Economics 620, Lecture 8: Asymptotics I

Nicholas M. Kiefer

Cornell University

We are interested in the properties of estimators as

$$n \longrightarrow \infty.$$

Consider a sequence of random variables

$$\{X_n, n \geq 1\}.$$

Often X_n is an estimator such as a sample mean or $\widehat{\beta_n}$

Often it is convenient to center the sequence: $\{\widehat{\beta_n} - \beta\}$

and sometimes to scale $\{(\widehat{\beta_n} - \beta)/\sigma_n\}$

Plim

Definition: (*convergence in probability*)

A sequence of random variables $\{X_n, n \geq 1\}$ is said to *converge weakly* to a constant c if

$$\lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 0$$

for every given $\varepsilon > 0$.

This is written $p\lim X_n = c$ or $X_n \xrightarrow{p} c$

Some properties of plim:

1. plim XY = plim X plim Y
2. plim $(X + Y)$ = plim X + plim Y
3. *Slutsky's theorem*: If the function g is continuous at plim X, then plim $g(X) = g(\text{plim } X)$.

A.S. convergence

Definition: (*Strong convergence*)

A sequence of random variables is said to *converge strongly* to a constant c if

$$P(\lim_{n\to\infty} X_n = c) = 1$$

or

$$\lim_{N\to\infty} P(\sup_{n>N} |x_n - c| > \varepsilon) = 0.$$

Strong convergence is also called *almost sure convergence* or *convergence with probability one* and is written $X_n \to c$ w.p. 1 or $X_n \overset{a.s.}{\to} c$.

Difference betwen convergence a.s. and plim

plim involves probabilities on each element of the sequence, and limits of these probabilities.

$$\lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 0$$

Strong convergence involves probabilities on the entire sequence.

$$P(\lim_{n\to\infty} X_n = c) = 1$$

Sequence of marginal probabilities vs. joint probability over infinite sequences.

Note a.s. convergence implies plim.
Difference usually doesn't matter in applications and plim is easier to establish.

Laws of Large Numbers:

Let $\{X_n, n \geq 1\}$ be observations and suppose we look at the sequence

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

when does $\bar{X}_n \xrightarrow{p} \xi$ where ξ is some parameter?

Weak Law of Large Numbers: $(WLLN)$ Let $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $cov(X_i X_j) = 0$.
Then $\bar{X}_n - \mu \to 0$ in probability.

Proof of WLLN

Lemma: Chebyshev's Inequality:

$P(|X-\mu| \geq k) \leq \sigma^2/k^2$ where $E(X)=\mu$ and $V(X)=\sigma^2$.

Proof of Chebshev's inequality

$$\sigma^2 = \int (x-\mu)^2 dF$$

$$= \int_{-\infty}^{\mu-k} (x-\mu)^2 dF + \int_{\mu-k}^{\mu+k} (x-\mu)^2 dF$$

$$+ \int_{\mu+k}^{\infty} (x-\mu)^2 dF$$

Proof of WLLN (cont'd)

Put in the largest value of x in the first and smallest in the last integral, and drop the middle to get:

$$\sigma^2 \geq k^2 P(|x - \mu| \geq k)$$

Proof of WLLN: Since we are interested in $\overline{X_n}$, note that

$$E(\overline{X_n}) = \mu \text{ and } V(\overline{X_n}) = \sigma^2/n.$$

Consequently,

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) \leqq \lim_{n\to\infty} \sigma^2/n\varepsilon^2 = 0.$$

Notes:

1. $E(X_i) = \mu_i$ is okay. Consider

$$\bar{X}_n - \bar{\mu}_n \text{ with } \bar{\mu}_n = n^{-1} \sum \mu_i.$$

2. $V(X_i) = \sigma_i^2$ is okay. As long as $\lim \sum \sigma_i^2 / n^2 = 0$, our proof applies.

3. Existence of σ^2 can be dropped if we assume independent and identically distributed observations.

In this case, the proof is different and is based on Markov's inequality

$$P(|X| \geq k) \leq E|X|/k$$

from which Chebyshev's inequality follows.

Strong Law of Large Numbers:

If X_i are *independent* with $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2$ and $\sum \sigma_i^2 / i^2 < \infty$. Then $\bar{X}_n - \bar{\mu}_n \to 0$ almost surely (a.s.).

We can drop the existence of σ_i^2 if we assume independent and identically distributed observations.

Example (Shiryayev): let the probability space be [0,1) with Lebesgue measure (length of intervals). To each element ω of [0,1), there is a sequence $\{x_i\}$where $x_i$is the ith element in the dyadic expansion of ω,i.e. ω=0.$x_1 x_2 ...$, $x_i \in \{0, 1\}$. Then
$P(X_1 = x_1, ..., X_n = x_n) = P(x_1/2 + x_2/2^2 + ...x_n/2^n \leq \omega <$
$x_1/2 + x_2/2^2 + ...x_n/2^n + 1/2^n) = 1/2^n$.Thus
$P(X = 1) = P(X = 0) = 1/2$ and the obs are iid. By the SLLN,
$\sum X_i/n \to 1/2$.
Interpretation? Borel result on normal numbers.

Weak Convergence in Distribution

Definition: (*Convergence in distribution*):

A sequence of random variables $\{X_n, n \geq 1\}$ with distribution functions $\{F_n(x) = P(X_n \leq x), n \geq 1\}$ is said to *converge in distribution* to a random variable X with distribution function $F(x)$ if and only if $\lim_{n\to\infty} F_n(x) = F(x)$ at all points of continuity of $F(x)$.

Notation: $X_n \xrightarrow{D} X$.
plim is a special case in which F is a degenerate distribution.

More on Convergence in Distribution

An equivalent characterization is:

$$Ef(X_n) \to Ef(X)$$

for all bounded continuous functions f.Another is

$$P(X_n \in B) \to P(X \in B)$$

for all sets B with $P(\partial B) = 0$.

We have $X_n \xrightarrow{as} \Rightarrow X_n \xrightarrow{p} \Rightarrow X_n \xrightarrow{d}$

Continuous Mapping Theorem

Convergence in distribution is used to approximate the distribution of estimators.

If an estimator is consistent (plim=true value), studying the limiting distribution nontrivially requires norming.

CMT: Let $g(x)$ be continuous on a set which has probability one. Then

$$
\begin{aligned}
X_n \xrightarrow{d} X &\Rightarrow g(X_n) \xrightarrow{d} g(X) \\
X_n \xrightarrow{p} X &\Rightarrow g(X_n) \xrightarrow{p} g(X) \\
X_n \xrightarrow{as} X &\Rightarrow g(X_n) \xrightarrow{as} g(X)
\end{aligned}
$$

The CMT is extremely useful. Why?

Some properties of convergence in probability (plim) and convergence in distribution:

1. X_n and Y_n are random variable sequences. If $\text{plim}(X_n - Y_n) = 0$ and $Y_n \xrightarrow{D} Y$, then $X_n \xrightarrow{D} Y$ as well. This is an extremely useful device.

2. If $Y_n \xrightarrow{D} Y$ and $X_n \to c$ in probability (i.e., $\text{plim} X_n = c$), then

   a. $X_n + Y_n \xrightarrow{D} c + Y$

   b. $X_n Y_n \xrightarrow{D} cY$

   c. $Y_n / X_n \xrightarrow{D} Y/c$, $c \neq 0$.

"Big O and little o"

This notation is used to denote relative orders of magnitude of sequences in the limit.

Sequences $\{x_i\}, \{b_i\}$ (nonstochastic, for now)

$$x_n = O(b_n) \Rightarrow \lim_{n\to\infty} x_n/b_n = -\infty < c < \infty$$
$$x_n = o(b_n) \Rightarrow \lim_{n\to\infty} x_n/b_n = 0$$

Thus

$$x_n = o(1) \Rightarrow x_n \to 0; x_n = o(n) \Rightarrow x_n/n \to 0$$
$$x_n = O(1) \Rightarrow x_n \to c; x_n = O(n) \Rightarrow x_n/n \to c$$

Stochastic Versions

For stochastic sequences $\{X_i\}$ we have

$$X_n = O_p(b_n) \Rightarrow \forall \epsilon \exists C \text{ such that } \lim_{n \to \infty} P(|X_n/b_n| < C) > 1 - \epsilon$$

This says that the ratio remains bounded in probability. Also

$$X_n = o_p(b_n) \Rightarrow p \lim X_n/b_n = 0$$

Thus for example (using results above) if

$$X_n - Y_n = o_p(1) \text{ and } X_n \xrightarrow{d} X$$

Then $Y_n \xrightarrow{d} X$

Further Properties of O_p and o_p

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

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

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

$$(1 + o_p(1))^{-1} = O_p(1)$$

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

$$O_p(b_n) = b_n O_p(1)$$