An Introduction to Large Sample Asymptotics

[Hansen, chapter 6]

Simon Lee

ECON G6411

Fall 2019

Outline

Limit Superior and Limit Inferior of a Sequence of Real Numbers

- The limit superior and limit inferior of a sequence x_n of real numbers are defined as

$$\limsup_{n\to\infty} x_n = \inf_{n\geq 1} \sup_{k\geq n} x_k,$$

$$\liminf_{n\to\infty} x_n = \sup_{n\geq 1} \inf_{k\geq n} x_k.$$

- The limit superior always exists as a real number if x_n is bounded from above. Likewise, the limit inferior exists as a real number whenever x_n is bounded from below.
- Note that we always have

$$\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n.$$

- The limit $\lim_{n\to\infty} x_n$ exists if and only if

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n.$$

- We have that $\liminf_{n\to\infty} x_n = -\limsup_{n\to\infty} -x_n$.

Limit Superior and Limit Inferior of a Sequence of Sets

- The limit superior and limit inferior of a sequence E_n of sets are defined as

$$\limsup_{n\to\infty} E_n = \cap_{n\geq 1}^{\infty} \cup_{k\geq n}^{\infty} E_k,$$
$$\liminf_{n\to\infty} x_n = \cup_{n\geq 1}^{\infty} \cap_{k\geq n}^{\infty} E_k.$$

- A point belongs to $\limsup_{n\to\infty} E_n$ if and only if it belongs to infinitely many terms of the sequence $\{E_n, n \geq 1\}$. Hence, sometimes we write $\limsup_{n\to\infty} E_n = E_n$ *i.o.* ("infinitely often").
- A point belongs to $\liminf_{n\to\infty} E_n$ if and only if it belongs to all terms of the sequence from a certain term on.
- We have that $\liminf_{n\to\infty} E_n = (\limsup_{n\to\infty} E_n^c)^c$.
- We also have that

$$P(\limsup_{n\to\infty} E_n) = \lim_{n\to\infty} P(\cup_{k\geq n}^{\infty} E_k),$$
$$P(\liminf_{n\to\infty} x_n) = \lim_{n\to\infty} P(\cap_{k\geq n}^{\infty} E_k).$$

References

- The following materials on stochastic convergence are closely based on Sections 2.1 and 2.4 of A. W. van der Vaart (1998), Asymptotic Statistics, Cambridge University Press.
- Other useful references are:
  - P. Billingsley (1995), Probability and Measure, 3rd ed., Wiley.
  - K.L. Chung (2001), A course in Probability Theory, 3rd ed., Academic Press.

Convergence in Distribution

- A sequence of random vectors X_n is said to *converge in distribution* to a random vector X if

$$P(X_n \leq x) \to P(X \leq x)$$

  for every x at which the limit distribution function $x \mapsto P(X \leq x)$ is continuous.
- Alternatively, they are called *weak convergence* and *convergence in law*.
- Weak convergence is denoted by $X_n \to_d X$.
- Note that the convergence only depends on the induced laws (but not on the probability spaces on which they are defined). Hence, sometimes we write $X_n \to_d L$ if X has distribution L (e.g. $X_n \to_d N(0, 1)$).

Convergence in Distribution: Example

- If $X_1, X_2, \ldots$ are iid uniform(0,1) and $X_{(n)} = \max_{1 \le i \le n} X_i$.
- Note that for every $t > 0$,

$$P\left(X_{(n)} \le 1 - \frac{t}{n}\right) = \left[1 - \frac{t}{n}\right]^n \to e^{-t},$$

equivalently

$$P\left(n(1 - X_{(n)}) \le t\right) \to 1 - e^{-t},$$

which implies that $n(1 - X_{(n)})$ converges in distribution to an exponential(1) random variable.

Properties of Weak Convergence

- (Portmanteau Theorem) The following statements are equivalent.
  - $X_n \to_d X_n$.
  - $Ef(X_n) \to Ef(X)$ for all bounded, continuous real-valued functions f.
  - $Ef(X_n) \to Ef(X)$ for all bounded, Lipschitz continuous real-valued functions f.
  - $\liminf Ef(X_n) \geq Ef(X)$ for all nonnegative, continuous functions f.
  - $\liminf P(X_n \in G) \geq P(X \in G)$ for every open set G.
  - $\limsup P(X_n \in F) \leq P(X \in F)$ for every closed set F.
  - $P(X_n \in B) \to P(X \in B)$ for all Borel sets B with $P(X \in \delta B) = 0$, where δB is the boundary of B.
- See Van der Vaart (1998) for the proof.

Convergence in Probability

- A sequence of random vectors X_n is said to *converge in probability* to a random vector X if for every $\epsilon > 0$,
$$P(\|X_n - X\| > \epsilon) \to 0, \tag{1}$$
where $\|\cdot\|$ is the Euclidean norm.
- An alternative expression to (1) is
$$P(\|X_n - X\| \leq \epsilon) \to 1 \tag{2}$$
for every $\epsilon > 0$.
- Convergence in probability is denoted by $X_n \to_p X$. Equivalently, $\|X_n - X\| \to_p 0$.

Almost Sure Convergence

- A sequence of random vectors X_n is said to *converge almost surely (a.s.)* to a random vector X if for every $\epsilon > 0$,

$$\lim_{m\to\infty} P(\|X_n - X\| > \epsilon \text{ for some } n \geq m) = 0. \tag{3}$$

- An alternative expression to (3) is

$$\lim_{m\to\infty} P(\|X_n - X\| \leq \epsilon \text{ for all } n \geq m) = 1 \tag{4}$$

  for every $\epsilon > 0$.
- Almost sure is denoted by $X_n \to_{a.s.} X$.

Almost Sure Convergence (Cont.)

- Almost sure convergence is similar to pointwise convergence of a sequence of functions, except that the convergence need not occur on a set with probability 0 (hence, the "almost sure" or sometimes we say "almost everywhere").
- Given the definition, it is obvious that $X_n \rightarrow_{a.s.} X$ implies that $X_n \rightarrow_p X$. (The converse is not true)
- Note that convergence in probability and almost sure convergence only make sense if X_n and X are defined on the same probability space for each n (recall that this was not necessary for convergence in distribution).

Continuous Mapping Theorems

- Let $g : \mathbf{R}^k \mapsto \mathbf{R}^m$ be continuous at every point of a set C such that $P(X \in C) = 1$.
  (i) If $X_n \to_d X$, then $g(X_n) \to_d g(X)$.
  (ii) If $X_n \to_p X$, then $g(X_n) \to_p g(X)$.
  (iii) If $X_n \to_{a.s.} X$, then $g(X_n) \to_{a.s.} g(X)$.
- This is a simple, but very useful result.

Continuous Mapping Theorems: Proof

- Proof: part (iii) is obvious.
- To prove part (ii), fix arbitrary $\epsilon > 0$. Then for any $\delta > 0$, note that

$$
\begin{aligned}
& P(\|g(X_n) - g(X)\| > \epsilon) \\
& = P(\|g(X_n) - g(X)\| > \epsilon \text{ and } \|X_n - X\| < \delta) \\
& \quad + P(\|g(X_n) - g(X)\| > \epsilon \text{ and } \|X_n - X\| \geq \delta) \\
& \leq P(\|g(X_n) - g(X)\| > \epsilon \text{ and } \|X_n - X\| < \delta) \\
& \quad + P(\|X_n - X\| \geq \delta).
\end{aligned}
$$

- The second term above converges to zero as $n \to \infty$ for every fixed $\delta > 0$ since $X_n \to_p X$.
- The first term converges to zero as $\delta \downarrow 0$. (In fact, this is a loose statement. A more rigorous proof requires more careful treatment here).
- Therefore, we proved part (ii).

Continuous Mapping Theorems: Proof (Cont.)

- To prove part (i), note that the event $\{g(X_n) \in F\}$ is identical to the event $\{X_n \in g^{-1}(F)\}$. For every closed set F,

$$g^{-1}(F) \subset \overline{g^{-1}(F)} \subset g^{-1}(F) \cup C^c.$$

- Then by the portmanteau theorem,

$$\begin{aligned}\limsup P(g(X_n) \in F) &= \limsup P(X_n \in g^{-1}(F)) \\ &\leq \limsup P(X_n \in \overline{g^{-1}(F)}) \\ &\leq P(X \in \overline{g^{-1}(F)}).\end{aligned}$$

  Furthermore, since $P(X \in C) = 1$,

$$P(X \in \overline{g^{-1}(F)}) = P(X \in g^{-1}(F)) = P(g(X) \in F).$$

- Then the desired result follows from the portmanteau theorem again.

Almost-Sure Representation

- Convergence in distribution does not imply convergence in probability or almost surely.
- However, for a given sequence $X_n \to_d X$, we can always find a sequence $\tilde{X}_n \to_{a.s.} \tilde{X}$ such that $X_n =_d \tilde{X}_n$ and $X =_d \tilde{X}$ for each n.
- Here, $X =_d Y$ means X and Y have the same marginal distributions.

Skorohod-Dudley-Wichura theorem

- (Skorohod-Dudley-Wichura theorem) Suppose that $X_n \to_d X_0$. Then there exists a probability space such that
  - $\tilde{X}_n$ and $\tilde{X}_0$ defined on this probability space;
  - $\tilde{X}_n$ has a probability distribution that is the same as that of X_n for each $n \geq 0$;
  - $\tilde{X}_n \to_{a.s.} \tilde{X}_0$.
- In view of Skorohod-Dudley-Wichura theorem, it is obvious that the continuous mapping theorem holds for all three modes of convergence. See Theorems 25.6 and 25.7 of Billingsley (1995, 3rd ed.).

Relationships among the Modes of Convergence

(i) $X_n \to_{a.s.} X$ implies that $X_n \to_p X$.

- Proof: this follows from the definitions of the modes of convergence. Note that

$$\begin{aligned} P(\|X_n - X\| > \epsilon) &\leq P(\cup_{m \geq n} \|X_m - X\| > \epsilon) \\ &= P(\|X_m - X\| > \epsilon \text{ for some } m \geq n) \\ &\to 0 \end{aligned}$$

as $n \to \infty$.

Relationships among the Modes of Convergence (Cont.)

(ii) $X_n \to_p X$ implies that $X_n \to_d X$.

- Proof:
  - Note that for every f with range $[0, 1]$ and Lipschitz norm at most 1 and every $\epsilon > 0$,

$$|Ef(X_n) - Ef(X)| \leq \epsilon E1\{\|X_n - X\| \leq \epsilon\} + 2E1\{\|X_n - X\| > \epsilon\}.$$

  - The second term converges to zero as $n \to \infty$.
  - The first term can be arbitrarily small by choosing a sufficiently small ϵ.
  - Thus, $Ef(X_n)$ converges to $Ef(X)$ for all bounded, Lipschitz continuous functions f.
  - Now the result follows from the portmanteau theorem.

Relationships among the Modes of Convergence (Cont.)

(iii) $X_n \to_p c$ for a constant c if and only if $X_n \to_d c$.

- Proof: it suffices to show that $X_n \to_d c$ implies that $X_n \to_p c$.
- To show this, let $B(c, \epsilon)$ be a open ball of radius ϵ around c. Then

$$P(\|X_n - c\| \geq \epsilon) = P(X_n \in B(c, \epsilon)^c).$$

- Note that

$$\limsup P(X_n \in B(c, \epsilon)^c) \leq P(c \in B(c, \epsilon)^c) = 0$$

by the portmanteau theorem.

Relationships among the Modes of Convergence (Cont.)

(iv) If $X_n \to_d X$ and $\|X_n - Y_n\| \to_p 0$, then $Y_n \to_d X$.

- Proof: this can be proved as in the proof of part (ii).

Relationships among the Modes of Convergence (Cont.)

(v) If $X_n \to_p X$ and $Y_n \to_p Y$, then $(X_n, Y_n) \to_p (X, Y)$.

- Proof: this follows from

$$\|(X_n, Y_n) - (X, Y)\| \leq \|X_n - X\| + \|Y_n - Y\| .$$

- This states that convergence in probability of a sequence of random vectors is equivalent to convergence of each component separately.

Relationships among the Modes of Convergence (Cont.)

(vi) If $X_n \to_d X$ and $Y_n \to_p c$ for a constant c, then $(X_n, Y_n) \to_d (X, c)$.

- Proof: We can prove this by applying part (iv). First of all,

$$\|(X_n, Y_n) - (X_n, c)\| = \|Y_n - c\| \to_p 0.$$

- Now it suffices to show that $(X_n, c) \to_d (X, c)$. For every bounded, continuous function $(x, y) \mapsto f(x, y)$, the function $x \mapsto f(x, c)$ is continuous and bounded.
- Hence, $Ef(X_n, c) \to Ef(X, c)$ if $X_n \to X$ (by the portmanteau theorem).
- Then the theorem follows again from the portmanteau theorem.

Joint vs. Marginal Convergence in Distribution

- In general, $X_n \rightarrow_d X$ and $Y_n \rightarrow_d Y$ would not imply that $(X_n, Y_n) \rightarrow_d (X, Y)$.
- Convergence in the distribution of the sequence (X_n, Y_n) is stronger than convergence of X_n and Y_n separately.
- The implication of part (vi) is that if $X_n \rightarrow_d X$ and $Y_n \rightarrow_d c$ for a constant c, then $(X_n, Y_n) \rightarrow_d (X, c)$.
- It follows from the continuous mapping theorem that if $X_n \rightarrow_d X$ and $Y_n \rightarrow_d c$ for a constant c, then

$$g(X_n, Y_n) \rightarrow_d g(X, c)$$

for every continuous map g.

Slutsky's Lemma

- As special cases to the previous result, we have the following result.
- (Slutsky's Lemma) If $X_n \to_d X$ and $Y_n \to_d c$ for a constant c, then
  - $X_n + Y_n \to_d X + c$.
  - $Y_n X_n \to_d cX$.
  - $Y_n^{-1} X_n \to_d c^{-1} X$ provided that $c \neq 0$.

Convergence in L_p

- A sequence of random vectors X_n is said to *converge in L_p* to a random vector X if

$$\lim_{n\to\infty} E(\|X_n - X\|^p) \to 0.$$

- Convergence in L_p is denoted by $X_n \to_{L_p} X$.
- (Theorem) If X_n converges in L_p, then it converges to 0 in probability.
- Proof: By Chebyshev's inequality,

$$P(\|X_n - X\| \geq \epsilon) = P(\|X_n - X\|^p \geq \epsilon^p) \leq \frac{E(\|X_n - X\|^p)}{\epsilon^p},$$

which proves the theorem if we let $n \to \infty$.

When is convergence in L_p implied by convergence in probability?

- First, assume that X belongs to L_p (an obvious condition to think about the convergence in L_p).
- Suppose that X_n is dominated by some Y that belongs to L_p, that is $\|X_n\| \leq Y$ a.s. with $EY^p < \infty$.
- Note that

$$\begin{aligned} E(\|X_n - X\|^p) &\leq E\left[1(\|X_n - X\| < \epsilon)\,\|X_n - X\|^p\right] \\ &\quad + E\left[1(\|X_n - X\| \geq \epsilon)\,\|X_n - X\|^p\right] \\ &\leq \epsilon^p + E\left[1(\|X_n - X\| \geq \epsilon)(Y + \|X\|)^p\right]. \end{aligned}$$

- Since $Y + \|X\|$ is in L_p, we have shown that $X_n \to_p X$ implies that $X_n \to_{L_p} X$ by letting first $n \to 0$ and then $\epsilon \to 0$.

Borel-Cantelli Lemma

- For any sequence of events $\{E_n : n = 1, \ldots\}$, if
$$\sum_{n=1}^{\infty} P(E_n) < \infty,$$
then
$$P(\limsup_{n\to\infty} E_n) = 0.$$
- Proof: the lemma follows from the fact that
$$P(\limsup_{n\to\infty} E_n) = \lim_{n\to\infty} P(\cup_{k\geq n}^{\infty} E_k) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(E_k).$$
- Thus, to show that $X_n \to_{a.s.} X$, it suffices to show that there exists an $\epsilon > 0$ satisfying
$$\sum_{n=1}^{\infty} P(\|X_n - X\| > \epsilon) < \infty.$$

Lévy's Continuity Theorem

- Let X_n and X be random vectors in $\mathbf{R}^k$. Then $X_n \to_d X$ if and only if $Ee^{it'X_n} \to Ee^{it'X}$ for every $t \in \mathbf{R}^k$.
- Moreover, if $Ee^{it'X_n}$ converges to a function $\phi(t)$ that is continuous at zero, then ϕ is the characteristic function of a random vector X and $X_n \to_d X$.

Laws of Large Numbers

- Let $\bar{X}_n$ denote sample average of iid random vectors $X_1, \ldots, X_n$.
- Suppose that $E \|X_1\| < \infty$.
- Then $\bar{X}_n \to_{a.s.} EX_1$ (the strong law of large numbers (SLLN)).
- Obviously, this implies that $\bar{X}_n \to_p EX_1$ (the weak law of large numbers (WLLN)).
- The WLLN can be proved easily by Chebyshev's inequality, if we make a stronger assumption that $E \|X_1\|^2 < \infty$ (prove this by yourself).

Laws of Large Numbers (Cont.)

- We now prove WLLN using the characteristic function.
- Let $X_1, \ldots, X_n$ be iid random variables with characteristic function ϕ.
- Then $\bar{X}_n \to_p \mu$ if ϕ is differentiable at zero with $\phi'(0) = i\mu$.
- Proof: Since $\phi(0) = 1$, $\phi(t) = 1 + t\phi'(0) + o(t)$ as $t \to 0$.
- Note that

$$
\begin{aligned}
Ee^{it\bar{X}_n} &= \prod_{i=1}^{n} Ee^{itX_i/n} = [\phi(t/n)]^n \\
&= \left[1 + \frac{t}{n}\phi'(0) + o\left(n^{-1}\right)\right]^n \to e^{it\mu},
\end{aligned}
$$

  which is the characteristic function of the constant variable μ.
- A sufficient condition for $\phi(t)$ to be differentiable at zero is that $E|X_1| < \infty$. Also, in this case, $\mu = EX_1$.

Central Limit Theorem

- Let $X_1, \ldots, X_n$ be iid random variables with $EX_1 = 0$ and $EX_1^2 = 1$.
- Then $n^{1/2}\bar{X}_n \to_d N(0,1)$.
- Proof: Since $EX_1^2 < \infty$, ϕ is twice differentiable. Also, note that $\phi'(0) = iEX_1 = 0$ and $\phi''(0) = i^2 EX_1^2 = -1$. By Taylor expansion, $\phi(t) = 1 - (t^2/2) + o(t^2)$ as $t \to 0$.
- Now note that

$$
\begin{aligned}
Ee^{itn^{1/2}\bar{X}_n} &= \prod_{i=1}^{n} Ee^{itX_i/n^{1/2}} = [\phi(t/n^{1/2})]^n \\
&= \left[1 - \frac{t^2}{2n} + o\left(n^{-1}\right)\right]^n \to e^{-\frac{1}{2}t^2},
\end{aligned}
$$

which is the characteristic function of $N(0,1)$.

Cramér-Wold Device

- (Cramér-Wold device) $X_n \to_d X$ if and only if $t'X_n \to t'X$ for all $t \in \mathcal{R}^k$.
- In other words, the Cramér-Wold device allows to reduce higher-dimensional problems to the one-dimensional case.
- Why? (prove this by yourself).

Delta Method

- Let Y_n be a sequence of random variables that satisfies
$$\sqrt{n}(Y_n - \theta) \to_d N(0, \sigma^2).$$
- For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not zero.
- Then
$$\sqrt{n}(g(Y_n) - g(\theta)) \to_d N(0, \sigma^2[g'(\theta)]^2).$$
- Why? $[g(Y_n) - g(\theta)] \approx g'(\theta)(Y_n - \theta)$ (prove this rigorously by yourself).

Delta Method: Example

- Let $X_1, \ldots, X_n$ be iid random variables with $EX_1 = \mu \neq 0$ and $EX_1^2 < \infty$.
- Then
$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right) \to_d N\left(0, \mu^{-4}\text{Var}(X_1)\right).$$