DISCLAIMERS:

1. There is absolutely no expectation for you to read these notes prior to math camp. Maximize utility as you see fit.

2. This is intended to provide a brief refresher on some basic concepts and preview some material that will be covered in the first year econometrics sequence. If some of the material is unfamiliar, do not worry.

3. These notes contain more content than we will have time to cover during math camp. This is intentional. Hopefully these notes can be a reference material for you throughout the year.

Contents

Types of Convergence 2
$O_p$ and $o_p$ Notation 6
The Law of Large Numbers and Central Limit Theorem 7
Stationarity and Martingales 9
The Delta Method 10
The world is very complex and we often do not want to make strict parametric assumptions in our econometric models.\(^2\) Can we still say something about the behavior of our estimators without these strict assumptions? It turns out that we can in large samples.

We ask the question: How would my estimator behave in very large samples?\(^3\) We then use the limiting behavior of our estimator in infinitely large samples to approximate its behavior in finite samples.

Of course this approach has its advantages and disadvantages. As the sample size gets infinitely large, the behavior of most estimators becomes very simple. In most cases, we can apply some version of the central limit theorem and so, our estimator behaves as if its sampling distribution were normal in large samples. However, this is only an approximation for the true, finite-sample distribution of the estimator and so, this approximation can be really bad.

In this note, we will summarize the basic tools necessary for asymptotic statistics. A large portion of econometrics revolves around deriving these asymptotic approximations, finding out when these approximations are poor and what to do about it.

**Types of Convergence**

Recall the definition of convergence for a non-stochastic sequence of real numbers. Let \( \{x_n\} \) be a sequence of real numbers. We say

\[
\lim_{n \to \infty} x_n = x
\]

if for all \( \epsilon > 0 \), there exists some \( N \) such that for all \( n > N \), \( |x_n - x| < \epsilon \). We want to generalize this to the convergence of random variables. That is, under what conditions does the sequence of random variables \( \{X_n\} \) “converge” to another random variable \( X \)? There are several notions of stochastic convergence.\(^4\)

**Definition 0.1.** The sequence of random variables \( \{X_n\} \) converges to the random variable \( X \) almost surely if

\[
P(\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}) = 1.
\]

We write

\[
X_n \xrightarrow{a.s.} X.
\]

**Remark 0.1.** What does almost sure convergence mean? For a given outcome \( \omega \) in the sample space \( \Omega \), we can ask whether

\[
\lim_{n \to \infty} X_n(\omega) = X(\omega)
\]

holds using the definition of non-stochastic convergence. If the set of outcomes for which this holds has probability one then \( X_n \xrightarrow{a.s.} X \).

\(^2\) For instance, it’s often not realistic in empirical applications to assumption that the error term in a linear regression is normally distributed.

\(^3\) As the sample size \( n \) goes to infinity.

\(^4\) All of these random variables are defined on the same sample space.
Definition 0.2. The sequence of random variables \( \{X_n\} \) **converges to the random variable** \( X \) **in probability** if for all \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P(|X_n - X| > \epsilon) \to 0.
\]

We write

\[ X_n \overset{p}{\to} X. \]

Remark 0.2. What does convergence in probability mean? Fix an \( \epsilon > 0 \) and compute

\[ P_n(\epsilon) = P(|X_n - X| > \epsilon). \]

This is just a number and so, we can check whether \( P_n(\epsilon) \to 0 \) using the definition of non-stochastic convergence. If \( P_n(\epsilon) \to 0 \) for all values \( \epsilon > 0 \), then \( X_n \overset{p}{\to} X \).

Definition 0.3. The sequence of random variables \( \{X_n\} \) **converges in mean to the random variable** \( X \) if

\[
\lim_{n \to \infty} E[|X_n - X|] = 0.
\]

We write

\[ X_n \overset{m}{\to} X. \]

\( \{X_n\} \) **converges in mean-square to** \( X \) if

\[
\lim_{n \to \infty} E[(X_n - X)^2] = 0.
\]

We write

\[ X_n \overset{m.s.}{\to} X. \]

Remark 0.3. \( m_n = E[|X_n - X|] \) is just a number. \( X_n \overset{m}{\to} X \) if and only if \( m_n \to 0 \) using the definition of non-stochastic convergence. Similarly, \( m_{n.s.} = E[(X_n - X)^2] \) is also just a number and we can think about mean-square convergence in the same way.

Definition 0.4. Let \( \{X_n\} \) be a sequence of random variables and \( F_n(\cdot) \) is the cdf of \( X_n \). Let \( X \) be a random variable with cdf \( F(\cdot) \). \( \{X_n\} \) **converges in distribution, weakly converges or converges in law** to \( X \) if

\[
\lim_{n \to \infty} F_n(x) = F(x)
\]

for all points \( x \) at which \( F(\cdot) \) is continuous. There are many ways of writing this

\[ X_n \overset{d}{\to} X, \quad X_n \overset{L}{\to} X, \quad X_n \overset{\mu}{\to} X. \]

We’ll use \( X_n \overset{d}{\to} X \).
Remark 0.4. Convergence in distribution describing the convergence of the cdfs. It does not mean that the realizations of the random variables will be close to each other. Recall that \( F(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) \). As a result, \( F_n(x) \to F(x) \) does not make any statement about \( X_n(\omega) \) getting close to \( X(\omega) \) for any \( \omega \in \Omega \).

Remark 0.5. Why is convergence in distribution restricted to the continuity points of \( F(x) \)? An example may help.

Let \( X_n \) be a degenerate random variable defined by \( X_n = 1/n \) with probability 1 and let \( X \) be a degenerate random variable defined by \( X = 0 \) with probability one. Then, \( F_n(x) = 1(x \geq 1/n) \) and \( F(x) = 1(x \geq 0) \) with \( F_n(0) = 0 \) for all \( n \) while \( F(0) = 1 \).

However, as \( n \to \infty \), \( X_n \) is getting closer and closer to \( X \) in the sense that for all \( x \neq 0 \), \( F_n(x) \) is well approximated by \( F(x) \). Alternatively, if we did not restrict convergence in distribution to the continuity points, we would have the strange case where a non-stochastic sequence \( \{X_n\} \) converges to \( X \) under the non-stochastic definition of convergence but not converge in distribution.

We can extend each of these definitions to random vectors. For example, the sequence of random vectors \( \{X_n\} \overset{a.s.}{\to} X \) if each element of \( X_n \) converges almost surely to each element of \( X \). The extension is analogous for convergence in probability. A sequence of random vectors converges into distribution to a random vector if we apply the definition above to the joint cumulative distribution function. Alternatively, the following theorem provides another characterization of multivariate convergence in distribution.

Theorem 0.1. Cramer-Wold Device

Let \( \{Z_n\} \) be a sequence of \( k \)-dimensional random vectors. Then, \( Z_n \overset{d}{\to} Z \) if and only if \( \lambda'Z_n \overset{d}{\to} \lambda'Z \) for all \( \lambda \in \mathbb{R}^k \).

How do these different definitions of stochastic convergence relate to each other? The next set of propositions lay out the relationships.\(^5\)

Proposition 0.1. Convergence in mean-square implies convergence in mean

Suppose \( X_n \overset{ms}{\to} X \). Then, \( X_n \overset{m}{\to} X \).

Proof. This follows from Jensen’s inequality. Recall that if \( h(\cdot) \) is a convex function, then

\[
E[h(Y)] \geq h(E[Y]).
\]

Set \( h(z) = z^2 \) and \( Y = |X_n - X| \). It follows that

\[
0 \leq E[|X_n - X|^2] \leq E[|X_n - X|^2].
\]

with \( E[|X_n - X|^2] \to 0 \). The result follows. \( \square \)
Proposition 0.2. Convergence in mean-square implies convergence in probability
Suppose $X_n \overset{ms}{\to} X$. Then, $X_n \overset{p}{\to} X$.

Proof. This follows from Markov's inequality. Recall that for all $c > 0$,
$$P(Y \geq c) \leq E[Y] / c.$$
Fix $\epsilon > 0$. Set $c = \epsilon^2$ and $Y = |X_n - X|^2$. We have that
$$0 \leq P(|X_n - X|^2 \geq \epsilon^2) = P(|X_n - X| \geq \epsilon) \leq E[|X_n - X|^2] / \epsilon^2.$$
Taking limits of both sides, we get that
$$0 \leq \lim_{n \to 0} P(|X_n - X| \geq \epsilon) = \lim_{n \to \infty} E[|X_n - X|^2] / \epsilon^2.$$
and the result follows. $\square$

Proposition 0.3. Convergence in mean implies convergence probability
Suppose $X_n \overset{m}{\to} X$. Then, $X_n \overset{p}{\to} X$.

Proof. This follows from Markov’s inequality analogously. $\square$

Proposition 0.4. Almost sure convergence implies convergence in probability
Suppose $X_n \overset{a.s}{\to} X$. Then, $X_n \overset{p}{\to} X$.

Remark 0.6. To prove convergence in probability, it is often easiest to prove convergence in mean-square. How do you show convergence in mean-square? Note that
$$E[|X_n - X|^2] = \text{Var}(X_n - X) + (E[X_n] - E[X])^2.$$
Therefore, $X_n \overset{ms}{\to} X$ if $\text{Var}(X_n - X) \to 0$ and $E[X_n] - E[X] \to 0$.

Proposition 0.5. Convergence in probability implies convergence in distribution
Suppose $X_n \overset{p}{\to} X$. Then, $X_n \overset{d}{\to} X$.

Exercise 0.1. Let $Y \sim N(0, 1)$ and $Y_n = (-1)^n Y$. Does $Y_n \overset{d}{\to} Y$? Does $Y_n \overset{p}{\to} Y$?

We conclude this section with two theorems that are very useful in deriving asymptotic distributions.

Theorem 0.2. Slutsky’s theorem
Let $c$ be a constant. Suppose that $X_n \overset{d}{\to} X$ and $Y_n \overset{p}{\to} Y$. Then,
1. $X_n + Y_n \overset{d}{\to} X + c$.
2. $X_n Y_n \overset{d}{\to} Xc$. 

3. \( X_n/Y_n \overset{d}{\rightarrow} X/c \) provided that \( c \neq 0 \).

If \( c = 0 \), then \( X_nY_n \overset{p}{\rightarrow} 0 \).

**Theorem 0.3.** Continuous mapping theorem

Let \( g \) be a continuous function. Then,

1. If \( X_n \overset{d}{\rightarrow} X \), then \( g(X_n) \overset{d}{\rightarrow} g(X) \).
2. If \( X_n \overset{p}{\rightarrow} X \), then \( g(X_n) \overset{p}{\rightarrow} X \).

**Proof.** We provide the proof for (2) and the case where \( X = a \in \mathbb{R} \).

Let \( \epsilon > 0 \). Since \( g(\cdot) \) is continuous at \( a \), there exists some \( \delta > 0 \) such that

\[
|x - a| < \delta \implies |g(x) - g(a)| < \epsilon.
\]

The contrapositive of this is

\[
|g(x) - g(a)| \geq \epsilon \implies |x - a| \geq \delta.
\]

Substituting in \( X_n \), it follows that

\[
P(|g(X_n) - g(a)| \geq \epsilon) \leq P(|X_n - a| \geq \delta) \rightarrow 0.
\]

The result follows. \( \square \)

**\( O_p \) and \( o_p \) Notation**

Recall big-\( O \) and little-\( o \) notation for sequences of real numbers. Let \( \{a_n\} \) and \( \{g_n\} \) be sequences of real numbers. We have that

\[
a_n = o(g_n) \quad \text{if} \quad \lim_{n \to \infty} \frac{a_n}{g_n} = 0
\]

and

\[
a_n = O(g_n) \quad \text{if} \quad \frac{a_n}{g_n} < M \quad \forall n.
\]

Just like we extended the definition of non-stochastic convergence to sequences of random variable, we also extend big-\( O \) and little-\( o \) notation.

**Definition 0.5.** Suppose \( \{A_n\} \) is a sequence of random variables. We write

\[
A_n = o_p(G_n) \quad \text{if} \quad \frac{A_n}{G_n} \overset{p}{\rightarrow} 0
\]

and

\[
A_n = O_p(G_n)
\]

if for all \( \epsilon > 0 \), there exists \( M \in \mathbb{R} \) such that \( P(\frac{|A_n|}{G_n} < M) = 1 - \epsilon \) for all \( n \).

**Remark 0.7.** You’ll often see someone write \( X_n = X + o_p(1) \) to denote \( X_n \overset{p}{\rightarrow} X \).
Proposition 0.6. If $X_n \xrightarrow{d} X$, then $X_n = O_p(1)$.

Proof. Since $X$ is a random variable, there exists some $M > 0$ such that $F_X$ is continuous at $-M, M$ and $P(|X| > M) = F_X(-M) + (1 - F_X(M)) < \varepsilon/2$. Since $X_n \xrightarrow{d} X$, for all $n$ large enough

$$|F_{X_n}(-M) - F_X(-M)| < \varepsilon/4, \quad |F_{X_n}(M) - F_X(M)| < \varepsilon/4.$$

And so, for all $n$ large enough, we have that $P(|X_n| > M) < \varepsilon$. □

Example 0.1. Let $X_n \sim N(0, n)$. Then,

$$X_n = O_p(n^{1/2}).$$

Why? We have that $X_n/n^{1/2} \sim N(0, 1)$ for all $n$. For any $\epsilon > 0$, we can choose an $M$ such that $P(|N(0, 1)| < M) > 1 - \epsilon$. We also have that

$$X_n = O_p(n).$$

Why? We have that $X_n/n \sim N(0, 1/n)$. Note that

$$P(|N(0, 1/n)| > \epsilon) = P(|N(0, 1)| > n^{1/2}\epsilon) \to 0.$$

Alternatively, note that

$$E[(X_n/n - 0)^2] = V(X_n/n) = 1/n \to 0$$

and so, $X_n \xrightarrow{ms} 0$.

The Law of Large Numbers and Central Limit Theorem

There are two basic building blocks that we use to construct all asymptotic results. The first set of building blocks are laws of large numbers (LLNs). These show that sample averages converge to expectations under certain conditions. The second set of building blocks are central limit theorems (CLTs). These show that properly centered sample averages will converge in distribution to normal random variables. In this section, we provide several LLNs and CLTs that appear regularly.

Theorem 0.4. Weak law of large numbers

Let $X_1, \ldots, X_n$ be a sequence of random variables with $E[X_i] = \mu, V(X_i) = \sigma^2 < \infty$ for all $i$ and $\text{Cov}(X_i, X_j) = 0$ for all $i \neq j$. Then,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} \mu.$$
Proof. By Chebyshev’s inequality,
\[ P(|\bar{X}_n - \mu| > \epsilon^2) \leq \frac{E[(\bar{X}_n - \mu)^2]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0. \]

Alternatively,
\[ V(\bar{X}_n) = E[(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n} \to 0, \quad E[\bar{X}_n] = \mu \]
and so, \( \bar{X}_n \overset{ms}{\to} \mu \) and the result follows. \( \square \)

**Theorem 0.5.** Chebyshev’s weak law of large numbers

Let \( X_1, X_2, \ldots \) be a sequence of random variables with \( E[X_i] = \mu_i, V(X_i) = \sigma_i^2 \) and \( \text{Cov}(X_i, X_j) = 0 \) for all \( i \neq j \). Define
\[ \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \bar{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \mu_i, \quad \bar{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \]
and assume that \( \bar{\sigma}_n^2 / n \to 0 \). Then,
\[ \bar{X}_n - \bar{\mu}_n \overset{p}{\to} 0. \]

Proof. First, we have that
\[ E[\bar{X}_n - \bar{\mu}_n] = 0. \]

Second, we have that
\[ V(\bar{X}_n - \bar{\mu}_n) = V(\bar{X}_n) = \frac{1}{n^2} \sum_{i,j} \text{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_i \sigma_i^2 = \frac{\bar{\sigma}_n^2}{n} \to 0. \]

Therefore, \( \bar{X}_n - \bar{\mu}_n \overset{ms}{\to} 0 \) and so, \( \bar{X}_n - \bar{\mu}_n \overset{p}{\to} 0. \) \( \square \)

**Theorem 0.6.** Strong law of large numbers

If \( X_1, X_2, \ldots \) are i.i.d with \( E[X_i] = \mu < \infty \), then
\[ \bar{X}_n \overset{a.s.}{\to} \mu. \]

**Remark 0.8.** Note that for the weak law of large numbers, we only required the sequence of \( X_i \)'s to be uncorrelated and also required finite second moments. For the strong law of large numbers, we required the \( X_i \)'s to be i.i.d. but did not require any assumptions about second moments.

**Theorem 0.7.** Central limit theorem I

Let \( Y_1, Y_2, \ldots \) be a sequence of random variables with \( E[Y_i] = 0, V(Y_i) = 1 \) for all \( i \) and that each MGF, \( \mu_{Y_i}(t) \), exists for \( t \in (-h, h) \) for some \( h > 0 \). Then,
\[ \sqrt{n}\bar{Y} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \overset{d}{\to} N(0,1). \]
**Theorem 0.8. Central limit theorem II**

Let \( X_1, X_2, \ldots \) be a sequence of i.i.d random variables with mean \( \mu \) and variance \( \sigma^2 \). Suppose each MGF, \( \mu_{X_i}(t) \), exists for \( t \in (-h, h) \) for some \( h > 0 \). Then,

\[
\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).
\]

This generalizes to random vectors. If \( X_1, X_2, \ldots \) are i.i.d random vectors with mean vector \( \mu \) and covariance matrix \( \Sigma \). Then,

\[
\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \Sigma).
\]

**Exercise 0.2.** Let \( W_i \sim \chi^2_{10} \) i.i.d and define \( \bar{W}_n = \frac{1}{n} \sum_{i=1}^{n} W_i \).

1. Show that \( E[\bar{W}_n] = 10 \).
2. Show that \( \bar{W}_n \xrightarrow{p} 10 \).
3. Show that \( \frac{1}{n} \sum_{i=1}^{n} (W_i - \bar{W}) \xrightarrow{p} V(W_i) \).
4. Does \( E[\frac{1}{n} \sum_{i=1}^{n} (W_i - \bar{W})] = V(W_i) \)?

**Exercise 0.3.** Suppose \( X_i \sim N_p(0, \Sigma) \) for \( i = 1, \ldots, n \). Let \( \alpha \in \mathbb{R}^p \) and define

\[
Y_n = \frac{\alpha'X_n'X_n\alpha}{\frac{1}{n-1} \sum_{i=1}^{n-1} \alpha'X_i'X_i\alpha}.
\]

1. Show that \( Y_n \sim F_{1,n-1} \).
2. Show that \( Y_n \xrightarrow{d} \chi^2_1 \).

**Stationarity and Martingales**

So far, each of the LLNs and CLTs we discussed relied on the random variables in the sequence to be uncorrelated or even independent. In this sub-section, we briefly introduce a LLN-type result and CLT for dependent data. The presentation in this sub-section closely follows Chapter 2 of Hayashi’s *Econometrics*.  

A **stochastic process** is a sequence of random variables. A **time series** is a stochastic process whose indices are time measurements.

**Definition 0.6.** A stochastic process is **strictly stationary** if the probability distribution of

\[
(X_t, X_{t+1}, \ldots, X_{t+k})
\]

is the same as the probability distribution of

\[
(X_{\tau}, X_{\tau+1}, \ldots, X_{\tau+k})
\]

for all \( t, \tau, k \).

\*This subsection is purely optional. Feel free to skip it.*
Definition 0.7. A strictly stationary stochastic process is **ergodic** if for any two bounded functions $f : \mathbb{R}^k \rightarrow \mathbb{R}$ and $g : \mathbb{R}^k \rightarrow \mathbb{R}$, the following holds:

$$\lim_{n \to \infty} E[f(X_t, \ldots, X_{t+k})g(X_{t+n}, \ldots, X_{t+n+k})] = E[f(X_t, \ldots, X_{t+k})]E[g(X_t, \ldots, X_{t+k})]$$

That is, sub-sequences separated by $n$ time periods become independent as $n$ grows large.

**Theorem 0.9.** Ergodic law of large numbers

Suppose $\{X_t\}$ is a strictly stationary, ergodic stochastic process with $E[X_1] = \mu$. Then,

$$\bar{X}_n \xrightarrow{a.s.} \mu.$$

Definition 0.8. A stochastic process $\{Z_i\}$ is a **martingale** if

$$E[Z_i|Z_{i-1}, \ldots, Z_1] = Z_{i-1}$$

for all $i \geq 2$. A stochastic process $\{Z_i\}$ with $E[Z_i] = 0$ for all $i$ is a **martingale difference sequence** if

$$E[Z_i|Z_{i-1}, \ldots, Z_1] = 0$$

for all $i \geq 2$.

**Theorem 0.10.** Ergodic stationary marginale difference CLT

Let $\{X_t\}$ be a martingale difference sequence that is stationary and ergodic with

$$E[X_iX_i'] = \Sigma.$$

Then,

$$\sqrt{n}\bar{X}_n \xrightarrow{d} N(0, \Sigma).$$

**The Delta Method**

Suppose we have some estimator $T_n$ of a parameter $\theta$. We know that

$$T_n \xrightarrow{p} \theta$$

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2).$$

However, we are interested in estimating and conducting inference on $g(\theta)$, where $g$ is some continuously differentiable function. A natural estimator is $g(T_n)$ and by the continuous mapping theorem, we know that

$$g(T_n) \xrightarrow{p} g(\theta).$$

Can we construct the asymptotic distribution of $g(T_n)$? That is,

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d}?$$

The **delta method** provides a general technique for constructing this asymptotic distribution.
Theorem 0.11. The delta method

Let $Y_n$ be a sequence of random variables and let $X_n = \sqrt{n}(Y_n - a)$ for some constant $a$. Let $g(\cdot)$ be a continuously differentiable function. Suppose that

$$X_n = \sqrt{n}(Y_n - a) \xrightarrow{d} X \sim N(0, \sigma^2).$$

Then,

$$\sqrt{n}(g(Y_n) - g(a)) \xrightarrow{d} g'(a)N(0, \sigma^2).$$

Proof. By the mean value theorem, 7

$$g(Y_n) = g(a) + (Y_n - a)g'({\bar{Y}_n})$$

where $\bar{Y}_n$ is some value between $Y_n$ and $a$. Since $X_n \xrightarrow{d} X$, it follows that $Y_n \xrightarrow{p} a$. Since $g$ is continuously differentiable, it follows that $g'(\bar{Y}_n) \xrightarrow{p} g'(a)$ by the continuous mapping theorem. So, it follows that

$$\sqrt{n}(g(Y_n) - g(a)) = g'(\bar{Y}_n)\sqrt{n}(Y_n - a)$$

$$= g'(\bar{Y}_n)X_n \xrightarrow{d} g'(a)X$$

by Slutsky’s theorem.

We can prove a similar result for random vectors. In the theorem above, replace everything with vectors. The result becomes

$$\sqrt{n}(g(Y_n) - g(a)) \xrightarrow{d} GN(0, \Sigma)$$

where

$$G = \frac{\partial g(a)}{\partial a'}.$$

Example 0.2. Suppose $X_i$ i.i.d for $i = 1, \ldots, n$ with mean 2 and variance 1. Then,

$$\sqrt{n}(\bar{X} - 2) \xrightarrow{d} N(0, 1).$$

Let $g(z) = z^2$. The delta method tells us that

$$\sqrt{n}(\bar{X}^2 - 4) \xrightarrow{d} g'(2)N(0, 1) = N(0, 16).$$
References


