

## Chapter 6

# A Review of Large Sample Asymptotics

### 6.1 Introduction

The most widely-used tool in sampling theory is large sample asymptotics. By “asymptotics” we mean approximating a finite-sample sampling distribution by taking its limit as the sample size diverges to infinity. In this chapter we provide a brief review of the main results of large sample asymptotics. It is meant as a reference, not as a teaching guide. Asymptotic theory is covered in detail in Chapters 7-9 of *Introduction to Econometrics*. If you have not previously studied asymptotic theory in detail you should study these chapters before proceeding.

### 6.2 Modes of Convergence

**Definition 6.1** A random vector  $Z_n \in \mathbb{R}^k$  **converges in probability** to  $Z$  as  $n \rightarrow \infty$ , denoted  $Z_n \xrightarrow{p} Z$  or alternatively  $\text{plim}_{n \rightarrow \infty} Z_n = Z$ , if for all  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P} [\|Z_n - Z\| \leq \delta] = 1. \quad (6.1)$$

We call  $Z$  the **probability limit** (or **plim**) of  $Z_n$ .

The above definition treats random variables and random vectors simultaneously using the vector norm. It is useful to know that for a random vector, (6.1) holds if and only if each element in the vector converges in probability to its limit.

**Definition 6.2** Let  $Z_n$  be a random vector with distribution  $F_n(u) = \mathbb{P} [Z_n \leq u]$ . We say that  $Z_n$  **converges in distribution** to  $Z$  as  $n \rightarrow \infty$ , denoted  $Z_n \xrightarrow{d} Z$ , if for all  $u$  at which  $F(u) = \mathbb{P} [Z \leq u]$  is continuous,  $F_n(u) \rightarrow F(u)$  as  $n \rightarrow \infty$ . We refer to  $Z$  and its distribution  $F(u)$  as the **asymptotic distribution**, **large sample distribution**, or **limit distribution** of  $Z_n$ .

### 6.3 Weak Law of Large Numbers

**Theorem 6.1 Weak Law of Large Numbers (WLLN)**

If  $Y_i \in \mathbb{R}^k$  are i.i.d. and  $\mathbb{E} \|Y\| < \infty$ , then as  $n \rightarrow \infty$ ,

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{p} \mathbb{E}[Y].$$

The WLLN shows that the sample mean  $\bar{Y}$  converges in probability to the true population expectation  $\mu$ . The result applies to any transformation of a random vector with a finite mean.

**Theorem 6.2** If  $Y_i \in \mathbb{R}^k$  are i.i.d.,  $h(y) : \mathbb{R}^k \rightarrow \mathbb{R}^q$  and  $\mathbb{E} \|h(Y)\| < \infty$ , then  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n h(Y_i) \xrightarrow{p} \mu = \mathbb{E}[h(Y)]$  as  $n \rightarrow \infty$ .

An estimator which converges in probability to the population value is called **consistent**.

**Definition 6.3** An estimator  $\hat{\theta}$  of  $\theta$  is **consistent** if  $\hat{\theta} \xrightarrow{p} \theta$  as  $n \rightarrow \infty$ .

### 6.4 Central Limit Theorem

**Theorem 6.3 Multivariate Lindeberg-Lévy Central Limit Theorem (CLT).** If  $Y_i \in \mathbb{R}^k$  are i.i.d. and  $\mathbb{E} \|Y\|^2 < \infty$ , then as  $n \rightarrow \infty$

$$\sqrt{n}(\bar{Y} - \mu) \xrightarrow{d} N(0, V)$$

where  $\mu = \mathbb{E}[Y]$  and  $V = \mathbb{E}[(Y - \mu)(Y - \mu)']$ .

The central limit theorem shows that the distribution of the sample mean is approximately normal in large samples. For some applications it may be useful to notice that Theorem 6.3 does not impose any restrictions on  $V$  other than that the elements are finite. Therefore this result allows for the possibility of singular  $V$ .

The following two generalizations allow for heterogeneous random variables.

**Theorem 6.4 Multivariate Lindeberg CLT.** Suppose that for all  $n$ ,  $Y_{ni} \in \mathbb{R}^k$ ,  $i = 1, \dots, r_n$ , are independent but not necessarily identically distributed with expectations  $\mathbb{E}[Y_{ni}] = 0$  and variance matrices  $V_{ni} = \mathbb{E}[Y_{ni} Y_{ni}']$ . Set  $\bar{V}_n = \sum_{i=1}^{r_n} V_{ni}$ . Suppose  $v_n^2 = \lambda_{\min}(\bar{V}_n) > 0$  and for all  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{v_n^2} \sum_{i=1}^{r_n} \mathbb{E}[\|Y_{ni}\|^2 \mathbb{1}_{\{\|Y_{ni}\|^2 \geq \epsilon v_n^2\}}] = 0. \quad (6.2)$$

Then as  $n \rightarrow \infty$

$$\bar{V}_n^{-1/2} \sum_{i=1}^{r_n} Y_{ni} \xrightarrow{d} N(0, \mathbf{I}_k).$$

**Theorem 6.5** Suppose  $Y_{ni} \in \mathbb{R}^k$  are independent but not necessarily identically distributed with expectations  $\mathbb{E}[Y_{ni}] = 0$  and variance matrices  $V_{ni} = \mathbb{E}[Y_{ni} Y_{ni}']$ . Suppose

$$\frac{1}{n} \sum_{i=1}^n V_{ni} \rightarrow V > 0$$

and for some  $\delta > 0$

$$\sup_{n,i} \mathbb{E} \|Y_{ni}\|^{2+\delta} < \infty. \quad (6.3)$$

Then as  $n \rightarrow \infty$

$$\sqrt{n} \bar{Y} \xrightarrow{d} N(0, V).$$

## 6.5 Continuous Mapping Theorem and Delta Method

Continuous functions are limit-preserving. There are two forms of the continuous mapping theorem, for convergence in probability and convergence in distribution.

**Theorem 6.6 Continuous Mapping Theorem (CMT).** Let  $Z_n \in \mathbb{R}^k$  and  $g(u) : \mathbb{R}^k \rightarrow \mathbb{R}^q$ . If  $Z_n \xrightarrow{p} c$  as  $n \rightarrow \infty$  and  $g(u)$  is continuous at  $c$  then  $g(Z_n) \xrightarrow{p} g(c)$  as  $n \rightarrow \infty$ .

### Theorem 6.7 Continuous Mapping Theorem

If  $Z_n \xrightarrow{d} Z$  as  $n \rightarrow \infty$  and  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$  has the set of discontinuity points  $D_g$  such that  $\mathbb{P}[Z \in D_g] = 0$  then  $g(Z_n) \xrightarrow{d} g(Z)$  as  $n \rightarrow \infty$ .

Differentiable functions of asymptotically normal random estimators are asymptotically normal.

**Theorem 6.8 Delta Method**

Let  $\mu \in \mathbb{R}^k$  and  $g(u) : \mathbb{R}^k \rightarrow \mathbb{R}^q$ . If  $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \xi$ , where  $g(u)$  is continuously differentiable in a neighborhood of  $\mu$ , then as  $n \rightarrow \infty$

$$\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} \mathbf{G}'\xi \quad (6.4)$$

where  $\mathbf{G}(u) = \frac{\partial}{\partial u} g(u)'$  and  $\mathbf{G} = \mathbf{G}(\mu)$ . In particular, if  $\xi \sim N(0, \mathbf{V})$  then as  $n \rightarrow \infty$

$$\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} N(0, \mathbf{G}'\mathbf{V}\mathbf{G}). \quad (6.5)$$

## 6.6 Smooth Function Model

The smooth function model is  $\theta = g(\mu)$  where  $\mu = \mathbb{E}[h(Y)]$  and  $g(\mu)$  is smooth in a suitable sense. The parameter of interest  $\theta$  is a smooth function of a population mean  $\mu$ .

The parameter  $\theta = g(\mu)$  is not a population moment so it does not have a direct moment estimator. Instead, it is common to use a **plug-in estimator** formed by replacing the unknown  $\mu$  with its point estimator  $\hat{\mu}$  and then “plugging” this into the expression for  $\theta$ . The first step is the sample mean  $\hat{\mu} = n^{-1} \sum_{i=1}^n h(Y_i)$ . The second step is the transformation  $\hat{\theta} = g(\hat{\mu})$ . The hat “^” indicates that  $\hat{\theta}$  is a sample estimator of  $\theta$ . The smooth function model includes a broad class of estimators including sample variances and the least squares estimator.

**Theorem 6.9** If  $Y_i \in \mathbb{R}^m$  are i.i.d.,  $h(u) : \mathbb{R}^m \rightarrow \mathbb{R}^k$ ,  $\mathbb{E}\|h(Y)\| < \infty$ , and  $g(u) : \mathbb{R}^k \rightarrow \mathbb{R}^q$  is continuous at  $\mu$ , then  $\hat{\theta} \xrightarrow{p} \theta$  as  $n \rightarrow \infty$ .

**Theorem 6.10** If  $Y_i \in \mathbb{R}^m$  are i.i.d.,  $h(u) : \mathbb{R}^m \rightarrow \mathbb{R}^k$ ,  $\mathbb{E}\|h(Y)\|^2 < \infty$ ,  $g(u) : \mathbb{R}^k \rightarrow \mathbb{R}^q$ , and  $\mathbf{G}(u) = \frac{\partial}{\partial u} g(u)'$  is continuous in a neighborhood of  $\mu$ , then as  $n \rightarrow \infty$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathbf{V}_\theta)$$

where  $\mathbf{V}_\theta = \mathbf{G}'\mathbf{V}\mathbf{G}$ ,  $\mathbf{V} = \mathbb{E}[(h(Y) - \mu)(h(Y) - \mu)']$ , and  $\mathbf{G} = \mathbf{G}(\mu)$ .

Theorem 6.9 establishes the consistency of  $\hat{\theta}$  for  $\theta$  and Theorem 6.10 establishes its asymptotic normality. It is instructive to compare the conditions. Consistency requires that  $h(Y)$  has a finite expectation; asymptotic normality requires that  $h(Y)$  has a finite variance. Consistency requires that  $g(u)$  be continuous; asymptotic normality requires that  $g(u)$  is continuously differentiable.

## 6.7 Best Unbiased Estimation

This section presents efficiency bounds for the mean and the smooth function model. The results are finite-sample rather than asymptotic, but are convenient to introduce at this point since the bounds are identical to the asymptotic variance matrices.

Our first result is for the mean.

**Theorem 6.11** Suppose  $Y_i$  are i.i.d.,  $\mu = \mathbb{E}[h(Y)]$ , and  $\mathbb{E}\|h(Y)\|^2 < \infty$ . If  $\tilde{\mu}$  is unbiased for  $\mu$  then  $\text{var}[\tilde{\mu}] \geq n^{-1} \mathbf{V}$  where  $\mathbf{V} = \mathbb{E}[(h(Y) - \mu)(h(Y) - \mu)']$ .

The second result is for the smooth function model.

**Theorem 6.12** Suppose  $Y_i$  are i.i.d.,  $\mu = \mathbb{E}[h(Y)]$ ,  $\theta = g(\mu)$ ,  $\mathbb{E}\|h(Y)\|^2 < \infty$ , and  $\mathbf{G}(u) = \frac{\partial}{\partial u} g(u)'$  is continuous in a neighborhood of  $\mu$ . If  $\tilde{\theta}$  is unbiased for  $\theta$  then

$$\text{var}[\tilde{\theta}] \geq n^{-1} \mathbf{V}_\theta$$

where  $\mathbf{V}_\theta = \mathbf{G}' \mathbf{V} \mathbf{G}$ ,  $\mathbf{V} = \mathbb{E}[(h(Y) - \mu)(h(Y) - \mu)']$  and  $\mathbf{G} = \mathbf{G}(\mu)$ .

For details and proofs see Section 11.6 of *Introduction to Econometrics*. Theorems 6.11 and 6.12 are analogs and extensions of the Cramér-Rao lower bound theory to semiparametric estimation. The results show that the asymptotic variances from Theorems 6.3 and 6.10 are the best possible in any finite sample among unbiased estimators. Theorem 6.11 is sharp, since the sample mean has the finite sample variance  $n^{-1} \mathbf{V}$ .

## 6.8 Stochastic Order Symbols

It is convenient to have simple symbols for random variables and vectors which converge in probability to zero or are stochastically bounded. In this section we introduce some of the most commonly found notation.

Let  $Z_n$  and  $a_n$ ,  $n = 1, 2, \dots$  be sequences of random variables and constants. The notation

$$Z_n = o_p(1)$$

(“small oh-P-one”) means that  $Z_n \xrightarrow{p} 0$  as  $n \rightarrow \infty$ . We also write

$$Z_n = o_p(a_n)$$

if  $a_n^{-1} Z_n = o_p(1)$ .

Similarly, the notation  $Z_n = O_p(1)$  (“big oh-P-one”) means that  $Z_n$  is bounded in probability. Precisely, for any  $\epsilon > 0$  there is a constant  $M_\epsilon < \infty$  such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}[|Z_n| > M_\epsilon] \leq \epsilon.$$

Furthermore, we write

$$Z_n = O_p(a_n)$$

if  $a_n^{-1} Z_n = O_p(1)$ .

$O_p(1)$  is weaker than  $o_p(1)$  in the sense that  $Z_n = o_p(1)$  implies  $Z_n = O_p(1)$  but not the reverse. However, if  $Z_n = O_p(a_n)$  then  $Z_n = o_p(b_n)$  for any  $b_n$  such that  $a_n/b_n \rightarrow 0$ .

A random sequence with a bounded moment is stochastically bounded.

**Theorem 6.13** If  $Z_n$  is a random vector which satisfies

$$\mathbb{E} \|Z_n\|^\delta = O(a_n)$$

for some sequence  $a_n$  and  $\delta > 0$ , then

$$Z_n = O_p(a_n^{1/\delta}).$$

Similarly,  $\mathbb{E} \|Z_n\|^\delta = o(a_n)$  implies  $Z_n = o_p(a_n^{1/\delta})$ .

There are many simple rules for manipulating  $o_p(1)$  and  $O_p(1)$  sequences which can be deduced from the continuous mapping theorem. For example,

$$o_p(1) + o_p(1) = o_p(1)$$

$$o_p(1) + O_p(1) = O_p(1)$$

$$O_p(1) + O_p(1) = O_p(1)$$

$$o_p(1)o_p(1) = o_p(1)$$

$$o_p(1)O_p(1) = o_p(1)$$

$$O_p(1)O_p(1) = O_p(1).$$

## 6.9 Convergence of Moments

We give a sufficient condition for the existence of the mean of the asymptotic distribution, define uniform integrability, provide a primitive condition for uniform integrability, and show that uniform integrability is the key condition under which  $\mathbb{E}[Z_n]$  converges to  $\mathbb{E}[Z]$ .

**Theorem 6.14** If  $Z_n \xrightarrow{d} Z$  and  $\mathbb{E} \|Z_n\| \leq C$  then  $\mathbb{E} \|Z\| \leq C$ .

**Definition 6.4** The random vector  $Z_n$  is **uniformly integrable** as  $n \rightarrow \infty$  if

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} [\|Z_n\| \mathbf{1}_{\{\|Z_n\| > M\}}] = 0.$$

**Theorem 6.15** If for some  $\delta > 0$ ,  $\mathbb{E} \|Z_n\|^{1+\delta} \leq C < \infty$ , then  $Z_n$  is uniformly integrable.

**Theorem 6.16** If  $Z_n \xrightarrow{d} Z$  and  $Z_n$  is uniformly integrable then  $\mathbb{E}[Z_n] \rightarrow \mathbb{E}[Z]$ .

## 6.10 Uniform Stochastic Bounds

**Theorem 6.17** If  $|Y_i|^r$  is uniformly integrable, then as  $n \rightarrow \infty$

$$n^{-1/r} \max_{1 \leq i \leq n} |Y_i| \xrightarrow{p} 0. \quad (6.6)$$

Equation (6.6) says that if  $Y$  has  $r$  finite moments then the largest observation will diverge at a rate slower than  $n^{1/r}$ . The higher the moments, the slower the rate of divergence.