UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 7

Ichiro Hashimoto

University of Toronto

July 21, 2025

Recap

Learnt in last module:

- Covariance
  - Covariance as an inner product
  - Correlation
  - Cauchy-Schwarz inequality
  - Uncorrelatedness and Independence
- Concentration
  - Markov's inequality
  - Chebyshev's inequality
  - Chernoff bounds

Outline

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

Stochastic Convergence

Recall: Convergence

Convergence of a sequence of numbers

A sequence $a_1, a_2, \cdots$ converges to a limit a if

$$\lim_{n \to \infty} a_n = a.$$

That is, for any $\epsilon > 0$, there exists an $N(\epsilon)$ such that

$$|a_n - a| < \epsilon, \quad \forall n > N(\epsilon).$$

Stochastic Convergence

Recall: Convergence

Convergence of a sequence of numbers

A sequence $a_1, a_2, \cdots$ converges to a limit a if

$$\lim_{n\to\infty} a_n = a.$$

That is, for any $\epsilon > 0$, there exists an $N(\epsilon)$ such that

$$|a_n - a| < \epsilon, \quad \forall n > N(\epsilon).$$

a_n is concentrated near a.

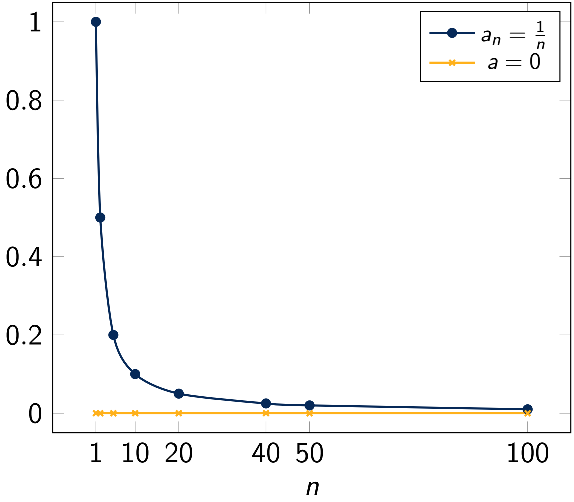
Example: $a_n = \frac{1}{n}$, $\forall \epsilon > 0$, take $N(\epsilon) = \lceil \frac{1}{\epsilon} \rceil$, then for $n > N(\epsilon)$,

$n > \lceil \frac{1}{\epsilon} \rceil \geq \frac{1}{\epsilon}$

$\therefore \frac{1}{n} < \epsilon$

$$|a_n - 0| = a_n < \epsilon, \quad \lim_{n\to\infty} a_n = 0.$$

Stochastic Convergence



- Capture the property of a series as $n \to \infty$;
- The limit is something where the series concentrate for large n;
- $|a_n - a|$ quantifies the closeness of the series and the limit.

UNIVERSITY OF TORONTO

Stochastic Convergence

Observation: closeness of random variables

Sample mean of i.i.d. random variables

For i.i.d. random variables $X_i, i = 1, \cdots, n$ with $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$, then for the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{X}$ depends on n.

independent and identically distributed

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

Proof: $\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \overset{\text{linearity of } \mathbb{E}}{=} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i \overset{\text{i.i.d.}}{=} \frac{1}{n}\cdot n\mu = \mu.$

$$Var(\bar{X}) = \mathbb{E}(\bar{X}-\mu)^2 = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2$$

$$= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} (X_i - \mu)\right)^2.$$

$$= \frac{1}{n^2} \mathbb{E} \sum_{i=1}^{n} (X_i - \mu)^2 + \frac{1}{n^2} \mathbb{E} \sum_{i \neq j} (X_i - \mu)(X_j - \mu)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{\mathbb{E}(X_i - \mu)^2}_{\substack{= \mathrm{Var}(X_i) \\ = \sigma^2}} + \frac{1}{n^2} \sum_{i \neq j} \underbrace{\mathbb{E}(X_i - \mu)(X_j - \mu)}_{=0}$$

$= 0$ since X_i's are independent and $\mathbb{E} X_i = \mu$

$$= \frac{1}{n^2} \cdot n\sigma^2 + 0 = \frac{\sigma^2}{n}.$$

Stochastic Convergence

Example:

Further suppose $X_i, i = 1, \cdots, n$ i.i.d. with distribution $\mathcal{N}(\mu, \sigma^2)$, then $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, so we can draw the probability density plot of $\bar{X}$.

Stochastic Convergence

Example:

Further suppose $X_i, i = 1, \cdots, n$ i.i.d. with distribution $\mathcal{N}(\mu, \sigma^2)$, then $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, so we can draw the probability density plot of $\bar{X}$.

Variance gets smaller as n increases

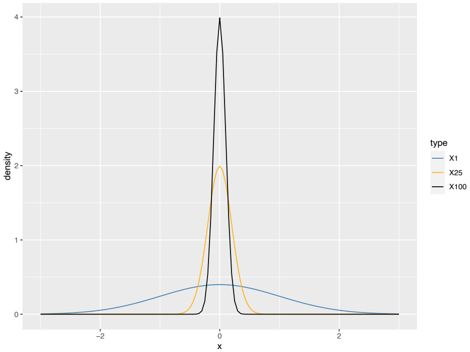


Figure: Probability density curve of sample mean of normal distribution

Stochastic Convergence

Intuition:

- Series of numbers $a_n \quad \Rightarrow \quad$ Series of random variables X_n;
- Limit $a \quad \Rightarrow \quad$ Limit X;
- How to quantify the closeness? ($|X_n - X|$?)

Stochastic Convergence

Intuition:

- Series of numbers $a_n \quad \Rightarrow \quad$ Series of random variables X_n;
- Limit $a \quad \Rightarrow \quad$ Limit X;
- How to quantify the closeness? $(|X_n - X|?)$

Pointwise convergence / Sure convergence

Suppose random variables X_n and X are defined over the same probability space, then we say X_n converges to X pointwise if

$$\lim_{n \to \infty} X_n(\omega) = X(\omega), \ \forall \omega \in \Omega.$$

For fixed w, $\{X_n(w)\}$ is a sequence.

Stochastic Convergence

Intuition:

- Series of numbers $a_n \quad \Rightarrow \quad$ Series of random variables X_n;
- Limit $a \quad \Rightarrow \quad$ Limit X;
- How to quantify the closeness? $(|X_n - X|?)$

Pointwise convergence / Sure convergence

Suppose random variables X_n and X are defined over the same probability space, then we say X_n converges to X pointwise if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \ \forall \omega \in \Omega.$$

Remark:

Incorporate probability measure in some sense.

Stochastic Convergence

Alternatives of describing the closeness:

- Utilize CDF: $F_{X_n}(x) - F_X(x)$;
- Utilize probability of an event: $\mathbb{P}(|X_n - X| > \epsilon)$;
- Utilize the probability over all ω: $\mathbb{P}(\lim_{n\to\infty} X_n(\omega) = X(\omega))$;
- Utilize mean/moments: $\mathbb{E}|X_n - X|^p$.

Stochastic Convergence

Use CDF to quantify the closeness of X_n and X.

Convergence in distribution

A sequence $X_1, X_2, \cdots$ of real-valued random variables is said to converge in distribution, or converge weakly to a random variable X if

$$\lim_{n \to \infty} F_n(x) = F(x),$$

for every number $x \in \mathbb{R}$ at which $F(\cdot)$ is continuous. Here, $F_n(\cdot)$ and $F(\cdot)$ are the cumulative distribution functions of the random variables X_n and X, respectively.

Notation:

$X_n \xrightarrow{d} X, \quad X_n \xrightarrow{\mathcal{D}} X, \quad X_n \Rightarrow X.$

UNIVERSITY OF TORONTO

Stochastic Convergence

Convergence in distribution

A sequence $X_1, X_2, \cdots$ of real-valued random variables is said to converge in distribution, or converge weakly to a random variable X if

$$\lim_{n \to \infty} F_n(x) = F(x),$$

for every number $x \in \mathbb{R}$ at which $F(\cdot)$ is continuous. Here, $F_n(\cdot)$ and $F(\cdot)$ are the cumulative distribution functions of the random variables X_n and X, respectively.

Notation:

$X_n \xrightarrow{d} X, \quad X_n \xrightarrow{\mathcal{D}} X, \quad X_n \Rightarrow X.$

Remark:

X_n and X do not need to be defined on the same probability space.

since this convergence is about closeness of distributions,

Stochastic Convergence

Example:
Let $X_n = Z + \frac{1}{n}$, where $Z \sim \mathcal{N}(0,1)$, then

- $X_n \xrightarrow{d} Z$,
- $X_n \xrightarrow{d} -Z$,
- $X_n \xrightarrow{d} Y$, $\boxed{Y \sim \mathcal{N}(0,1).}$

X_n can converge to multiple random variables at the same time.

Here, Z and $-Z$ both are $\sim \mathcal{N}(0,1)$

A new random variable which could be defined on on a different probability space.

Proof:

(.)

$$\underbrace{\mathbb{P}(X_n \leq x)}_{\text{def. of CDF}} = \mathbb{P}\left(Z \leq x - \frac{1}{n}\right)$$

$$= \Phi\left(x - \frac{1}{n}\right),$$ where Φ is the CDF of $\mathcal{N}(0,1)$.

$\xrightarrow{n\to\infty}$ $\Phi(x) = \mathbb{P}(X \leq x)$

$\uparrow$.

due to continuity Φ.

Φ is continuous since $N(0,1)$ is a continuous distribution

or in other words, there exists density.

$$\Phi(x) = \underbrace{\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du}_{\text{continuous in } x.}$$

2) Since $N(0,1)$ is symmetric,

$$\mathbb{P}(-Z \leq x) = \mathbb{P}(Z \geq -x)$$

$$= \mathbb{P}(Z \leq x).$$

due to symmetry

3) Since $Y \sim N(0,1)$,

$$\mathbb{P}(Y \leq x) = \mathbb{P}(Z \leq x)$$

Stochastic Convergence

quantifying closeness of X_i and X using probability of an event

Convergence in probability

A sequence X_n of random variables converges in probability towards the random variable X if for all $\epsilon > 0$,

$$\lim_{n\to\infty} \boxed{\mathbb{P}(|X_n - X| > \epsilon)} = 0.$$

Notation: $X_n \xrightarrow{p} X, \quad X_n \xrightarrow{P} X.$

Remark:
X_n and X need to be defined on the same probability space.

so $\mathbb{P}(|X_n - X| > \varepsilon)$ makes sense

Stochastic Convergence

Examples:

- Let $X_n = Z + \frac{1}{n}$, where $Z \sim \mathcal{N}(0,1)$, then $X_n \xrightarrow{P} Z$.

Proof: Let $\forall \varepsilon > 0$. $\mathbb{P}(|X_n - Z| > \varepsilon) = \mathbb{P}(\frac{1}{n} > \varepsilon) = 0$ if $\frac{1}{n} \leq \varepsilon$ $\Leftrightarrow n \geq \frac{1}{\varepsilon}$

Thus, $\lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \varepsilon) = 0$

- Let $X_n = Z + Y_n$, where $Z \sim \mathcal{N}(0,1)$, $\mathbb{E}(|Y_n|) = \frac{1}{n}$, then $X_n \xrightarrow{P} Z$.

Proof: $\mathbb{P}(|X_n - Z| > \varepsilon) = \mathbb{P}(|Y_n| > \varepsilon)$

by Markov $\leq \varepsilon^{-1} \mathbb{E}|Y_n| = \frac{1}{n\varepsilon} \to 0$ as $n \to \infty$.

Thus $\lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \varepsilon) = 0$. for $\forall \varepsilon > 0$.

Stochastic convergence

Probability of pointwise convergence

almost all

Convergence almost surely

A sequence X_n of random variables converges almost surely or almost everywhere or with probability 1 or strongly towards X means that

$$\boxed{\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right)} = \mathbb{P}\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

Notation: $X_n \xrightarrow{a.s.} X$. almost sure

$X_n \xrightarrow{a.e.} X$, $X_n \xrightarrow{a.a.} X$, $X_n \longrightarrow X$ w.p.1.

Remark:
X_n and X need to be defined on the same probability space.

a.s. convergence allows $\lim_{n\to\infty} X_n(\omega) \neq X(\omega)$ with probability zero

this makes significant difference from pointwise convergence.

Stochastic convergence

Examples:

- Let $X_n = Z + \frac{1}{n}$, where $Z \sim \mathcal{N}(0,1)$, then $X_n \xrightarrow{a.s.} Z$.

  Proof: Pointwise argument is necessary for a.s. convergence.

  For any $\omega \in \Omega$, $\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} \left(Z(\omega) + \frac{1}{n}\right) = Z(\omega) + 0 = Z(\omega)$.

  Thus, $\mathbb{P}\left(\lim_{n\to\infty} X_n(\omega) = Z(\omega)\right) = 1$.

- Let $X_n = Z + Y_n$, where $Z \sim \mathcal{N}(0,1)$, $\underline{\mathbb{E}(|Y_n|) = \frac{1}{n}}$, do we have $X_n \xrightarrow{a.s.} Z$?

  Proof: we already know $X_n \xrightarrow{P} Z$.

  No, you can construct a counter example.

< Counterexample >

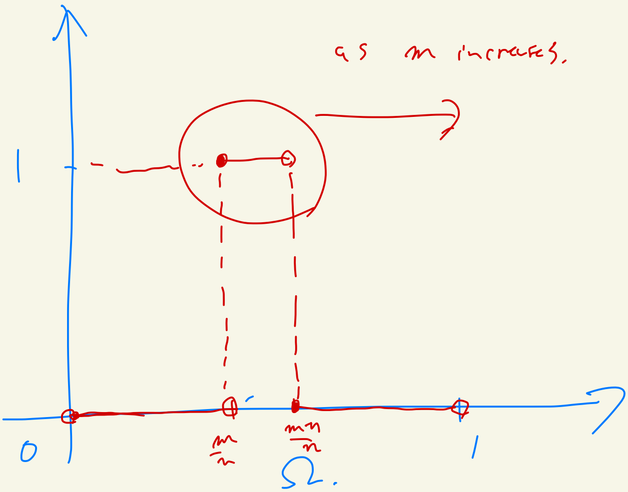
$$\Omega = (0,1], \quad \mathbb{P} \sim \mathrm{Unif}(0,1]$$

Define. $$Y_{m,n}(\omega) := \begin{cases} 1 & \text{if } \omega \in [\frac{m}{n}, \frac{m+1}{n}) \\ 0 & \text{otherwise} \end{cases}$$

$0 \leq m \leq n-1$

$$\mathbb{P}(Y_{m,n} = 1) = \frac{1}{n}.$$

and $$\mathbb{E}|Y_{m,n}| = \frac{1}{n}.$$



For each n
it moves from left to right.

that means

$\lim_{m,n\to\infty} Y_{m,n}(\omega)$ does not exist.

Since $Y_{m,n}$ is not converging to 0 everywhere,

$X_n(\omega) \not\to Z(\omega)$ everywhere.

Stochastic convergence

Use pth moment to quantify the closeness of r.v.s

Convergence in L^p

A sequence $\{X_n\}$ of random variables converges in L_p to a random variable X, $p \geq 1$, if

$$\lim_{n\to\infty} \boxed{\mathbb{E}|X_n - X|^p} = 0 \iff \lim_{n\to\infty} \|X_n - X\|_{L^p} = 0$$

Notation: $X_n \xrightarrow{L^p} X$.

Remark:
X_n and X need to be defined on the same probability space.

Stochastic convergence

Examples:

- Let $X_n = Z + \frac{1}{n}$, where $Z \sim \mathcal{N}(0,1)$, then $X_n \xrightarrow{L^p} Z$.

  Proof: $\mathbb{E}\,|X_n - Z|^p = \mathbb{E}\,\frac{1}{n^p} = \frac{1}{n^p} \longrightarrow 0$ as $n \to \infty$

- Let $X_n = Z + Y_n$, where $Z \sim \mathcal{N}(0,1)$, $\mathbb{E}(|Y_n|^p) = \frac{1}{n}$, then $X_n \xrightarrow{L^p} Z$.

  Proof:

  $\mathbb{E}\,|X_n - Z|^p = \mathbb{E}\,|Y_n|^p = \frac{1}{n} \longrightarrow 0$ as $n \to \infty$

Stochastic convergence

L^p norm is indeed a norm when $p \ge 1$.

Recall: A random variable $X \in L^p$ if $\|X\|_{L^p} = (E|X|^p)^{1/p} < \infty$.
$X_n \to X$ in L^p if $\lim_{n\to\infty} \|X_n - X\|_{L^p} = 0$

Monotonicity of L^p Convergence

If $q > p > 0$, L^q convergence implies L^p convergence

Proof: Lyapunov inequality.

$$\underbrace{(E|X|^p)^{1/p}}_{=\|X\|_{L^p}} \le \underbrace{(E|X|^q)^{1/q}}_{=\|X\|_{L^q}} \quad \text{if } 0<p<q$$

By Lyapunov inequality

$$\left(\mathbb{E}|X_n - X|^p\right)^{1/p} \leq \left(\mathbb{E}|X_n - X|^q\right)^{1/q} \longrightarrow 0 \text{ as } n \to \infty$$

Thus, $X_n \xrightarrow{L^p} X.$

Stochastic convergence

Recall: X_n converges to X in probability if for any $\epsilon > 0$ $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$.

L^p convergence implies Convergence in Probability

If $X_n \to X$ in L^p, then $X_n \to X$ in probability.

Proof: By Markov inequality

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p)$$

Markov

$$\leq \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p} \to 0$$

since $X_n \to X$ in L^p

UNIVERSITY OF TORONTO

Stochastic convergence

Recall: X_n converges to X in probability if for any $\epsilon > 0$ $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$.

a.s. Convergence implies Convergence in Probability

If $X_n \to X$ almost surely, then $X_n \to X$ in probability.

Proof:

Stochastic convergence

Recall: X_n converges to X in distribution if for any continuity point x of $P(X \leq x)$, $\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x)$ holds.

Convergence in Probability implies Convergence in Distribution

If $X_n \to X$ in probability, then $X_n \to X$ in distribution.

Proof: Omitted

Stochastic convergence

Relationship between convergences (on complete probability space):

strong

$$\begin{array}{ccccc} \xrightarrow{L^s} & \underset{s>r\geq 1}{\Rightarrow} & \xrightarrow{L^r} & & \\ & & \Downarrow & & \\ \xrightarrow{\text{a.s.}} & \Rightarrow & \xrightarrow{p} & \Rightarrow & \xrightarrow{d} \end{array}$$

weakest

Figure: relationship between convergences

Stochastic convergence

Highlights:

- Almost sure convergence implies convergence in probability:

$$X_n \xrightarrow{\text{a.s.}} X \quad \Rightarrow \quad X_n \xrightarrow{P} X;$$

- Convergence in probability implies convergence in distribution:

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{d} X;$$

- If X_n converges in distribution to a constant c, then X_n converges in probability to c:

$$X_n \xrightarrow{d} c \quad \Rightarrow \quad X_n \xrightarrow{P} c, \quad \text{provided } c \text{ is a constant.}$$

Problem Set

Problem 1: Prove that on a complete probability space, if $X_n \xrightarrow{L^p} X$, then $X_n \xrightarrow{P} X$. (Hint: use Markov's inequality)

Problem 2: Let $X_1, \cdots, X_n$ be i.i.d. random variables with $Bernoulli(p)$ distribution, and $X \sim Bernoulli(p)$ is defined on the same probability space, independent with X_i's. Does X_n converge in probability to X?

Problem 3: Give an example where X_n converges in distribution to X, but not in probability.