UNIVERSITY OF TORONTO

Statistical Sciences

DoSS Summer Bootcamp Probability Module 8

Miaoshiqi (Shiki) Liu

University of Toronto

July 27, 2022

Recap

Learnt in last module:

- Stochastic convergence
  - Convergence in distribution
  - Convergence in probability
  - Convergence almost surely
  - Convergence in L^p
  - Relationship between convergences

Outline

- Convergence of functions of random variables
  - Slutsky's theorem
  - Continuous mapping theorem
- Laws of large numbers
  - WLLN
  - SLLN
  - Glivenko-Cantelli theorem
- Central limit theorem

Convergence of functions of random variables

Recall: Stochastic convergence If $X_n \to X$, $Y_n \to Y$ in some sense, how is the limiting property of $f(X_n, Y_n)$? $a_n X_n + b_n Y_n$

Convergence of functions of random variables

Recall: Stochastic convergence If $X_n \to X$, $Y_n \to Y$ in some sense, how is the limiting property of $f(X_n, Y_n)$?

Convergence of functions of random variables (a.s.)

Suppose the probability space is complete, if $X_n \xrightarrow{a.s.} X$, $Y_n \xrightarrow{a.s.} Y$, then for any real numbers a, b,

- $aX_n + bY_n \xrightarrow{a.s.} aX + bY$;
- $X_nY_n \xrightarrow{a.s.} XY$.

$\{\omega : \lim_{n\to\infty} aX_n(\omega) + bY_n(\omega) = aX(\omega) + bY(\omega)\}$

Remark:

- Still require all the random variables to be defined on the same probability space

Convergence of functions of random variables

Convergence of functions of random variables (probability)

Suppose the probability space is complete, if $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, then for any real numbers a, b,

- $aX_n + bY_n \xrightarrow{P} aX + bY$;
- $X_n Y_n \xrightarrow{P} XY$.

Remark:

- Still require all the random variables to be defined on the same probability space

Convergence of functions of random variables

Convergence of functions of random variables (L^p)

Suppose the probability space is complete, if $X_n \xrightarrow{L^p} X$, $Y_n \xrightarrow{L^p} Y$, then for any real numbers a, b,

- $aX_n + bY_n \xrightarrow{L^p} aX + bY$;

$X_n Y_n \xrightarrow{L^p} XY$?

$E|X_n - X|^p \longrightarrow 0 \quad n \to \infty.$

$E|X_n|^p < \infty$

Remark:

- Still require all the random variables to be defined on the same probability space

$X_n \quad E|X_n| < \infty.$

$Y_n \quad E|Y_n| < \infty$

$X_n Y_n. \quad E|X_n Y_n| < \infty$?

Convergence of functions of random variables

Remark: Convergence in distribution is different.

Slutsky's theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ (c is a constant), then

- $X_n + Y_n \xrightarrow{d} X + c$;
- $X_n Y_n \xrightarrow{d} cX$;
- $X_n / Y_n \xrightarrow{d} X/c$, where $c \neq 0$.

$Y_n \xrightarrow{d} c.$

Convergence of functions of random variables

Remark: Convergence in distribution is different.

Slutsky's theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ (c is a constant), then

- $X_n + Y_n \xrightarrow{d} X + c$;
- $X_n Y_n \xrightarrow{d} cX$;
- $X_n / Y_n \xrightarrow{d} X/c$, where $c \neq 0$.

Remark:

- The theorem remains valid if we replace all the convergence in distribution with convergence in probability.

Convergence of functions of random variables

Remark: The requirement that $Y_n \xrightarrow{P} c$ (c is a constant) is necessary.

Convergence of functions of random variables

Remark: The requirement that $Y_n \xrightarrow{P} c$ (c is a constant) is necessary.

Examples:

$X_n \sim \mathcal{N}(0,1)$, $Y_n = -Xn$, then

$X_1 \quad Y_1 = -X_1$

$X_2 \quad Y_2 = -X_2.$

- $X_n \xrightarrow{d} Z \sim \mathcal{N}(0,1)$, $Y_n \xrightarrow{d} Z \sim \mathcal{N}(0,1)$;
- $X_n + Y_n \xrightarrow{d} 0$; $\longrightarrow$ $X_n + Y_n \xrightarrow{d} 2Z.$ ✗
- $X_n Y_n = -X_n^2 \xrightarrow{d} -\chi^2(1)$; $\longrightarrow$ $X_n Y_n \xrightarrow{d} Z^2 \sim \chi^2(1)$ ✗
- $X_n / Y_n = -1.$

$\longrightarrow$ $X_n / Y_n \xrightarrow{d} Z/Z = 1$ ✗.

Convergence of functions of random variables

Continuous mapping theorem

Let X_n, X be random variables, if $g(\cdot) : \mathbb{R} \to \mathbb{R}$ satisfies $\mathbb{P}(X \in D_g) = 0$, then

- $X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad g(X_n) \xrightarrow{a.s.} g(X)$;
- $X_n \xrightarrow{P} X \quad \Rightarrow \quad g(X_n) \xrightarrow{P} g(X)$;
- $X_n \xrightarrow{d} X \quad \Rightarrow \quad g(X_n) \xrightarrow{d} g(X)$;

where D_g is the set of discontinuity points of $g(\cdot)$.

$X_n Y_n \xrightarrow{L^p} XY$.

$g(X_n) = X_n^2$

Convergence of functions of random variables

Continuous mapping theorem

Let X_n, X be random variables, if $g(\cdot) : \mathbb{R} \to \mathbb{R}$ satisfies $\underline{\mathbb{P}(X \in D_g) = 0}$, then

- $X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad g(X_n) \xrightarrow{a.s.} g(X)$;
- $X_n \xrightarrow{P} X \quad \Rightarrow \quad g(X_n) \xrightarrow{P} g(X)$;
- $X_n \xrightarrow{d} X \quad \Rightarrow \quad g(X_n) \xrightarrow{d} g(X)$;

where D_g is the set of discontinuity points of $g(\cdot)$.

Remark:

- If $g(\cdot)$ is continuous, then ...
- If X is a continuous random variable, and D_g only include countably many points, then ...

$P(X = x) = 0.$ $\qquad D_g = \{x_1, x_2, \ldots \ldots x \ldots\}$

$P(X \in D_g) = \sum_{i=1}^{\infty} \underbrace{P(X = x_i)}_{=0} = 0$

Law of large numbers

$\mu = E(X_i) \quad E(|X_i|) < \infty.$

Weak Law of Large Numbers (WLLN)

If $X_1, X_2, \cdots, X_n$ are i.i.d. random variables, $\mu = \mathbb{E}(|X_i|) < \infty$, then

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

$\bar{X} - \mu \xrightarrow{P} 0.$

$$\bar{X} - \mu = \frac{\sum_{i=1}^{n} X_i - n\mu}{n} \xrightarrow{P} 0.$$

Remark:

A more easy-to-prove version is the L^2 weak law, where an additional assumption $Var(X_i) < \infty$ is required.

Sketch of the proof:

$Var(X_i) = \sigma^2 < \infty.$

$E(\bar{X}) = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n}.$

$E|\bar{X} - \mu|^2 = Var(\bar{X}) = \frac{\sigma^2}{n} \xrightarrow{n \to \infty} 0$

$\forall \varepsilon > 0$

$P(|\bar{X} - \mu| > \varepsilon) \le \frac{Var(\bar{X} - \mu)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0$

Law of large numbers

A generalization of the theorem: triangular array

Triangular array

A triangular array of random variables is a collection $\{X_{n,k}\}_{1\le k\le n}$.

$$X_{1,1}$$
$$X_{2,1}, X_{2,2}$$
$$X_{3,1}, X_{3,2}, X_{3,3}$$
$$\vdots$$
$$X_{n,1}, X_{n,2}, \cdots, X_{n,n}$$

$n=1$

$n=2.$ X_1 X_2

$$\frac{S_n - n\mu}{n}$$

$$\frac{S_n - \mu_n}{b_n}$$

$$S_n = \sum_{i=1}^{n} X_i$$

$$S_n = \sum_{k=1}^{n} X_{n,k}.$$

Remark: We can consider the limiting property of the row sum S_n.

$$\mu_n = E(S_n)$$
$$\neq n\mu$$

Law of Large Numbers

not require.

L^2 weak law for triangular array

Suppose $\{X_{n,k}\}$ is a triangular array, $n = 1, 2, \cdots, k = 1, 2, \cdots, n$. Let $S_n = \sum_{k=1}^{n} X_{n,k}$, $\mu_n = \mathbb{E}(S_n)$, if $\sigma_n^2 / b_n^2 \to 0$, where $\sigma_n^2 = Var(S_n)$ and b_n is a sequence of positive real numbers, then

$$\frac{S_n - \mu_n}{b_n} \xrightarrow{P} 0.$$

Remark:

The L^2 weak law for i.i.d. random variables is a special case of that for triangular array.

$X_{n,1} \cdots X_{n,n}$. i.i.d. $E(X_{n,k}) = \mu$. $Var(X_{n,k}) = \sigma^2 < \infty$

$S_n = n\bar{X}$ $\quad \mu_n = E(S_n) = n\mu$.

$\sigma_n^2 = Var(S_n) = Var(n\bar{X}) = n^2 \frac{\sigma^2}{n} = n\sigma^2$

$b_n^2 = n^2 \quad \sigma_n^2 / b_n^2 = \frac{\sigma^2}{n}$

Law of large numbers

$\forall \varepsilon > 0.$

Proof:

$$P\left(\left|\frac{S_n - \mu_n}{b_n}\right| > \varepsilon\right) = P\left(\left(\frac{S_n - \mu_n}{b_n}\right)^2 > \varepsilon^2\right)$$

$$\leq \frac{E\left(\frac{S_n - \mu_n}{b_n}\right)^2}{\varepsilon^2} = \frac{E(S_n - E(S_n))^2}{b_n^2 \varepsilon^2}$$

$$= \frac{Var(S_n)}{b_n^2 \varepsilon^2} = \frac{\sigma_n^2}{b_n^2 \varepsilon^2}$$

UNIVERSITY OF TORONTO

Law of large numbers

Proof:

$P(|X_n| > b_n)$

$X_n \mathbb{1}(|X_n| \le b_n)$

Remark:
A more generalized version incorporates truncation, then the second-moment constraint is relieved.

Law of large numbers

Strong Law of Large Numbers (SLLN)

Let $X_1, X_2, \cdots$ be an i.i.d. sequence satisfying $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}(|X_i|) < \infty$, then $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{a.s.} \mu$.

Remark: The proof needs Borel-Cantelli lemma.

Law of large numbers

Strong Law of Large Numbers (SLLN)

Let $X_1, X_2, \cdots$ be an i.i.d. sequence satisfying $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}(|X_i|) < \infty$, then $\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{a.s.} \mu$.

Remark: The proof needs Borel-Cantelli lemma.

$X_1, \cdots X_n \sim X \sim F_X(\cdot)$

Glivenko-Cantelli theorem

Let $X_i, i = 1, \cdots, n$ i.i.d. with distribution function $F(\cdot)$, and let $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leq x)$, then as $n \to \infty$,

$F(x) = P(X \leq x)$

Empirical CDF

$$\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| \to 0, \quad a.s.$$

Law of large numbers

Proof:

$\forall x \in \mathbb{R}.\quad F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \le x) \xrightarrow{a.s} F(x).$

$Y_i \quad E(Y_i) = E(1(X_i \le x))$
$= P(X_i \le x)$
$= F(x). \in [0,1]$

$A_x = \{w: \lim_{n\to\infty} \underbrace{F_n(x)} \ne F(x)\}. \quad P(A_x) = 0$

on A_x^c. $\quad \lim_{n\to\infty} \underbrace{F_n(x)} = \underbrace{F(x)} \quad \forall \varepsilon > 0. \quad \exists N(\varepsilon, x)$

$\forall n > N(\varepsilon, x). \quad |F_n(x) - F(x)| < \varepsilon. \quad \to P(\bigcup_{x\in\mathbb{R}} A_x) = 0.$

$\sup_{x\in\mathbb{R}} \underbrace{|F_n(x) - F(x)|}_{\underbrace{|F_n(x) - F(x)|}} <$ a.s. / limit.

$\underbrace{\bigcup_{x\in\mathbb{R}} A_x.}_{N(\varepsilon)} \quad \underbrace{(\bigcup_{x\in\mathbb{R}} A_x)^c}$

$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| < \varepsilon.$

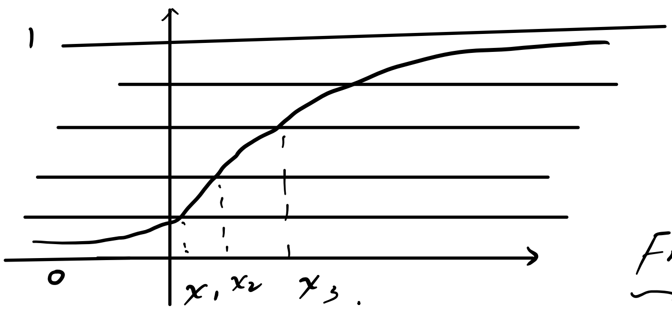
$\boxed{F(x) \text{ is continuous.}}$ $\to$ simplest.

Law of large numbers

$\forall m, \quad -\infty = x_0 < x_1 < x_2 \cdots < x_m = \infty.$

Proof:

$F(x_j) - F(x_{j-1}) = \frac{1}{m}. \quad \forall j = 1, \dots, m.$



$\forall x \in \mathbb{R}. \quad \exists j. \quad x \in [x_{j-1}, x_j)$

$$F_n(x) - F(x) \le F_n(x_j) - F(x_{j-1})$$

$$= F_n(x_j) - F(x_j) + \frac{1}{m}.$$

$$F_n(x) - F(x) \ge F_n(x_{j-1}) - F(x_{j-1}) - \frac{1}{m}.$$

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \le \max_{j \in \{1, \dots, m\}} |F_n(x_j) - F(x_j)| + \frac{1}{m}.$$

Denote $A_i = \{w : \lim_{n \to \infty} F_n(x_i) \ne F(x_i)\} \quad P(A_i) = 0.$

$A = \bigcup_{i=1}^{m} A_i \quad P(A) = 0. \quad$ on $A^c \quad \lim_{n \to \infty} F_n(x_i) = F(x_i) \quad \forall i = 1, \dots, m$

$\forall \varepsilon > 0.$ choose $m \quad \frac{1}{m} < \varepsilon, \quad N(\varepsilon, i) \quad \forall n > N(\varepsilon, i)$

$$|F_n(x_i) - F(x_i)| < \varepsilon - \frac{1}{m}.$$

Law of large numbers

Proof: choose $N(\varepsilon) = \max\limits_{1\le i\le m} N(\varepsilon, i) < \infty$

$\forall\ n > N(\varepsilon)$

$$\max_{1\le i\le m} |F_n(x_i) - F(x_i)| < \varepsilon - \frac{1}{m}.$$

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| < \varepsilon - \frac{1}{m} + \frac{1}{m} = \varepsilon.$$

$A.$ $P(A) = 0.$ on A^c. $\forall\ \varepsilon > 0.$ $\exists\ N(\varepsilon)$ $\forall\ n > N(\varepsilon)$

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| < \varepsilon.$$

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \longrightarrow 0$$

on A^c.

Central Limit Theorem

What is the limiting distribution of the sample mean?

Classic CLT

Suppose $X_1, \cdots X_n$ is a sequence of i.i.d. random variables with $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2 < \infty$, then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

$\frac{1}{\sqrt{n}}$

$\bar{X}_n \xrightarrow{p.\ a.s} \mu.$

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

$$\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} \frac{1}{\sqrt{n}} N(0,1) = 0$$

Remark:

- The proof involves characteristic function.
- A more generalized CLT is referred to as "Lindeberg CLT".

Central Limit Theorem

Example:

Suppose $X_i \sim Bernoulli(p)$, i.i.d., consider $Z_n = \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}}$, then by CLT, $Z_n \sim \mathcal{N}(0,1)$ asymptotically.

$$\frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$$

Problem Set

Problem 1: Prove that on a complete probability space, if $X_n \xrightarrow{a.s.} X$, $Y_n \xrightarrow{a.s.} Y$, then $X_n + Y_n \xrightarrow{a.s.} X + Y$.

Problem 2: Prove that on a complete probability space, if $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, then $X_n + Y_n \xrightarrow{P} X + Y$.

Problem 3: A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $\mathbb{E}(X_i) = 2$ (minutes) and $Var(X_i) = 1$. We assume that service times for different bank customers are independent. Let Y be the total time the bank teller spends serving 50 customers. Find $\mathbb{P}(90 < Y < 110)$.