Estimation Theory

2102523

Suwichaya Suwanwimolkul[1]

[1]Electrical Engineering
Chulalongkorn University

Semester I, 2023

Lecture1/Lecture1.tex

- H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*.
  Kappa Research, LLC, 2014.
- A. C. Robert Hogg, Joseph McKean, *Introduction to Mathematical Statistics*.
  Pearson; 7th edition, 2012.
- D. Romik, “Lecture Notes on Probability Theory, math 235a,” Fall 2009.
- L. Wasserman, “Lecture note 15 in intermediate statistics. stats 705,” Fall 2020.
- Z. Fan, “Lecture 3: Consistency and asymptotic normality of the MLE. stats 200,” Autumn 2016.
- D. Panchenko, “Lecture 3: Properties of MLE: consistency, asymptotic normality, Fisher information. 18.650,” Fall 2006.

Estimation Theory

Part II: Probabilistic Convergences

Suwichaya Suwanwimolkul[1]

[1]Electrical Engineering
Chulalongkorn University

Lecture II
Semester I, 2023

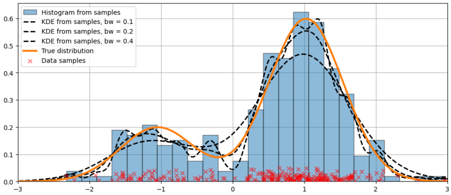
Table of Contents

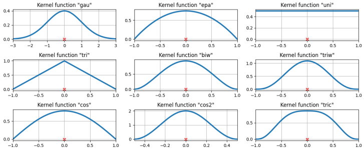
Contents

Why probabilistic convergence?

In many applications, we want the estimated results (distribution/ probability/ parameters/ samples) to be similar to GT.

For example ...



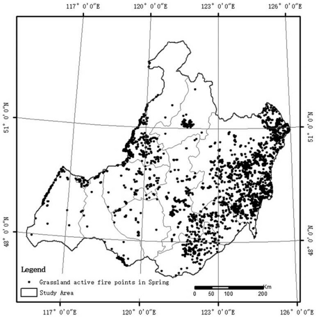


Images are taken from www.statsmodels.org

Kernel density estimation

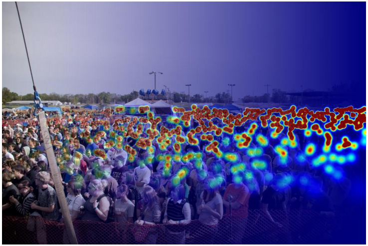
Why probabilistic convergence?

In many applications, we want the estimated results (distribution/ probability/ parameters/ samples) to be similar to GT.



The example is from https://www.degruyter.com

Find the active fire points



The example is from Crowd-counting, cvlab, EPFL

Count # people in the crowd

Kernel density estimation applications

Why probabilistic convergence?

In many applications, we want the estimated results (distribution/ probability/ parameters/ samples) to be similar to GT.

For example ...novel-view synthesizing

Pictures/video.png

How it works $p_{model}(\mathbf{x})$ $p_{data}(\mathbf{x})$

Images are taken from Real-Time Radiance Fields for Single-Image Portrait View Synthesis. SIGGRAPH 2023

Want $p_{model}(\mathbf{x})$ to be similar to $p_{data}(\mathbf{x})$ (generative model)

Why probabilistic convergence?

Today's topic is about how can we check if our estimator can converge ? and what are 4 different types of (probabilistic) convergences.

- Convergence in distribution.
- Convergence in probability.
- Convergence in mean.
- Almost sure convergence.

Settings

We want to see if a sequence of random variables $X_1, X_2, ..., X_n, ...$ converges to a random variable X.

Example 1: Applying probabilistic convergence for the estimators

We can partially observe the behavior of X...

- At first, we give an estimate of X and call it X_1.
- Given more observations, we update our estimation and call it X_2.
- Then, more observations, we update our estimation and call it X_3.
- and so on ...

As n increases, our estimator should provide X_n that is closer to X.

Why probabilistic convergence?

Today's topic is about how can we check if our estimator can converge ? and what are 4 different types of (probabilistic) convergences.

Convergence in distribution	$\Rightarrow$	CDF/CMF of X_n becomes X's.
Convergence in probability	$\Rightarrow$	Random variable X_n becomes X or 'x'.
Convergence in mean	$\Rightarrow$	Bounded differences btw. X_n and X's.
Almost sure convergence	$\Rightarrow$	Actual value X_n becomes X.

Why probabilistic convergence?

Today's topic is about how can we check if our estimator can converge ? and what are 4 different types of (probabilistic) convergences.

Convergence in distribution $\Rightarrow$ CDF/CMF of X_n becomes X's.

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

Convergence in probability $\Rightarrow$ Random variable X_n becomes X or 'x'.

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

Convergence in mean $\Rightarrow$ Bounded differences btw. X_n and X's.

$$\lim_{n\to\infty} \mathrm{E}[|X_n - X|^r] = 0$$

Almost sure convergence $\Rightarrow$ Actual value X_n becomes X.

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Convergence in Deterministic sequences

Convergence of deterministic sequences

Definition 1: Convergence of a deterministic sequence [1]

A sequence $x_1, x_2, ..., x_n$, converges to a limit L if

$$lim_{n\to\infty} x_n = L. \tag{1}$$

Then, there exists an $N \in \mathbb{N}$, such that

$$|x_n - L| \leq \epsilon, \quad \forall n > N. \tag{2}$$

Example 2: Examples of deterministic series

$$1,\ \frac{1}{2},\ \frac{1}{4},\ \frac{1}{8}, ..., \frac{1}{2^n}, ... \quad \Rightarrow \lim_{n\to\infty} \frac{1}{2^n} = 0 \qquad \Rightarrow\ |x_n - 0| \leq \epsilon = 1$$

$$\frac{1}{2},\ \frac{2}{3},\ \frac{3}{4}, ..., \frac{n}{n+1}, ... \quad \Rightarrow \lim_{n\to\infty} \frac{n}{n+1} = 1 \qquad \Rightarrow\ |x_n - 1| \leq \epsilon = \frac{1}{2}$$

Convergence in Probabilistic sequences

Convergence in distribution — Definition

Definition 2: Convergence in distribution

Random variables $X_1, X_2, X_3, \dots$ **converge in distribution** to a random variable X, shown by $X_n \xrightarrow{d} X$, if the CDF $\mathrm{F}_{X_n}(x)$ converges to $\mathrm{F}_X(x)$ as $n \to \infty$, or

$$\lim_{n \to \infty} \mathrm{F}_{X_n}(x) = \mathrm{F}_X(x) \tag{3}$$

for all x at which $F_X(x)$ is continuous.

Example 3: Example of Convergence in distribution

Let $X_1, X_2, \ldots$ be a sequence of random variables such that

$$F_{X_n}(x) = 1 - \left(1 - \frac{1}{n}\right)^{nx}, \quad \text{for } x > 0$$

$F_{X_n}(x) = 0$, if otherwise. Show that X_n converges in distribution to the same distribution as $X \sim p_X(x) = \mathrm{e}^{-x}$ whose CDF is defined as

$$P_X(x) = \begin{cases} 1 - \mathrm{e}^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Note that $\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{nx} = \mathrm{e}^{-x}$

Solution.

1. For $x \leq 0$,

2. For $x > 0$,

From 1. & 2., we can conclude that ...

Example 3 (Cont'): Example of Convergence in distribution

Let $X_1, X_2, ...$ be a sequence of random variables such that

$$F_{X_n}(x) = 1 - \left(1 - \frac{1}{n}\right)^{nx}, \quad \text{for } x > 0$$

$F_{X_n}(x) = 0$, if otherwise. Show that X_n converges in distribution to the same distribution as $X \sim p_X(x) = e^{-x}$ whose CDF is defined as

$$P_X(x) = \begin{cases} 1 - e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Solution.

1. For $x \leq 0$, we have $F_{X_n}(x) = 0$. Thus, $\lim_{n\to\infty} F_{X_n}(x) = 0 = P_X(x)$.

2. For $x > 0$, we have $\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} 1 - \left(1 - \frac{1}{n}\right)^{nx}$
$= 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{nx} = 1 - exp(-x) = P_X(x)$.

From 1. & 2., we can conclude that X_n converges in distribution to the same distribution as X.

Convergence in distribution & moment-generating function

Definition 3: Continuity Theorem

Let $M_{X_n}(t)$ denote the moment-generating function corresponding to a sequence of RVs X_n, and X is a random variable with the moment-generating function $M_X(t)$.

$$\text{If} \quad \lim_{n\to\infty} M_{X_n}(t) = M_X(t), \text{ then } \lim_{n\to\infty} \mathsf{F}_{X_n}(x) = \mathsf{F}_X(x) \tag{4}$$

for all the continuity points of F. That is $X_n \xrightarrow{d} X$.

Application: Lindeberg–Lévy CLT

Concept: *sum of iid RVs. will converge to a normal distribution.*

Let's look at the simulation Github Tutorial 2

Application: Lindeberg–Lévy CLT

Concept: *sum of iid RVs. will converge to a normal distribution.*

Example 4: Lindeberg–Lévy Central Limit Theorem (CLT) [2]

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with mean $\mathrm{E}[X_n] = \mu \leq \infty$ and variance $\mathrm{Var}[X_n] = \sigma^2 \leq \infty$. Then, the random variable Z_n, defined as,

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \tag{5}$$

converges **in distribution** to the standard normal random variable X, *i.e.*, $Z_n \xrightarrow{d} X$ where $X \sim \mathcal{N}(0, 1^2)$. That is, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.

Bonus 1. Show that this theorem is true (using Continuity Theorem). Hint in Slide 150

Notes about Lindeberg–Lévy CTL

1. Note that $X_1, X_2, \ldots$ can have any distribution. The sequence of RVs X_n can be, *e.g.*, discrete, continuous, or mixed random variables.

2. Unlike the estimator settings, $X_1, X_2, \ldots$ is **unrelated** to X.

3. Note that if $\lim_{n\to\infty} \mathsf{F}_{X_n}(x) = \mathsf{F}_X(x)$, this is equivalent to
   2.1 $\quad P_{X_n}(X_n \le x) \quad = \Phi_X(x)$, and
   2.2 $\quad P(a \le X_n \le b) = P(a \le X \le b)$
   where $\Phi_X(x)$ is the CDF of $\mathcal{N}(0, 1^2)$

4. If $X_n \xrightarrow{D} X$, means that they only share the same CDF. However,they are still not the same random variable, *i.e.*, $\lim_{n\to\infty} |X_n - X| = 0$, nor $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$.

5. **Convergence in distribution** is the weakest form of convergence. **The convergences in probability and mean** are the **stronger** conditions, so they will imply $X_n \xrightarrow{D} X$.

How to apply Lindeberg–Lévy CLT

- *CLT: sum of the iid RVs. will converge to a normal distribution*.
- *CDF of a normal distribution can be easily calculated*.

How to apply Lindeberg–Lévy CLT

- *CLT: Sum of the iid RVs. will converge to a normal distribution.*
- *CDF of a normal distribution can be easily calculated.*

Definition 4: How to apply the CTL [1]

1. Identify the random variable of interest Y, as the sum of n iid random variable X_i's: $Y = X_1 + X_2 + \ldots + X_n$.
2. Calculate the mean $\mathrm{E}[Y]$ and $\mathrm{Var}[Y]$, by noting that $\mathrm{E}[Y] = n\mathrm{E}[X_i]$ and $\mathrm{Var}[Y] = n\mathrm{Var}[X_i]$.
3. $Z_n = \frac{Y - \mathrm{E}[Y]}{\sqrt{\mathrm{Var}[Y]}}$ is approximately standard normal; thus, we calculate

$$\begin{aligned} P(Y \leq y_{margin}) &= P(Z_n \leq z_{n\,margin}) \\ &\approx \Phi(z_{n\,margin}) \end{aligned}$$

where $z_{n\,margin} = \frac{y_{margin} - \mathrm{E}[Y]}{\sqrt{\mathrm{Var}[Y]}}$

Example of using Lindeberg–Lévy CLT

Example 5: Calculation example

A data packet consists of 1000 independent bits. Each bit has some error with a probability of 0.2. What is the probability that there are more than 22.5% errors in a data packet?

Goal $\Rightarrow$: Find $P(Y \geq y_n)$ where y_n denotes the 22.5% errors.

Write the random variable of interest. Number of bits $:= Y = \sum_i X$.
$Y_{margin} = 1000 \cdot 22.5\% = 225$

Calculate mean and variance $\mathrm{E}[Y]$ and $\mathrm{Var}[Y]$.
$\mathrm{E}[Y] = n\mathrm{E}[X_i] = np = 1000 * 0.2 = 200$ and
$\mathrm{Var}[Y] = n\mathrm{Var}[X_i] = np(1-p) = 1000(0.2)(1-0.2) = 160$.

Find scaling Y to Z_n, *i.e.*, $z_n = \frac{y - \mathrm{E}[Y]}{\mathrm{Var}[Y]}$.
$z_n = \frac{y-200}{\sqrt{160}}$ and $z_{n\,margin} = \frac{225-200}{\sqrt{160}} = \frac{25}{\sqrt{160}}$.

Find $P(Y \geq Y_{margin}) = P(Z_n \geq z_{n\,margin}) = 1 - \Phi(z_{n\,margin})$
$P(Y \geq 225) = 1 - \Phi(\frac{25}{\sqrt{160}})$

Example of using Lindeberg–Lévy CLT

Example 5 (Cont'): Calculation example

A data packet consists of 1000 independent bits. Each bit has some error with a probability of 0.2. What is the probability that there are more than 22.5% errors in a data packet?

Goal $\Rightarrow$: Find $P(Y \geq y_n)$ where y_n denotes the 22.5% errors.

---

1. Write the random variable of interest. Number of bits $:= Y = \sum_i X$.

   $Y_{margin} = 1000 \cdot 22.5\% = 225$

2. Calculate mean and variance $\mathrm{E}[Y]$ and $\mathrm{Var}[Y]$.

   $\mathrm{E}[Y] = n\mathrm{E}[X_i] = np = 1000 * 0.2 = 200$ and

   $\mathrm{Var}[Y] = n\mathrm{Var}[X_i] = np(1-p) = 1000(0.2)(1-0.2) = 160.$

3. Find scaling Y to Z_n, *i.e.*, $z_n = \frac{y - \mathrm{E}[Y]}{\mathrm{Var}[Y]}$.

   $z_n = \frac{y-200}{\sqrt{160}}$ and $z_{n\,margin} = \frac{225-200}{\sqrt{160}} = \frac{25}{\sqrt{160}}$.

4. Find $P(Y \geq Y_{margin}) = P(Z_n \geq z_{n\,margin}) = 1 - \Phi(z_{n\,margin})$

   $P(Y \geq 225) = 1 - \Phi(\frac{25}{\sqrt{160}})$.

Example of using Lindeberg–Lévy CLT

Example 5 (Cont'): Calculation example

A data packet consists of 1000 independent bits. Each bit has some error with a probability of 0.2. What is the probability that there are more than 22.5% errors in a data packet?
Goal $\Rightarrow$: Find $P(Y \geq y_n)$ where y_n denotes the 22.5% errors.

---

1. Write the random variable of interest. Number of bits := $Y = \sum_i X$.
   $Y_{margin} = 1000 \cdot 22.5\% = 225$.
2. Calculate mean and variance $\mathrm{E}[Y]$ and $\mathrm{Var}[Y]$.
   $\mathrm{E}[Y] = n\mathrm{E}[X_i] = np = 1000 * 0.2 = 200$ and
   $\mathrm{Var}[Y] = n\mathrm{Var}[X_i] = np(1-p) = 1000(0.2)(1-0.2) = 160$.
3. Find scaling Y to Z_n, *i.e.*, $z_n = \frac{y - \mathrm{E}[Y]}{\mathrm{Var}[Y]}$.
   $z_n = \frac{y-200}{\sqrt{160}}$ and $z_{n\,margin} = \frac{225-200}{\sqrt{160}} = \frac{25}{\sqrt{160}}$.
4. Find $P(Y \geq Y_{margin}) = P(Z_n \geq z_{n\,margin}) = 1 - \Phi(z_{n\,margin})$
   $P(Y \geq 225) = 1 - \Phi(\frac{25}{\sqrt{160}})$.

2. Convergence in probability

Convergence in probability— Definition

Definition 5: Convergence in probability

Random variables $X_1, X_2, X_3, ...$ converges in probability to a random variable X, if the probability of X_n diverges from X is getting close to 0 as $n \to \infty$

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0, \quad \epsilon > 0, \tag{6}$$

which is equivalent to $\lim_{n\to\infty} P(|X_n - X| \leq \epsilon) = 1$. [2]
This can also be written as $X_n \xrightarrow{p} X$.

Application I $\Rightarrow$ consistency of an estimator $\Rightarrow$ Lecture 1
Application II $\Rightarrow$ weak law of large numbers $\Rightarrow$ Next Slide 110

Example - Noise

Example 6: Example $X_n \xrightarrow{p} X$

Let X be a random variable and $X_n = X + Y_n$ where $\text{E}[Y_n] = \frac{1}{n}$ and $\text{Var}[Y_n] = \frac{\sigma^2}{n}$; $\sigma^2 > 0$ is a positive constant.
Question: show that X_n converges to X in probability. ....

- **Hint!**: Show that $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.
- This means we show $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = P(|Y_n| \geq \epsilon) = 0$.
- Have a quick look. ....
- Cherbyshev's inequality: $P(|Z - \mu_z| \geq \epsilon) \leq \frac{1}{\epsilon}\sigma_z^2$.

Example 6 (Cont'): Example $X_n \xrightarrow{p} X$

Let X be a random variable and $X_n = X + Y_n$ where $\mathrm{E}[Y_n] = \frac{1}{n}$ and $\mathrm{Var}[Y_n] = \frac{\sigma^2}{n}$; $\sigma^2 > 0$ is a positive constant.
Question: show that X_n converges to X in probability. ....

- **Hint!**: Show that $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.
- This means we show $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = P(|Y_n| \geq \epsilon) = 0$.
- Have a quick look ...
- Cherbyshev's inequality: $P(|Z - \mu_z| \geq \epsilon) \leq \frac{1}{\epsilon}\sigma_z^2$.
- Also, notice that we don't want to touch X_n which is the combination of Y_n and X, and X is an unknown R.V. ($\mathrm{E}[X]$ and $\mathrm{Var}[X]$ are unknown.)

But we have to write $P(|Y_n| \geq \epsilon)$ in a form of $P(|Y_n - \mathrm{E}[Y_n]| \geq \epsilon)$ to use Cherbyshev's inequality.
How?

Example - Noise

Example 6 (Cont'): Example $X_n \xrightarrow{p} X$

Let X be a random variable and $X_n = X + Y_n$ where $\mathrm{E}[Y_n] = \frac{1}{n}$ and $\mathrm{Var}[Y_n] = \frac{\sigma^2}{n}$; $\sigma^2 > 0$ is a positive constant.
Question: show that X_n converges to X in probability. ....

- **Hint!**: Show that $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.
- This means we show $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = P(|Y_n| \geq \epsilon) = 0$.

But we have to write $P(|Y_n| \geq \epsilon)$ in a form of $P(|Y_n - \mathrm{E}[Y_n]| \geq \epsilon)$ to use Cherbyshev's inequality $P(|Z - \mu_z| \geq \epsilon) \leq \frac{1}{\epsilon}\sigma_z^2$.
How ?

$$\begin{aligned} |Y_n| &= |Y_n + \mathrm{E}[Y_n] - \mathrm{E}[Y_n]| \\ &\leq |Y_n - \mathrm{E}[Y_n]| + |\mathrm{E}[Y_n]| \end{aligned}$$

$$|Y_n| - \epsilon \leq |Y_n - \mathrm{E}[Y_n]| + |\mathrm{E}[Y_n]| - \epsilon$$

Example - Noise

Example 6 (Cont'): Example $X_n \xrightarrow{p} X$

Let X be a random variable and $X_n = X + Y_n$ where $\mathrm{E}[Y_n] = \frac{1}{n}$ and $\mathrm{Var}[Y_n] = \frac{\sigma^2}{n}$; $\sigma^2 > 0$ is a positive constant.
Question: show that X_n converges to X in probability. ....

- **Hint!**: Show that $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.
- This means we show $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = P(|Y_n| \geq \epsilon) = 0$. From

$$|Y_n| - \epsilon \leq |Y_n - \mathrm{E}[Y_n]| + |\mathrm{E}[Y_n]| - \epsilon$$

Left hand-side has the tighter bound.

$$\begin{aligned} P(|Y_n| \geq \epsilon) &\leq P(|Y_n - \mathrm{E}[Y_n]| + |\mathrm{E}[Y_n]| \geq \epsilon) \\ &= P(|Y_n - \mathrm{E}[Y_n]| \geq \epsilon - |\mathrm{E}[Y_n]|) \end{aligned}$$

- Apply Cherbyshev's inequality: $P(|Y_n - \mathrm{E}[Y_n]| \geq \epsilon') \leq \frac{1}{\epsilon'}\mathrm{Var}[Y_n]$.
- Simulation Github Tutorial 2

Application: weak law of large numbers

Definition 6: Weak law of large numbers (WLLN)

Let $X_1, X_, ..., X_n$ be i.i.d. random variables with a finite expected value and variance $\mathrm{E}[X_i] = \mu < \infty$ and $\mathrm{Var}[X_i] = \sigma < \infty$. Then, for any $\epsilon > 0$, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges to μ in probability, *i.e.*,

$$\lim_{n\to\infty} P\left(|\bar{X} - \mu| \geq \epsilon\right) = 0 \qquad (7)$$

Note:
The weak law of large numbers states that the average of a large number of i.i.d. random variables converges to the expected value.

Meanwhile, **the central limit theorem (Slide 97)** states that the sum of a large number of random variables will asymptotically normally distributed.

Implication to convergence in distribution

Convergence in probability **is stronger**, *i.e.*, it implies the convergence in distribution: $X_n \xrightarrow{p} X \xRightarrow{\textit{Imply}} X_n \xrightarrow{d} X$.

But the reverse **is not true**: $X_n \xrightarrow{d} X \overset{\text{Not Imply}}{\not\Longrightarrow} X_n \xrightarrow{p} X$.

Example - Implication

Example 7: Example $X_n \xrightarrow{d} X \overset{\text{Not Imply}}{\not\Longrightarrow} X_n \xrightarrow{p} X$

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables Bernoulli(p).
Question. Show that these $X_1, X_2, X_3, \ldots$ converges to $X \sim$ Bernoulli(p) in distribution, but not in probability.

---

1. Show that $X_n \xrightarrow{d} X \ldots$
2. Verify if $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$? ....

Example

Example 7 (Cont'): Example $X_n \xrightarrow{d} X \xRightarrow[/]{\text{Not Imply}} X_n \xrightarrow{p} X$

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables Bernoulli(p).

Question. Show that these $X_1, X_2, X_3, \ldots$ converges to $X \sim$ Bernoulli(p) in distribution, but not in probability.

---

1. Show that $X_n \xrightarrow{d} X \ldots$

Because $X_1, X_2, X_3, \ldots \sim$ Bernoulli(p) and $X \sim$ Bernoulli(p), CDF of the same pdf distribution is the same: $F_{X_n}(x) = F_X(x)$ for all n.

Therefore, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, or $X_n \xrightarrow{d} X$.

Example

Example 7 (Cont'): Example $X_n \xrightarrow{d} X \overset{\text{Not Imply}}{\not\Longrightarrow} X_n \xrightarrow{p} X$

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables Bernoulli(p).
Question. Show that these $X_1, X_2, X_3, \ldots$ converges to $X \sim$ Bernoulli(p) in distribution, but not in probability.

---

1. Show that $X_n \xrightarrow{d} X \ldots$

Because $X_1, X_2, X_3, \ldots \sim$ Bernoulli(p) and $X \sim$ Bernoulli(p), CDF of the same pdf distribution is the same: $F_{X_n}(x) = F_X(x)$ for all n.
Therefore, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, or $X_n \xrightarrow{d} X$.

Note that this is the assumption of the sampling from distribution.
So, $X_n \xrightarrow{d} X$ is true regardless of the chosen pdf.

Example 7 (Cont'): Example $X_n \xrightarrow{d} X \overset{\text{Not Imply}}{\not\Longrightarrow} X_n \xrightarrow{p} X$

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables Bernoulli(p).

Question. Show that these $X_1, X_2, X_3, \ldots$ converges to $X \sim$ Bernoulli(p) in distribution, but not in probability.

---

2. Verify if $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$?

Since $|X_n - X| \sim$ Bernoulli(p), $P(|X_n - X| > 0) = p$.

Example

Example 7 (Cont'): Example $X_n \xrightarrow{d} X \overset{\text{Not Imply}}{\not\Longrightarrow} X_n \xrightarrow{p} X$

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables Bernoulli(p).
Question. Show that these $X_1, X_2, X_3, \ldots$ converges to $X \sim$ Bernoulli(p) in distribution, but not in probability.

---

2. Verify if $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$?

Since $|X_n - X| \sim$ Bernoulli(p), $P(|X_n - X| > 0) = p$.
Likewise, because $\epsilon > 0$, $P(|X_n - X| \geq \epsilon) = p > 0$.
Thus, $X_n \overset{p}{\not\to} X$.

Exception

Convergence in distribution does not imply the convergence in probability, *i.e.*, $X_n \xrightarrow{d} X \overset{\text{Not Imply}}{\not\Longrightarrow} X_n \xrightarrow{p} X$.

Except the following case...

Definition 7: Converge to a constant in distribution ⇒ probability

If random variables $X_1, X_2, X_3, \ldots$ converge in distribution to a constant c, then it also implies the convergence in probability to c.

$$\text{If } \; X_n \xrightarrow{d} c, \;\; \text{then } X_n \xrightarrow{p} c. \tag{8}$$

Example - Exception

Example 8: Converge to **a constant** in distribution/probability

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables Bernoulli$(1/n)$.
Question. Verify if $X_1, X_2, X_3, \ldots$ converges to a constant 0 in distribution ? and in probability?

---

1. Verify if $X_n \xrightarrow{d} 0$, *i.e.*, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x = 0)$?
2. Verify if $\lim_{n\to\infty} P(|X_n - 0| \geq \epsilon) = 0$? ....

Homework 1.

Definition 8: Convergence in mean

Let $r \geq 1$ be a fixed number. A sequence of random variables $X_1, X_2, X_3, \ldots$ converges in the r^{th} mean or in the L^r norm to a random variable X, if

$$\lim_{n \to \infty} \mathrm{E}[|X_n - X|^r] = 0. \tag{9}$$

This can also be shown as $X_n \xrightarrow{L^r} X$.

If $r = 2$, it is called the **mean-square convergence**, and it is shown by $X_n \xrightarrow{m.s.} X$.

Example 9: Example of convergence in mean

Let $\bar{X}$ denote the sample mean of the random sample $X_1, X_2, ..., X_n$ that came from the same distribution as X with a finite mean and variance, *i.e.*, $\mu, \sigma^2 \leq \infty$. Show that $\bar{X}$ converges to μ in mean with $r = 2$.

Because $\mathrm{Var}[\bar{X}] = \mathrm{E}[|\bar{X} - \mu|^2] = \frac{\sigma^2}{n}$. So, $\lim_{n \to \infty} \frac{\sigma^2}{n} = 0$.

Implication to convergence in probability

Example 10: Convergence in mean $\xRightarrow{Imply}$ convergence in probability

If $X_n \xrightarrow{L^r} X$ for some $r \geq 1$, then $X_n \xrightarrow{p} X$.

Homework 2. Show that $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$. Hint! Since $|X_n - X|$ is an RV $\in \mathbb{R}^+$, this can be done easily using Markov Inequality.

Example 11: Implication of convergence

Consider a sequence $\{X_n, n = 1, 2, 3, ...\}$ such that

$$F_{X_n}(n) = \begin{cases} \frac{1}{n} & n = 1, 2, 3... \\ 1 - \frac{1}{n} & n = 0 \end{cases}$$

Show that

- Q1. $X_n \xrightarrow{p} X$
- Q2. X_n does not converge to 0 in the r^{th} mean for any $r \geq 1$.

Q1. Show that $X_n \xrightarrow{p} 0$, *i.e.*, $\lim_{n\to\infty} P(|X_n| \geq \epsilon) = 0$.

Example 12: Implication of convergence

Consider a sequence $\{X_n, n = 1, 2, 3, ...\}$ such that

$$F_{X_n}(x) = \begin{cases} \frac{1}{n} & x = n^2 ... \\ 1 - \frac{1}{n} & x = 0 \end{cases}$$

Show that

- Q1. $X_n \xrightarrow{p} 0$
- Q2. X_n does not converge to 0 in the r^{th} mean for any $r \geq 1$.

Q1. Show that $X_n \xrightarrow{p} 0$, *i.e.*, $\lim_{n\to\infty} P(|X_n| \geq \epsilon) = 0$.
Because $\epsilon > 0, \lim_{n\to\infty} P(|X_n| \geq \epsilon) = \lim_{n\to\infty} \frac{1}{n} = 0$. So, $X_n \xrightarrow{p} 0$.

Example 13: Implication of convergence

Consider a sequence $\{X_n, n = 1, 2, 3, ...\}$ such that

$$F_{X_n}(x) = \begin{cases} \frac{1}{n} & x = n^2 ... \\ 1 - \frac{1}{n} & x = 0 \end{cases}$$

Show that

- Q1. $X_n \xrightarrow{p} 0$
- Q2. X_n does not converge to 0 in the r^{th} mean for any $r \geq 1$.

Q2. Show that X_n does not converge to 0 in the r^{th} mean for any $r \geq 1$.

Example 14: Implication of convergence

Consider a sequence $\{X_n, n = 1, 2, 3, \ldots\}$ such that

$$F_{X_n}(x) = \begin{cases} \frac{1}{n} & x = n^2 \text{ for } n > 1 \\ 1 - \frac{1}{n} & x = 0 \end{cases}$$

Show that

- Q1. $X_n \xrightarrow{p} 0$
- Q2. X_n does not converge to 0 in the r^{th} mean for any $r \geq 1$.

Q2. Show that X_n does not converge in the r^{th} mean for any $r \geq 1$.
Firstly, we plugin the value:
$\lim_{n\to\infty} \mathrm{E}[|X_n|^r] = \lim_{n\to\infty} \sum_n |n^2|^r \frac{1}{n} = \lim_{n\to\infty} \sum_n n^{2r-1}$.

Because $r \geq 1$, then $2r - 1 \geq 1$. Thus, $\lim_{n\to\infty} \sum_n n^{2r-1} = \infty$. X_n does not converge in the r^{th} mean for any $r \geq 1$.

Almost sure convergence

Definition 9: Almost sure convergence

The sequence X_n **converges almost surely** (converges with probability 1) towards X, if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{10}$$

Nevertheless, the above notion can also be applied to the sample space of $X_n(\cdot)$ which is a function of random variable, as follows:

We say that $X_n(\cdot)$ converges almost surely to $X(\cdot)$, if there exists a sample space $\mathbb{S} := \{s_1, s_2, ..., s_k\}$ where the random variable $X_n(\cdot)$ mapping $\mathbb{S}$ to a real space $\mathbb{R}$, or

$$P\left(\{s \in \mathbb{S} : \lim_{n\to\infty} X_n(s) = X(s)\}\right) = 1. \tag{11}$$

Example 15: Convergence of function of random variables

Consider a sequence of $X_1, X_2, \ldots$ where $X_n(s) := \frac{n}{n+1}s + (1-s)^n$ mapping $s \in S$ to a real space, and the sample space $S = [0,1]$ has a uniform distribution. Let the random variable $X(s) := s$.

Question. Check whether $X_n(s) \xrightarrow{a.s.} X(s)$?

Goal: $P\left(\{s \in \mathbb{S} : \lim_{n\to\infty} X_n(s) = X(s)\}\right) = 1$?
Or is there a sample space $\mathbb{S} \subset S$ such that $\lim_{n\to\infty} X_n(s) = X(s)$.

Example 15 (Cont'): Convergence of function of random variables

Consider a sequence of $X_1, X_2, \ldots$ where $X_n(s) := \frac{n}{n+1}s + (1-s)^n$ mapping $s \in S$ to a real space, and the sample space $S = [0,1]$ has a uniform distribution. Let the random variable $X(s) := s$.

Question. Check whether $X_n(s) \xrightarrow{a.s.} X(s)$?

---

1. What is $\lim_{n\to\infty} X_n(s)$?
2. Is $P(\{s \in S : \lim_{n\to\infty} X_n(s) = X(s)\}) = 1$?

Example 15 (Cont'): Convergence of function of random variables

Consider a sequence of $X_1, X_2, \ldots$ where $X_n(s) := \frac{n}{n+1}s + (1-s)^n$ mapping $s \in S$ to a real space, and the sample space $S = [0, 1]$ has a uniform distribution. Let the random variable $X(s) := s$.

Question. Check whether $X_n(s) \xrightarrow{a.s.} X(s)$?

---

1. What is $\lim_{n\to\infty} X_n(s)$?

Example 15 (Cont'): Convergence of function of random variables

Consider a sequence of $X_1, X_2, \ldots$ where $X_n(s) := \frac{n}{n+1}s + (1-s)^n$ mapping $s \in S$ to a real space, and the sample space $S = [0, 1]$ has a uniform distribution. Let the random variable $X(s) := s$.

Question. Check whether $X_n(s) \xrightarrow{a.s.} X(s)$?

---

1. What is $\lim_{n\to\infty} X_n(s)$?
$\lim_{n\to\infty} X_n(s) = \lim_{n\to\infty} \frac{n}{n+1}s + (1-s)^n$.
Because $s \in [0, 1]$, $\lim_{n\to\infty}(1-s)^n$ could take two possible values , *i.e.*, $\lim_{n\to\infty}(1-s)^n = 0$ for all $s \in (0, 1]$, and $\lim_{n\to\infty}(1-s)^n = 1$ for $s = 0$.
Thus,

$$\lim_{n\to\infty} X_n(s) = \frac{n}{n+1}s + (1-s)^n = \begin{cases} s+1 & s = 0; \\ s & s \in (0, 1]. \end{cases}$$

Example 15 (Cont'): Convergence of function of random variables

Consider a sequence of $X_1, X_2, \ldots$ where $X_n(s) := \frac{n}{n+1}s + (1-s)^n$ mapping $s \in S$ to a real space, and the sample space $S = [0, 1]$ has a uniform distribution. Let the random variable $X(s) := s$.

Question. Check whether $X_n(s) \xrightarrow{a.s.} X(s)$?

---

2. Is $P(\{s \in S : \lim_{n\to\infty} X_n(s) = X(s)\}) = 1$?

Example 15 (Cont'): Convergence of function of random variables

Consider a sequence of $X_1, X_2, \ldots$ where $X_n(s) := \frac{n}{n+1}s + (1-s)^n$ mapping $s \in S$ to a real space, and the sample space $S = [0,1]$ has a uniform distribution. Let the random variable $X(s) := s$.

Question. Check whether $X_n(s) \xrightarrow{a.s.} X(s)$?

---

2. Is $P(\{s \in S : \lim_{n\to\infty} X_n(s) = X(s)\}) = 1$?
Previously,

$$\lim_{n\to\infty} X_n(s) = \frac{n}{n+1}s + (1-s)^n = \begin{cases} s+1 & s = 0; \\ s & s \in (0,1]. \end{cases}$$

$\lim_{n\to\infty} X_n(s)$ will converge to $X(s)$ in the sample space $(0,1]$.
Because the uniform distribution is a continuous function.
Let $\mathbb{S} = (0,1]$. $P(s \in \mathbb{S}) = P(S/\{0\}) = 1$ (only $\{0\}$ is excluded).
Therefore, $X_n(s) \xrightarrow{a.s.} X(s)$.

Definition 10: Convergences almost completely

For all $\epsilon > 0$, we say that X_n **converges almost completely** towards X, *i.e.*,

$$\sum_n P(|X_n - X| > \epsilon) < \infty; \tag{12}$$

then, $X_n \xrightarrow{a.s.} X$.

In other words, if X_n converges in probability to X sufficiently quickly (*i.e.*, the above sequence of tail probabilities is summable) for all $\epsilon > 0$, then X_n converges almost surely to X.

This is the implication from Borel–Cantelli lemma [3].

Definition 11: Implication to Convergence in Probability

If $P\left(\lim_{n\to\infty} X_n - X = 0\right) = 1$, then

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0, \quad \epsilon > 0. \tag{13}$$

In other words, if X_n converges almost surely to X, then X_n also converges in probability to the same random variable.

Application: strong law of large numbers (SLLN)

Definition 12: Strong law of large numbers (SLLN)

Let $X_1, X_2, ...X_n$ be i.i.d. random variables with a finite expected value $\mathrm{E}[X_i] = \mu \leq \infty$. Then, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1 \tag{14}$$

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

Strong vs. Weak law of large number

Definition 13: Meaning of limit operations ($\lim_{n\to\infty}$) in SLLN

Strong law of large number (SLLN)
The meaning of $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$

$$\text{At first,} \quad \lim_{n\to\infty} \bar{X}_n = \mu$$
$$\Downarrow$$
$$\text{Then,} \ P(\cdot)$$

The limit of is $\bar{X}_n$ as $n \to \infty$ **always** equal to μ where **always** signifies $P(\cdot) = 1$

On the other words...

- Generally, for any $n < \infty$, $\bar{X}_n \neq \mu$. But as $n \to \infty$ $\bar{X}_n = \mu$.
- Then we can also say, for any $\epsilon > 0$, we will always have $|\lim_{n\to\infty} \bar{X}_n - \mu| \leq \epsilon$.

Strong vs. Weak law of large number

Definition 14: These **limit operations** ($\lim_{n\to\infty}$) have different meaning

Strong law of large number (see previous slide)	Weak law of large number (see Slide 110)
$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$	$\lim_{n\to\infty} P(\|\bar{X}_n - \mu\| < \epsilon) = 1$ s

- If $\lim_{n\to\infty} \bar{X}_n = \mu$, then we will always have $|\lim_{n\to\infty} \bar{X}_n - \mu| \leq \epsilon$.
- SLLN consider the probability of $|\lim_{n\to\infty} \bar{X}_n - \mu| \leq \epsilon$. That is the probability that $|\lim_{n\to\infty} \bar{X}_n - \mu|$ is always bounded.
- On the other hand, WLLN put the limit on the probability of $|\bar{X}_n - \mu|$ being bounded. That mean, for some n, $P(|\bar{X}_n - \mu| < \epsilon) \neq 1$. Or, there is a chance that $|\bar{X}_n - \mu| \geq \epsilon$.

Example 16: Converge to **a constant** in distribution/probability

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables Bernoulli$(1/n)$.
Question. Verify if $X_1, X_2, X_3, \ldots$ converges to 0 in distribution ? and in probability?

---

1. Verify if $X_n \xrightarrow{d} 0$, *i.e.*, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x=0)$?
2. Verify if $\lim_{n\to\infty} P(|X_n - 0| \geq \epsilon) = 0$? ....

These are the questions from Slide 118

---

3. Verify if $P(\lim_{n\to\infty} X_n = 0) = 1$? .... $\Rightarrow$ **Homework 3.** Hint it is easier to use **Converges almost completely**....

$\Rightarrow$ Simulation Github Tutorial 2

Summary

Cond.	Types	Applications
Weak	$X_n \xrightarrow{d} X$ $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$	Central Limited Theorem
Inbetween	$X_n \xrightarrow{p} X$ $\lim_{n\to\infty} P(\|X_n - X\| \geq \epsilon) = 0$	Weak law of large number (WLLN) Consistency of estimator
Strong	$X_n \xrightarrow{r^{th}} X$ $\lim_{n\to\infty} \mathrm{E}[\|X_n - X\|^r] = 0$	MSE of estimator (Efficiency)
Strong	$X_n \xrightarrow{a.s} X$ $P\left(\lim_{n\to\infty} X_n = X\right) = 1$	Strong law of large number (SLLN)

6 Homework

- Homework 2.1 (scores=5x3=15)
- Homework 2.2 (scores=5)
- Bonus: Proof of Lindeberg–Lévy CLT (10 scores)

Homework 2.1 (scores=5x3=15)

HW 1: Converge to a constant in distribution/probability

If random variables $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. Bernoulli($1/n$).

Questions:

1. Verify if $X_n \xrightarrow{d} 0$, *i.e.*, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x=0)$? (5 scores)
2. Verify if $X_n \xrightarrow{p} 0$, *i.e.*, $\lim_{n\to\infty} P(|X_n - 0| \geq \epsilon) = 0$? (5 scores)
3. Verify if $X_n \xrightarrow{a.s.} 0$, *i.e.*, $P(\lim_{n\to\infty} X_n = 0) = 1$? To get a full score, you have to provide: 1.theoretical proof & 2. simulation....(5 scores)

```python
# Provide histrogram density of $|X_n| > \epsilon $
N_list = [1,10,1000] # < You have to change this ... Just look at the formular .. what is the range of n..
P_count_n = []
P_count_temp = 0
N_size=1000

for N_select in N_list:
    XN = np.random.binomial(size=N_size, p=1/N_select, n=1)
    counts = 1*(XN > 0)

    P_count = counts.sum()/N_size
    P_count_temp += P_count
    P_count_n.append(P_count_temp)


P_count_n
plt.figure(figsize=(12,6))
plt.plot(N_list, P_count_n, color="blue", linewidth=2, label="Uniform")
plt.ylabel("$P_n$")
plt.xlabel("Iterations")
plt.grid()
plt.legend()
plt.show()
```

Homework 2.2 (scores=5)

HW 2: Convergence in mean $\xRightarrow{\mathit{Imply}}$ convergence in probability

If $X_n \xrightarrow{L^r} X$ for some $r \geq 1$, then $X_n \xrightarrow{p} X$.

Show that $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$. Hint! Since $|X_n - X|$ is an RV $\in \mathbb{R}^+$, this can be done easily using Markov Inequality.

Bonus: Proof of Lindeberg–Lévy CLT (10 scores)

Bonus 1: Lindeberg–Lévy Central Limit Theorem (CLT) [2]

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with mean $\mathrm{E}[X_n] = \mu \leq \infty$ and variance $\mathrm{Var}[X_n] = \sigma^2 \leq \infty$. Then, the random variable Z_n, defined as,

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \tag{15}$$

converges **in distribution** to the standard normal random variable X, *i.e.*, $Z_n \xrightarrow{d} X$ where $X \sim \mathcal{N}(0, 1^2)$. That is, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.

Note that $X_1, X_2, \ldots$ can have any distribution.
Unlike the estimator settings, $X_1, X_2, \ldots$ is unrelated to X.

Show that this theorem is true (using Continuity Theorem). Hint in Slide 150

Soln. Bonus 1: Sketch of how you may do this & given info ...

- If $X \sim \mathcal{N}(0, 1^2)$, $M_X(t) = \mathrm{E}[e^{tX}] = e^{\frac{t^2}{2}}$
- Use the fact that if $\lim_{n\to\infty} M_{Z_n}(t) = M_X(t) = e^{\frac{t^2}{2}}$, then $Z_n \xrightarrow{d} X$.
- Taylor series expansion centering around a: $f(q) = \sum_i \frac{f^i(q=a)(q-a)^i}{i!}$
- $\lim_{n\to\infty} \left(1 + \frac{t^2}{2n}\right)^n = e^{\frac{t^2}{2}}$ (L'Hôpital's rule)

Goal: show that $\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}$.

Sketch the steps ...

1. $\Rightarrow$ Derive $M_{Z_n}(t) = \mathrm{E}[e^{tZ_n}]$ (4 scores)
2. $\Rightarrow$ Use Taylor expansion to calculate $\mathrm{E}[e^{tZ_n}]$ (4 scores)
3. $\Rightarrow$ Take the limit of $n \to \infty$ (2 scores)

- H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*.
  Kappa Research, LLC, 2014.
- A. C. Robert Hogg, Joseph McKean, *Introduction to Mathematical Statistics*.
  Pearson; 7th edition, 2012.
- D. Romik, "Lecture Notes on Probability Theory, math 235a," Fall 2009.
- L. Wasserman, "Lecture note 15 in intermediate statistics. stats 705," Fall 2020.
- Z. Fan, "Lecture 3: Consistency and asymptotic normality of the MLE. stats 200," Autumn 2016.
- D. Panchenko, "Lecture 3: Properties of MLE: consistency, asymptotic normality, Fisher information. 18.650," Fall 2006.