UNIVERSITY OF TORONTO | Statistical Sciences

# DoSS Summer Bootcamp Probability Module 8

Ichiro Hashimoto

University of Toronto

July 24, 2024

# 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 $f(X_n, Y_n)$?

e.g. $X_n + Y_n \to ?$

$X_n \cdot Y_n \to ?$

$X_n / Y_n \to ?$

$S_n = X_1 + \cdots + X_n$

$S_n \to ?$

# 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

# 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

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

(Proof) Recall def $\xrightarrow{P}$

$$X_n \xrightarrow{P} X \iff \forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

By triangle inequality

$$|(X_n + Y_n) - (X+Y)| \leq |X_n - X| + |Y_n - Y|.$$

So,

$$\underbrace{\{|(X_n+Y_n) - (X+Y)| > \varepsilon\} \subset \{|X_n - X| > \tfrac{\varepsilon}{2}\} \cup \{|Y_n - Y| > \tfrac{\varepsilon}{2}\}}_{\text{Union bound argument}}$$

Hence.

$$\mathbb{P}(|(X_n+Y_n) - (X+Y)| > \varepsilon) \leq \underbrace{\mathbb{P}(|X_n - X| > \tfrac{\varepsilon}{2})}_{\to 0} + \underbrace{\mathbb{P}(|Y_n - Y| > \tfrac{\varepsilon}{2})}_{\to 0}$$

since $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$.

Thus

$$\lim_{n\to\infty} \mathbb{P}(|(X_n+Y_n) - (X+Y)| > \varepsilon) = 0 \quad \text{for } \forall \varepsilon > 0$$

//

# Convergence of functions of random variables

$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] = 0$

## 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

$$X_n + Y_n \to X + Y$$

(pf)

Fact If $p \geq 1$, then $L^p$ space has triangle inequality

i.e.

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

$$\|X\|_{L^p} = \left( \mathbb{E}|X|^p \right)^{1/p}$$

By triangle inequality

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

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

# Convergence of functions of random variables

**Remark**: Convergence in distribution is different.

## Slutsky's theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ $=0$ ($c$ 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:**

$\sim \mathcal{N}(0,1)$

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

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

# 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 $g(\cdot)$.

g is essentially continuous w.r.t. X

UNIVERSITY OF TORONTO

# 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, $\mu = \cancel{\mathbb{E}(|X_i|)} \cancel{< \infty}$, then

$\mathbb{E} X_i$, $\mathbb{E}|X_i| < \infty$

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

**Remark:**

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

**Sketch of the proof:**

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

$$= Var\left(\frac{\sum X_i}{n}\right)$$

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

$$= \frac{n\,Var(X_1)}{n^2} = \frac{Var(X_1)}{n} \to 0$$

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 $\{X_{n,k}\}_{1\leq k\leq n}$.

$$\begin{array}{lll}
n=1 \to & X_{1,1} & \xrightarrow{\text{sum}} S_1 \\
n=2 \to & X_{2,1}, X_{2,2} & \xrightarrow{\text{sum}} S_2 \\
n=3 \to & X_{3,1}, X_{3,2}, X_{3,3} & \xrightarrow{\text{sum}} S_3 \\
 & \vdots & \\
n \to & X_{n,1}, X_{n,2}, \cdots, X_{n,n} & \xrightarrow{\text{sum}} S_n
\end{array}$$

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

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

# 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:**
$$\mathbb{E}\left|\frac{S_n-\mu_n}{b_n}\right|^2 = \frac{\sigma_n^2}{b_n^2}. \to 0$$

So, $\frac{S_n-\mu_n}{b_n} \to 0$ in $L^2$,

and $\to 0$ 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 $F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$, then as $n \to \infty$,

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

$F_n(x)$ is random

it's hard to prove with supremum.

# ~~Law of large numbers~~

**Proof:** Weaker version.

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

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

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

finite.

By SLLN, $F_n(x) \to F(x)$ a.s.

UNIVERSITY OF TORONTO

# 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**

# 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

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

convergen in "distribution"

**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.

# Monotone Convergence Theorem

**Monotone Convergence Theorem**

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

**Usage:**

Let $X_n$ be $\mathbb{P}(X_n = \frac{1}{n^2}) = p = 1 - \mathbb{P}(X_n = 0)$

Note $0 \leq X_n \leq \frac{1}{n^2}$, and $\mathbb{E} X_n = \frac{p}{n^2}$

Let $S_n = \sum_{i=1}^{n} X_i$. Then $S_n \geq 0$ and is monotonically increasing.

Furthermore, $S_n \leq \sum_{i=1}^{n} \frac{1}{i^2} \leq \frac{\pi^2}{6} < \infty$

Therefore $S_n \nearrow^{?} S \leq \frac{\pi^2}{6}$

S is random

Q. $\mathbb{E}\xi_n \to \mathbb{E}\xi$ ?

A. Yes by monotone convergence theorem.

Therefore $\mathbb{E}\xi = \lim\limits_{n\to\infty} \mathbb{E}\xi_n$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{i^2} = 1 \cdot \frac{\pi^2}{6}$$

we used

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

# Dominate Convergence Theorem

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

$\parallel$

**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 $EX_n \to EX$

all $X_n$ must be dominated integrable $Y$.

**Usage:**

We can show if $M(t) < \infty$ for any $t \in [-\varepsilon, \varepsilon]$,

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

when $M(t) = \mathbb{E}\exp(Xt)$, moment generating function of $X$.

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

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

Note that $\displaystyle\lim_{h\to 0} \frac{e^{hX}-1}{h} = X$.

Note also that $\displaystyle\left|\frac{e^{hX}-1}{h}\right| = \left|\frac{hX\cdot e^{\xi X}}{h}\right| = |X|\, e^{\xi X}$

by mean value theorem, where $\xi$ is between $0$ and $h$.

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

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

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

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

between $0 \sim h$, $|h| < \frac{\varepsilon}{2}$

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

$$\le \underbrace{\frac{2}{\varepsilon}\left(e^{\varepsilon X} + e^{-\varepsilon X}\right)}_{\text{integrable by assumption.}}$$

By dominated convergence theorem,

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

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

# 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 mean value theorem, there exists $c_n$ s.t.

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

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

By SLLN, $\lim_{n\to\infty} \bar{X} = 0$ a.s.

Since $C_n$ is between $0$ and $\bar{X}$, $\lim_{n\to\infty} C_n = 0$ a.s.

Since $g$ is continuously differentiable.

$$\lim_{n\to\infty} g'(C_n) = \underline{g'(0)}$$

$\underline{\text{constant}}$

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

Therefore, by Slutsky's theorem,

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

$$\xrightarrow{d} g'(0)\cdot N(0, \sigma^2)$$

$$= N(0, \sigma^2 g'(0)^2)$$

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