ST5215 Advanced Statistical Theory, Lecture 16

HUANG Dongming

National University of Singapore

13 Oct 2020

Overview

Last time

- Convergence modes
- Stochastic orders

Today

- Continuous mapping
- Slutsky's theorem
- δ-method
- Strong Law of Large Number (SLLN)

Recap: Convergence Modes

- $X_n \overset{a.s.}{\to} X$ if $\quad P(X_n \to X) = 1$
- $X_n \overset{P}{\to} X$ if $\quad P(|X_n - X| > \epsilon) \to 0$, for all $\epsilon > 0$
- $X_n \overset{L^p}{\to} X$ if $\quad E|X_n - X|^p \to 0$
- $X_n \overset{D}{\to} X$ if $\quad F_{X_n}(x) \to F_X(x)$ at every continuity point x of F_X

Relations between different modes of convergence

$$
\begin{array}{ccccc}
\xrightarrow{L^p} & \underset{p>q\geq 0}{\Longrightarrow} & \xrightarrow{L^q} & & \\
 & & \Downarrow & & \\
\xrightarrow{a.s.} & \Longrightarrow & \xrightarrow{P} & \Longrightarrow & \xrightarrow{D}
\end{array}
$$

- If $X_n \overset{D}{\to} c$ for a constant c, then $X_n \overset{P}{\to} c$.
- If $X_n \overset{\mathcal{P}}{\to} X$ then there is a subsequence s.t. $X_{n_k} \overset{a.s.}{\to} X$
- If $X_n \overset{D}{\to} X$, then we can find $Y_n \overset{\mathcal{D}}{=} X_n$ and $Y \overset{\mathcal{D}}{=} X$ s.t. $Y_n \overset{a.s.}{\to} Y$

Properties and relations

Theorem (Continuous mapping)

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random k-vectors and X is a random k-vector in the same probability space.
Let $g : \mathcal{R}^k \to \mathcal{R}$ be continuous. Then

- *if $X_n \overset{a.s.}{\to} X$, then $g(X_n) \overset{a.s.}{\to} g(X)$;*
- *if $X_n \overset{P}{\to} X$, then $g(X_n) \overset{P}{\to} g(X)$;*
- *if $X_n \overset{D}{\to} X$, then $g(X_n) \overset{D}{\to} g(X)$.*

1. Uniqueness of the limit
   - If $X_n \xrightarrow{*} X$ and $X_n \xrightarrow{*} Y$, then $X = Y$ a.s., where $*$ could be *a.s.*, P or L^p
   - If $F_n \Rightarrow F$ and $F_n \Rightarrow G$, and $F(t) = G(t)$ for all t
2. Concatenation:
   - If $X_n \xrightarrow{*} X$ and $Y_n \xrightarrow{*} Y$, then $(X_n, Y_n) \xrightarrow{*} (X, Y)$ where $*$ is either P or *a.s.*
   - If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$, then $(X_n, Y_n) \xrightarrow{D} (X, c)$ for a constant c
   - This is NOT true: $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} Y$, then $(X_n, Y_n) \xrightarrow{D} (X, Y)$
3. Linearity
   - If $X_n \xrightarrow{*} X$ and $Y_n \xrightarrow{*} Y$, then $aX_n + bY_n \xrightarrow{*} aX + bY$, where $*$ could be *a.s.*, P or L^p, and a and b are real numbers
   - This statement is NOT true for convergence in distribution
4. Cramér-Wold device: $X_n \xrightarrow{D} X$ iff $c^\top X_n \xrightarrow{D} c^\top X$ for every $c \in \mathcal{R}^k$

Theorem (Slutsky's theorem)

If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$ for a constant c, then

- $X_n + Y_n \xrightarrow{D} X + c$
- $X_n Y_n \xrightarrow{D} cX$
- $X_n / Y_n \xrightarrow{D} X/c$ *if* $c \neq 0$

- Slutsky's theorem is a consequence of continuous mapping theorem and concatenation property
- This result is very useful in statistics. For example, for i.i.d. samples X_i's with finite variance,
  - By CLT, the sample mean $\bar{X}_n$ satisfies $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} N(0, \mathrm{Var}(X))$
  - By SLLN, $S^2 \xrightarrow{\mathcal{D}} \mathrm{Var}(X)$. Therefore,

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{S^2}} \xrightarrow{\mathcal{D}} N(0,1)$$

δ-Method

If we have an approximate distribution of $\hat{\theta}$ (often by CLT), what is the approximate distribution of $g(\hat{\theta})$ for a smooth function g?

- Suppose $a_n(\hat{\theta}_n - \theta) \xrightarrow{D} Z$, where $a_n \to \infty$
- When $\hat{\theta}_n \approx \theta$, and since g is differentiable, then by Taylor expansion

$$\frac{g(\hat{\theta}_n) - g(\theta)}{\hat{\theta}_n - \theta} \approx g'(\theta) \tag{1}$$

or

$$\frac{g(\hat{\theta}_n) - g(\theta)}{g'(\theta)} \approx \hat{\theta}_n - \theta \tag{2}$$

and further

$$a_n \frac{g(\hat{\theta}_n) - g(\theta)}{g'(\theta)} \approx a_n(\hat{\theta}_n - \theta) \xrightarrow{D} Z \tag{3}$$

δ-method, Univariate

Theorem

Let $X_1, X_2, \ldots,$ Y be random variables, and $\{a_n\}$ is a sequence of positive numbers with $\lim_{n\to\infty} a_n = \infty$ satisfying

$$a_n (X_n - c) \xrightarrow{D} Y,$$

where $c \in \mathcal{R}$. Let g be a function from $\mathcal{R}$ to $\mathcal{R}$.

(i) *If g is differentiable at c, then*

$$a_n [g(X_n) - g(c)] \xrightarrow{D} g'(c) Y$$

where $g'(x)$ is the derivatives of g at x

(ii) *Suppose that g has continuous derivatives of order $m > 1$ in a neighborhood of c, s.t. $g^{(j)}(c) = 0$ for all $1 \leq j \leq m-1$, and $g^{(m)}(c) \neq 0$. Then*

$$a_n^m [g(X_n) - g(c)] \xrightarrow{D} \frac{1}{m!} g^{(m)}(c) Y^m$$

Example

Suppose $X_1, \ldots, X_n$ are i.i.d. sample from P_λ with p.d.f.

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \in [0, \infty), \tag{4}$$

where the parameter $\lambda > 0$ is called the rate

- $\mu = EX = 1/\lambda$, or $\lambda = \mu^{-1}$
- $\mathrm{Var}(X) = \mu^2$
- Let $\hat{\mu}_n = \bar{X}_n$ and $\hat{\lambda}_n = \hat{\mu}_n^{-1} = 1/\bar{X}_n$
- CLT says that $\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{D} Z \sim N(0, \mu^2)$
- Apply δ-method with $c = \mu$, $g(\mu) = \mu^{-1} = \lambda$
- Since $g'(\mu) = -\mu^{-2} = -\lambda^2$, we have

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{D} -\lambda^2 Z \sim N(0, \lambda^2)$$

Examples

Suppose $X_1, \ldots, X_n$ IID with $\mathrm{Var}(X_1) = 1$, $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$, $c = EX_1$, $a_n = \sqrt{n}$, and $Z \sim N(0,1)$

- If $g(x) = x^2$,
  - if $c \neq 0$ then $\sqrt{n}(\overline{X}_n^2 - c^2) \xrightarrow{D} N(0, 4c^2)$ since $g'(c) = 2c$;
  - if $c = 0$, then $g'(c) = 0$ but $g''(c) = 2 \neq 0$, so we have $(\sqrt{n})^2(\overline{X}_n^2 - 0) \xrightarrow{D} Z^2 \sim \chi_1^2$
- If $g(x) = x^{-1}$ and $c \neq 0$, then $\sqrt{n}(\overline{X}_n^{-1} - c^{-1}) \xrightarrow{D} N(0, 1/c^4)$, since $g'(c) = -c^{-2}$.
  - What if $c = 0$ in this case? (Left for exercise)

Proof of (i)

Let

$$Z_n = a_n \left[g\left(X_n\right) - g(c)\right] - a_n g'(c)\left(X_n - c\right) \tag{5}$$

If we can show that $Z_n = o_p(1)$, then by the convergency of $a_n\left(X_n - c\right)$ and Slutsky's theorem, we conclude the proof.

- The differentiability of g at c implies that for any $\epsilon > 0$, there is a $\delta_\epsilon > 0$ such that

$$\left|g(x) - g(c) - g'(c)(x - c)\right| \leq \epsilon|x - c| \tag{6}$$

  whenever $|x - c| < \delta_\epsilon$
- On the event $\{|X_n - c| < \delta_\epsilon\}$, we have $|Z_n| < \epsilon a_n |X_n - c|$
- Consider any $\eta > 0$.
  If $\eta < |Z_n|$, then either $|X_n - c| \geq \delta_\epsilon$, or $\eta < \epsilon a_n |X_n - c|$

- For any $\eta > 0$, $\epsilon > 0$, we have

$$P\left(|Z_n| \geq \eta\right) \leq P\left(|X_n - c| \geq \delta_\epsilon\right) + P\left(a_n\left|X_n - c\right| \geq \eta/\epsilon\right) \tag{7}$$

- Since $a_n \to \infty$, by Slutsky's theorem, $X_n = \frac{1}{a_n} a_n(X - c) + c \xrightarrow{P} c$
- By continuous mapping, $a_n\left|X_n - c\right| \xrightarrow{D} |Y|$
- Fixed η. Choose ϵ sufficiently small such that η/ϵ is a continuity point of $F_{|Y|}$ and $P(|Y| \geq \eta/\epsilon)$ is smaller than η
  - For a monotone function, its discontinuity points are at most countably many
- From Eq (7), we have

$$\limsup_n P\left(|Z_n| \geq \eta\right) \leq 0 + P\left(|Y| \geq \eta/\epsilon\right) < \eta \tag{8}$$

- Since η is arbitrary, we conclude that $Z_n = o_p(1)$

Theorem (δ-method, multivariate)

Let $X_1, X_2, \ldots,$ Y be random k-vectors, and $\{a_n\}$ is a sequence of positive numbers with $\lim_{n\to\infty} a_n = \infty$ satisfying

$$a_n (X_n - c) \xrightarrow{D} Y,$$

where $c \in \mathcal{R}^k$. Let g be a function from $\mathcal{R}^k$ to $\mathcal{R}$.

(i) *If g is differentiable at c, then*

$$a_n [g(X_n) - g(c)] \xrightarrow{D} [\nabla g(c)]^\tau Y$$

where $\nabla g(x)$ is the partial derivatives of g at x

(ii) *Suppose that g has continuous partial derivatives of order $m > 1$ in a neighborhood of c, with all the partial derivatives of order $j, 1 \le j \le m-1$, vanishing at c, but with the mth-order partial derivatives not all vanishing at c. Then*

$$a_n^m [g(X_n) - g(c)] \xrightarrow{D} \frac{1}{m!} \sum_{i_1=1}^{k} \cdots \sum_{i_m=1}^{k} \left. \frac{\partial^m g}{\partial x_{i_1} \cdots \partial x_{i_m}} \right|_{x=c} Y_{i_1} \cdots Y_{i_m}$$

Strong Law of Large Numbers

Theorem (SLLN)

Let $X_1, X_2, \ldots$ be i.i.d. random variables. A necessary and sufficient condition for the existence of a constant c for which

$$\frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{a.s.}{\to} c$$

is that $E|X_1| < \infty$, in which case $c = EX_1$

Remark

- "Strong" refers to the a.s. convergence
- The necessity is simple
- The proof of sufficiency is harder

Lemma (A)

$E|X| \leq \sum_{n=0}^{\infty} P(|X| \geq n) \leq 1 + E|X|$

Necessity of SLLN

- Suppose that $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} c$
- Let $T_n = \sum_{i=1}^{n} X_i$
- Then
$$\frac{X_n}{n} = \frac{T_n}{n} - c - \frac{n-1}{n}\left(\frac{T_{n-1}}{n-1} - c\right) + \frac{c}{n} \xrightarrow{a.s.} 0$$
- Therefore $P\left(\{|X_n|/n > 1\} i.o.\right) = 0$
- Since X_n's are independent, by second Borel-Cantelli lemma, if $\sum_{n=1}^{\infty} P\left(|X_n| \geq n\right) = \infty$, then $P\left(\{|X_n|/n > 1\} i.o.\right) = 1$, which contradicts the last bullet point
- Therefore
$$\sum_{n=1}^{\infty} P\left(|X_n| \geq n\right) = \sum_{n=1}^{\infty} P\left(|X_1| \geq n\right) < \infty,$$
which implies $E|X_1| < \infty$ by Lemma A

Proof of Sufficiency

Suppose $E|X_n| < \infty$. Let $Y_n = X_n I_{\{|X_n| \le n\}}, n = 1, 2, \ldots$.

- Since $\sum_n P(X_n \neq Y_n) = \sum_n P(|X_n| > n) < \infty$, by the first Borel-Cantelli lemma, $P(X_n \neq Y_n, i.o.) = 0$.
  That is, with probability 1, $X_n = Y_n$ for all n sufficiently large, and
  $$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i \overset{a.s.}{\to} 0$$
- By dominated convergence theorem, $EY_n = E\left(X_1 I_{|X_1| \le n}\right) \to EX_1$, and thus $\frac{1}{n}\sum_{i=1}^{n} EY_i \to EX_1$
- We only need to show $\frac{1}{n}\sum_{i=1}^{n}(Y_i - EY_i) \overset{a.s.}{\to} 0$

Lemma (Kronecker's lemma)

Suppose $\{x_n\}$ is a sequence of real numbers, and $a_n \uparrow \infty$ and are nonnegative. If $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ converges, then $a_n^{-1}\sum_{i=1}^{n} x_i \to 0$

We only need to show $\sum_{n=1}^{\infty} \frac{(Y_n - EY_n)}{n}$ converges almost surely

Kolmogorov's inequality

Lemma (B)

Suppose $Z_1, Z_2, \ldots$ are independent and have 0 means and finite variances. Let $S_j = \sum_{i=1}^{j} Z_i$.
Then

$$P\left(\max_{1\le j\le n} |S_j| > t\right) \le \frac{1}{t^2}\sum_{i=1}^{n} \mathrm{Var}(Z_i) \tag{9}$$

- Let $S_0 = 0$, $A_k = \{\max_{1\le j<k} |S_j| < t \le |S_k|\}$. A_k's are disjoint
- Note that $S_k I_{A_k}$ is $\sigma(Z_1, \ldots, Z_k)$-measurable and $S_n - S_k$ is $\sigma(Z_{k+1}, \ldots, Z_n)$-measurable, they are independent
- $$\int_{A_k} S_n^2 \mathrm{d}P = \int_{A_k} (S_k + S_n - S_k)^2 \mathrm{d}P = \int_{A_k} S_k^2 \mathrm{d}P + \int_{A_k} (S_n - S_k)^2 \mathrm{d}P$$
  $$\geqslant \int_{A_k} S_k^2 \mathrm{d}P \geqslant t^2 P(A_k)$$
- Summing over $k = 1, \cdots, n$, we obtain $ES_n^2 \ge t^2 P(\bigcup_k A_k)$

Prove: $\sum_{n=1}^{\infty} \frac{(Y_n - EY_n)}{n}$ converges a.s.

- Let $Z_n = \frac{(Y_n - EY_n)}{n}$. Let $S_n = \sum_{j=1}^{n} Z_j$
- We later show $\sum_{n=1}^{\infty} \mathrm{Var}(Z_n) < \infty$
- For any $\epsilon > 0$ and any $m \in \mathcal{N}$, we apply Kolmogorov's inequaltity to Z_j for $j = m+1, \ldots, k$, so that

$$P\left(\max_{m+1 \leq n \leq k} |S_n - S_m| > \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{n=m+1}^{k} \mathrm{Var}(Z_n) \tag{10}$$

- Take $k \to \infty$, then $m \to \infty$, we have

$$P\left(\lim_{n \leq m,k} |S_n - S_m| > \epsilon\right) = 0. \tag{11}$$

- Since ϵ is arbitrary, we have S_n is a Cauchy sequence (and thus converge) almost surely

Show $\sum_{n=1}^{\infty} \mathrm{Var}(Z_n) < \infty$

$$\begin{aligned}\sum_{n=1}^{\infty} \mathrm{Var}(Z_n) &\leq \sum_{n=1}^{\infty} \frac{1}{n^2} E X_n^2 I_{|X_n| \leq n} \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} E X^2 I_{k-1<|X|\leq k} \\ &= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{1}{n^2} E X^2 I_{k-1<|X|\leq k}\end{aligned}$$

Note that $\sum_{n=k}^{\infty} \frac{1}{n^2} \leq 1/k^2 + \sum_{n=k+1}^{\infty} \frac{1}{n(n-1)} = 1/k^2 + 1/k < 2/k$.

$$\begin{aligned}\sum_{n=1}^{\infty} \mathrm{Var}(Z_n) &\leq \sum_{k=1}^{\infty} \frac{2}{k} E X^2 I_{k-1<|X|\leq k} \\ &\leq \sum_{k=1}^{\infty} \frac{2}{k} k E|X| I_{k-1<X\leq k} = 2E|X| < \infty\end{aligned}$$

Tutorial

1. Suppose X is a nonnegative random variable. Show that

$$EX \leq \sum_{i=0}^{\infty} P(X \geq n) \leq 1 + EX.$$

   Therefore, X is integrable if and only if $\sum_{i=0}^{\infty} P(X \geq n) < \infty$

2. If $X_n \xrightarrow{D} c$ for a constant c, then $X_n \xrightarrow{P} c$.

3. Let $X_1, X_2, \ldots$ be a sequence of identically distributed random variables with $E\,|X_1| < \infty$ and let $Y_n = n^{-1} \max_{1 \leq i \leq n} |X_i|$. Show that $\lim_n E\,(Y_n) = 0$ and $\lim_n Y_n = 0$ a.s.

4. (Postponed) Suppose that $X_n \xrightarrow{\mathrm{D}} X$. Then, for any $r > 0$

$$\lim_{n\to\infty} E|X_n|^r = E|X|^r < \infty$$

   if and only if $\left\{|X_n|^r\right\}$ is uniformly integrable in the sense that

$$\lim_{t\to\infty} \sup_n E\left(|X_n|^r \, I_{\{|X_n|>t\}}\right) = 0$$

Exercise 1

Suppose X is a nonnegative random variable. Show that

$$EX \leq \sum_{i=0}^{\infty} P(X \geq n) \leq 1 + EX.$$

Therefore, X is integrable if and only if $\sum_{i=0}^{\infty} P(X \geq n) < \infty$

Proof: By Fubini's theorem,

$$\begin{aligned}
EX &= E\left(\int_0^{\infty} I_{X \geq t} \, \mathrm{d}m(t)\right) \\
&= \int_0^{\infty} P(X \geq t) \, \mathrm{d}m(t) \\
&= \sum_{n=0}^{\infty} \int_n^{n+1} P(X \geq t) \, \mathrm{d}t \\
&\leq \sum_{n=0}^{\infty} P(X \geq n)
\end{aligned}$$

The last step can also be $\geq \sum_{n=0}^{\infty} P(X \geq n+1) = \sum_{n=1}^{\infty} P(X \geq n)$

Exercise 2

If $X_n \stackrel{D}{\to} c$ for a constant c, then $X_n \stackrel{P}{\to} c$.

Proof:

- Note that the cumulative distribution function of $X = c$ has only one discontinuity point c.
- For any $\epsilon > 0$

$$\begin{aligned} P\left(|X_n - X| > \epsilon\right) &= P\left(|X_n - c| > \epsilon\right) \\ &\leq P\left(X_n > c + \epsilon\right) + P\left(X_n < c - \epsilon\right) \\ &\to P(X > c + \epsilon) + P(X < c - \epsilon) \\ &= 0 \end{aligned}$$

as $n \to \infty$. Thus, $X_n \to_p X$.

Exercise 3

Let $X_1, X_2, \ldots$ be a sequence of identically distributed random variables with $E|X_1| < \infty$ and let $Y_n = n^{-1} \max_{1 \le i \le n} |X_i|$. Show that $\lim_n E(Y_n) = 0$ and $\lim_n Y_n = 0$ a.s.

Proof: Part (i)

- Let $g_n(t) = n^{-1} P\left(\max_{1 \le i \le n} |X_i| > t\right)$
- Then $\lim_n g_n(t) = 0$ for any $t$and

$$0 \le g_n \le \frac{1}{n} \sum_{i=1}^{n} P(|X_i| > t) = P(|X_1| > t)$$

- Since $E|X_1| < \infty$, $\int_0^\infty P(|X_1| > t)\, dt < \infty$
- By the dominated convergence theorem,

$$\lim_n E(Y_n) = \lim_n \int_0^\infty g_n(t) dt = \int_0^\infty \lim_n g_n(t) dt = 0$$

Part (ii)

- Since $E|X_1| < \infty$

$$\sum_{n=1}^{\infty} P\left(|X_n|/n > \epsilon\right) = \sum_{n=1}^{\infty} P\left(|X_1| > \epsilon n\right) < \infty,$$

  which implies that $\lim_n \frac{|X_n|}{n} = 0$ a.s.
- Let $\Omega_0 = \{\omega : \lim_n |X_n(\omega)|/n = 0\}$. Then $P(\Omega_0) = 1$
- Let $\omega \in \Omega_0$. For any $\epsilon > 0$, there exists an $N_{\epsilon,\omega}$ such that if $n \geq N_{\epsilon,\omega}$, then $|X_n(\omega)| < n\epsilon$
- Note that $\max_{1 \leq i \leq N_{c\omega}} |X_i(\omega)|$ is fixed, we can find a sufficiently large integer $M_{\epsilon,\omega} > N_{\epsilon,\omega}$ such that

$$\max_{1 \leq i \leq N_{c\omega}} |X_i(\omega)| \leq n\epsilon$$

  whenever $n > M_{\epsilon,\omega}$
- Therefore, if $n > M_{\epsilon,\omega}$, we have

$$\max_{1 \leq i \leq n} |X_i(\omega)| \leq n\epsilon, \text{ i.e. } Y_n(\omega) < \epsilon. \tag{12}$$

  Hence, $\lim_n Y_n(\omega) = 0$