Econ 508B: Lecture 9

Stochastic Convergence and Orders

Hongyi Liu

Washington University in St. Louis

August 2, 2017

Outline

Outline

Motivation:Asymptotic Arguments in Statistics

Good large sample performance is a basic requirement of any statistical procedure. What is more,

- Approximating solutions to intractable problems, where asymptotic arguments allow us to understand the error properties;
- Distributional approximations, especially asymptotically pivotal quantities;
- Studying power and accuracy of tests;
- Breakdown points of estimators, etc.

Stochastic Convergence

Definition 1.1 (Convergence in probability)

Let X_n and X be random variables on a common probability space.The sequence X_n **converges** in X **in probability**, is defined by $X_n \xrightarrow{p} X$ if, for any $\epsilon > 0, \exists n > N(\epsilon)$

$$P(|X_n - X| > \epsilon) \to 0 \text{ as } n \to \infty$$

The above probability is a joint distribution.

Definition 1.2 (Converges almost surely)

Let X_n and X be defined on the same probability space. The sequence X_n converges **almost surely** to X, denoted by $X_n \xrightarrow{a.s.} X$ if

$$P(\omega : X_n(\omega) \to X) = 1.$$

Continued

A characterization of almost sure convergence is that, for any $\epsilon > 0$,

$$\lim_{m \to \infty} P(|X_n - X| \leq \epsilon \ \forall n \geq m) = 1,$$

which makes clear that *convergence almost surely* is stronger than, and implies, convergence in probability.

Example 1.1

Consider $X_1, X_2, ...$ is an infinite sequence of i.i.d. $U[0,1]$ random variables and let $X_{(n)} = \max\{X_1, ..., X_n\}$. It is reasonable to speculate $X_{(n)} \xrightarrow{a.s.} 1$ as n goes to ∞. Actually

$$\begin{aligned}
P(|X_{(n)} - 1| \leq \epsilon \; \forall n \geq m) &= P(1 - X_{(n)} \leq \epsilon \; \forall n \geq m) \\
&= P(X_{(n)} \geq 1 - \epsilon \; \forall n \geq m) \\
&= P(X_{(m)} \geq 1 - \epsilon) \\
&= 1 - (1 - \epsilon)^m \\
&\to 1 \text{ as } m \to \infty \; ,
\end{aligned}$$

Hence, $X_{(n)} \xrightarrow{a.s.} 1$.

LLNs for sums of i.i.d. random variables

Let $\{X_i\}$ indicate a sequence of random variables, $i = 1, 2, ...$, $\{S_n\}$ is a sequence of sums, where $S_n = \sum_{i=1}^{n} X_i$, and $\{\bar{X}_n\}$ is a sequence of sample means, where $\bar{X}_n = S_n/n$.

Theorem 1.1 (Khintchine's Weak LLN)

Let $\{X_i\}$ be a sequence of i.i.d. random variables with $E(X_i) = \mu$ and $E(|X_i|) < +\infty$. Then $\bar{X}_n \xrightarrow{p} \mu$.

Theorem 1.2

Let $\{X_i\}$ be a sequence of i.i.d. random variables with $E(X_i) = \mu$ and $E(|X_i|) < +\infty$. Then $\bar{X}_n \xrightarrow{a.s.} \mu$.

Outline

Convergence in Law/Distribution

Definition 2.1

Let X_n and X be real-valued random variables defined on a common probability space. The sequence X_n **converges in distribution** to X if

$$F_n(x) = P(X_n \leq x) \to P(X \leq x) = F(x) \text{ as } n \to \infty$$

for all x at which $F(x)$ is continuous.

Convergence in Law/Distribution

Definition 2.1

Let X_n and X be real-valued random variables defined on a common probability space. The sequence X_n **converges in distribution** to X if

$$F_n(x) = P(X_n \leq x) \rightarrow P(X \leq x) = F(x) \text{ as } n \rightarrow \infty$$

for all x at which $F(x)$ is continuous.

Question: why only for continuity points?

Response by an example

Let $X_1, X_2, ...$ be a sequence of random variables such that

$$P(X_n = 1/n) = 1,\ n = 1, 2, ...$$

Let F_n be the distribution function of X_n,

$$F_n(x) = \begin{cases} 0,\ x < 1/n, \\ 1,\ x \geq 1/n. \end{cases}$$

Fix x, then

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0,\ x < 0 \\ 1,\ x > 0. \end{cases}$$

Continued

Consider the sequence $F_n(0), n = 1, 2, ...$ and $0 < 1/n, \forall n = 1, 2, ...$ Therefore,

$$\lim_{n \to \infty} F_n(0) = 0,$$

Hence,

$$\lim_{n \to \infty} F_n(x) = F(x) \equiv \begin{cases} 0, & x \leq 0 \\ 1, & x > 0. \end{cases}$$

Clearly, $F(x)$ is not *right-continuous*, hence it is not the CDF of any random variable.

Remark 2.1

Therefore, $\lim_{n \to \infty} F_n(x) = F(x)$ for all $x \neq 0$, namely all x at F is continuous.

Comments

(1) : The definition of *convergence in probability* only makes sense if, for each n, X_n and X are defined on the same underlying sample space. However, it is not required for *convergence in distribution*.

Comments

(1) : The definition of *convergence in probability* only makes sense if, for each n, X_n and X are defined on the same underlying sample space. However, it is not required for *convergence in distribution*.

(2) : *Convergence in distribution* only depends on the marginal distribution functions of X_n and X.

Comments

(1) : The definition of *convergence in probability* only makes sense if, for each n, X_n and X are defined on the same underlying sample space. However, it is not required for *convergence in distribution*.

(2) : *Convergence in distribution* only depends on the marginal distribution functions of X_n and X.

(3) : *Convergence in probability*, is a statement about the joint distribution of $|X_n - X|$, namely which depends on the joint distribution of X_n and X.

In general,

(i) *convergence in probability* **implies** *convergence in distribution*,

(ii) but the converse is not necessarily true unless $X_n \xrightarrow{d} X$ and $P(X = c) = 1$, where c is a constant.

Proof of (i): it is essentially important!!!

Let F_n, F denote the c.d.f of X_n, X, respectively. For $n \geq 0$, fix $x \in C(F)$. Then, for any $\epsilon > 0$, first of all,

$$P(X_n \leq x) \leq P(X \leq x + \epsilon) + P(X_n \leq x, X > x + \epsilon)$$
$$\leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon)$$

Then $\forall \epsilon > 0$,

$$\lim_{n\to\infty} \sup F_n(x) \leq F(x + \epsilon) \quad (*)$$

Likewise, $P(X \leq x - \epsilon) - P(|X_n - X| > \epsilon) \leq P(X_n \leq x)$. so that $\forall \epsilon > 0$,

$$\lim_{n\to\infty} \inf F_n(x) \geq F(x - \epsilon) \quad (**)$$

Combine $(*)\&(**)$, we have

$$F(x - \epsilon) \leq \lim_{n\to\infty} \inf F_n(x) \leq \lim_{n\to\infty} \sup F_n(x) \leq F(x + \epsilon)$$

Since F is continuous at x, $F(x + \epsilon) - F(x - \epsilon) \to 0$ completes the proof. Actually $(*)\&(**)$ come from **Portmanteau theorem**.

Proof of (ii)

Suppose that $X_n \xrightarrow{d} X$ and $P(X = c) = 1$, then X has the c.d.f

$$F(x) = \begin{cases} 0, & x < c, \\ 1, & x \geq c. \end{cases}$$

Since F is not continuous at $x = c$, it follows that

$$\lim_{n \to \infty} F_n(x) = \begin{cases} 0, & x < c, \\ 1, & x > c. \end{cases}$$

Fix $\epsilon > 0$, then

$$\begin{aligned} P(|X_n - c| \geq \epsilon) &= P(X_n \leq c - \epsilon) + P(X_n \geq c + \epsilon) \\ &\leq F_n(c - \epsilon) + 1 - F_n(c + \epsilon/2) \to 0 \text{ as } n \to \infty. \end{aligned}$$

It holds for all $\epsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - c| \geq \epsilon) = 0, \ \forall \epsilon > 0$$

Therefore, $X_n \xrightarrow{p} c$ as $n \to \infty$.

Outline

Review of Limits

Definition 3.1

Say a sequence $a_n, n = 1, 2, ...,$ if for any $\epsilon > 0, \exists\ N = N(\epsilon)$ such that,

$$|a_n - a| < \epsilon$$

for all $n > N$, then $a_n \to a$ as $n \to \infty$.

Example 3.1

Fix a finite constant c,

$$(1 + \frac{c}{n})^n \to e^c \text{ as } n \to \infty$$

Asymptotically equivalent

Consider two sequences $\{a_n\}$ and $\{b_n\}$ are **asymptotically equivalent** (as $n \to \infty$), written as $a_n \sim b_n$ if

$$a_n/b_n \to 1.$$

If b_n has a *finite* limit and it is non-zero, then a_n tends to the same limit. Note that if the limit of b_n is 0 or $\pm\infty$, saying that $a_n \sim b_n$ is informative. Especially note that **this does not say that** $\lim a_n / \lim b_n = 1$.

Example 3.2

Consider $a_n = \frac{1}{n} + \frac{3}{n^2} + \frac{1}{n^3} \to 0, b_n = \frac{1}{n} \to 0$, the ratio

$$\frac{a_n}{b_n} = 1 + \frac{3}{n} + \frac{1}{n^2} \to 1,$$

hence $a_n \sim b_n$.

Asymptotically equivalent

Consider two sequences $\{a_n\}$ and $\{b_n\}$ are **asymptotically equivalent** (as $n \to \infty$), written as $a_n \sim b_n$ if

$$a_n/b_n \to 1.$$

If b_n has a *finite* limit and it is non-zero, then a_n tends to the same limit. Note that if the limit of b_n is 0 or $\pm\infty$, saying that $a_n \sim b_n$ is informative. Especially note that **this does not say that** $\lim a_n / \lim b_n = 1$.

Example 3.2

Consider $a_n = \frac{1}{n} + \frac{3}{n^2} + \frac{1}{n^3} \to 0, b_n = \frac{1}{n} \to 0$, the ratio

$$\frac{a_n}{b_n} = 1 + \frac{3}{n} + \frac{1}{n^2} \to 1,$$

hence $a_n \sim b_n$.

Question: What if we replace the complicated a_n with the simpler b_n?

Absolute error v.s. Relative error:

Let $a_n = n + n^2, b_n = n^2$. Let $n = 100$, then

$$a_n = 10100,\ b_n = 10000$$

absolute error: $|a_n - b_n| = 100$, but the **relative error:**

$$|\frac{a_n - b_n}{a_n}| = \frac{100}{10100} < 0.01.$$

Proposition 3.1

$a_n \sim b_n$ *iff relative error* $\to 0$ as $n \to \infty$.

Order relations

We say $a_n = o(b_n)$ as $n \to \infty$ if

$$a_n/b_n \to 0.$$

When both tend to infinity, this means a_n tends to infinity slower than b_n. When both converge to zero, it means a_n converges to zero faster than b_n.

Example 3.3

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

Since $n^{-2}/n^{-1} = \frac{n}{n^2} \to 0.$

Order of a remainder

Suppose

$$a_n = \frac{1}{n} - \frac{2}{n^2} + \frac{4}{n^3}.$$

Let $b_n = 1/n$, $a_n \sim b_n$, though this approximation has error. Write R_n so that

$$a_n = \frac{1}{n} + R_n,$$

where

$$R_n = \frac{-2}{n^2} + \frac{4}{n^3} = o(1/n).$$

These two statements are usually summarized by

$$a_n = \frac{1}{n} + o(\frac{1}{n}),$$

Big 'O'

Say $a_n = O(b_n)$ as $n \to \infty$ if $\exists M > 0$ and $N > 0$ such that $|a_n/b_n| < M$ for all $n > N$. And naturally,

$$a_n = o(b_n) \Rightarrow \; a_n = O(b_n).$$

It is tempting to take $a_n = o(b_n)$ and try to apply a function to each side and claim $f(a_n) = o[f(b_n)]$. Unfortunately it is not true in general unless the following theorem:

Theorem 3.1

Suppose $a_1, a_2, ...$ and $b_1, b_2, ...$ are sequences of real numbers such that $a_n \to \infty, b_n \to \infty$ and $a_n = o(b_n)$; and $f(x)$ is a convex function such that $f(x) \to \infty$ as $x \to \infty$. Then $f(a_n) = o[f(b_n)]$.

Some growth rates have names as they are so common. Suppose α, β, γ are arbitrary positive constants:

- logarithmic: $(\log n)^{\alpha}$,
- polynomial: n^{β},
- exponential: $(1+\gamma)^{n}$.

Thus polynomial growth is *always* faster than logarithmic growth and exponential growth is always faster than polynomial growth. In practice, you can see that by **L'Hospital's Rule**.

Outline

Let $\{X_n\}, n = 1, 2, ...$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$.

Definition 4.1 (Convergence in probability)

The sequence $\{X_n\}$ is said to **convergence in probability** to a random variable X if for each $\delta > 0$,

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

or equivalently we write it as $\text{plim}_{n\to\infty} X_n = X, X_n = X + o_p(1)$.

Definition 4.2

Let θ be an unknown parameter, and denote $\hat{\theta}$ as an estimator of θ based on sample size n. Then $\hat{\theta}_n$ is a weakly consistent estimator of θ if $\hat{\theta}_n - \theta \xrightarrow{p} 0$.

Definition 4.3 (Stochastic orders)

We say $X_n = O_p(f_n)$ (or X_n/f_n is bounded in probability), if $\forall \epsilon > 0$, $\exists C < \infty$ and $n_0 > 0$ such that $P(|X_n| > Cf_n) < \epsilon$, $\forall n \geq n_0$.

Definition 4.4

We say $X_n = o_p(f_n)$ if $X_n/f_n \xrightarrow{p} 0$. In particular, $X_n = o_p(1)$ is synonymous with $X_n \xrightarrow{p} 0$.

For a finite non-zero constant c,

$$X_n \xrightarrow{p} c \Rightarrow X_n = O_p(1)$$

For a finite non-zero constant c,

$$X_n \xrightarrow{p} c \Rightarrow X_n = O_p(1)$$

Proof:
For any $\delta > 0$, $\exists$ $C = |c| + \delta$ such that

$$\begin{aligned} P(|X_n| > C) &= P(|X_n| - |c| > \delta) \\ &\leq P(|X_n - c| > \delta) \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Remark 4.1

$P(|X_n| > C) \to 0$ for <u>some</u> $C \equiv X_n = O_p(1)$.
$P(|X_n| > C) \to 0$ for <u>all</u> $C \equiv X_n = o_p(1)$.

Useful theorems

Theorem 4.1

(i) $X_n = o_p(f_n) \Rightarrow X_n = O_p(f_n)$.

(ii) $X_n = O_p(f_n) \Rightarrow X_n = o_p(g_n)$ if $\frac{f_n}{g_n} \to 0$.

(iii) $X_n = O_p((E|X_n|^r)^{1/r})$ for $r > 0$.

Proof

(i) Replace X_n by X_n/f_n and apply previous remark.

(ii) For any $\epsilon > 0$,

$$P(|X_n| > \epsilon g_n) = P(|X_n| > \epsilon f_n g_n/f_n) \leq P(|X_n| > C f_n) < \epsilon.$$

as $n \to \infty$ for all $\epsilon > 0$, for some $C < \infty$ (since $\epsilon g_n/f_n > C$ for n large enough).

Proof

(i) Replace X_n by X_n/f_n and apply previous remark.

(ii) For any $\epsilon > 0$,

$$P(|X_n| > \epsilon g_n) = P(|X_n| > \epsilon f_n g_n / f_n) \leq P(|X_n| > C f_n) < \epsilon.$$

as $n \to \infty$ for all $\epsilon > 0$, for some $C < \infty$ (since $\epsilon g_n / f_n > C$ for n large enough).

(iii) By *Markov's inequality*, for any $\epsilon > 0, \exists\, C = C(\epsilon) < \infty$ such that

$$P\{|X_n| > C(E|X_n|^r)^{1/r}\} \leq \frac{E|X_n|^r}{C^r E|X_n|^r} = C^{-r} < \epsilon$$

Algebra of stochastic orders

Theorem 4.2 (Algebra of stochastic orders)

(i) If $X_n = O_p(f_n), Y_n = O_p(g_n)$, then
  (a) $X_n Y_n = O_p(f_n g_n)$
  (b) $X_n + Y_n = O_p(\max(f_n, g_n))$. Alternatively, $O_p(f_n + g_n)$.

(ii) Replace "O" by "o" everywhere in (i).

(iii) $X_n = O_p(f_n), Y_n = o_p(g_n) \Rightarrow X_n Y_n = o_p(f_n g_n)$.

Proof: Homework.

Stochastic Orders of Powers of n

Especially, consider $f_n = n^\alpha$

Definition 4.5

A sequence of random variables X_n is **asymptotically of order** $o(n^\alpha)$ **in probability**, denoted by $o_p(n^\alpha)$, if X_n/n^α converges in probability to 0, i.e. if for every $\delta > 0$,

$$\lim_{n\to\infty} P\{|\frac{X_n}{n^\alpha}| > \delta\} = 0.$$

Example 4.1

- For instance, $X_n \sim N(0, 1/n)$ is of order $o_p(1)$, but not of order $o_p(n^{-1/2})$.
- If $X_n \sim \chi^2_n$, then $X_n = o_p(n^{3/2})$ but not $o_p(n)$.

Bounded in probability

Definition 4.6

A sequence of random variables X_n is **asymptotically of order** $O(n^\alpha)$ **in probability**, denoted by $O_p(n^\alpha)$, if for each $\epsilon > 0$,$\exists$ a real number $C = C_\epsilon > 0$ and natural number $\bar{n} = \bar{n}_\epsilon$ such that for every $n \geq \bar{n}$,

$$P\{|\frac{X_n}{n^\alpha}| < C\} > 1 - \epsilon$$

- If $X_n = O_p(1)$, then it is called **bounded in probability**.
- If $X_n \xrightarrow{d} X$, then X_n is bounded in probability.
- Say $X_n = O_p(n^\alpha)$ if X_n/n^α is bounded in probability.

For example, $X_n \sim N(0, 1/n)$ is $O_p(n^{-1/2})$ and $X_n \sim \chi^2_n$ is $O_p(n)$. (*why???*)