chapter10b_limit_theorems

Monday, August 17, 2026 12:43 PM

chapter10b_limit_the...

$a_1, a_2, a_3, \dots \in \mathbb{R}$ $(a_n)_{n \in \mathbb{N}} \to a$

$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n \geq N \ |a_n - a| < \varepsilon$

Chapter 10B: Limit Theorems

MATH/STAT 394: Probability I

$f(x) = x^n \quad 0 < x < 1$

$n \to \infty$

Arman Jahangiri

University of Washington Department of Mathematics

$f_1(x), f_2(x), \dots$

$X_1, X_2, X_3, \dots \to X$

Summer 2026

(weak) pointwise

uniform convergence

$f_n(x) \to f(x)$

Learning Goals

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

$\bar{X}_n \to \mu$

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.

$\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \to N(0,1)$

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

$Var(\bar{X}_n) = \sigma^2/n$

Law of Large Numbers (LLN)

$\forall \varepsilon > 0 \ \left(\lim_{n\to\infty} \underbrace{\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon)}_{a_n} = 0 \right) \equiv \left(\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| < \varepsilon) = 1 \right)$ WLLN

$\mathbb{P}\left(\left\{ \omega \in \Omega \ \middle| \ \lim_{n\to\infty} \bar{X}_n(\omega) = \mu \right\} \right) = 1$ SLLN

a number

$\omega \in S$

$\bar{X}_n : S \text{ (or } \Omega) \to \mathbb{R}$

sample space

$\bar{X}_n \xrightarrow[a.e.]{a.s.} \mu$ $\quad x = 1 \quad \lim_{n\to\infty} f_n(x) = \dots$

$\mathbb{P}(A) = 1 \not\Rightarrow A = S$

$\mathbb{P}(A) = 0 \not\Rightarrow A = \phi$

Weak Law of Large Numbers

$a_n \to a \iff |a_n - a| \to 0 \iff |a_n - a| < \varepsilon$

Theorem (WLLN)

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

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

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

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

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

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

$\bar{X}_n = \dfrac{\sum X_i}{n}$ $\quad \bar{X}_1 = X_1$ $\quad P(|\bar{X}_1 - \mu| > \varepsilon), P(|\bar{X}_2 - \mu| > \varepsilon) \dots \to 0$

$0 <$ $\quad < 1$ $\quad 0 <$ $\quad < 1$

$\bar{X}_2 = \dfrac{X_1 + X_2}{2}$

Proof of WLLN Using Chebyshev's Inequality

Recall:

$$\mathbb{E}(\bar{X}_n) = \mu, \quad = E\left(\frac{\sum X_i}{n}\right) = \frac{\sum_{i=1}^{n} E X_i}{n} = \frac{n \cdot \mu}{n} = \mu$$

and

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}. \quad = Var\left(\frac{\sum X_i}{n}\right) = \left(\frac{1}{n^2}\right) \sum_{i=1}^{n} Var X_i = \frac{1}{n^2} \cdot n\sigma^2 =$$

Applying Chebyshev:

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

Thus:

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}. \xrightarrow{n\to\infty} 0$$

As

$$n \to \infty,$$

the right-hand side goes to

$$0.$$

$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$

Coin Toss Example

Suppose:

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

$E X_i = p$

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.

$\bar{X}_n \xrightarrow{P} \mu = p$

LLN's Connection to Relative Frequency (Probability's Naive Definition)

The coin toss example in the previous slide gives a mathematical justification for the frequency interpretation of probability (naive probability) introduced earlier in the course.

Let $A \subseteq \mathcal{S}$ be an event. Define the indicator random variables

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

Then $X_i \sim \text{Bernoulli}(\mathbb{P}(A))$, so $\mathbb{E}(X_i) = \mathbb{P}(A)$. The sample mean becomes

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

which is the relative frequency of the event A (exactly what we previously defined as (naive) definition of probability). Hence, by the LLN,

$$\boxed{\bar{X}_n \xrightarrow{P} \mathbb{P}(A)},$$

or less formally put,

$$\boxed{\text{naive definition of probability}} \xrightarrow[\text{number of times we repeat the experiment} \to \infty]{P} \boxed{\text{Axiomatic Definition of Probability}}$$

Strong Law of Large Numbers

$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| < \varepsilon) = 1$

Strong Law

Under the same assumptions of WLLN, we have

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

$\{\lim_{n\to\infty} \bar{X}_n = \mu\} = \{s \in S \mid \lim_{n\to\infty} \bar{X}_n(s) = \mu\}$ $\quad A \quad P(A) = 1$

Equivalently,

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

$a_1 = P(|\bar{X}_1 - \mu| < \varepsilon)$

$a_2 = P(|\bar{X}_2 - \mu| < \varepsilon)$

$\vdots$

$a_n = P(|\bar{X}_n - \mu| < \varepsilon) \to 1$

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).

$X_n \xrightarrow[a.s.]{a.e.} a \implies X_n \xrightarrow{P} a$

LLN Application: Monte Carlo Integration

Monte Carlo Integration

Suppose we want to compute

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

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

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

$g(U_1), g(U_2), \dots, g(U_n)$

$x_{i+1} = x_i + \Delta$

Hence,

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

$\mathbb{E}(\text{sample avg.}) = \text{pop. avg}$ $\quad \mathbb{E}(\bar{X}_n) = \mu$

$\mathbb{E}(g(U_i))$ $\quad U_i \sim \text{Uniform}(0,1)$

By the WLLN,

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

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

Empirical CDF

$F(1) \approx \frac{3}{5}$ $\quad F(2) = 1 \left(= \frac{5}{5}\right)$

$F_X(x) = \mathbb{P}(X \leq x)$

Given observations:

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

define the empirical CDF:

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

$\xrightarrow[\text{(WLLN)}]{P} \mathbb{E}[\mathbb{1}(X_j \leq x)] = \mathbb{P}(X_j \leq x) = \mathbb{P}(X \leq x) = F_X(x)$

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:

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

The empirical CDF approximates the true CDF.

Empirical CDF is a "consistent" estimator for the true CDF $F(x)$.

Central Limit Theorem

$(X_n \xrightarrow{P} a)$

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

$Y_n = (\bar{X}_n - \mu)/(\sigma/\sqrt{n})$ $\quad$ convergence in prob. $\quad P(|\bar{X}_n - \mu| > \varepsilon) \to 0$ (WLLN)

$\mathbb{P}(\lim \bar{X}_n = \mu) = 1$ (SLLN) $\quad$ convergence almost surely (a.s.) almost everywhere (a.e.) $\quad \bar{X}_n \xrightarrow[(a.e.)]{a.s.} \mu$

Central Limit Theorem

Theorem (CLT)

Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then for every $z \in \mathbb{R}$,

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

$F_{Y_n}(z) \xrightarrow{n\to\infty} F_{N(0,1)}(z)$

$\dfrac{\bar{X}_n - E(\bar{X}_n)}{std(\bar{X}_n) = \sqrt{Var(\bar{X}_n)}} = \dfrac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$

Interpretation

We previously learned that $\mathbb{E}[\bar{X}_n] = \mu$ and $\mathrm{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$, 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 $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$.

$\{W_1, W_2, \dots\}$ $\quad W_n, \ W$

$\forall w \in \mathbb{R} \ \lim_{n\to\infty} F_{W_n}(w) = F_W(w) \Leftarrow$ Definition of $W_n \xrightarrow{D} W$ (Convergence in Distribution)

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 $X_1, \dots, X_n \sim Ber(p)$.
Since:

$$\mu = \mathbb{E}[X_i] = p \qquad \sigma^2 = \mathrm{Var}[X_i] = p(1-p).$$

$E(X) = np < 10$ $\quad n(1-p) < 10$

the CLT gives:

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

$P(a \leq \ \leq b) \quad \Phi(b) - \Phi(a)$

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.

Poisson to Normal

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.