Econometrics – Asymptotic Theory

William M Volckmann II

January 3, 2017

We will be considering a sequence of random variables $\{X_n\}$ that gets "close" to another random variable X as $n \to \infty$, and the implications thereof.

1 Convergence in Probability

Definition 1. Let $\{X_n\}$ be a sequence of random variables and let X be a random variable defined on a sample space. We say that X_n **converges in probability** to X if, for all $\epsilon > 0$,

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

Equivalently, we can write the convergence condition as

$$\lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1.$$

We denote such convergence as $X_n \xrightarrow{P} X$ or $\operatorname{plim} X_n = X$.

Intuitively, it means the probability of an element in sequence being "far" away from X, as determined by ϵ, approaches zero. Many times the limiting random variable X is a constant, i.e. $X_n \xrightarrow{P} a$.

Theorem 1 (Weak Law of Large Numbers). *Let $\{X_n\}$ be a sequence of iid random variables having common mean μ and finite variance σ^2. Let $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Then*

$$\overline{X}_n \xrightarrow{P} \mu.$$

Proof. The mean of $\overline{X}_n = \mu$ and the standard deviation is σ^2/n. By Chebyshev's

inequality, we have

$$P(|\overline{X}_n - \mu| > \epsilon) = P\left(|\overline{X}_n - \mu| > \epsilon\frac{\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}}\right) \leq \frac{\sigma^2}{ne^2}.$$

As $n \to \infty$, this will approach zero. $\quad\square$

The preceding theorem is sometimes called *Khinchine's Theorem.*

As you might have suspected, there are also strong laws of large numbers. In such a case, convergence is "almost surely." This requires measure theory and is more advanced than what is covered here. *Markov's Strong Law of Large Numbers* is one notable example because it allows us to work with an independent but not identically distributed sequence of random variables.

Convergent sequences of random variables have nice properties. In particular, suppose $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, and $a \in \mathbb{R}$. Then

(a) $X_n + Y_n \xrightarrow{P} X + Y$

(b) $aX_n \xrightarrow{P} aX$

(c) $X_n Y_n \xrightarrow{P} XY$

Furthermore, suppose $X_n \xrightarrow{P} a$ and the real function g is continuous at a. Then $g(X_n) \xrightarrow{P} g(a)$.

Definition 2. Let X be a random variable with cdf $F(x; \theta)$. Let $X_1, \ldots, X_n$ be a sample from the distribution of X, and let T_n denote a statistic. We say that T_n is a **consistent** estimator of θ if

$$T_n \xrightarrow{P} \theta.$$

For example, the Weak Law of Large Numbers suggests that $\overline{X}_n$ is a consistent estimator of μ under the assumed conditions. As $n \to \infty$, the statistic $\overline{X}$ gets closer and closer to μ, arbitrarily so.

Now consider the experiment of flipping a (by all impressions, fair) coin. The coin shows tails with probability $p = 0.5$, and we wish to estimate p. Suppose we make our estimate $\hat{p}_n = 1$ if the first toss is tails and $\hat{p}_n = 0$ if the first toss is heads. Now we should expect a consistent estimator to satisfy $\hat{p}_n \xrightarrow{P} 0.5$ since the coin is fair. However, our estimator will either remain at $\hat{p}_n = 1$ or $\hat{p}_n = 0$, regardless of how many times the coin is flipped. Thus, this is *not* a consistent estimator; if we choose

2

$\epsilon = 0.25$, for instance, it will *never* be the case that $|\hat{p}_n - 0.50| < \epsilon$. This estimate is unbiased, however, because

$$E[\hat{p}_n] = 1(0.5) + 0(0.5) = 0.5 = p.$$

One more illustration. Suppose we estimate the mean with

$$\hat{\mu}_n = \overline{X}_n + \frac{1}{n}.$$

We know that $\overline{X}$ will converge in probability to μ by the Weak Law of Large Numbers. Furthermore, $1/n$ will converge to zero. Thus, as $n \to \infty$, we have $\hat{\mu}_n \xrightarrow{P} \mu$. Thus, $\hat{\mu}_n$ is a consistent estimator for μ. However, it is a biased estimator because

$$E[\hat{\mu}_n] = E\left[\overline{X}_n\right] + E\left[\frac{1}{n}\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i] + \frac{1}{n}$$
$$= \mu + \frac{1}{n}.$$

A more useful result is that the sample variance, given as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

is a consistent estimator.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$
$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}^2 + \overline{X}^2)\right]$$
$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n}X_i + \overline{X}^2\frac{1}{n}\sum_{i=1}^{n}\right]$$
$$= \frac{n}{n-1}\left[\overline{X^2} - 2\overline{X}^2 + \overline{X}^2\right]$$
$$= \frac{n}{n-1}\left[\overline{X^2} - \overline{X}^2\right].$$

As $n \to \infty$, the fraction $n/(n-1)$ approaches one; the term $\overline{X^2}$ converges in probability to $E[X^2]$; and the term $\overline{X}^2$ converges in probability to $E[X]^2$. Thus, we have convergence in probability to $1 \cdot (E[X^2] - E[X]^2) = var(X)$.

There are alternative modes of convergence, although convergence in probability is the most important one for us to understand. That said, I will only briefly mention **mean square convergence**, defined as satisfying

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

The form of the expectation should look familiar. In particular, if $E[X_n] = X$, then mean square convergence holds if $Var(X_n) \to 0$. Furthermore, if $E[X_n] \neq X$, then mean square convergence holds if $Bias[X_n] \to 0$ and $Var(X_n) \to 0$. Convergence in mean square implies convergence in probability. (Note: this actually forms a metric, and the Cauchy criterion is a sufficient and necessary condition for such convergence.)

2 Convergence in Distribution

It's nice that we can say a statistic converges to a parameter. It's also worth asking how close the statistic is to the estimator. To do so, we need some sense of the distribution of a random variable as it converges.

Definition 3. Let $\{X_n\}$ be a sequence of random variables and let X be a random variable. Let F_{X_n} and F_x be, respectively, the cdfs of X_n and X. Let $C(F_x)$ denote the set of all points where F_X is continuous. We say that X_n **converges in distribution** to X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x) \quad \text{for all } x \in C(F_x).$$

We denote this convergence by

$$X_n \xrightarrow{D} X.$$

Often we say that the distribution of X is the **asymptotic distribution** or the **limiting distribution** of the sequence $\{X_n\}$. We might write something like

$$X_n \xrightarrow{D} \mathcal{N}(0, 1),$$

even though it doesn't make sense for a random variable to converge to a distribution – technically it converges to another random variable with said distribution. But as long as the meaning is understood, it should be okay. I will certainly be using that notation, in any case.

The problem is that F_{X_n} is usually a complicated function, and it can be difficult or impossible to actually prove convergence in distribution. That said, one can usually prove convergence in distribution using moment generating functions. In particular, if $M_{X_n}(t) = M_X(t)$ for $|T| \leq h$, then $X_n \xrightarrow{D} X$.

An alternative approach is to rely on the Central Limit Theorem. For instance, suppose each X_i has a mean μ and variance σ^2. We know that

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, 1).$$

We can multiply "both sides" by the denominator and add μ to "both sides" to get

$$\overline{X}_n \overset{a}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

In this case, we say "$\overline{X}_n$ is asymptotically distributed as $\mathcal{N}(\mu, \sigma^2/n)$ for large n. " Unlike convergence in distribution, we aren't taking the limit as $n \to \infty$ – we are instead noting its approximate distribution for "large" n.

As you might suspect, convergence in probability and convergence in distribution are related concepts and have implications with respect to each other.

Theorem 2. *If X_n converges to X in probability, then X_n converges to X in distribution.*

Theorem 3. *If X_n converges to the constant b in distribution, then X_n converges to b in probability.*

Furthermore, convergence in distribution has "nice" properties that can make it easy to work with.

Theorem 4 (Continuous Mapping Theorem)**.** *Suppose X_n converges to X in distribution and g is a continuous function on the support of X. Then $g(X_n)$ converges to $g(X)$ in distribution.*

Theorem 5. *If X_n converges in probability to X and Y_n converges in probability to the constant b, then*

(a) $X_n + Y_n \xrightarrow{D} X + b$

(b) $X_n Y_n \xrightarrow{D} X_n b$

(c) $X_n / Y_n \xrightarrow{D} X/b$, *provided $b \neq 0$.*

3 Δ-Method

Theorem 6. *Let $\{X_n\}$ be a sequence of random variables such that*

$$\sqrt{n}(X_n - \theta) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2\right).$$

Suppose the function $g(x)$ is differentiable at θ and $g'(\theta) \neq 0$. Then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{D} \mathcal{N}\left(0, \sigma^2 g'(\theta)^2\right).$$

Alternatively, we can suppose that $X_n \xrightarrow{D} \mathcal{N}\left(\theta, \sigma^2/n\right)$, from which it follows via the preceding theorem that

$$g(X_n) \xrightarrow{D} \mathcal{N}\left(\theta, \frac{\sigma^2}{n} g'(\theta)^2\right).$$

This is called the **Δ-method** and is used frequently.