भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Stochastic Processes

Lecture 07

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

03 February 2026

Weak Law of Large Numbers (WLLN)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Theorem (Weak Law of Large Numbers)

Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of **independent and identically distributed (i.i.d.)** real-valued RVs with $\mathbb{E}[|X_1|] < +\infty$. Further, let $\mathbb{E}[X_1] = \mu \in \mathbb{R}$. For each $n \in \mathbb{N}$, let

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

denote the partial sum of random variables up to time n. Then,

$$\frac{S_n}{n} \xrightarrow{\text{p.}} \mu. \qquad \lim_{n \to \infty} \mathbb{P}\left(\left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\}\right) = 0 \qquad \forall\, \varepsilon > 0.$$

Proof of WLLN – Using Finite Variance Assumption

- Suppose that $\mathrm{Var}(X_1) = \sigma^2 < +\infty$
- Then, for any choice of $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\}\right) &\leq \frac{1}{\varepsilon^2}\,\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] \qquad \text{(Chebyshev's inequality)} \\ &= \frac{\sigma^2}{n\varepsilon^2} \end{aligned}$$

- It follows that

$$\forall\, \varepsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}\left(\left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\}\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0,$$

thus proving that $X_n \xrightarrow{\text{p.}} \mu$

Two Important Results

+

Proof of WLLN Without Finite Variance Assumption

Reverse Implication $\mathrm{d.} \implies \mathrm{p.}$

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Lemma (Reverse Implication $\mathrm{d.} \implies \mathrm{p.}$)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued RVs
For any $c \in \mathbb{R}$,

$$X_n \xrightarrow{\mathrm{d.}} c \quad \implies \quad X_n \xrightarrow{\mathrm{p.}} c$$

Proof of Reverse Implication $\mathrm{d.} \implies \mathrm{p.}$

- Notice that

$$X_n \xrightarrow{\mathrm{d.}} c \qquad \iff \qquad \lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0, & x < c, \\ 1, & x > c. \end{cases}$$

  Convergence of CDFs for $x = c$ may or may not hold (we will not make any assumption)

- **To prove:** For every choice of $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\{|X_n - c| > \varepsilon\}) \; = \; 0.$$

- Fixing an arbitrary $\varepsilon > 0$, we note that for each $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(\{|X_n - c| > \varepsilon\}) \quad &= \quad \mathbb{P}(\{X_n > c + \varepsilon\}) + \mathbb{P}(\{X_n < c - \varepsilon\}) \\ &= \quad 1 - \mathbb{P}(\{X_n \le c + \varepsilon\}) + \mathbb{P}(\{X_n < c - \varepsilon\}) \\ &\le \quad 1 - \mathbb{P}(\{X_n \le c + \varepsilon\}) + \mathbb{P}(\{X_n \le c - \varepsilon\}) \\ &= \quad 1 - F_{X_n}(c + \varepsilon) + F_{X_n}(c - \varepsilon). \end{aligned}$$

- Taking limits as $n \to \infty$, we get

$$\forall\, \varepsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}(\{|X_n - c| > \varepsilon\}) \; \le \; 1 \; - \; \lim_{n\to\infty} F_{X_n}(c+\varepsilon) \; + \; \lim_{n\to\infty} F_{X_n}(c-\varepsilon) \; = \; 0.$$

Characteristic Function

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be a random variable w.r.t. $\mathscr{F}$.

Definition (Characteristic Function)

The **characteristic function** of X is a function $C_X : \mathbb{R} \to \mathbb{C}$, defined as

$$C_X(s) = \mathbb{E}[e^{jsX}] = \mathbb{E}[\cos sX] + j\,\mathbb{E}[\sin sX], \qquad s \in \mathbb{R}.$$

- Observe that
$$\left|C_X(s)\right| \leq 1 \qquad \forall\, s \in \mathbb{R}.$$
- If X is extended real-valued, then its characteristic function is not defined

Taylor Expansion for Characteristic Functions

Lemma (Taylor Expansion for Characteristic Functions)

Suppose that X is a random variable such that $\mathbb{E}[|X|^k] < +\infty$ for some $k \in \mathbb{N}$. Then,

$$C_X(s) = \sum_{\ell=0}^{k} \frac{\mathbb{E}[X^\ell]}{\ell!} (js)^\ell + o(s^k), \qquad s \in \mathbb{R}.$$

For a proof, see [KT08].

- Given two sequences $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$, we say that $a_n = o(b_n)$ if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

- For instance,

$$\sqrt{n} = o(n), \qquad \log n = o(n), \qquad \frac{1}{n^2} = o\left(\frac{1}{n}\right), \qquad \frac{1}{n} = o(1).$$

Characteristic Functions and Convergence in Distribution

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Definition (Characteristic Functions and Convergence in Distribution)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued RVs, and let X an extended real-valued RV. Then,

$$X_n \xrightarrow{\text{d.}} X \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} C_{X_n}(s) = C_X(s) \quad \forall\, s \in \mathbb{R}.$$

That is, **pointwise convergence of characteristic functions** is equivalent to convergence in distribution.

Proof is based on Skorokhod's representation theorem [GS20, Section 7.2].

Proof of WLLN – Without Finite Variance Assumption

- We need to prove that if $\{X_n\}_{n\in\mathbb{N}}$ is an i.i.d. sequence with finite mean μ, then $\frac{S_n}{n} \xrightarrow{\text{p.}} \mu$
- For this, we will show that

$$\lim_{n\to\infty} C_{\frac{S_n}{n}}(s) = e^{js\mu} \qquad \forall\, s \in \mathbb{R}.$$

  This will then imply that $\frac{S_n}{n} \xrightarrow{\text{d.}} \mu$, which in turn implies that $\frac{S_n}{n} \xrightarrow{\text{p.}} \mu$
- Fixing an arbitrary $s \in \mathbb{R}\setminus\{0\}$, we have

$$\begin{aligned} C_{\frac{S_n}{n}}(s) &= \left(C_{X_1}\left(\frac{s}{n}\right)\right)^n \\ &= \left(1 + \frac{\mathbb{E}[X]}{1!}\cdot\frac{js}{n} + o\left(\frac{s}{n}\right)\right)^n \qquad \text{(using Taylor expansion formula).} \end{aligned}$$

- Taking limits as $n \to \infty$, we get

$$\forall\, s \in \mathbb{R}\setminus\{0\}, \qquad \lim_{n\to\infty} C_{\frac{S_n}{n}}(s) = \lim_{n\to\infty}\left(1 + \frac{\mu}{1!}\cdot\frac{js}{n} + o\left(\frac{s}{n}\right)\right)^n \overset{(a)}{=} \lim_{n\to\infty}\left(1 + \frac{\mu}{1!}\cdot\frac{js}{n}\right)^n = e^{js\mu}.$$

Proof of WLLN – Without Finite Variance Assumption

- Justification for (a):
  - For a fixed $s \in \mathbb{R} \setminus \{0\}$, let $\{g_n(s)\}_{n\in\mathbb{N}}$ be any sequence such that

$$g_n(s) = o\left(\frac{s}{n}\right), \qquad n \in \mathbb{N}.$$

  - Then, we have

$$\lim_{n\to\infty} \frac{g_n(s)}{s/n} = 0 \quad \iff \quad \lim_{n\to\infty} \frac{g_n(s)}{1/n} = 0 \quad \iff \quad g_n(s) = o(1/n).$$

Linear Regression in Large Language Models (LLMs)

Consider the token embeddings generated at some intermediate layer (say L) of an LLM. Each token embedding is a vector in $\mathbb{R}^d$ for some large d. Let $\{X_n\}_{n\in\mathbb{N}}$ denote the first coordinate readings of the vector embeddings across various contexts.

Model: $X_1, X_2, \cdots \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

Let Y denote a downstream performance metric (e.g., ROUGE/BLEU scores), and suppose that

$$Y_i = \beta X_i + \varepsilon_i,$$

where $\beta \in \mathbb{R}$ **is unknown**, and $\varepsilon_1, \varepsilon_2 \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, independent of $X_1, X_2, \ldots$, and σ^2 **is known**

- Considering n data points $\{(X_i, Y_i)\}_{i=1}^n$, write down the expression for the ℓ_2 loss (a.k.a. squared loss) between input and output assuming a guess θ for the unknown β.
- Give a closed-form expression for $\widehat{\beta}_n$, the value of θ that minimizes the ℓ_2 loss.
- Prove that

$$\widehat{\beta}_n \xrightarrow{\text{a.s.}} \beta.$$

Strong Law of Large Numbers (SLLN)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Theorem (Strong Law of Large Numbers)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of **independent and identically distributed (i.i.d.)** real-valued RVs with $\mathbb{E}[|X_1|] < +\infty$. Further, let $\mathbb{E}[X_1] = \mu \in \mathbb{R}$. For each $n \in \mathbb{N}$, let

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

denote the partial sum of random variables up to time n. Then,

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu. \qquad \mathbb{P}\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\}\right) = 0 \qquad \forall\, \varepsilon > 0.$$

Consequences of SLLN

- **Estimating the unknown bias of a coin:**

  Suppose that $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(p)$, where p is **unknown**
  According to SLLN, we have
  $$\frac{1}{n}\sum_{\ell=1}^{n} X_\ell \xrightarrow{\text{a.s.}} p.$$

- **Empirical risk minimization:**

  Suppose $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$, where μ is **unknown**
  According to SLLN, we have
  $$\frac{1}{n}\sum_{\ell=1}^{n} X_\ell \xrightarrow{\text{a.s.}} \mu.$$

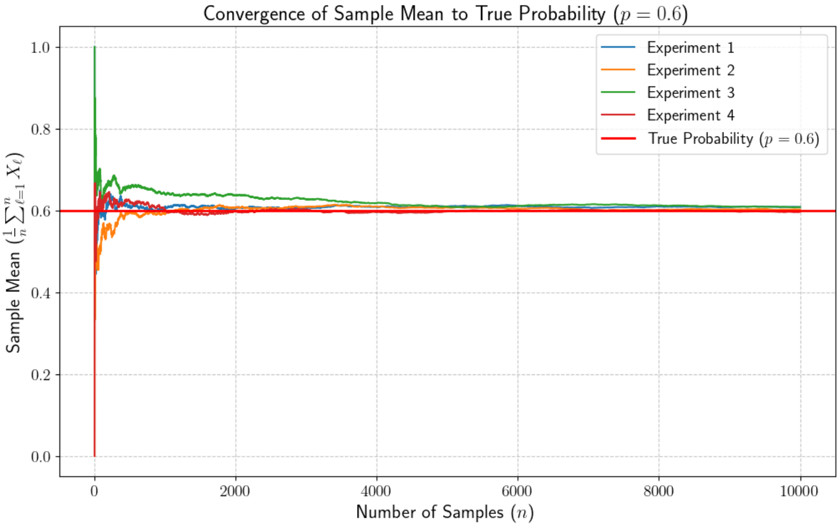
- **Linear regression in LLMs:**

  Suppose $Y_i = \beta X_i + \varepsilon_i$, where $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\varepsilon_1, \varepsilon_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ are independent of $\{X_n\}_{n\in\mathbb{N}}$. Here, $\beta \in \mathbb{R}$ is **unknown**, while $\sigma > 0$ is **known**. According to SLLN, we have
  $$\frac{\sum_{\ell=1}^{n} X_\ell Y_\ell}{\sum_{\ell=1}^{n} X_\ell^2} \xrightarrow{\text{a.s.}} \frac{\mathbb{E}[X_1 Y_1]}{\mathbb{E}[X_1^2]} = \beta.$$

SLLN Example: Estimating the Unknown Bias of a Coin

```
Experiment 1: First 10 samples: [0 0 1 1 1 1 1 1 1 0]
Experiment 1: Sample mean = 0.5934
Experiment 2: First 10 samples: [1 1 0 0 1 1 1 1 0 1]
Experiment 2: Sample mean = 0.6004
Experiment 3: First 10 samples: [1 1 0 0 0 0 0 0 0 0]
Experiment 3: Sample mean = 0.5951
Experiment 4: First 10 samples: [0 1 0 1 0 0 0 1 1 0]
Experiment 4: Sample mean = 0.5957
Experiment 5: First 10 samples: [0 0 1 0 0 0 1 1 1 0]
Experiment 5: Sample mean = 0.6009
Experiment 6: First 10 samples: [1 1 1 0 1 1 1 0 0 1]
Experiment 6: Sample mean = 0.6002
Experiment 7: First 10 samples: [1 0 1 1 1 0 1 0 1 1]
Experiment 7: Sample mean = 0.5993
Experiment 8: First 10 samples: [1 0 1 1 1 0 1 0 1 0]
Experiment 8: Sample mean = 0.6005
Experiment 9: First 10 samples: [1 0 1 1 1 1 0 1 1 0]
Experiment 9: Sample mean = 0.6037
Experiment 10: First 10 samples: [0 0 1 1 0 0 1 1 1 1]
Experiment 10: Sample mean = 0.6031
Experiment 11: First 10 samples: [0 0 1 0 0 0 0 0 0 0]
Experiment 11: Sample mean = 0.6015
Experiment 12: First 10 samples: [0 0 1 0 1 1 0 0 0 1]
Experiment 12: Sample mean = 0.6095
Experiment 13: First 10 samples: [0 0 0 0 0 1 1 0 1 0]
Experiment 13: Sample mean = 0.6013
Experiment 14: First 10 samples: [1 0 0 1 1 1 1 1 1 1]
Experiment 14: Sample mean = 0.5989
Experiment 15: First 10 samples: [1 0 0 0 0 0 0 1 0 0]
```

SLLN Example: Estimating the Unknown Bias of a Coin



References

Geoffrey Grimmett and David Stirzaker.
Probability and random processes.
Oxford university press, 2020.

John Frank Charles Kingman and Samuel James Taylor.
Introdction to Measure and Probability.
Cambridge University Press, 2008.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Stochastic Processes

Lecture 08

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

06 February 2026

Weak Law of Large Numbers (WLLN)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Theorem (Weak Law of Large Numbers)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of **independent and identically distributed (i.i.d.)** real-valued RVs with $\mathbb{E}[|X_1|] < +\infty$. Further, let $\mathbb{E}[X_1] = \mu \in \mathbb{R}$. For each $n \in \mathbb{N}$, let

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

denote the partial sum of random variables up to time n. Then,

$$\frac{S_n}{n} \xrightarrow{\text{p.}} \mu. \qquad \lim_{n\to\infty} \mathbb{P}\left(\left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\}\right) = 0 \qquad \forall\, \varepsilon > 0.$$

Reverse Implication $\mathrm{d.} \implies \mathrm{p.}$

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Lemma (Reverse Implication $\mathrm{d.} \implies \mathrm{p.}$)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued RVs
For any $c \in \mathbb{R}$,

$$X_n \xrightarrow{\mathrm{d.}} c \implies X_n \xrightarrow{\mathrm{p.}} c$$

Characteristic Function

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be a random variable w.r.t. $\mathscr{F}$.

Definition (Characteristic Function)

The **characteristic function** of X is a function $C_X : \mathbb{R} \to \mathbb{C}$, defined as

$$C_X(s) = \mathbb{E}[e^{jsX}] = \mathbb{E}[\cos sX] + j\,\mathbb{E}[\sin sX], \qquad s \in \mathbb{R}.$$

- Observe that
$$\left|C_X(s)\right| \leq 1 \qquad \forall\, s \in \mathbb{R}.$$
- If X is extended real-valued, then its characteristic function is not defined

Taylor Expansion for Characteristic Functions

Lemma (Taylor Expansion for Characteristic Functions)

Suppose that X is a random variable such that $\mathbb{E}[|X|^k] < +\infty$ for some $k \in \mathbb{N}$. Then,

$$C_X(s) = \sum_{\ell=0}^{k} \frac{\mathbb{E}[X^\ell]}{\ell!} (js)^\ell + o(s^k), \qquad s \in \mathbb{R}.$$

For a proof, see [KT08].

- Given two sequences $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$, we say that $a_n = o(b_n)$ if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

- For instance,

$$\sqrt{n} = o(n), \qquad \log n = o(n), \qquad \frac{1}{n^2} = o\left(\frac{1}{n}\right), \qquad \frac{1}{n} = o(1).$$

Characteristic Functions and Convergence in Distribution

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Definition (Characteristic Functions and Convergence in Distribution)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued RVs, and let X an extended real-valued RV. Then,

$$X_n \xrightarrow{\text{d.}} X \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} C_{X_n}(s) = C_X(s) \quad \forall\, s \in \mathbb{R}.$$

That is, pointwise convergence of characteristic functions is equivalent to convergence in distribution.

Proof is based on Skorokhod's representation theorem [GS20, Section 7.2].

Strong Law of Large Numbers (SLLN)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Theorem (Strong Law of Large Numbers)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of **independent and identically distributed (i.i.d.)** real-valued RVs with $\mathbb{E}[|X_1|] < +\infty$. Further, let $\mathbb{E}[X_1] = \mu \in \mathbb{R}$. For each $n \in \mathbb{N}$, let

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

denote the partial sum of random variables up to time n. Then,

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu. \qquad \mathbb{P}\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\}\right) = 0 \qquad \forall\, \varepsilon > 0.$$

Consequences of SLLN

- **Estimating the unknown bias of a coin:**

  Suppose that $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$, where p is **unknown**
  According to SLLN, we have

  $$\frac{1}{n}\sum_{\ell=1}^{n} X_\ell \xrightarrow{\text{a.s.}} p.$$

- **Empirical risk minimization:**

  Suppose $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$, where μ is **unknown**
  According to SLLN, we have

  $$\frac{1}{n}\sum_{\ell=1}^{n} X_\ell \xrightarrow{\text{a.s.}} \mu.$$

- **Linear regression in LLMs:**

  Suppose $Y_i = \beta X_i + \varepsilon_i$, where $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\varepsilon_1, \varepsilon_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ are independent of $\{X_n\}_{n\in\mathbb{N}}$. Here, $\beta \in \mathbb{R}$ is **unknown**, while $\sigma > 0$ is **known**. According to SLLN, we have

  $$\frac{\sum_{\ell=1}^{n} X_\ell Y_\ell}{\sum_{\ell=1}^{n} X_\ell^2} \xrightarrow{\text{a.s.}} \frac{\mathbb{E}[X_1 Y_1]}{\mathbb{E}[X_1^2]} = \beta.$$

Proof of SLLN Under Regularity Assumptions

- We will present a proof of SLLN under finite fourth moment assumption: $\mathbb{E}[X_1^4] < +\infty$
- Without loss of generality, let $\mathbb{E}[X_1] = 0$
- We want to show

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0.$$

- Fixing an arbitrary $\varepsilon > 0$, we note that

$$\begin{aligned}
\mathbb{P}\left(\left\{\left|\frac{S_n}{n}\right| > \varepsilon\right\}\right) &= \mathbb{P}\left(\{|S_n| > n\varepsilon\}\right) \\
&\leq \frac{\mathbb{E}[(S_n)^4]}{n^4\,\varepsilon^4} \\
&= \frac{\mathbb{E}[(X_1 + \cdots + X_n)^4]}{n^4\,\varepsilon^4} \\
&= \frac{n\,\mathbb{E}[X_1^4] + 6\binom{n}{2}\left(\mathrm{Var}(X_1)\right)^2}{n^4\,\varepsilon^4} \\
&\leq \frac{1}{\varepsilon^4}\left(\frac{\mathbb{E}[X_1^4]}{n^3} + 3\left(\mathrm{Var}(X_1)\right)^2\frac{1}{n^2}\right)
\end{aligned}$$

Proof of SLLN Under Regularity Assumptions

- Summing both sides over n, we get

$$\forall\, \varepsilon > 0, \qquad \sum_{n\in\mathbb{N}} \mathbb{P}\left(\left\{\left|\frac{S_n}{n}\right| > \varepsilon\right\}\right) \le \frac{\mathbb{E}[X_1^4]}{\varepsilon^4}\left(\sum_{n\in\mathbb{N}} \frac{1}{n^3}\right) + \frac{3\mathrm{Var}(X_1))^2}{\varepsilon^4}\left(\sum_{n\in\mathbb{N}} \frac{1}{n^2}\right) < +\infty.$$

- The result follows from an application of Borel–Cantelli lemma

Central Limit Theorem (CLT)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Theorem (Strong Law of Large Numbers)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of **independent and identically distributed (i.i.d.)** real-valued RVs. Let $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ and $\mathrm{Var}(X_1) = \sigma^2 < +\infty$. For each $n \in \mathbb{N}$, let

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

denote the partial sum of random variables up to time n. Then,

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{\text{d.}} \mathcal{N}(0,1).$$

More formally,

$$\forall x \in \mathbb{R}, \qquad \lim_{n\to\infty} \mathbb{P}\left(\left\{\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \leq x\right\}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t.$$

Proof of CLT

$$Z_i = \frac{X_i - \mathbb{E}[X_i]}{\sqrt{\mathrm{Var}(X_1)}}, \qquad U_n = \frac{\sum_{i=1}^{n} Z_i}{\sqrt{n}}$$

$$\begin{aligned}
C_{Z_i}(s) &= 1 - \frac{s^2}{2} + o(s^2), \qquad s \in \mathbb{R}, \\
C_{U_n}(s) &= \left(C_{Z_1}\left(\frac{s}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{s^2}{2n} + o\left(\frac{s^2}{n}\right)\right)^n
\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty} C_{U_n}(s) &= \lim_{n\to\infty} \left(1 - \frac{s^2}{2n} + o\left(\frac{s^2}{n}\right)\right)^n \\
&= \exp\left(\frac{-s^2}{2}\right) \qquad \forall\, s \in \mathbb{R}.
\end{aligned}$$

Some Notes about CLT

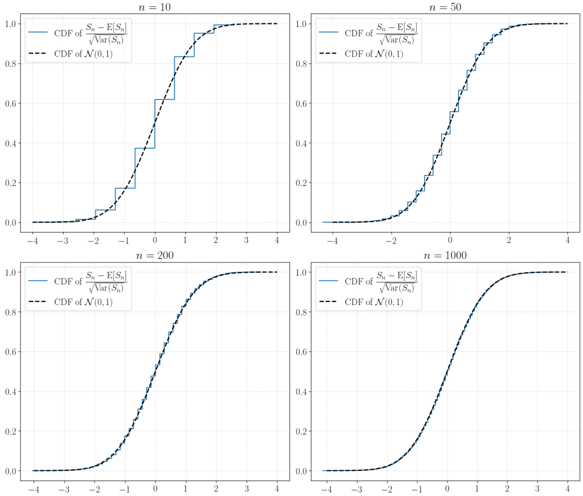
- From SLLN, we know that

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

- Thus, the random variable $S_n - \mathbb{E}[S_n]$, when **divided by n**, is degenerate for large n
- **CLT**: the random variable $S_n - \mathbb{E}[S_n]$, when **divided by $\sqrt{n}$**, is non-degenerate for large n
- According to CLT, for large n,

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} > t\right) = \mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sigma\sqrt{n}} > t\right) \approx \mathbb{P}(X > t), \qquad X \sim \mathcal{N}(0,1)$$

Demonstration of CLT



What CLT is Not

- CLT is not a statement about convergence of the PDFs of $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$ to Gaussian PDF
- If $X_1, X_2, \ldots$ are discrete random variables, then $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$ is a discrete random variable, and hence does not admit any PDF
- Even if $X_1, X_2, \ldots$ are continuous random variables, and $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$ admits a PDF, **CLT does not make any claim about the convergence of these PDFs to the Gaussian PDF**
- In many practical examples where $X_1, X_2, \ldots$ are continuous, one may observe convergence of PDFs of $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$ to the Gaussian PDF, but this is NOT to be interpreted as a consequence of the CLT
  This may be a consequence of some stronger property playing in hindsight

CLT $\neq$ Convergence of PDFs to Standard Normal PDF

- [GK68] Suppose that $X_1, X_2, \ldots$ are i.i.d. with common PDF

$$f(x) = \begin{cases} \dfrac{1}{2|x| \left(\log(1/|x|)\right)^2}, & 0 < |x| < e^{-1}, \\ 0, & |x| \geq e^{-1}. \end{cases}$$

- It can be shown that $\mathbb{E}[X_1] = 0$ and $\mathrm{Var}(X_1) = \sigma^2 < +\infty$
- Also, with a bit of cumbersome calculations, one can show that

$$\lim_{n\to\infty} f_{\frac{S_n}{\sigma\sqrt{n}}}(0) = +\infty \neq \frac{1}{\sqrt{2\pi}} \qquad \text{(value of standard normal PDF evaluated at 0).}$$

Local CLT – Convergence of PDFs to Standard Normal PDF

Theorem (Local Central Limit Theorem)

Suppose that $X_1, X_2, \cdots \overset{\text{i.i.d.}}{\sim} f$.
Without loss of generality, let $\mathbb{E}[X_1] = 0$ and $\mathrm{Var}(X_1) = 1$.
Suppose that there exists $r \in \mathbb{N}$ such that

$$\int_{-\infty}^{\infty} |C_{X_1}(s)|^r \, \mathrm{d}s \; < \; +\infty.$$

Then,

$$\lim_{n \to \infty} f_{\frac{S_n}{\sqrt{n}}}(x) \; = \; \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad \forall\, x \in \mathbb{R}.$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Fatou's Lemma,
Dominated Convergence Theorem,
Monotone Convergence Theorem

Interchange of Limit/Limsup/Liminf and Expectations

Monotone Convergence Theorem

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Theorem (Monotone Convergence Theorem)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued RVs, and let X, Y be two other extended real-valued RVs.

1. *Let $X_n \geq Y$ for all $n \in \mathbb{N}$ with $\mathbb{E}[Y] > -\infty$.*
   Furthermore, suppose that

   $$Y \leq X_1 \leq X_2 \leq X_3 \leq \cdots \leq X, \qquad X_n \xrightarrow{\text{pointwise}} X \qquad \left(X = \lim_{n\to\infty} X_n\right).$$

   Then, we have

   $$\mathbb{E}[Y] \leq \mathbb{E}[X_1] \leq \mathbb{E}[X_2] \leq \mathbb{E}[X_3] \leq \cdots \leq \mathbb{E}[X], \qquad \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right].$$

2. *An exact analogue of the above statement holds when $X_1, X_2, \ldots$ are non-increasing and $X_n \leq Y$ for all $n \in \mathbb{N}$, with $\mathbb{E}[Y] < +\infty$*

Fatou's Lemma

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All random variables are assumed to be defined on this space.

Lemma (Fatou's Lemma)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued RVs, and let Y be another real-valued RV.

1. *If $X_n \geq Y$ for all $n \in \mathbb{N}$ and $\mathbb{E}[Y] > -\infty$, then*
$$\mathbb{E}[Y] \;\leq\; \mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \;\leq\; \liminf_{n\to\infty}\; \mathbb{E}[X_n].$$
2. *If $X_n \leq Y$ for all $n \in \mathbb{N}$ and $\mathbb{E}[Y] < +\infty$, then*
$$\mathbb{E}[Y] \;\geq\; \mathbb{E}\left[\limsup_{n\to\infty} X_n\right] \;\geq\; \limsup_{n\to\infty}\; \mathbb{E}[X_n].$$
3. *If $|X_n| \leq Y$ for all $n \in \mathbb{N}$, with $|\mathbb{E}[Y]| < +\infty$, then*
$$-\mathbb{E}[Y] \;\leq\; \mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \;\leq\; \liminf_{n\to\infty}\; \mathbb{E}[X_n] \;\leq\; \limsup_{n\to\infty}\; \mathbb{E}[X_n] \;\leq\; \mathbb{E}\left[\limsup_{n\to\infty} X_n\right] \;\leq\; \mathbb{E}[Y].$$

Proof of Lemma 2, Part 1

- For each $n \in \mathbb{N}$, let

$$Y_n = \inf_{k \geq n} X_k.$$

- Because $X_k \geq Y$ for all k, it follows that

$$Y_n \geq Y \qquad \forall\, n \in \mathbb{N}$$

- Furthermore, note that

$$Y \leq Y_1 \leq Y_2 \leq Y_3 \leq \cdots, \qquad \lim_{n\to\infty} Y_n = \sup_{n\geq 1} Y_n = \liminf_{n\to\infty} X_n.$$

- Applying MCT, we get

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] = \mathbb{E}\left[\lim_{n\to\infty} Y_n\right] \overset{\text{MCT}}{=} \lim_{n\to\infty} \mathbb{E}[Y_n] = \liminf_{n\to\infty} \mathbb{E}[Y_n] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

References

Boris Vladimirovich Gnedenko and Andreĭ Nikolaevich Kolmogorov.
Limit distributions for sums of independent random variables, volume 2420.
Addison-Wesley, 1968.

Geoffrey Grimmett and David Stirzaker.
Probability and random processes.
Oxford university press, 2020.

John Frank Charles Kingman and Samuel James Taylor.
Introdction to Measure and Probability.
Cambridge University Press, 2008.