UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 8

Ichiro Hashimoto

University of Toronto

July 22, 2026

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 $f(X_n, Y_n)$?

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

UNIVERSITY OF TORONTO

Convergence of functions of random variables

Proof of $X_n + Y_n \xrightarrow{a.s.} X + Y$:

Since $X_n \xrightarrow{a.s.} X$ there exists $E_X \subset \Omega$ s.t.

$$\mathbb{P}(E_X) = 1 \quad \text{and} \quad \lim_{n\to\infty} X_n(\omega) = X(\omega) \text{ for any } \omega \in E_X$$

Since $Y_n \xrightarrow{a.s.} Y$ there exists $E_Y \subset \Omega$ s.t.

$$\mathbb{P}(E_Y) = 1 \quad \text{and} \quad \lim_{n\to\infty} Y_n(\omega) = Y(\omega) \text{ for any } \omega \in E_Y$$

Then for any $\omega \in E_X \cap E_Y$, we have

$$\lim_{n\to\infty} (X_n(\omega) + Y_n(\omega)) = X(\omega) + Y(\omega).$$

$$\mathbb{P}(E_X \cap E_Y) = 1 - \mathbb{P}(E_X^c \cup E_Y^c) \underset{\text{union bound}}{\geq} 1 - (\underbrace{\mathbb{P}(E_X^c)}_{=0} + \underbrace{\mathbb{P}(E_Y^c)}_{=0}) = 1$$

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

Convergence of functions of random variables

Proof of $X_n + Y_n \xrightarrow{P} X + Y$:

Let $\forall \varepsilon > 0$.

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

by the triangle inequality

due to

$a + b > \varepsilon \Rightarrow a > \frac{\varepsilon}{2}$ or $b > \frac{\varepsilon}{2}$

$$\leq \mathbb{P}\left(|X_n - X| > \frac{\varepsilon}{2} \text{ or } |Y_n - Y| > \frac{\varepsilon}{2} \right)$$

by union bound

$$\leq \underbrace{\mathbb{P}\left(|X_n - X| > \frac{\varepsilon}{2}\right)} + \underbrace{\mathbb{P}\left(|Y_n - Y| > \frac{\varepsilon}{2}\right)}$$

$\to 0$ since $X_n \xrightarrow{P} X$

$\to 0$ since $Y_n \xrightarrow{P} Y$.

$\to 0$.

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,

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

but you cannot show $X_n Y_n \xrightarrow{L^p} XY$

Remark:

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

Convergence of functions of random variables

Proof of $X_n + Y_n \xrightarrow{L^p} X + Y$**:**

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

$$\|X_n + Y_n - (X+Y)\|_{L^p} = \|(X_n - X) + (Y_n - Y)\|_{L^p}$$

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

Convergence of functions of random variables

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 $c \neq 0$.

You cannot allow $Y_n \xrightarrow{d} Y$ for arbitrary Y.

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

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

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 taking value at discontinuity points of g

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

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

L^p convergence fails in general.

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

Let $g(x) = x^2$

Then $g(X_n) = g(X) \notin L^p$ since $X^2 \notin L^p$ since $X \notin L^{2p}$.

X takes values at continuity points of g with probability one

"

g is essentially continuous w.r.t X.

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

Weak Law of Large Numbers (WLLN)

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

Law of large numbers

A generalization of the theorem: triangular array

Triangular array

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

$$
\begin{aligned}
n=1 &\longrightarrow X_{1,1}\\
n=2 &\longrightarrow X_{2,1}, X_{2,2}\\
n=3 &\longrightarrow X_{3,1}, X_{3,2}, X_{3,3}\\
&\vdots\\
n &\longrightarrow X_{n,1}, X_{n,2}, \cdots, X_{n,n} \xrightarrow{\text{sum}} S_n
\end{aligned}
$$

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 $\boxed{S_n = \sum_{k=1}^{n} X_{n,k}}$, $\mu_n = \mathbb{E}(S_n)$, 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

sum of nth row

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

Remark:

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

by $\left(\begin{array}{l} X_{n,k} \quad \text{i.i.d with} \quad \sigma^2 = Var(X_i) \\ b_n = n \end{array} \right.$

$\sigma_n^2 = n\sigma^2$ so $\frac{\sigma_n^2}{b_n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0$

Law of large numbers

We first prove L^2 convergence.

Proof: $\mathbb{E}\left|\frac{S_n-\mu_n}{b_n}\right|^2 = \frac{\mathbb{E}(S_n-\mu_n)^2}{b_n^2} = \frac{Var(S_n)}{b_n^2} = \frac{\sigma_n^2}{b_n^2} \to 0$

by assumption.

Therefore $\frac{S_n-\mu_n}{b_n} \to 0$ in L^2.

Since L^2 convergence implies convergence in probability,

we have $\frac{S_n-\mu_n}{b_n} \xrightarrow{p} 0$.

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

no need to assume finite variance.

stronger conclusion

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 $F_n(x) = \frac{1}{n} \sum_{i=1}^n \boxed{I(X_i \leq x)}$, then as $n \to \infty$,

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

F_n is the proportion of X_i's which satisfies $X_i \leq x$

$$I(X_i \leq x) = \begin{cases} 1 & \text{if } X_i \leq x \\ 0 & \text{otherwise.} \end{cases}$$

Law of large numbers

Proof of Glivenko-Cantelli theorem:

weaker version : $|F(x) - F_n(x)| \to 0$ a.s. for any $x \in \mathbb{R}$.

Note that $0 \le I(X_i \le x) \le 1$.

So, $0 \le \mathbb{E}\, I(X_i \le x) = P(X_i \le x) = \underline{F(x)} \le 1$.

Now we apply SLLN to $I(X_i \le x)$.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le x) \xrightarrow{a.s} \underline{\mathbb{E}\, I(X_i \le x)} = F(x)$$

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

$\frac{S_n}{n}$ diverges with probability one.

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

only distribution converge

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

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

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

Informal Implication of CLT

$S_n \stackrel{\sim}{\approx} N(n\mu, n\sigma^2)$

$\bar{X}_n \stackrel{\sim}{\approx} N(\mu, \frac{\sigma^2}{n})$

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.

$Z_n \to N(0,1)$ in distribution

$E X_i = p$

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

Monotone Convergence Theorem

Monotone Convergence Theorem

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

Usage: let $X_n = \begin{cases} \frac{1}{n^2} & \text{w.p } p \\ 0 & \text{w.p } 1-p \end{cases}$ and $S_n = \sum_{i=1}^{n} X_i$

Then $0 \leq S_n \nearrow \lim_{n\to\infty} S_n = S = \sum_{i=1}^{\infty} X_i$.

By the monotone convergence theorem, we have.

$$\mathbb{E} S = \lim_{n\to\infty} \mathbb{E} S_n = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E} X_i = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{p}{i^2} = p \sum_{i=1}^{\infty} \frac{1}{i^2}$$

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

Dominat~~e~~ Convergence Theorem

$\mathbb{E}|Y| < \infty$

Dominated Convergence Theorem

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

X_n's are "dominated" by integrable Y

Usage:

Prop

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

$\varepsilon > 0$ is arbitrary

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

(Pf) $\frac{M(h) - M(0)}{h} = \mathbb{E}\,\frac{e^{hX} - 1}{h}$ $\to$ Note that $\lim_{h\to 0} \frac{e^{hX}-1}{h} = X$

By the mean value theorem, there exists ξ between 0 and h.

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

Let $h \in (-\varepsilon/2, \varepsilon/2)$.

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

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

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

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

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

Note that $|\xi| \le |h| \le \frac{\varepsilon}{2}$

$$\le \underbrace{\frac{2}{\varepsilon}\left(e^{\varepsilon x} + e^{-\varepsilon x}\right)}_{\text{doesn't depend on } h}.$$

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

Thus $\frac{2}{\varepsilon}\left(e^{\varepsilon x} + e^{-\varepsilon x}\right)$ is integrable.

Thus $\frac{e^{hx}-1}{h}$ is dominated by an integrable random variable.

By the dominated convergence theorem applied to

$$\frac{e^{hX}-1}{h} \xrightarrow{a.s.} X \quad (h \to 0),$$

we obtain

$$\frac{d}{dt} M(t)\Big|_{t=0} = \lim_{h\to 0} \mathbb{E}\, \frac{e^{hX}-1}{h} = \mathbb{E} \lim_{h\to 0} \frac{e^{hX}-1}{h} = \mathbb{E} X$$

Delta Method

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*.

By MVT, there exists c_n s.t.

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

where c_n is between $\bar{X}$ and 0.

Delta Method

Proof Cont'd:

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

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

Since g is continuously differentiable

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

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

Then

$$\sqrt{n}\left(g(\bar{X}) - g(0)\right) = \sqrt{n} \cdot g'(c_n)\, \bar{X}$$

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

$$\longrightarrow N(0, g'(0)^2 \sigma^2)$$ 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 $\mathbb{P}(90 < Y < 110)$.