STA 131A: Introduction to Probability Theory

Lecture 25: The Central Limit Theorem

Dogyoon Song

Spring 2026, UC Davis

Announcements

Final exam: Thu, June 11, 1:00–3:00 PM in Wellman Hall 226 (=classroom)

- **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" $\times$ 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.)

Homework 7 is due tomorrow (Tue, June 2, 11:59 pm)

Course evaluation: Please share your feedback comments by Thu, June 4

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: What WLLN tells us

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

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

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.

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:

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

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

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

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,

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

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

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

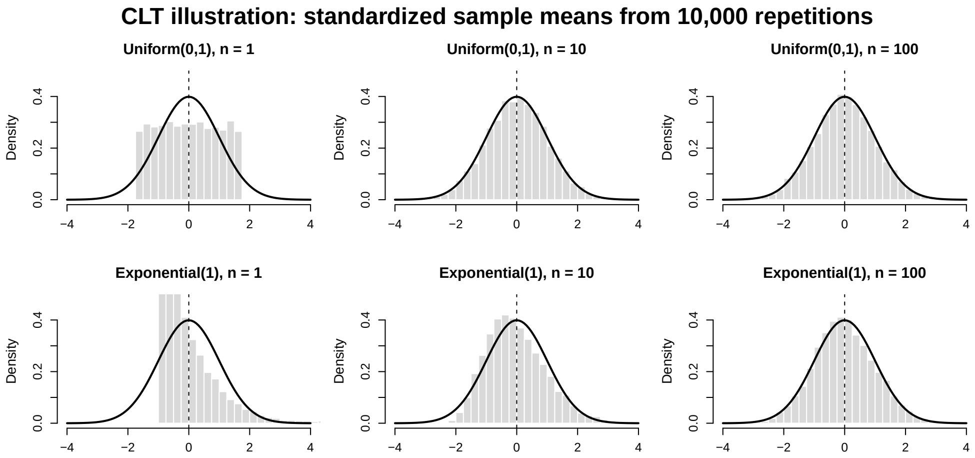
For sample averages,

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

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: accuracy generally improves as *n* grows, but the required sample size depends on the shape of the summand distribution.

Intuition: Why the normal distribution appears

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

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

If $M_X(t)$ is the MGF of X_i, then 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,

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

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:

$$\text{many independent small effects} \quad \Longrightarrow \quad \text{approximately normal aggregate.}$$

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

Why WLLN separately if CLT gives a stronger conclusion?

When the CLT applies, it gives more detailed information about fluctuations:

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

But WLLN remains important and useful:

- WLLN states the basic consistency principle:

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

- WLLN is conceptually simpler and follows from Chebyshev under finite variance.
- Versions of WLLN can hold under weaker conditions than the CLT.
- CLT describes the shape of fluctuations, while WLLN explains why averaging works at all.

Takeaway: WLLN explains stabilization; CLT explains the remaining fluctuation.

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

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

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

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

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,

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

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 independent $\mathrm{Bernoulli}(p)$ random variables. Then

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

By the CLT, M_n is approximately distributed per 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)

For example, take

$$p = \frac{1}{2}, \qquad n = 100, \qquad \epsilon = 0.1.$$

Then

$$\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} = \frac{0.1\sqrt{100}}{\sqrt{1/4}} = 2.$$

Thus,

$$P(|M_n - p| \geq 0.1) \approx 2(1 - \Phi(2)) \approx 0.046.$$

By contrast, Chebyshev's inequality gives

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

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

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

Example (Polling sample size)

Question: How large should n be 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 $p(1-p) \leq 1/4$. Then it is sufficient to require

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

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 $X_1, \dots, X_{36}$ 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

$$S_n \sim \mathrm{Binomial}(n, p).$$

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

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

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**:

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

where

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

For example,

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

For one-sided events:

$$P(S_n \leq k) \approx P(G \leq k + 0.5), \qquad P(S_n \geq k) \approx P(G \geq k - 0.5).$$

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

Example: Binomial normal approximation (1/2)

Example

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

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

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

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

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,

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

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 $G \sim N(40, 24)$.

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

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 accuracy of normal approximation.

Pop-up quiz: Binomial normal approximation

Let

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

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 $S \sim \mathrm{Binomial}(n, p)$, use a normal random variable

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

  - Use continuity correction when approximating probabilities involving integer cutoffs.

Suggested reading: [BT08, Ch. 5.4]

References

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