AI 5090: Stochastic Processes

Lecture 12

Scribe: Anubhav Kumar

14 February 2025

Contents

In this lecture, we discuss the Central Limit Theorem (CLT) in great detail, focusing on its formal statement, proof and implications. We discuss what CLT is about, and more importantly what it is not about. We also discuss the Local Central Limit Theorem, and the stronger conditions it imposes over the base CLT.

1 Recap

We have talked about the condition $\mathbb{E}[|X|] < +\infty$ in context of the Laws of Large Numbers, here we explore what the condition means in terms of random variables

1.1 Expectation

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$
Let X be any random variable with respect to $\mathscr{F}$

Definition 1 (Expectation)**.**

$$X_+ := \max\{X, 0\}, \quad X_- := \max\{-X, 0\}$$
$$X = X_+ - X_-.$$

Expectation of an arbitrary random variable X is defined as:

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-],$$

where X_+ and X_- are the non-negative random variables representative of the positive and the negative parts of X respectively. The expectation is not well defined when both $\mathbb{E}[X_+]$ and $\mathbb{E}[X_-]$ are $+\infty$.

Now consider the random variable $|X|$. As it is evident from the Figure 1 it is defined as:

$$|X| = X_+ + X_-.$$

Hence, by linearity of expectations, its expectation is defined as:

$$\mathbb{E}[|X|] = \mathbb{E}[X_+] + \mathbb{E}[X_-].$$

Because this is a non-negative random variable, its expectation is always well defined, and we do not run into the $\infty - \infty$ problem. Observe that $\mathbb{E}[|X|] < +\infty$ if and only if both $\mathbb{E}[X_+]$ and $\mathbb{E}[X_-]$ are finite valued and well defined.

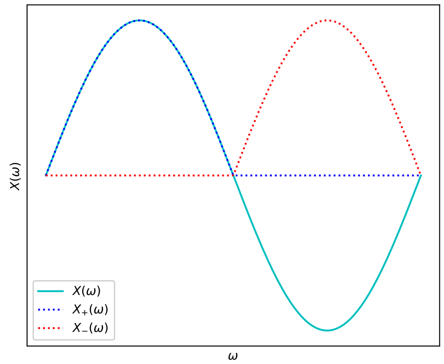

Figure 1: Representation of a Random Variable X with X_+ and X_-.

1.2 The Laws of Large Numbers

We provide here a quick recap of the laws of large numbers.
Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be defined w.r.t. $\mathscr{F}$.

Theorem 2 (Weak Law of Large Numbers)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with $\mathbb{E}[|X_1|] < +\infty$. Further, let $\mathbb{E}[X_1] = \mu$. Let*

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

Then

$$\frac{S_n}{n} \xrightarrow{\text{p.}} \mu.$$

More formally, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\big|\tfrac{S_n}{n} - \mu\big| \geq \varepsilon\Big) \;=\; 0.$$

Theorem 3 (Strong Law of Large Numbers)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with $\mathbb{E}[|X_1|] < +\infty$. Further, let $\mathbb{E}[X_1] = \mu$. Let*

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

Then

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu.$$

More formally,

$$\mathbb{P}\Big(\lim_{n\to\infty} \frac{S_n}{n} = \mu\Big) = 1.$$

What the laws of large numbers talk about is, when the mean is finite, the empirical average of a sequence of i.i.d random variables will converge to the true mean of the random variables in probability (in case of the weak law), and almost surely (with the strong law), as the number of random variables grows large.

2 The Central Limit Theorem

As we saw with the laws of large numbers, the sum of n numbers of random variables divided by n converged to the actual mean of the random variables as n grew. That is the sum of random variables converged to a constant random variable for large n. However, what happens when we divide the sum by $\sqrt{n}$ instead? This is what the Central Limit Theorem explores.

2.1 Statement of the CLT

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be defined w.r.t. $\mathscr{F}$.

Theorem 4 (Central Limit Theorem)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with mean $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ and* $\text{Var}(X_1) = \sigma^2 < +\infty$*. Let*

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

Then,

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{d.}} X, \quad X \sim \mathcal{N}(0,1).$$

More formally,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \quad \forall x \in \mathbb{R}.$$

That is, for a sequence of i.i.d. finite variance random variables, the random variable $S_n - \mathbb{E}[S_n]$, when divided by its standard deviation, which is of the order of $\sqrt{n}$, starts resembling a Standard Normal, in distribution. This happens regardless of the nature of X_n, they could be a discrete, continuous, singular or mixed. As long as they have finite variance, this random variable will resemble a standard normal in distribution.

2.2 Proof of CLT

Consider Z_i to be the normalized version of X_i, and consider the sum of these random variables divided by $\sqrt{n}$. That is, let

$$Z_i = \frac{X_i - \mathbb{E}[X_i]}{\sqrt{\text{Var}(X_1)}} \quad \text{and} \quad U_n = \frac{\sum_{i=1}^{n} Z_i}{\sqrt{n}}.$$

To prove convergence in distribution, we have to show pointwise convergence of the CDF of the random variables to that of a standard normal distribution. Equivalently, we can show the convergence of the characteristic function to that of a standard normal.

Because the random variables Z_i have a finite second absolute moment (because of normalization), their characteristic function admits a second order Taylor Series expansion. We can hence express the characteristic function of Z_i as

$$C_{Z_i}(s) = \sum_{k=0}^{2} \frac{\mathbb{E}[Z_i^k](sj)^k}{k!} + o(s^2), \quad s \in \mathbb{R}.$$

Because Z_i is mean centered, all its odd moments are 0, and because it is normalized, the second moment is 1. Therefore further simplifying:

$$\begin{aligned} C_{Z_i}(s) &= 1 + 0 + \frac{1(sj)^2}{2!} + o(s^2)C_{Z_i}(s) \\ &= 1 - \frac{s^2}{2} + o(s^2). \end{aligned}$$

Now, consider the characteristic function of U_n. Because this is a sum of n i.i.d. random variables $\frac{Z_i}{\sqrt{n}}$, we have

$$C_{U_n}(s) = \left(C_{Z_1}\left(\frac{s}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{s^2}{2n} + o\left(\frac{s^2}{n}\right)\right)^n.$$

As n grows large, the term $o\left(\frac{s^2}{n}\right)$ decays much faster than $\frac{s^2}{n}$. Hence, for large n,

$$C_{U_n}(s) \approx \left(1 - \frac{s^2}{2n}\right)^n$$

Taking the limit $n \to \infty$:

$$\lim_{n\to\infty} C_{U_n}(s) = \lim_{n\to\infty}\left(1 - \frac{s^2}{2n} + o\left(\frac{s^2}{n}\right)\right)^n = \lim_{n\to\infty}\left(1 - \frac{s^2}{2n}\right)^n = e^{-\frac{s^2}{2}}.$$

Because the characteristic function of U_n converges to that of a standard normal distribution, this shows

$$U_n \xrightarrow{\text{d.}} X, \qquad X \sim \mathcal{N}(0,1)$$

Note that, this does not imply the convergence of the probability density functions. All this implies is that for a large n, the CDF of the normalized sum divided by $\sqrt{n}$ approaches the CDF of a standard normal distribution.

2.3 Some Notes about the CLT

The **Strong Law of Large Numbers** tells us that

$$\frac{S_n - \mathbb{E}[S_n]}{n} \xrightarrow{\text{a.s.}} 0,$$

meaning that when scaled by n, the deviation $S_n - \mathbb{E}[S_n]$ collapses to zero almost surely, resulting in a degenerate (deterministic) limit. In contrast, the **Central Limit Theorem (CLT)** provides a non-degenerate scaling:

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}} \xrightarrow{\text{d.}} \mathcal{N}(0,\sigma^2),$$

where $\sigma^2 = \mathrm{Var}(X_1)$ for i.i.d. summands X_i. The $\sqrt{n}$-scaling retains the random fluctuations around the mean, yielding a meaningful Gaussian limit in distribution.

Interpretation of Scaling and Convergence:

- **Degenerate Scaling ($1/n$):** The scaling collapses the variability, leading to convergence to a constant.
- **CLT Scaling ($1/\sqrt{n}$):** This scaling preserves the scale of fluctuations, resulting in convergence in distribution to a non-trivial Gaussian random variable.

Note that while the individual realizations of the standardized sum are not Gaussian, their cumulative distribution functions (CDFs) converge to that of a normal variable.

Tail Probability Approximation:

For large n, the tail probability of the standardized sum can be approximated by that of a standard normal variable:

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sigma\sqrt{n}} > t\right) \approx \mathbb{P}(Z > t), \quad Z \sim \mathcal{N}(0,1).$$

A common upper bound for the tail of the standard normal is

$$\mathbb{P}(Z > t) \le e^{-t^2/2},$$

which implies

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sigma\sqrt{n}} > t\right) \lesssim e^{-t^2/2} \quad \text{for large } n.$$

CLT is not about PDF convergence:

The CLT guarantees that the CDFs of the standardized sums converge to the CDF of $\mathcal{N}(0,1)$. It does not imply that the probability density functions (PDFs) converge pointwise to the Gaussian PDF. Consider the cases:

- **Discrete Random Variables:** For discrete random variables, the standardized sum remains discrete and may not have a PDF in the traditional sense; however, its CDF still converges to that of $\mathcal{N}(0,1)$.
- **Continuous Random Variables:** Even when the X_i are continuous and the standardized sum has a PDF, the CLT only asserts convergence in distribution (i.e., of the CDFs) rather than convergence of the PDFs.

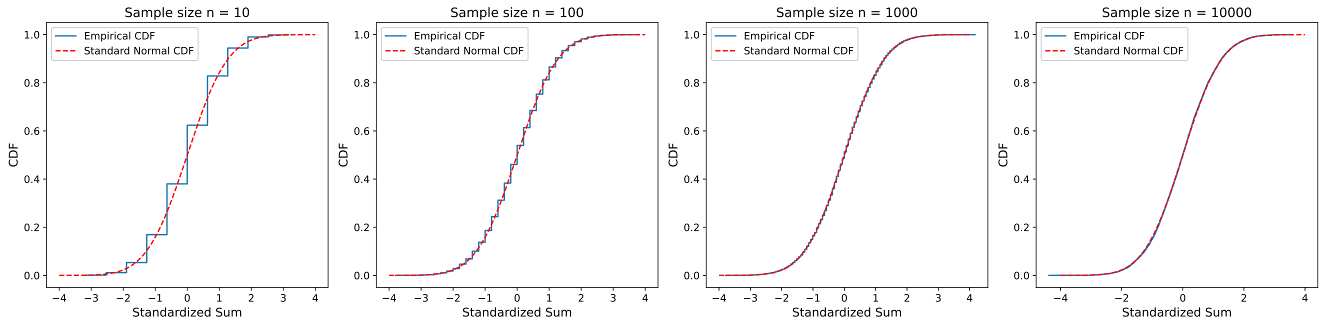

Figure 2: Demonstration of the Central Limit Theorem using Bern(0.5) samples. As the number of samples increases, the empirical CDF of the standardized sum approaches the standard normal CDF, despite the underlying distribution being discrete.

Example: Uniform$(-\sqrt{3}, \sqrt{3})$ Summands

Consider i.i.d. random variables $X_1, X_2, \ldots$ drawn from the uniform distribution on $[-\sqrt{3}, \sqrt{3}]$. Then:

$$\mathbb{E}[X_1] = 0, \quad \text{Var}(X_1) = 1, \quad \text{and} \quad C_{X_1}(s) = \frac{\sin(\sqrt{3}\, s)}{\sqrt{3}\, s}.$$

Let $S_n = X_1 + \cdots + X_n$ and define the standardized sum

$$U_n = \frac{S_n}{\sqrt{n}}.$$

Then the characteristic function of U_n is

$$C_{U_n}(s) = \left(\frac{\sin\left(\sqrt{3}\, \frac{s}{\sqrt{n}}\right)}{\sqrt{3}\, \frac{s}{\sqrt{n}}} \right)^n.$$

By the CLT, as $n \to \infty$,

$$C_{U_n}(s) \to e^{-s^2/2}, \quad \text{so} \quad U_n \xrightarrow{d} \mathcal{N}(0,1).$$

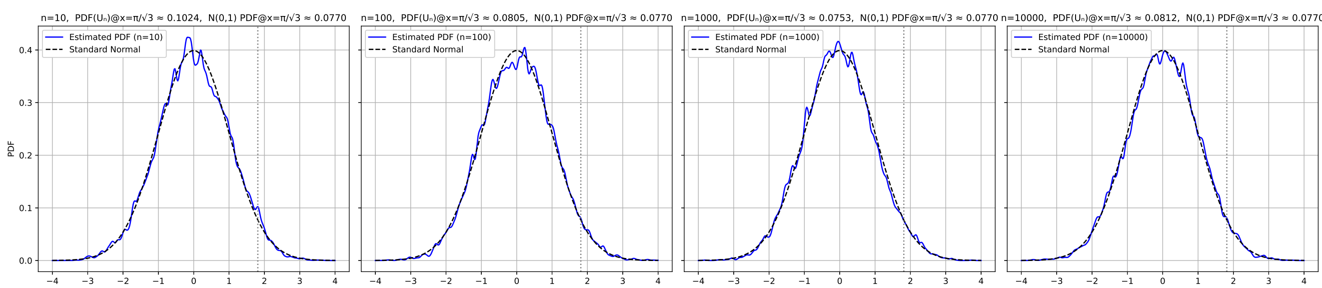

Figure 3: Demonstration of non convergence of PDFs of U_n despite increasing number of samples

However, as shown in Figure 3, the PDFs of the standardized sums U_n do not converge pointwise to the Gaussian PDF, even for large n. Consider $x = \frac{\pi}{\sqrt{3}}$, a point where the characteristic function of X_1 vanishes. At this location, the standard normal PDF is approximately 0.0770. Yet the estimated values of the PDF of U_n at this point fluctuate significantly: 0.1024 for $n = 10$, 0.0805 for $n = 100$, 0.0753 for $n = 1000$, and 0.0812 for $n = 10000$.
These persistent oscillations arise because the characteristic function of U_n has infinitely many zeros. Inverting it via Fourier transform leads to non-vanishing ripples in the PDF, especially near these zeros. Thus, even though $U_n \xrightarrow{\text{d.}} \mathcal{N}(0,1)$, the densities do not converge pointwise to the standard normal PDF. This highlights that convergence in distribution does not imply convergence of the PDFs.

Some more remarks about convergence:

The type of convergence matters greatly when it comes to expectations.

- **Almost Sure Convergence** ($X_n \xrightarrow{\text{a.s.}} X$) captures the idea that the random variables X_n start resembling X on sample paths that collectively belong to a set of probability 1. However, this does *not* imply convergence of expectations. Even if X_n looks like X almost everywhere, the areas under the curves—that is, the expected values—may differ significantly.
- **Mean Squared Convergence** ($X_n \xrightarrow{\text{m.s.}} X$) *does* imply that $\mathbb{E}[X_n] \to \mathbb{E}[X]$. This follows from a short and straightforward argument using standard inequalities like Cauchy–Schwarz or Jensen's inequality.
- **Convergence in Distribution** ($X_n \xrightarrow{\text{d.}} X$) means the CDFs of X_n converge pointwise to that of X. Although the CDF encodes all the information about a random variable—including moments—convergence of CDFs does *not* guarantee convergence of expectations.

To summarize:

1. $X_n \xrightarrow{\text{a.s.}} X \quad \not\Rightarrow \quad \mathbb{E}[X_n] \to \mathbb{E}[X]$.
2. $X_n \xrightarrow{\text{m.s.}} X \quad \Rightarrow \quad \mathbb{E}[X_n] \to \mathbb{E}[X]$.
3. $X_n \xrightarrow{d} X \quad \not\Rightarrow \quad \mathbb{E}[X_n] \to \mathbb{E}[X]$.

2.4 Local Central Limit Theorem

While the Central Limit Theorem (CLT) guarantees convergence in distribution, i.e., convergence of CDFs, it does **not** imply that the PDFs of standardized sums converge to the Gaussian PDF.
However, in many practical scenarios, we often observe such convergence of PDFs. This behavior is not a consequence of the CLT itself, but rather due to some *stronger conditions* holding implicitly.

Key Insight: If convergence of PDFs is observed, it must be due to some stronger property. One such sufficient condition is given by the Local CLT.

Theorem 5 (Local Central Limit Theorem)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. random variables with common density f_X, and without loss of generality, let $\mathbb{E}[X_1] = 0$ and* $\text{Var}(X_1) = 1$. *Define*

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

Suppose there exists an integer $r \in \mathbb{N}$ such that

$$\int_{-\infty}^{\infty} |C_{X_1}(s)|^r \, ds < +\infty,$$

where $C_{X_1}(s)$ denotes the characteristic function of X_1. Then, for all $x \in \mathbb{R}$,

$$f_{\frac{S_n}{\sqrt{n}}}(x) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Remarks:

- This is a much **stronger result** than the usual CLT. Convergence of PDFs implies convergence of CDFs, but not vice versa.
- The condition is technical: it asks for the r^{th} power of the characteristic function to be absolutely integrable. This is not automatically satisfied just because mean and variance are finite.
- The term "local" refers to the behavior of the characteristic function near the origin, linked to Taylor expansions and local regularity.
- This is only a **sufficient condition**. If the condition holds, PDF convergence is guaranteed. If it doesn't, we cannot conclude either way.