STA 131A: Introduction to Probability Theory

Lecture 24: The (Weak) Law of Large Numbers

Dogyoon Song

Spring 2026, UC Davis

Announcements

Final exam: Thu, June 11, 1:00–3:00 PM

- **Cumulative:** Lecture 1–Lecture 26
- **Arrive early**: The exam starts at 1:00 PM and ends at 3:00 PM sharp.
- **Three hand-written cheat sheets:** Letter-size (8.5" × 11"), double-sided, brief formulas/notes
- **Calculator**: A simple non-graphing scientific calculator is allowed
- **No other materials** beyond the three permitted cheat sheets are allowed (no textbooks, etc.)
- **SDC accommodations:** If you have “SDC administers exam” language, confirm scheduling with AES online ASAP

Agenda

Last time:

- Markov's inequality:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \qquad (X \geq 0)$$

- Chebyshev's inequality:

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

- Chebyshev bound for sample averages:

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Today:

- Convergence in probability
- Weak law of large numbers
- Sample proportions and Monte Carlo averages
- Preview: central limit theorem

Motivation: Do sample averages stabilize?

Suppose we observe independent repetitions:

$$X_1, X_2, \ldots, X_n.$$

The sample average is

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

If each X_i has mean μ, for large n, we expect

$$\bar{X}_n \approx \mu.$$

But $\bar{X}_n$ is a random variable.

Question: In what precise sense does $\bar{X}_n$ get close to μ?

First clue: Chebyshev bound

Suppose $X_1, \ldots, X_n$ are i.i.d. with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Then

$$\mathbb{E}[\bar{X}_n] = \mu, \qquad \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

By Chebyshev,

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

For any fixed $\epsilon > 0$,

$$\frac{\sigma^2}{n\epsilon^2} \to 0.$$

This suggests that $\bar{X}_n$ converges to μ *in probability*.

Convergence of deterministic sequence

Definition (Convergence of a sequence)

A deterministic sequence of real numbers a_n converges to a, written

$$\lim_{n\to\infty} a_n = a \qquad \text{or} \qquad a_n \to a,$$

if for every $\epsilon > 0$, there exists N_ϵ such that

$$|a_n - a| < \epsilon \qquad \text{for all } n \geq N_\epsilon.$$

Intuition: after some point, all terms of the sequence stay within any prescribed tolerance of the limit.

For random variables, such exact deterministic closeness is usually too much to ask.

- Instead, we ask that large deviations become unlikely:

$$P(|Y_n - a| \geq \epsilon) \to 0.$$

Convergence in probability

Definition (Convergence in probability)

A sequence of random variables Y_n *converges in probability* to a constant a, written

$$Y_n \xrightarrow{P} a,$$

if for every $\epsilon > 0$,

$$P(|Y_n - a| \geq \epsilon) \to 0 \qquad \text{as } n \to \infty.$$

- Equivalently, for every $\epsilon > 0$,
$$P(|Y_n - a| < \epsilon) \to 1.$$
- Recall the definition of convergence:
$$\lim_{n \to \infty} a_n = a \quad \iff \quad \forall \epsilon > 0,\ \exists N \text{ s.t. } |a_n - a| \leq \epsilon \text{ for all } n \geq N.$$

Interpretation: the probability of a noticeable deviation from a goes to zero.

Interpreting convergence in probability

To show

$$Y_n \xrightarrow{P} a,$$

we must show that for every fixed tolerance $\epsilon > 0$,

$$P(|Y_n - a| \geq \epsilon) \to 0.$$

Important:

- The tolerance ϵ is fixed first.
- Then $n \to \infty$.
- The random variable Y_n does not need to equal a.
- It only needs to be close to a with probability tending to 1.
- The probability $P(|Y_n - a| \geq \epsilon)$ converges to 0 in the standard deterministic sense.

Example: A simple convergence-in-probability check

Example

Let

$$Y_n = \begin{cases} n, & \text{with probability } 1/n, \\ 0, & \text{with probability } 1 - 1/n. \end{cases}$$

Question: Does $Y_n \to 0$ in probability?

For any $\epsilon > 0$, when $n > \epsilon$,

$$P(|Y_n - 0| \geq \epsilon) = P(Y_n = n) = \frac{1}{n} \to 0.$$

Thus,

$$Y_n \xrightarrow{P} 0.$$

Lesson: convergence in probability may allow large deviations, but they must be rare.

Pop-up quiz: Convergence in probability

Let

$$Y_n = \begin{cases} 1, & \text{with probability } 1/n, \\ 0, & \text{with probability } 1 - 1/n. \end{cases}$$

Question: Which statement is correct?

A) $Y_n \xrightarrow{P} 0$

B) $Y_n \xrightarrow{P} 1$

C) Y_n does not converge in probability

D) $Y_n = 0$ for all sufficiently large n

Answer: A.

For any $0 < \epsilon \leq 1$,

$$P(|Y_n - 0| \geq \epsilon) = P(Y_n = 1) = \frac{1}{n} \to 0.$$

Follow-up: Does Y_n equal 0 eventually with probability 1? That is a stronger question.

Weak law of large numbers

Theorem (Weak law of large numbers: finite-variance version)

Let $X_1, X_2, \ldots$ be i.i.d. random variables with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Then the sample average

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

satisfies

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

A more general version of WLLN holds under weaker assumptions, but the finite-variance version is sufficient for this course and follows directly from Chebyshev's inequality.

Proof of WLLN under finite variance

Assume $X_1, X_2, \ldots$ are i.i.d. with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Then

$$\mathbb{E}[\bar{X}_n] = \mu, \qquad \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

By Chebyshev's inequality,

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Since

$$\frac{\sigma^2}{n\epsilon^2} \to 0,$$

we conclude

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

What the WLLN says

The weak law says:

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

That is, for any fixed $\epsilon > 0$,

$$P(|\bar{X}_n - \mu| \geq \epsilon) \to 0.$$

Equivalently,

$$P(|\bar{X}_n - \mu| < \epsilon) \to 1.$$

It does not say:

- $\bar{X}_n = \mu$ exactly for large n
- every sample path behaves nicely
- the error has an approximately normal distribution

The CLT will describe the approximate distribution of the standardized error.

Example: Bernoulli trials and sample proportions

Example

Let $X_1, \ldots, X_n$ be i.i.d. $\mathrm{Bernoulli}(p)$. Then

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is the sample proportion of successes. Since

$$\mathbb{E}[X_i] = p,$$

the weak law says

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

Interpretation: long-run empirical frequencies stabilize near the underlying probability.

This connects probability theory back to the frequency interpretation of probability.

Example: Chebyshev bound for a sample proportion

Example

Let $X_i \sim \mathrm{Bernoulli}(p)$, independently, and let

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

Then

$$\mathbb{E}[\bar{X}_n] = p, \qquad \mathrm{Var}(\bar{X}_n) = \frac{p(1-p)}{n}.$$

By Chebyshev's inequality,

$$P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}.$$

Since $p(1-p) \leq \frac{1}{4}$, we get the distribution-free bound

$$P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}.$$

Example: Monte Carlo estimation

Example

Suppose we want to approximate

$$\mu = \mathbb{E}[g(U)],$$

where $U \sim \mathrm{Uniform}(0, 1)$.

Generate independent samples $U_1, \dots, U_n$, and estimate

$$\mu \approx \frac{1}{n} \sum_{i=1}^{n} g(U_i).$$

Let $X_i = g(U_i)$. Then the estimator is

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

By the weak law,

$$\bar{X}_n \xrightarrow{P} \mathbb{E}[X_1] = \mathbb{E}[g(U)].$$

Message: Monte Carlo averages work because of the weak law of large numbers.

Sample size from Chebyshev's inequality

Suppose $X_1, \dots, X_n$ are i.i.d. with

$$\mathrm{Var}(X_i) = \sigma^2.$$

To guarantee

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \delta,$$

Chebyshev gives the sufficient condition

$$\frac{\sigma^2}{n\epsilon^2} \leq \delta.$$

Thus, a sufficient condition is

$$n \geq \frac{\sigma^2}{\delta\epsilon^2}.$$

Important: this is usually conservative, but it requires only a variance bound.

Pop-up quiz: Weak law of large numbers

Let $X_1, X_2, \ldots$ be i.i.d. with

$$\mathbb{E}[X_i] = 5, \qquad \mathrm{Var}(X_i) = 9.$$

Let

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

Question: Which statement follows from Chebyshev's inequality?

A) $P(|\bar{X}_n - 5| \geq 1) \leq 9/n$
B) $P(|\bar{X}_n - 5| \geq 1) \leq 3/n$
C) $P(|\bar{X}_n - 5| \geq 1) = 9/n$
D) $\bar{X}_n = 5$ for all sufficiently large n

Answer: A. Since $\mathrm{Var}(\bar{X}_n) = 9/n$, Chebyshev gives

$$P(|\bar{X}_n - 5| \geq 1) \leq \frac{9/n}{1^2} = \frac{9}{n}.$$

What WLLN does not tell us

The WLLN tells us:

$$\bar{X}_n - \mu \to 0 \quad \text{in probability.}$$

But it does not describe the distribution of the error:

$$\bar{X}_n - \mu.$$

We know $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$. So the typical size of the error is roughly

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

This motivates the standardized error:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Next lecture: the central limit theorem describes the limiting distribution of this standardized error.

WLLN vs. CLT preview

	WLLN	**CLT**
Main object	$\bar{X}_n$	$\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$
Main message	$\bar{X}_n \to \mu$ in probability	standardized error is approximately normal
Scale	error goes to 0 eventually	error is of order $1/\sqrt{n}$

Big picture:

- WLLN explains why averages stabilize.
- CLT explains the approximate shape of the remaining fluctuation.

A note on the strong law

There is a stronger theorem called the **strong law of large numbers**.

Roughly speaking, it says that sample averages converge to the mean with probability 1, a stronger mode of convergence than convergence in probability.

We will not cover the strong law in detail in this course.

Instead, our focus is:

- convergence in probability,
- the weak law of large numbers,
- the central limit theorem.

These are the right tools for the level and goals of this course.

Wrap-up

Convergence in probability

$$Y_n \xrightarrow{P} a \quad \iff \quad P(|Y_n - a| \geq \epsilon) \to 0, \quad \forall \epsilon > 0.$$

Weak law of large numbers

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X_1].$$

- Under finite variance, Chebyshev gives a short proof.

Interpretation

- Sample averages stabilize near the true mean.
- Empirical proportions stabilize near true probabilities.
- Monte Carlo averages are justified by WLLN.

Suggested reading: [BT08, Ch. 5.2 & 5.3]

References

Dimitri Bertsekas and John N Tsitsiklis.
Introduction to probability, volume 1.
Athena Scientific, 2nd edition, 2008.