Asymptotics Review

Harvard Math Camp - Econometrics

Ashesh Rambachan

Summer 2018

Outline

Why Asymptotics?

Can we still say something about the behavior of our estimators without strong, parametrics assumptions (e.g. i.i.d. normal errors)? We can *in large samples*.

- How would my estimator behave in very large samples?
- Use the limiting behavior of our estimator in infinitely large samples to approximate its behavior in finite samples.

Advantage: As the sample size gets infinitely large, the behavior of most estimators becomes very simple.

- Use appropriate version of CLT...

Disadvantage: This is only an *approximation* for the true, finite-sample distribution of the estimator and this approximation may be quite poor.

- Two recent papers by Alwyn Young: "Channelling Fisher" and "Consistency without Inference."

Outline

Stochastic 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 and there are many ways to do so.

Outline

Almost sure convergence

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

$$X_n \xrightarrow{a.s} X.$$

Almost sure convergence: In English

For a given outcome ω in the sample space Ω, 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

$$X_n \xrightarrow{a.s.} X$$

.

Outline

Convergence in probability

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

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

Convergence in probability: In English

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 $X_n \xrightarrow{p} X$.

Outline

Convergence in mean and mean-square

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

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

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

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

We write

$$X_n \xrightarrow{m.s.} X.$$

Outline

Convergence in distribution

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 $F(x)$ is continuous.

There are many ways of writing this

$$X_n \xrightarrow{d} X$$

$$X_n \xrightarrow{\mathcal{L}} X$$

$$X_n \implies X.$$

We'll use $X_n \xrightarrow{d} X$.

Convergence in distribution: In English

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 $\omega \in \Omega$.

Convergence in distribution: Continuity?

Why is convergence in distribution restricted to the continuity points of $F(x)$?

Example: Let $X_n = 1/n$ with probability 1 and let $X = 0$ with probability one. Then,

$$F_n(x) = 1(x \geq 1/n)$$
$$F(x) = 1(x \geq 0)$$

with $F_n(0) = 0$ for all n while $F(0) = 1$.

- 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 $F(x)$ but NOT at $x = 0$!
- If we did not restrict convergence in distribution to the continuity points, strange case where a non-stochastic sequence $\{X_n\}$ converges to X under the non-stochastic definition of convergence but not converge in distribution.

Multivariate Convergence

We can extend each of these definitions to random vectors.

- The sequence of random vectors $\{X_n\} \xrightarrow{a.s} X$ if each element of X_n converges almost surely to each element of X. 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.

Cramer-Wold Device: Let $\{Z_n\}$ be a sequence of k-dimensional random vectors. Then, $Z_n \xrightarrow{d} Z$ if and only if $\lambda' Z_n \xrightarrow{d} \lambda' Z$ for all $\lambda \in \mathbb{R}^k$.

- Simpler characterization of convergence in distribution for random vectors.

Outline

How do they relate to each other?

How do these different definitions of stochastic convergence relate to each other? See picture below.

- We will skip the results but see the notes if you want more details.

$$
\begin{array}{ccc}
\text{almost surely} & & \text{in mean squared} \\
\Downarrow & & \Downarrow \\
\text{in probability} & \Longleftarrow & \text{in mean} \\
\Downarrow & & \\
\text{in distribution} & &
\end{array}
$$

Counter-examples

Almost sure convergence does not imply convergence in mean.

Example: Let X_n be a random variable with

$$P(X_n = 0) = 1 - \frac{1}{n^2}$$

$$P(X_n = 2^n) = \frac{1}{n^2}.$$

$X_n \xrightarrow{as} 0$ but $E[X_n]$ does converge in mean to 0.

Counter-examples

Almost sure convergence does not imply convergence in mean square.

Example: Let X_n be a random variable with

$$P(X_n = 0) = 1 - \frac{1}{n^2}$$

$$P(X_n = n) = \frac{1}{n^2}$$

Then, $X_n \xrightarrow{as} 0$ but $E[X_n^2] = 1$ for all n.

Outline

- Types of Convergence
  - Almost sure convergence
  - Convergence in probability
  - Convergence in mean and mean-square
  - Convergence in distribution
  - How do they relate to each other?
  - Slutsky's Theorem and the Continuous Mapping Theorem
- O_p and o_p Notation
- Law of Large Numbers
- Central Limit Theorem
- The Delta Method

Slutsky's Theorem

Slutsky's Theorem: Let c be a constant. Suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} Y$. Then,

1. $X_n + Y_n \xrightarrow{d} X + c$.
2. $X_n Y_n \xrightarrow{d} Xc$.
3. $X_n / Y_n \xrightarrow{d} X/c$ provided that $c \neq 0$.

If $c = 0$, then $X_n Y_n \xrightarrow{p} 0$.

Continuous Mapping Theorem

Continuous Mapping Theorem: Let g be a continuous function. Then,

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

Outline

big-O, little-o

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

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

We also extend big-O and little-o notation to random variables

O_p and o_p definition

Suppose $\{A_n\}$ is a sequence of random variables. We write

$$A_n = o_p(G_n) \quad \text{if} \quad \frac{A_n}{G_n} \xrightarrow{p} 0$$

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.

- Often see $X_n = X + o_p(1)$ to denote $X_n \xrightarrow{p} X$.

Simple Examples

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

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

Why?

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

Moreover,

$$X_n = o_p(n)$$

Why?

- $X_n/n \sim N(0, 1/n)$. So,

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

- Alternatively, note that

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

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

Outline

Law of Large Numbers

First building block of asymptotic results: Law of Large Numbers

Provides conditions under which sample averages converge to expectations.

We'll discuss three of them.

Weak Law of Large Numbers

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

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu.$$

<u>Proof</u>: By Chebyshev's inequality,

$$P(|\bar{X}_n - \mu| > \epsilon^2) =\leq E[(\bar{X}_n - \mu)^2]/\epsilon^2 = \sigma^2/n\epsilon^2 \rightarrow 0.$$

Alternatively,

$$V(\bar{X}_n) = E[(\bar{X}_n - \mu)^2] = \sigma^2/n \rightarrow 0, \quad E[\bar{X}_n] = \mu$$

and so, $\bar{X}_n \xrightarrow{ms} \mu$ and the result follows.

Chebyshev's Weak Law of Large Numbers

Chebyshev's WLLN: 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 $Cov(X_i, X_j) = 0$ 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,

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$$

.

Chebyshev's WLLN Proof

First,

$$E[\bar{X}_n - \bar{\mu}_n] = 0.$$

Second,

$$\begin{aligned} V(\bar{X}_n - \bar{\mu}_n) &= V(\bar{X}_n) \\ &= \frac{1}{n^2} \sum_{i,j} Cov(X_i, X_j) \\ &= \frac{1}{n^2} \sum_{i} \sigma_i^2 = \bar{\sigma}_n^2 / n \to 0. \end{aligned}$$

Therefore, $\bar{X}_n - \bar{\mu}_n \xrightarrow{ms} 0$ and so, $\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$.

Strong LLN

Strong LLN: If $X_1, X_2, \ldots$ are i.i.d with $E[X_i] = \mu < \infty$, then

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

Outline

Central Limit Theorem

Second building block of asymptotic results: Central Limit Theorems

Provides conditions under which properly centered sample averages will converge in distribution to normal random variables.

We'll discuss two of them.

Central Limit Theorem I

CLT I: Let $Y_1, Y_2, \dots$ be an i.i.d. sequence of random variables with $E[Y_i] = 0, V(Y_i) = 1$ for all i. Then,

$$\sqrt{n}\bar{Y} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \xrightarrow{d} N(0, 1).$$

Central Limit Theorem II

CLT II: Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with mean μ and variance σ^2. 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 μ and covariance matrix Σ. Then,

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

Exercise

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

Outline

Motivation

Suppose we have some estimator T_n of a parameter θ. We know that

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

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

We are interested in estimating and conducting inference on $g(\theta)$, where g is some continuously differentiable function.

- Natural estimator is $g(T_n)$ and by CMT, we know that

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

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

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

The Delta Method

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,

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

Multivariate Extension: he result becomes

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

where

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

Delta Method: Proof Sketch

By the mean value theorem,

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

where $\tilde{Y}_n$ is some value between Y_n and a.

- Recall mean value theorem. Let $g(\cdot)$ be a continuously differentiable function and WLOG, let $a < b$. There exists some $c \in (a, b)$ such that $g(b) = g(a) + g'(c)(b - a)$.

We have that $Y_n \xrightarrow{p} a$. Since g is continuously differentiable, it follows that $g'(\tilde{Y}_n) \xrightarrow{p} g'(a)$. Why? So, it follows that

$$\begin{aligned}\sqrt{n}(g(Y_n) - g(a)) &= g'(\tilde{Y}_n)\sqrt{n}(Y_n - a) \\ &= g'(\tilde{Y}_n)X_n \xrightarrow{d} g'(a)X\end{aligned}$$

by Slutsky's theorem.