STA 130A: Mathematical Statistics: Brief Course

Lecture 21: The Central Limit Theorem

Dogyoon Song

Fall 2026, UC Davis

Agenda

Last time:

- 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:

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

Today: Central limit theorem

- What WLLN tells us, and what it does not tell us
- Statement and intuition of the CLT
- Normal approximation for sums and averages
- De Moivre–Laplace approximation to the binomial

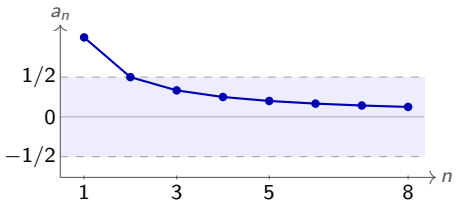
Recap: Convergence of values vs. probabilities

For every fixed tolerance $\epsilon > 0$ around a target a:

Ordinary convergence

$a_n \to a$: $\quad |a_n - a| < \epsilon$
for all sufficiently large n.

Example: $a_n = 1/n$.
Target $a = 0$; tolerance $\epsilon = 1/2$.

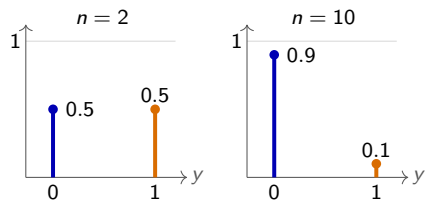


Shaded band: $|a_n| < 1/2$.

Convergence in probability

$Y_n \xrightarrow{P} a$: $\quad P(|Y_n - a| \geq \epsilon) \to 0$
as $n \to \infty$.

Example: $P(Y_n = 1) = 1/n$,
$P(Y_n = 0) = 1 - 1/n$.



Orange mass: $P(|Y_n| \geq 1/2) = 1/n$.

Distinction: Values eventually stay within the band; error probabilities tend to zero.

Recap: What WLLN tells us

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

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

The sample average is

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

The weak law of large numbers says

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

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

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

Message: Sample averages stabilize near the mean (consistency).

Recap: Monte Carlo means simulate, then average

Generate **independent runs from a probability model**; record one numerical output X_i per run.

Delivery reliability

$$X_i = \mathbf{1}_{\{T_i > 30\}}, \qquad \mu = P(T > 30).$$

Average the late-delivery indicators to estimate a probability.

Inventory planning

$$X_i = \max(50 - D_i, 0).$$

Average the leftover loaves to estimate $\mu = \mathbb{E}[\max(50 - D, 0)]$.

The common principle: By the WLLN,

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

Previously, we wrote $X_i = g(U_i)$, using i.i.d. $U_i \sim \mathrm{Uniform}(0, 1)$.

Caution: The limit is model-based; more simulations do not correct an inaccurate model.

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

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

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

Answer: B. Apply WLLN to $X_i = U_i^2$:

$$\widehat{\mu}_n \xrightarrow{P} \mathbb{E}[U^2] = \int_0^1 u^2\, du = \frac{1}{3}.$$

The target is $\mathbb{E}[U^2]$, not $(\mathbb{E}[U])^2$; convergence does not mean exact equality at a finite n.

What WLLN does not tell us

The WLLN does not, by itself, describe the approximate distribution of the error

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

We know

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

So the typical size of the error is roughly

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

This motivates studying the standardized error:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \quad \text{or equivalently} \quad \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \text{where} \quad S_n = \sum_{i=1}^{n} X_i$$

Question: What does this standardized error look like for large n?

Central limit theorem

Theorem (Central limit theorem)

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

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2, \qquad \sigma > 0.$$

Let

$$S_n = X_1 + \cdots + X_n.$$

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

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) \to \Phi(z),$$

where Φ is the standard normal CDF.

Key takeaway: After centering by $n\mu$ and scaling by $\sigma\sqrt{n}$, the sum behaves approximately like a standard normal random variable for large n.

Central limit theorem: Interpretation

The CLT says that for large n,

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) \approx \Phi(z).$$

Equivalently, for large n, probabilities involving S_n can be approximated by

$$P(S_n \leq a) \approx \Phi\left(\frac{a - n\mu}{\sigma\sqrt{n}}\right).$$

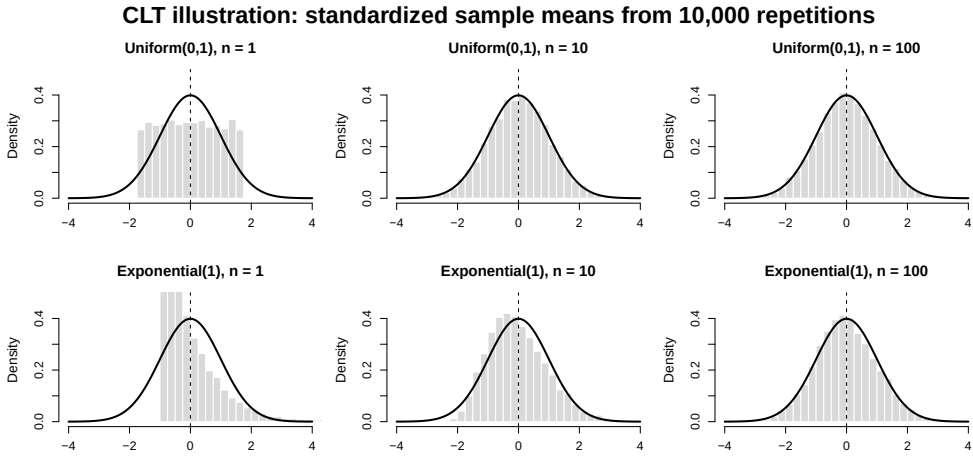
For sample averages,

$$P(\bar{X}_n \leq b) \approx \Phi\left(\frac{b - \mu}{\sigma/\sqrt{n}}\right).$$

Important:

- The individual X_i's do not need to be normal.
- Skewness, heavy tails, and discreteness of X_i's can affect approximation quality.
- The approximation generally becomes more accurate as n grows.

Quality of the normal approximation



The CLT is asymptotic: approximation quality depends on sample size and population shape. There is no single sample size that guarantees a good approximation for every population.

Intuition: Why the normal distribution appears

Assume for simplicity that $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1$. Let

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}.$$

For this MGF argument, suppose M_X is finite near 0. Independence gives

$$M_{Z_n}(t) = \left(M_X\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Because $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1$, a Taylor expansion around 0 gives

$$M_X(u) = 1 + \frac{u^2}{2} + o(u^2).$$

Thus,

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + o(1/n)\right)^n \to e^{t^2/2}.$$

Recall $e^{t^2/2}$ is the MGF of $N(0,1)$. This helps explain why the standard normal appears.

Why the normal distribution is important

The CLT helps explain why normal distributions appear so often.

Many quantities arise as sums of many small independent contributions:

$$S_n = X_1 + \cdots + X_n.$$

Even if the individual pieces X_i are not normally distributed, probabilities involving the standardized sum are often well approximated by standard normal probabilities:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1) \quad \text{in distribution as } n \to \infty.$$

Big picture:

many independent small effects $\implies$ approximately normal aggregate.

This underlies many statistical procedures based on standard errors and normal approximations.

Why study WLLN and CLT separately?

The WLLN and CLT answer different questions.

WLLN: stability of averages

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

- Explains why sample averages and sample proportions are reliable.
- Gives the basic consistency principle:

$$\text{large sample average} \approx \text{true mean}.$$

CLT: shape of the remaining fluctuation

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \approx N(0,1).$$

- Describes fluctuations on the standard-deviation scale $1/\sqrt{n}$.
- Gives normal approximations, standard errors, and margins of error.

Takeaway: WLLN says averages stabilize; CLT describes the remaining randomness.

WLLN vs. CLT

	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	unscaled error goes to 0	error is of order $1/\sqrt{n}$

Interpretation:

- WLLN explains why averages stabilize.
- CLT explains how the remaining error fluctuates.

Pop-up quiz: What does the CLT say?

Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance $\sigma^2 > 0$.

Question: Which statement is closest to the CLT?

A) The individual X_i's become normal as n grows.

B) The sample average $\bar{X}_n$ is approximately $N(0, 1)$.

C) The standardized average $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ is approximately $N(0, 1)$ for large n.

D) The sample average $\bar{X}_n$ equals μ for sufficiently large n.

Answer: C.

The CLT is about the centered and scaled sum or average, not the individual summands and not the unstandardized average.

Follow-up: What factors might affect the quality of the normal approximation?

Normal approximation for sums and averages

Suppose $X_1, \ldots, X_n$ are i.i.d. with mean μ and variance σ^2, with $\sigma > 0$.

For the sum $S_n = X_1 + \cdots + X_n$, the CLT suggests

$$P(S_n \leq c) \approx \Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right).$$

For the sample average $\bar{X}_n = S_n/n$, the CLT suggests

$$P(\bar{X}_n \leq a) \approx \Phi\left(\frac{a - \mu}{\sigma/\sqrt{n}}\right).$$

Workflow:

1. Identify μ, σ^2, and n.
2. Center and scale.
3. Convert the probability into a standard normal probability.

Example: Total package weight

Example

Suppose a plane carries 100 packages whose weights are independent and uniformly distributed between 5 and 50 pounds.

Question: What is the approximate probability that the total weight exceeds 3000 pounds?

For a single package $X_i \sim \mathrm{Uniform}(5, 50)$,

$$\mu = \mathbb{E}[X_i] = \frac{5+50}{2} = 27.5, \qquad \text{and} \qquad \sigma^2 = \mathrm{Var}(X_i) = \frac{(50-5)^2}{12} = \frac{2025}{12} = 168.75.$$

Let $S_{100} = X_1 + \cdots + X_{100}$. Then

$$S_{100} \approx N(100 \cdot 27.5,\ 100 \cdot 168.75).$$

Thus,

$$P(S_{100} > 3000) \approx 1 - \Phi\left(\frac{3000 - 2750}{\sqrt{16875}}\right) \approx 1 - \Phi(1.92) \approx 0.027.$$

Example: Polling error and sample size (1/3)

Example (Polling)

We poll n voters and record the fraction M_n who support a candidate. If p is the true support probability, write

$$M_n = \frac{X_1 + \cdots + X_n}{n},$$

where the X_i's are modeled as independent $\mathrm{Bernoulli}(p)$ variables, $0 < p < 1$. Then

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

By the CLT, M_n is approximately normal $N\left(p, \frac{p(1-p)}{n}\right)$.
Therefore,

$$P(|M_n - p| \geq \epsilon) \approx 2\left[1 - \Phi\left(\frac{\epsilon}{\sqrt{p(1-p)/n}}\right)\right] = 2\left[1 - \Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)\right].$$

Example: Polling error and sample size (2/3)

Example (Polling, continued)

Take $p = 1/2$, $n = 100$, and $\epsilon = 0.1$. Then

$$\mathrm{SD}(M_{100}) = \sqrt{\frac{(1/2)(1/2)}{100}} = 0.05.$$

The normal approximation, without continuity correction, gives

$$P(|M_{100} - 0.5| \geq 0.1) \approx 2\left[1 - \Phi\left(\frac{0.1}{0.05}\right)\right] = 2(1 - \Phi(2)) \approx 0.046.$$

By contrast, Chebyshev gives

$$P(|M_{100} - 0.5| \geq 0.1) \leq \frac{0.25}{100(0.1)^2} = 0.25.$$

Message: Chebyshev is robust but conservative; the CLT often gives sharper approximations.

Self-study: Polling error and sample size (3/3)

Example (Polling sample size)

Question: What sample size targets approximately 95% accuracy within 0.01, so that

$$P(|M_n - p| \leq 0.01) \approx 0.95?$$

Using the CLT, we want approximately

$$\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}} \geq 1.96.$$

If p is unknown, use the worst-case bound $p(1-p) \leq 1/4$. The normal-approximation planning rule is then

$$\frac{0.01\sqrt{n}}{1/2} \geq 1.96 \qquad \Longrightarrow \qquad n \geq 9604.$$

For comparison, using Chebyshev's inequality would give a more conservative conclusion:

$$\frac{1}{4n(0.01)^2} \leq 0.05, \qquad \Longrightarrow \qquad n \geq 50000.$$

Pop-up quiz: CLT standardization

Let $X_1, \ldots, X_{36}$ be i.i.d. with $\mathbb{E}[X_i] = 10$ and $\mathrm{Var}(X_i) = 9$. Let

$$\bar{X}_{36} = \frac{1}{36} \sum_{i=1}^{36} X_i.$$

Question: Which CLT approximation is correct?

A) $P(\bar{X}_{36} > 11) \approx 1 - \Phi(1/3)$

B) $P(\bar{X}_{36} > 11) \approx 1 - \Phi(2)$

C) $P(\bar{X}_{36} > 11) \approx \Phi(2)$

D) $P(\bar{X}_{36} > 11) \approx 1 - \Phi(6)$

Answer: B. Here

$$\mathrm{SD}(\bar{X}_{36}) = \frac{3}{\sqrt{36}} = \frac{1}{2}.$$

Thus

$$P(\bar{X}_{36} > 11) \approx P\left(Z > \frac{11 - 10}{1/2}\right) = 1 - \Phi(2).$$

De Moivre–Laplace approximation to the binomial

We consider a special case of CLT. Let

$$S_n \sim \mathrm{Binomial}(n, p), \qquad 0 < p < 1.$$

We can write

$$S_n = X_1 + \cdots + X_n,$$

where $X_i \sim \mathrm{Bernoulli}(p)$ independently.

Since

$$\mathbb{E}[X_i] = p, \qquad \mathrm{Var}(X_i) = p(1-p),$$

the CLT gives the normal approximation

$$P(S_n \le c) \approx \Phi\left(\frac{c - np}{\sqrt{np(1-p)}}\right).$$

This is called the **de Moivre–Laplace normal approximation** to the binomial.

Continuity correction

A binomial random variable is discrete, but the normal approximation is continuous.

To approximate probabilities involving integer values, use a **continuity correction**:

$$P(k \le S_n \le \ell) \approx P(k - 0.5 \le G \le \ell + 0.5),$$

where

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

For example,

$$P(10 \le S_n \le 20) \approx P(9.5 \le G \le 20.5).$$

For one-sided events:

$$P(S_n \le k) \approx P(G \le k + 0.5), \qquad P(S_n \ge k) \approx P(G \ge k - 0.5).$$

Message: the adjustment by 0.5 better aligns discrete bars with continuous area.

Example: Binomial normal approximation (1/2)

Example

Let $S \sim \text{Binomial}(100, 0.4)$.

Question: Approximate $P(35 \leq S \leq 45)$.

Since S is a sum of 100 independent Bernoulli(0.4) random variables,

$$\mathbb{E}[S] = np = 100(0.4) = 40, \qquad \text{Var}(S) = np(1-p) = 100(0.4)(0.6) = 24.$$

Thus, for probability calculations, we use $G \sim N(40, 24)$ as a normal approximation to S.

Without continuity correction, $P(35 \leq S \leq 45) \approx P(35 \leq G \leq 45)$. Therefore,

$$\begin{aligned} P(35 \leq S \leq 45) &\approx \Phi\left(\frac{45-40}{\sqrt{24}}\right) - \Phi\left(\frac{35-40}{\sqrt{24}}\right) \\ &\approx \Phi(1.02) - \Phi(-1.02) \\ &= 2\Phi(1.02) - 1 \approx 2(0.8461) - 1 = 0.6922. \end{aligned}$$

Example: Binomial normal approximation (2/2)

Example

With continuity correction, the integer event $35 \leq S \leq 45$ is approximated by the continuous interval $34.5 \leq G \leq 45.5$, where $G \sim N(40, 24)$.

$$\begin{aligned} P(35 \leq S \leq 45) &\approx P(34.5 \leq G \leq 45.5) \\ &= \Phi\left(\frac{45.5 - 40}{\sqrt{24}}\right) - \Phi\left(\frac{34.5 - 40}{\sqrt{24}}\right) \\ &\approx \Phi(1.12) - \Phi(-1.12) = 2\Phi(1.12) - 1 \approx 2(0.8686) - 1 = 0.7372. \end{aligned}$$

For comparison, the exact binomial probability is

$$P(35 \leq S \leq 45) = \sum_{k=35}^{45} \binom{100}{k} 0.4^k 0.6^{100-k} \approx 0.7386.$$

Takeaway: The continuity correction improves the normal approximation in this example.

Self-study check: Binomial normal approximation

Let

$$S \sim \text{Binomial}(100, 0.4).$$

Question: With continuity correction, which expression approximates $P(S \leq 45)$?

A) $\Phi\left(\dfrac{45 - 40}{\sqrt{24}}\right)$

B) $\Phi\left(\dfrac{45.5 - 40}{\sqrt{24}}\right)$

C) $1 - \Phi\left(\dfrac{45.5 - 40}{\sqrt{24}}\right)$

D) $\Phi\left(\dfrac{45 - 40}{24}\right)$

Answer: B.

The event $S \leq 45$ is approximated by the continuous event $Z \leq 45.5$, where $Z \sim N(40, 24)$.

Follow-up: What continuity correction would you use for $P(S \geq 45)$?

Wrap-up

Central limit theorem

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) \approx \Phi(z) \qquad \text{for large } n.$$

- CLT describes the approximate shape of the remaining fluctuations.
- Approximation quality depends on n, distribution shape, and discreteness.

Normal approximation

- For sums, compare S_n to $G_n \sim N(n\mu, n\sigma^2)$ in probability calculations.
- For averages, compare $\bar{X}_n$ to $H_n \sim N(\mu, \sigma^2/n)$ in probability calculations.
- Example: Binomial approximation
  - For $S \sim \text{Binomial}(n, p)$, use a normal random variable
    $$G \sim N(np, np(1-p)).$$
  - Use continuity correction when approximating probabilities involving integer cutoffs.

Suggested reading: [BT08, Sec. 5.4]

Next lecture: From random samples to estimation and sampling distributions.

References


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