xrightarrow{p} 0$.

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 have some kind of dependence in the observations.

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

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

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.

Theorem 11 Lindeberg-Levy CLT *If* Y_i *are iid and* $E\left[\|Y_i\|^2\right] < \infty$ *then*

$$\sqrt{n}(\overline{Y} - E\left[\overline{Y}\right]) \xrightarrow{d} N(0, Var(Y_i)) \quad (6)$$

This is the basic CLT for iid data, and should be sufficient for many cross-section or panel data applications. However, when the regressors are not iid and we want to show that the t- and F-statistics are valid asymptotically we need to weaken some of the assumptions.

Theorem 12 Lindeberg-Feller CLT *Let* $S_1 = X_1 + \ldots + X_n$ *where* X_k *are independent with* $E(X_k) = \mu_k$ *and* $Var(X_k) = \sigma_k^2$. *Define* $c_n^2 = Var(S_n) = \sum_{k=1}^n \sigma_k^2$. *If for every* $\varepsilon > 0$

$$\frac{1}{c_n^2} \sum_{k=1}^n E\left[(X_k - \mu_k)^2 \; 1\{|X_k - \mu_k| > \varepsilon c_n\}\right] \to 0 \text{ as } n \to \infty \quad (7)$$

[2] In this section I follow a handout prepared by W. Newey that Victor handed out in 14.382, Spring 2003.

then

$$\frac{S_n - E(S_n)}{c_n} \xrightarrow{d} N(0,1) \tag{8}$$

Condition (7) is called the Lindeberg condition. It puts restriction on how influential the tails can be when computing the "total" variance; it states that their weight has to go to 0 as the sample size grows to inifnity.

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

Theorem 13 Cramer-Wold Device*: If* $c'Y_n \xrightarrow{d} c'Y$ *for all* c *with* $\|c\| = 1$ *then* $Y_n \xrightarrow{d} Y$*. Where* c *is a* $(q \times 1)$ *vector,* q *being the dimension of* 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 *Lindeberg-Feller CLT* to a vector, need to prove first that in fact jointly converges, and as it was 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 still need 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”.

A mapping to the book's therminology

Notice that the book states two different Central Limit Theorem (the "Central Limit Theorem" and the "Stronger form of the Central Limit Theorem"), both of them refer to *independent and identically distributed* (*iid*) data. The "stronger form" corresponds to the *Lindeberg-Levy CLT* while the "central limit theorem" is a less powerful version of the *Lindeberg-Levy CLT* because it requires existences of all the moments (or existence of the moment generating function). The version we have here puts conditions on the first two moments.