

Chapter 10B: Limit Theorems

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington
Department of Mathematics

Summer 2026

By the end of this lecture, students should be able to:

- ▶ understand the Law of Large Numbers;
- ▶ distinguish convergence in probability from exact equality;
- ▶ apply the Weak Law of Large Numbers;
- ▶ understand Monte Carlo integration;
- ▶ define the empirical CDF;
- ▶ understand the Central Limit Theorem;
- ▶ standardize sums and averages;
- ▶ apply Normal approximations to Binomial and Poisson distributions.

Why Limit Theorems Matter

Order Emerging from Randomness

Individual random outcomes are unpredictable.

But large collections of random variables often display remarkable regularity.

Examples:

- ▶ sample averages stabilize;
- ▶ proportions settle down;
- ▶ histograms become bell-shaped;
- ▶ randomness averages out.

Main message

Large-scale randomness becomes highly structured.

Two Fundamental Theorems

The two most important limit theorems are:

① Law of Large Numbers (LLN)

averages converge to expectations;

② Central Limit Theorem (CLT)

standardized sums become approximately Normal.

Together, these form the foundation of modern statistics.

Law of Large Numbers (LLN)

Theorem (WLLN)

Let $X_1, X_2, \dots$ be i.i.d. random variables with finite mean μ (and finite variance σ^2)

Then

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0, \quad \text{or equivalently,} \quad \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| \leq \varepsilon) = 1.$$

Interpretation

As the sample size grows, the sample mean $\bar{X}_n$ becomes arbitrarily close to the population mean μ with probability approaching 1.

Notation (Convergence in Probability): This phenomena is also known as *convergence in probability*, and is denoted by

$$\bar{X}_n \xrightarrow{P} \mu.$$

Proof of WLLN Using Chebyshev's Inequality

Recall:

$$\mathbb{E}(\bar{X}_n) = \mu,$$

and

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Applying Chebyshev:

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2}.$$

Thus:

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

As

$$n \rightarrow \infty,$$

the right-hand side goes to

$$0.$$

Suppose:

$$X_i = \begin{cases} 1, & \text{Heads,} \\ 0, & \text{Tails.} \end{cases}$$

Then $\mathbb{E}(X_i) = p$. The sample mean is:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \text{proportion of Heads.}$$

The LLN says $\bar{X}_n \xrightarrow{P} p$. Thus, observed proportions stabilize near the true probability p .

LLN and the Relative-Frequency Interpretation of Probability

The Law of Large Numbers gives a mathematical justification for the **long-run relative-frequency interpretation** of probability: if an experiment is repeated many times, the proportion of trials on which an event occurs approaches its probability.

Let $A \subseteq \mathcal{S}$ be an event, and suppose the experiment is repeated independently under the same conditions. Define

$$X_i = \begin{cases} 1, & \text{if } A \text{ occurs on trial } i, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$X_i \sim \text{Bernoulli}(\mathbb{P}(A)), \quad \mathbb{E}(X_i) = \mathbb{P}(A).$$

Moreover,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{\#\{\text{times } A \text{ occurs in the first } n \text{ trials}\}}{n},$$

so $\bar{X}_n$ is exactly the **relative frequency** of A . By the Law of Large Numbers, $\bar{X}_n \xrightarrow{P} \mathbb{P}(A)$. Thus,

$$\text{relative frequency of } A \xrightarrow[\text{number of trials} \rightarrow \infty]{P} \text{theoretical probability } \mathbb{P}(A).$$

Strong Law

Under the same assumptions of WLLN, we have

$$\mathbb{P}(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1.$$

Equivalently,

$$\mathbb{P}(\{s \in \mathcal{S} : \lim_{n \rightarrow \infty} X(s) = \mu\}) = 1$$

Notation (Almost-Sure Convergence): This phenomenon is also known as *almost-sure convergence* or *almost-everywhere convergence*, and is denoted by

$$\bar{X}_n \xrightarrow{a.s.} \mu, \quad \text{or} \quad \bar{X}_n \xrightarrow{a.e.} \mu$$

Remark: Almost-sure convergence implies (i.e., is stronger than) convergence in probability. Thus, SLLN is stronger than WLLN. The proof requires graduate-level background of Probability (Borel-Canteli Lemma).

LLN Application: Monte Carlo Integration

Monte Carlo Integration

Suppose we want to compute

$$\int_0^1 g(x) dx.$$

Let $U_1, \dots, U_n \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, 1)$.

Then

$$\mathbb{E}(g(U_i)) = \int_0^1 g(x) dx.$$

Hence,

$$\mathbb{E}\left[\frac{1}{n} \sum_{j=1}^n g(U_j)\right] = \int_0^1 g(x) dx.$$

By the WLLN,

$$\frac{1}{n} \sum_{j=1}^n g(U_j) \xrightarrow{P} \int_0^1 g(x) dx, \quad \text{and thus,} \quad \frac{1}{n} \sum_{j=1}^n g(U_j) \approx \int_0^1 g(x) dx \quad \text{for large } n.$$

This is called Monte Carlo estimation of integrals. Draw n random points on the interval, evaluate the function g on them, and then average. The larger the n the closer the average to the true integral.

LLN Application: Empirical Distribution Function

Given observations:

$$X_1, \dots, X_n \sim F$$

define the empirical CDF:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^n I(X_j \leq x).$$

Interpretation:

$$\hat{F}_n(x) = \text{fraction of observations} \leq x.$$

For fixed x , the indicators $I(X_j \leq x)$ are i.i.d. Bernoulli random variables with mean $F(x) = \mathbb{P}(X \leq x)$. Thus, by the LLN:

$$\hat{F}_n(x) \rightarrow F(x).$$

The empirical CDF approximates the true CDF.

Central Limit Theorem

Theorem (CLT)

Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$.

Then for every $z \in \mathbb{R}$,

$$\forall z \in \mathbb{R}, \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z),$$

Interpretation

We previously learned that $\mathbb{E}[\bar{X}_n] = \mu$ and $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$, which means that the fraction $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ is the standardized version of $\bar{X}_n$. CLT says that for large n , the sample mean $\bar{X}_n$ is approximately normal: $\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$. This is especially important because we do not assume any distribution for $X_1, \dots, X_n$ yet the sample mean follows a normal distribution for large n . This offers that Normality emerges universally from averaging.

Notation (Convergence in Distribution): This phenomenon is also known as *convergence in distribution*, and is denoted by $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$.

CLT Application: Normal's Approximation to Binomial

We have previously seen that a binomial r.v. can be written as a sum of n i.i.d. Bernouli r.v.s. Let $X \sim \text{Bin}(n, p)$, and $X = X_1 + \cdots + X_n$, where $X_1, \dots, X_n \sim \text{Ber}(p)$.

Since:

$$\mu = \mathbb{E}[X_i] = p, \quad \sigma^2 = \text{Var}[X_i] = p(1 - p),$$

the CLT gives:

$$\frac{\bar{X}_n - \mathbb{E}[\bar{X}_n]}{\sqrt{\text{Var}[\bar{X}_n]}} = \frac{X/n - p}{\sqrt{p(1 - p)/n}} = \frac{X - np}{\sqrt{np(1 - p)}} \approx N(0, 1).$$

Continuity Correction

A subtle point here is that Binomial distributions are discrete, while Normal distributions are continuous.

A better approximation uses a continuity correction:

$$\mathbb{P}(X \leq k) \approx \mathbb{P}\left(Y \leq k + \frac{1}{2}\right),$$

where

$$Y \sim N(np, np(1 - p)).$$

Idea

The correction accounts for the width of discrete bins.

Suppose:

$$X \sim \text{Pois}(\lambda).$$

For large

$$\lambda,$$
$$\frac{X - \lambda}{\sqrt{\lambda}} \approx N(0, 1).$$

Equivalently:

$$X \approx N(\lambda, \lambda).$$

Why the CLT Is Remarkable

The original distribution can be:

- ▶ discrete;
- ▶ continuous;
- ▶ skewed;
- ▶ highly non-Normal.

Yet averages still become approximately Normal.

Universality

The CLT explains why Normal distributions appear everywhere.

Summary

- ▶ The LLN says averages stabilize.
- ▶ Monte Carlo integration is justified by the LLN.
- ▶ The empirical CDF approximates the true CDF.
- ▶ The CLT says standardized sums/averages become approximately Normal.
- ▶ Standard errors decrease like

$$\frac{1}{\sqrt{n}}.$$

- ▶ Normal approximations apply to Binomial and Poisson distributions.

Next lecture

Chi-Square and Student t distributions.

Temporary page!

$\text{\LaTeX}$ was unable to guess the total number of pages correctly. As there was some unprocessed data have been added to the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page will go away, because $\text{\LaTeX}$ now knows the number of pages to expect for this document.