

# Basic Asymptotic Theory

## ECON 628

In this note, we discuss basic asymptotic theory. The results that we discuss here will be crucial to understanding the large sample properties of estimation methods such as OLS, MLE and GMM (which includes OLS). Most results will simply be stated as rigorous proofs are beyond the scope of this course. For those of you who are ambitious enough, I refer you to *Asymptotic Statistics* by Van der Vaart which has all of the  $\varepsilon$ 's and  $\delta$ 's that you could possibly want.

The following are useful.

**Theorem 1 (Markov's Inequality)** *Let  $X$  be a random variable such that  $P(X \geq 0) = 1$ . Then for all  $t > 0$ ,  $P(X \geq t) \leq \frac{E[X]}{t}$ .*

A proof is provided in DeGroot and Schervish (DS). We can now use Markov's Inequality to prove another important inequality: Chebyshev's Inequality.

**Theorem 2 (Chebyshev's Inequality)** *Let  $X$  be a random variable such that  $Var(X) < \infty$ . Then for all  $t > 0$ ,  $P(|X - E[X]| \geq t) \leq \frac{Var(X)}{t^2}$*

**Proof.** *Let  $Y = (X - E[X])^2$ . Then  $P(Y \geq 0) = 1$  and  $E[Y] = Var(X)$ . We can now apply Markov's Inequality and get*

$$P(|X - E[X]| \geq t) = P(Y \geq t^2) \leq \frac{Var(X)}{t^2}$$

■

Chebyshev's Inequality will be important later on when talking about the Law of Large Numbers. For now, let's talk about the sample mean:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

where the sample  $X_i$  is distributed independently and identically or i.i.d. Clearly,

$$E[\bar{X}_n] = \mu$$

and

$$Var(\bar{X}_n) = \frac{Var(X_i)}{n}.$$

Note that to derive these formulae, we used both the independence assumption in  $Var(\bar{X}_n)$  and the identical assumption in  $Var(\bar{X}_n)$  and  $E[\bar{X}_n]$ .

Now, let's formalize a notion of convergence of random variables so that we can derive the limit of the sample mean.

**Definition 3**  $X_n$  converges in probability to  $X$  if for all  $\varepsilon > 0$ ,  $\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$ .  
We say  $X_n \xrightarrow{p} X$ .

Typically, for our purposes,  $X$  will be a constant, but it can also be a random variable. We can now prove the Weak Law of Large Numbers.

**Theorem 4 (Law of Large Numbers)** Suppose that  $X_1, \dots, X_n$  is an i.i.d. sample from a distribution with mean  $\mu$  and a finite variance,  $\sigma^2$ . Then

$$\bar{X}_n \xrightarrow{p} \mu.$$

**Proof.**  $P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon} \Leftrightarrow P(|\bar{X}_n - \mu| < \varepsilon) > 1 - \frac{\sigma^2}{n\varepsilon} \rightarrow 1 \blacksquare$

We are now going to state some more definitions of convergence and discuss their relationship to convergence in probability.

**Definition 5**  $X_n$  converges in distribution to  $X$  if  $P(X_n \leq x) \rightarrow P(X \leq x)$  for all  $x$  for which the limiting distribution function is continuous. We say  $X_n \rightsquigarrow X$ .

**Definition 6**  $X_n$  converges almost surely to  $X$  if  $P(\lim_{n \rightarrow \infty} |X_n - X| = 0) = 1$ . We say  $X_n \xrightarrow{as} X$ .

The next theorem relates each definition of convergence.

**Theorem 7 (Relationship of Modes of Convergence)**

- (i)  $X_n \xrightarrow{as} X \Rightarrow X_n \xrightarrow{p} X$
- (ii)  $X_n \xrightarrow{p} X \Rightarrow X_n \rightsquigarrow X$
- (iii)  $X_n \xrightarrow{p} c$  where  $c$  is a constant  $\Leftrightarrow X_n \rightsquigarrow c$ .

This theorem states that convergence almost surely is the strongest definition of convergence that we have talked about. Convergence in distribution is the weakest.

Another useful theorem concerns continuous maps of statistics.

**Theorem 8 (Continuous Mapping Theorem)** Let  $g(\cdot)$  be a continuous function. Then, we will have the following:

- (i)  $X_n \xrightarrow{as} X \Rightarrow g(X_n) \xrightarrow{as} g(X)$
- (ii)  $X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$
- (iii)  $X_n \rightsquigarrow X \Rightarrow g(X_n) \rightsquigarrow g(X)$

Finally, we have the Slutsky Theorem.

**Theorem 9 (Slutsky Theorem)** Suppose that  $X_n \rightsquigarrow X$  and  $Y_n \rightsquigarrow c$  where  $c$  is a constant.

- (i)  $X_n + Y_n \rightsquigarrow X + c$
- (ii)  $X_n Y_n \rightsquigarrow Xc$
- (iii)  $X_n/Y_n \rightsquigarrow c^{-1}X$  for  $c \neq 0$ .

Now, let's give a simple example to illustrate these ideas. Let

$$y_i = x_i\beta + \varepsilon_i$$

where  $E[\varepsilon_i x_i] = E[\varepsilon_i] = 0$  and  $(x_i, y_i)$  is a random sample. There is no constant in the regression. Let's estimate  $\beta$  with OLS so that

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}.$$

Then, we will have that

$$\sqrt{n}(\hat{\beta} - \beta) = \frac{\frac{1}{\sqrt{n}} \sum x_i \varepsilon_i}{\frac{1}{n} \sum x_i^2}.$$

We know that

$$\frac{1}{n} \sum x_i^2 \xrightarrow{p} E[x_i^2]$$

by the Law of Large Numbers. Because  $E[x_i^2]$  is a constant, (iii) of Theorem 7 gives us that

$$\frac{1}{n} \sum x_i^2 \rightsquigarrow E[x_i^2]$$

Also, we will learn that

$$\frac{1}{\sqrt{n}} \sum x_i \varepsilon_i \rightsquigarrow N(0, E[x_i^2 \varepsilon_i^2]).$$

Then, by Slutsky, we will have that

$$\sqrt{n}(\hat{\beta} - \beta) \rightsquigarrow N(0, E[x_i^2]^{-2} E[x_i^2 \varepsilon_i^2]).$$

If  $Var(\varepsilon_i | x_i) = \sigma^2$ , we will have that

$$\sqrt{n}(\hat{\beta} - \beta) \rightsquigarrow N(0, \sigma^2 E[x_i^2]^{-1}).$$