UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 8

Ichiro Hashimoto

University of Toronto

July 22, 2025

Recap

$q \geq p$

in L^q $\Longrightarrow$ in L^p

$\Downarrow$

a.s $\Longrightarrow$ in probability $\Longrightarrow$ in distribution

CDF

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 $f(X_n, Y_n)$?

could be
in prob.
in distribution
a.s
L^p

e.g.

$X_n + Y_n \to ?$

$X_n \cdot Y_n \to ?$

$X_n / Y_n \to ?$

$\zeta_n = X_1 + \cdots + X_n$

$\dfrac{\zeta_n}{n^c} \to ?$

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,

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

Remark:

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

Recall $X_n \to X$ a.s. if $\mathbb{P}(\lim X_n = X) = 1$

(pf) Since $X_n \to X$ a.s, there exists $\exists N_X \subset \Omega$ s.t.

$$\underbrace{X_n \to X}_{\text{pointwise}} \text{ on } N_X \quad \text{and} \quad \mathbb{P}(N_X) = 1.$$

Since $Y_n \to Y$ a.s, there exists $N_Y \subset \Omega$ s.t.

$$Y_n \to Y \text{ on } N_Y \quad \text{and} \quad \mathbb{P}(N_Y) = 1.$$

On $N_X \cap N_Y$, we know $X_n \to X$, $Y_n \to Y$ pointwise.

Thus, we have on $N_X \cap N_Y$,

$$aX_n + bY_n \to aX + bY \quad \text{pointwise}$$
$$X_n \cdot Y_n \to X \cdot Y \quad \text{pointwise.}$$

$$\mathbb{P}(N_X \cap N_Y) = 1 - \mathbb{P}((N_X \cap N_Y)^c)$$
$$= 1 - \underbrace{\mathbb{P}(N_X^c \cup N_Y^c)}_{\text{apply union bound}}$$
$$\geq 1 - (\underbrace{\mathbb{P}(N_X^c)}_{=0} + \underbrace{\mathbb{P}(N_Y^c)}_{=0}) = 1 - 0 = 1$$

$$\therefore \mathbb{P}(N_X \cap N_Y) = 1.$$

Thus we have confirmed pointwise convergence of

$$aX_n + bY_n \to X \quad \text{and} \quad X_n \cdot Y_n \to X \cdot Y$$

hold with probability 1.

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,

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

Remark:

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

Recall $X_n \xrightarrow{P} X$ if. $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$.

$X_n + Y_n \xrightarrow{p} X + Y$ if $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$

(pf.) Let $\forall \varepsilon > 0$.

$$\mathbb{P}\left(|X_n + Y_n - (X+Y)| > \varepsilon \right).$$

↳ by triangle inequality.

$$\{ |X_n + Y_n - (X+Y)| > \varepsilon \} \subset \{ |X_n - X| > \tfrac{\varepsilon}{2} \} \cup \{ |Y_n - Y| > \tfrac{\varepsilon}{2} \}.$$

$$\Rightarrow \quad \varepsilon < | X_n + Y_n - (X+Y) |$$

$$= | (X_n - X) + (Y_n - Y) |$$

$$\le \underbrace{| X_n - X | + | Y_n - Y |}_{\text{it's impossible to have both of them } \le \frac{\varepsilon}{2}}.$$

$$\le \mathbb{P}\left(\{ |X_n - X| > \tfrac{\varepsilon}{2} \} \cup \{ |Y_n - Y| > \tfrac{\varepsilon}{2} \} \right)$$

by union bound

$$\le \underbrace{\mathbb{P}\left(|X_n - X| > \tfrac{\varepsilon}{2} \right)}_{\to 0 \text{ since } X_n \xrightarrow{p} X} + \underbrace{\mathbb{P}\left(|Y_n - Y| > \tfrac{\varepsilon}{2} \right)}_{\to 0 \text{ since } Y_n \xrightarrow{p} Y}$$

$\to 0$ as $n \to \infty$

Therefore, $X_n + Y_n \xrightarrow{p} X + Y$.

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,

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

Remark:

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

$X_n + Y_n \xrightarrow{L^p} X + Y$ if $X_n \xrightarrow{L^p} X$, $Y_n \xrightarrow{L^p} Y$

(pf) Recall that $\|X\|_{L^p} = \left(\mathbb{E}|X|^p\right)^{1/p}$ is a norm if $p \geq 1$.

Therefore, we have triangle inequality, i.e.

$$\|X + Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p}$$

apply

$$\|(X_n+Y_n) - (X+Y)\|_{L^p} \leq \underbrace{\|X_n - X\|_{L^p}}_{\to 0 \text{ since } X_n \xrightarrow{L^p} X} + \underbrace{\|Y_n - Y\|_{L^p}}_{\to 0 \text{ since } Y_n \xrightarrow{L^p} Y} \longrightarrow 0.$$

Convergence of functions of random variables

Even $X_n + Y_n \xrightarrow{d} X + Y$ fails in general

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 $c \neq 0$.

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:

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

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

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 $g(\cdot)$.

g is essentially continuous w.r.t X

L^p convergence fail in general.

Let $X_n = X$ when $X \in L^p$ but $\notin L^{2p}$.

Let $g(x) = x^2$.

Then $g(X_n) \notin L^p$. So L^p convergence doesn't make sense.

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 $g(\cdot)$.

Remark:

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

Law of large numbers

$\mu = \mathbb{E}X$

Weak Law of Large Numbers (WLLN)

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

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

Remark:

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

$\bar{X} \xrightarrow{L^2} \mu$

Sketch of the proof:

$$\mathbb{E}|\bar{X} - \mu|^2 = Var(\bar{X})$$

$$= Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

$$= \frac{\sum_{i=1}^{n} Var(X_i)}{n^2}.$$ since. X_i's are independent.

$$= \frac{n \, Var(X_1)}{n^2}$$

$$= \frac{Var(X_1)}{n} \to 0 \quad \text{as } n \to \infty.$$

therefore. $\bar{X} \to \mu$ in L^2.

Law of large numbers

A generalization of the theorem: triangular array

Triangular array

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

$$
\begin{array}{lll}
n=1 \longrightarrow & X_{1,1} & \overset{\text{sum}}{\Longrightarrow} S_1 \\
n=2 \longrightarrow & X_{2,1}, X_{2,2} & \overset{\text{sum}}{\Longrightarrow} S_2 \\
n=3 \longrightarrow & X_{3,1}, X_{3,2}, X_{3,3} & \overset{\text{sum}}{\Longrightarrow} S_3 \\
 & \vdots & \\
n \longrightarrow & X_{n,1}, X_{n,2}, \cdots, X_{n,n} & \overset{\text{sum}}{\Longrightarrow} S_n = \sum_{k=1}^{n} X_{n,k}
\end{array}
$$

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

Law of Large Numbers

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.

Law of large numbers

Proof:

$$\mathbb{E}\left|\frac{\xi_n - \mu_n}{b_n}\right|^2 = \frac{\sigma_n^2}{b_n^2} \to 0$$

So, $\dfrac{\xi_n - \mu_n}{b_n} \to 0$ in L^2,

and hence $\to 0$ in probability.

Law of large numbers

Proof:

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.

Glivenko-Cantelli theorem

Let $X_i, i = 1, \cdots, n$ i.i.d. with distribution function $F(\cdot)$, and let $F_n(x) = \underbrace{\frac{1}{n}\sum_{i=1}^n I(X_i \le x)}_{\frac{\#\text{ of } X_i\text{'s} \le x}{n}}$, then as $n \to \infty$,

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

Law of large numbers

Weaker version : $|F(x) - F_n(x)| \to 0$ a.s. on $\mathbb{R}$.

Proof:

Note that $0 \leq I(X_i \leq x) \leq 1$

So, $0 \leq \underbrace{\mathbb{E}\, I(X_i \leq x) = \mathbb{P}(X_i \leq x) = F(x)}_{\text{finite}} \leq 1.$

By $\underbrace{\text{SLLN}}_{\text{applied to } Y_i = I(X_i \leq x)}$, $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$

$\xrightarrow{a.s.} \mathbb{E}\, I(X_i \leq x) = F(x)$

Limit Theorems and Counterexamples

Recall: For the law of large numbers to hold, the assumption $E|X| < \infty$ is crucial.

Law of Large Numbers fail for infinite mean i.i.d. random variables

If $X_1 X_2, \ldots$ are i.i.d. to X with $E|X_i| = \infty$, then for $S_n = X_1 + \cdots + X_n$, $P(\lim_{n\to\infty} S_n/n \in (-\infty, \infty)) = 0$.

Proof: Omitted

$\frac{S_n}{n} \to 0$ fails with probability 1.

Central Limit Theorem

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

$std(\bar{X}) = \frac{\sigma}{\sqrt{n}} \Leftrightarrow std(\sqrt{n}\,\bar{X}) = \sigma$

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

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

$$\parallel$$

$$\frac{\sqrt{n}}{\sigma} \cdot \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu) = \frac{\sum_{i=1}^{n} (X_i - \mu)}{\sigma\sqrt{n}}$$

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.

$Var(X_i) = p(1-p)$

Monotone Convergence Theorem

Monotone Convergence Theorem

If $X_n \geq c$ and $X_n \nearrow X$, then $EX_n \nearrow EX$

Usage: ↳ fails if X_n is not lower bounded.

Let $X_n = \begin{cases} \frac{1}{n^2} & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$ and $S_n = \sum_{k=1}^{n} X_k$.

Then $0 \leq S_n \nearrow \lim_{n\to\infty} S_n \stackrel{\text{def}}{=} S \leq \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} < \infty$.

By Monotone convergence theorem,

$\mathbb{E}\, S \stackrel{\downarrow}{=} \lim_{n\to\infty} \mathbb{E}\, S_n = \lim_{n\to\infty} \mathbb{E} \sum_{k=1}^{n} X_k = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\, X_k = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{p}{k^2}$

$= \frac{\pi^2}{6} \cdot p$

Dominated Convergence Theorem

Dominated Convergence Theorem

If $X_n \to X$ a.s. and $|X_n| \leq Y$ a.s. for all n and Y is integrable, then $EX_n \to EX$

independent of n

$\lim_{n\to\infty} EX_n = \mathbb{E} \lim_{n\to\infty} X_n$

Usage:

Prop Suppose $\underbrace{M(t) = \mathbb{E}\exp(tX)}_{\text{MGF of } X.} < \infty$ for any $t \in [-\varepsilon, \varepsilon]$.

Then, $\frac{d}{dt} M(t)\Big|_{t=0} = \mathbb{E}X$

(pf) For $h \in (-\varepsilon/2, \varepsilon/2)$

$$\frac{M(h) - M(0)}{h} = \mathbb{E}\,\frac{e^{hX} - 1}{h}$$

Note that $\lim_{h\to 0} \frac{e^{hx}-1}{h} = x$

By MVT, there exists ξ between 0 and h s.t.

$$\left|\frac{e^{hx}-1}{h}\right| = \left|\frac{hx \cdot e^{\xi x}}{h}\right| = |x|\, e^{\xi x}.$$

By $|u| \le e^{u}+e^{-u}$

$$\left|\frac{e^{hx}-1}{h}\right| = |x|\, e^{\xi x}$$

$$= \frac{2}{\varepsilon} \cdot \left(\frac{\varepsilon}{2}|x|\right) e^{\xi x}$$

$$\le \frac{2}{\varepsilon}\left(e^{\frac{\varepsilon}{2}x} + e^{-\frac{\varepsilon}{2}x}\right)\cdot e^{\xi x}$$

$$= \frac{2}{\varepsilon}\left(e^{(\frac{\varepsilon}{2}+\xi)x} + e^{-(\frac{\varepsilon}{2}-\xi)x}\right)$$

$$\le \frac{2}{\varepsilon}\left(e^{\varepsilon x} + e^{-\varepsilon x}\right) \quad \text{since } |\xi| \le |h| < \frac{\varepsilon}{2}$$

Now note that

$$\mathbb{E}\left(e^{\varepsilon x} + e^{-\varepsilon x}\right) = M_X(\varepsilon) + M_X(-\varepsilon) < \infty$$

by the assumption.

Therefore, $\dfrac{e^{hx}-1}{h}$ is dominated by integrable.

$$\underbrace{\frac{2}{\varepsilon}\left(e^{\varepsilon x}+e^{-\varepsilon x}\right)}_{\text{independent of } h.}$$

By the dominated convergence theorem,

$$\frac{d}{dt}M(t)\Big|_{t=0} = \lim_{h\to 0}\frac{M(h)-M(0)}{h}$$

$$= \lim_{h\to 0}\mathbb{E}\,\frac{e^{hx}-1}{h}$$

$$= \mathbb{E}\lim_{h\to 0}\frac{e^{hx}-1}{h}$$

$$= \mathbb{E}\,X.$$

Delta Method

CLT : $\sqrt{n}(\bar{X}-\mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

More about CLT: Delta method

Suppose X_n are i.i.d. random variables with $EX_n = 0$, $VAR(X_n) = \sigma^2 > 0$. Let g be a measurable function that is differentiable at 0 with $g'(0) \neq 0$. Then

$$\sqrt{n}\left(g\left(\frac{\sum_{k=1}^n X_k}{n}\right) - g(0)\right) \to N(0, \sigma^2 g'(0)^2) \quad \text{weakly.}$$

Proof under stronger assumption: Here, we suppose g is continuously differentiable on $\mathbb{R}$. If you are interested in a general proof refer to Robert Keener's *Theoretical Statistics*.

$\sqrt{n}\left(g(\bar{X}) - g(0)\right)$

By MVT, there exists c_n s.t.

$g(\bar{X}) - g(0) = g'(c_n) \cdot \bar{X}$, where

c_n is between 0 and $\bar{X}$,

By SLLN, $\bar{X} \to 0$ a.s.

Since c_n is between 0 and $\bar{X}$, we have $c_n \to 0$ a.s.

Since g is continuously differentiable,

$$\lim_{n\to\infty} g'(c_n) = \underbrace{g'(0)}_{\text{constant}} \quad \text{a.s.}$$

By CLT, $\sqrt{n}\,\bar{X} \xrightarrow{d} N(0, \sigma^2)$.

$$\sqrt{n}\left(g(\bar{X}) - g(0)\right) = \underbrace{g'(c_n)}_{\xrightarrow{a.s.} \underbrace{g'(0)}_{\text{constant}}} \cdot \underbrace{\sqrt{n}\,\bar{X}}_{\xrightarrow{d} N(0,\sigma^2)} \xrightarrow{d} N(0, g'(0)^2\sigma^2)$$

by Slutsky's theorem.

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 $\mathbb{P}(90 < Y < 110)$.