The Law of Large Numbers

Weak and strong forms through a measure-theoretic lens

Zhida

May 5, 2026

UCSB

Probability $\leftrightarrow$ Measure Theory Dictionary

Probability language	Measure-theoretic language
Probability space $(\Omega, \mathcal{F}, \mathbb{P})$	A measure space with total mass $\mathbb{P}(\Omega) = 1$
Event A	Measurable set $A \in \mathcal{F}$
Random variable X	Measurable map $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$
"Almost surely"	$\mathbb{P}$-almost everywhere: failures live inside a null set
Expectation $\mathbb{E}[X]$	Lebesgue integral $\displaystyle\int_\Omega X \, d\mathbb{P}$
Distribution of X	Pushforward measure $X_\# \mathbb{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$
$\mathbb{E}[g(X)]$	$\displaystyle\int_\Omega g(X) \, d\mathbb{P} = \int_\mathbb{R} g \, d(X_\# \mathbb{P})$
Convergence in probability	Convergence in measure under $\mathbb{P}$
L^p **random variable**	Element of $L^p(\Omega, \mathcal{F}, \mathbb{P})$

LLN as a measure statement

SLLN: $\overline{X}_n \to \int X_1 \, d\mathbb{P}$ $\mathbb{P}$-a.e.; WLLN: $\overline{X}_n \to \int X_1 \, d\mathbb{P}$ in measure.

Roadmap

Motivation

The Intuition

Empirical averages

- Coin flips: the proportion of heads approaches 1/2.
- Fair die rolls: the average outcome approaches 3.5.

Mathematical target

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \longrightarrow \quad \mu := \mathbb{E}[X_1].$$

The key question is the *mode* of convergence.

Historical landmarks: Bernoulli's theorem for Bernoulli trials, Borel's almost-sure version, and Kolmogorov's modern strong law.

Setup and Statements

Setup

Let $X_1, X_2, \dots$ be i.i.d. random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

Assume $\mathbb{E}|X_1| < \infty$, and write

$$\mu = \mathbb{E}[X_1] = \int_\Omega X_1 \, d\mathbb{P}.$$

Define partial sums and sample means by

$$S_n = X_1 + \cdots + X_n, \qquad \overline{X}_n = \frac{S_n}{n}.$$

Goal

Understand when and how $\overline{X}_n$ converges to μ as $n \to \infty$.

Weak Law of Large Numbers

Theorem 1 (WLLN)

If $X_1, X_2, \ldots$ *are i.i.d.,* $\mathbb{E}|X_1| < \infty$*, and* $\mu = \mathbb{E}[X_1]$*, then*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

Equivalently, for every $\varepsilon > 0$*,*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \longrightarrow 0.$$

Measure-theoretic translation

$S_n/n \to \mu$ in probability means convergence in measure on $(\Omega, \mathcal{F}, \mathbb{P})$.

Strong Law of Large Numbers

Theorem 2 (Kolmogorov)

Under the same hypotheses,

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

that is,

$$\mathbb{P}\left(\left\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\right\}\right) = 1.$$

Modes of convergence

$$\text{a.s. convergence} \quad \Longrightarrow \quad \text{convergence in probability.}$$

Thus SLLN implies WLLN; the converse is not true in general.

A Short Proof of the WLLN

Tools: Markov and Chebyshev

Lemma 3 (Markov's inequality)

For a non-negative random variable X and $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof.

$\mathbb{E}[X] \geq \mathbb{E}[X\mathbf{1}_{\{X \geq a\}}] \geq a\,\mathbb{P}(X \geq a)$. $\square$

Lemma 4 (Chebyshev's inequality)

If $\mathrm{Var}(X) < \infty$, *then for every* $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

Proof.

Apply Markov's inequality to $(X - \mathbb{E}[X])^2$ with threshold a^2. $\square$

WLLN via Chebyshev

Theorem 5 (Finite-variance WLLN)

If $X_1, X_2, \ldots$ *are i.i.d. and* $\sigma^2 := \mathrm{Var}(X_1) < \infty$, *then*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

Proof.

Linearity and independence give

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mu, \qquad \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

Chebyshev's inequality then yields, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0.$$

$\square$

Removing the Variance Assumption

The full WLLN assumes only $\mathbb{E}|X_1| < \infty$.

Truncation idea

Replace X_i by a bounded version, for example

$$X_i^{(n)} = X_i \mathbf{1}_{\{|X_i| \le n\}}.$$

Then:

- the truncated sum has finite variance, so Chebyshev applies;
- integrability controls the contribution from the tails.

Alternative route

Characteristic functions also prove the L^1 WLLN via $\varphi_{S_n/n}(t) = \varphi_{X_1}(t/n)^n$ and Levy's continuity theorem.

SLLN under Finite Variance

SLLN with Finite Variance

Theorem 6 (L^2 SLLN)

If $X_1, X_2, \dots$ are i.i.d., $\sigma^2 := \mathrm{Var}(X_1) < \infty$, and $\mu = \mathbb{E}[X_1]$, then

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

Why the WLLN proof does not directly upgrade

Chebyshev gives

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2},$$

but $\sum_n 1/n = \infty$, so Borel–Cantelli cannot be applied directly.

Proof Strategy

1. Reduce to non-negative random variables using $X = X^+ - X^-$.
2. Thin the sequence to $n_k = k^2$; now $\sum_k k^{-2} < \infty$.
3. Apply Chebyshev and Borel–Cantelli along k^2.
4. Fill in the gaps between k^2 and $(k+1)^2$ by monotonicity.

Key idea

The subsequence is sparse enough for Borel–Cantelli, but dense enough that the gaps can be controlled.

Steps 1–2: Reduction and Subsequence

Reduction. Write $X_i = X_i^+ - X_i^-$, where $X_i^+, X_i^- \geq 0$. Since

$$\mathbb{E}[(X_i^\pm)^2] \leq \mathbb{E}[X_i^2] < \infty,$$

it suffices to prove the theorem when $X_i \geq 0$.

Subsequence. By Chebyshev,

$$\mathbb{P}\left(\left|\frac{S_{k^2}}{k^2} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{k^2\varepsilon^2}.$$

Since $\sum_k k^{-2} < \infty$, Borel–Cantelli gives

$$\frac{S_{k^2}}{k^2} \xrightarrow{\text{a.s.}} \mu.$$

Step 3: Sandwich the Gaps

Assume $X_i \geq 0$. For $k^2 \leq n \leq (k+1)^2$, monotonicity gives

$$S_{k^2} \leq S_n \leq S_{(k+1)^2}.$$

Therefore

$$\frac{S_{k^2}}{(k+1)^2} \leq \frac{S_n}{n} \leq \frac{S_{(k+1)^2}}{k^2}.$$

Rewriting the two endpoints,

$$\frac{k^2}{(k+1)^2} \cdot \frac{S_{k^2}}{k^2} \leq \frac{S_n}{n} \leq \frac{(k+1)^2}{k^2} \cdot \frac{S_{(k+1)^2}}{(k+1)^2}.$$

Both endpoints converge to μ almost surely, hence $S_n/n \to \mu$ almost surely.

Beyond the Finite-Variance SLLN

- **Kolmogorov's SLLN:** assumes only $\mathbb{E}|X_1| < \infty$.
- **Etemadi's proof:** gives an elegant route under pairwise independence.
- **Martingale viewpoint:** laws of large numbers can be derived from martingale convergence theorems.

Perspective

The L^2 proof is not the most general theorem, but it makes the mechanism transparent: variance control plus Borel–Cantelli.

Closing

Why the LLN Matters

- **Statistics:** consistency of the sample mean as an estimator.
- **Monte Carlo integration:** $n^{-1} \sum_{i=1}^{n} f(X_i) \to \mathbb{E}[f(X)]$.
- **Frequentist probability:** long-run frequencies are stabilized by averaging.
- **Ergodic theory:** Birkhoff's theorem generalizes the SLLN to stationary processes.
- **Empirical processes:** uniform LLNs lead to Glivenko–Cantelli and VC theory.

Summary

Result	Convergence	Main proof tool
WLLN, finite variance	in probability	Chebyshev inequality
SLLN, finite variance	almost surely	Subsequence k^2 plus Borel–Cantelli
Full SLLN, L^1 only	almost surely	Kolmogorov, Etemadi, or martingale methods

One-line takeaway

LLN turns averages of random variables into an integral: empirical means converge to $\int X_1 \, d\mathbb{P}$.

Thank you

Questions?