STA 130A: Mathematical Statistics: Brief Course

Lecture 20: The Weak Law of Large Numbers

Dogyoon Song

Fall 2026, UC Davis

Announcements

Midterm 2 solutions and scores are posted online

- Check course webpage to review the exam problems and solutions
- Check Canvas for scores and the regrade request protocol
- You can review your graded exam on GradeScope
- Note: Only the **higher of your two midterm scores** counts toward the course grade

To submit regrade requests

- If you believe your work deserves additional credit or that there was an error in grading, you may submit a regrade request via GradeScope **by noon on Monday, November 16** with:
  - The specific problem(s) for which you are requesting a regrade
  - A clear explanation of why your answer merits more credit

Homework 6 is posted

- Due Thu, Nov 19, 11:59 PM PT
- Please submit on time and follow the submission instructions
- Review the problems early so you have time to ask questions

Agenda

Before Midterm 2: For an i.i.d. sample average,

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}, \qquad P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}, \quad \epsilon > 0.$$

New question: What does this imply as $n \to \infty$?

Today:

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

Motivation: Do sample averages stabilize?

Suppose we record i.i.d. observations, such as repeated delivery times:

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

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

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

Then

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

By Chebyshev's inequality,

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

For any fixed $\epsilon > 0$,

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

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

Reference: Convergence of a 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 N_ϵ, all terms of the sequence stay within any prescribed tolerance ϵ 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:
$$\lim_{n\to\infty} a_n = a \quad \iff \quad \forall \epsilon > 0,\ \exists N \text{ s.t. } |a_n - a| \leq \epsilon \text{ for all } n \geq N.$$
- Fix a tolerance ϵ first; then let the sample size grow.

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 "always."
- 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

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

For instance, when $n = 100$, the value is 0 with probability 0.99, but 100 with probability 0.01.

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

Suppose Y_n is a sequence of random variables such that

$$\mathbb{E}[Y_n] = a + \frac{1}{n}, \qquad \mathrm{Var}(Y_n) = \frac{1}{n}.$$

Question: Which statement is correct?

A) $Y_n \xrightarrow{P} a$

B) $Y_n \xrightarrow{P} a + 1$

C) Y_n does not converge in probability because $\mathbb{E}[Y_n] \neq a$

D) There is not enough information to determine whether $Y_n \xrightarrow{P} a$

Answer: A. For any fixed $\epsilon > 0$, when n is large enough so that $1/n < \epsilon/2$,

$$|Y_n - a| \geq \epsilon \quad \Longrightarrow \quad |Y_n - \mathbb{E}[Y_n]| \geq \frac{\epsilon}{2}.$$

Therefore, by Chebyshev's inequality,

$$P(|Y_n - a| \geq \epsilon) \leq P\left(|Y_n - \mathbb{E}[Y_n]| \geq \frac{\epsilon}{2}\right) \leq \frac{\mathrm{Var}(Y_n)}{(\epsilon/2)^2} = \frac{4}{n\epsilon^2} \to 0.$$

Weak law of large numbers

Theorem (Weak law of large numbers)

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

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

Then the sample average

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

satisfies

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

Interpretation: For any fixed tolerance, a large sample average is close to μ with probability approaching 1.

Note: 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 via Chebyshev

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,

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

However, WLLN does not say:

- $\bar{X}_n = \mu$ exactly for large n
- the running average gets closer to the mean with every new observation
- 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. $\text{Bernoulli}(p)$. 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 $X_i \sim \text{Bernoulli}(p)$, independently, and let

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

Then

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

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 a bound that does not depend on the unknown value of p

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

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 $U_1, \ldots, U_n$, 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 of large numbers,

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

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

Reference: Sample size from Chebyshev's inequality

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

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

To guarantee

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

For $\epsilon > 0$ and $0 < \delta < 1$, 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: Round up to an integer. This is sufficient, not necessarily minimal for the actual distribution.

Pop-up quiz: What does WLLN justify?

Let $U_1, U_2, \ldots$ be i.i.d. $\mathrm{Uniform}(0,1)$, and define

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} U_i^2.$$

Question: What does the weak law of large numbers imply?

A) $\widehat{\mu}_n \xrightarrow{P} U^2$, where $U \sim \mathrm{Uniform}(0,1)$

B) $\widehat{\mu}_n \xrightarrow{P} \mathbb{E}[U^2] = \dfrac{1}{3}$

C) $\widehat{\mu}_n \xrightarrow{P} \left(\mathbb{E}[U]\right)^2 = \dfrac{1}{4}$

D) $\widehat{\mu}_n = \dfrac{1}{3}$ for all sufficiently large n

Answer: B. By the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} U_i^2 \xrightarrow{P} \mathbb{E}[U^2] = \int_0^1 u^2\, du = \frac{1}{3}.$$

This is not equal to $(\mathbb{E}[U])^2 = 1/4$, and it does not mean the sample average is exactly $1/3$ for finite 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:

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

For $\sigma > 0$, 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.

Reference: A note on the strong law of large numbers

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

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

Weak law of large numbers: for i.i.d. observations

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

- Chebyshev gives a short proof using the variance of the average.

Interpretation

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

Suggested reading: [BT08, Secs. 5.2–5.3]

Next lecture: The central limit theorem and normal approximation.

References


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