

Statistics for Economists II

Week 1: Overview & the Central Limit Theorem

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Why Statistics Matters for Economists
- 2 Quick Review: Probability Essentials
- 3 Joint, Marginal & Conditional Distributions
- 4 Random Samples and Statistics
- 5 Convergence Concepts
- 6 The Central Limit Theorem
- 7 Looking Ahead

Important Note

These lecture slides are designed to facilitate and guide your learning, but they are **not a substitute for the textbooks**. To truly understand the material, you must read the relevant textbook chapters carefully. The slides highlight key ideas and intuition; the textbooks provide the depth, proofs, and additional examples necessary for mastery.

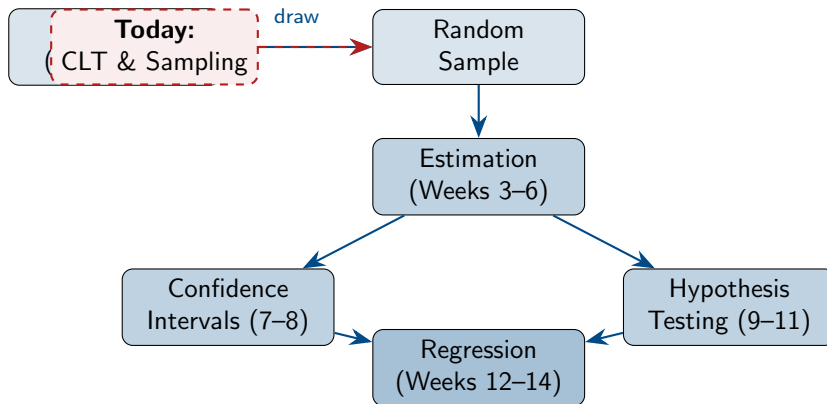
Why Should an Economist Care About Statistics?

- Economics is about **decisions under uncertainty**
- We *never* observe entire populations—only samples
- Statistics gives us tools to make **rigorous claims** from limited data
- Every empirical paper uses concepts from this course:
 - Point estimation, confidence intervals
 - Hypothesis testing, regression

Think About It

The Central Bank decides interest rates based on *sample* inflation data, not the price of every single good. How confident should they be?

Course Overview: The Big Picture



Random Variables: Discrete vs. Continuous

Discrete

Takes values from a **countable** set.

$$\mathbb{P}(X = x_i) \geq 0, \quad \sum_i \mathbb{P}(X = x_i) = 1$$

Examples:

- Number of unemployed in a survey
- Number of firms entering a market

Continuous

Takes values from an **uncountable** set.

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$$

Examples:

- GDP growth rate
- Household income

Expected Value: The “Center of Gravity”

Discrete:

$$\mathbb{E}(X) = \sum_x x p(x)$$

Continuous:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Key properties:

- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ (linearity)
- $\mathbb{E}(g(X)) = \sum_x g(x)p(x)$ or $\int g(x)f(x) dx$ (LOTUS)

Think of $\mathbb{E}(X)$ as the **balance point** of the distribution. If you placed the pdf on a seesaw, $\mathbb{E}(X)$ is where it balances.

Economically: the **long-run average** if the experiment is repeated many times.

Variance: Measuring Uncertainty

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

- Variance measures how **spread out** outcomes are around the mean
- **Higher variance = more uncertainty**
- In finance: variance of returns $\approx$ risk of an asset

Key properties:

- $\text{Var}(aX + b) = a^2 \text{Var}(X)$ (shifting does not change spread)
- $\text{Var}(X) \geq 0$, with equality only if X is constant
- Standard deviation: $\sigma = \sqrt{\text{Var}(X)}$ (same units as X)

Practice: Expected Value (Discrete)

Problem: Let X have pmf:

x	1	2	3	4
$p(x)$	4/10	1/10	3/10	2/10

Find (a) $\mathbb{E}(X)$, (b) $\mathbb{E}(X^2)$, (c) $\text{Var}(X)$.

Practice: Expected Value (Discrete)

Problem: Let X have pmf:

x	1	2	3	4
$p(x)$	$4/10$	$1/10$	$3/10$	$2/10$

Find (a) $\mathbb{E}(X)$, (b) $\mathbb{E}(X^2)$, (c) $\text{Var}(X)$.

Solution:

$$\mathbb{E}(X) = 1\left(\frac{4}{10}\right) + 2\left(\frac{1}{10}\right) + 3\left(\frac{3}{10}\right) + 4\left(\frac{2}{10}\right) = \frac{4+2+9+8}{10} = \boxed{2.3}$$

$$\mathbb{E}(X^2) = 1\left(\frac{4}{10}\right) + 4\left(\frac{1}{10}\right) + 9\left(\frac{3}{10}\right) + 16\left(\frac{2}{10}\right) = \frac{4+4+27+32}{10} = 6.7$$

$$\text{Var}(X) = 6.7 - (2.3)^2 = 6.7 - 5.29 = \boxed{1.41}$$

Practice: Expected Value (Continuous)

Problem: Let X have pdf $f(x) = 4x^3$ for $0 < x < 1$, zero elsewhere.

(a) Verify valid pdf. (b) Find $\mathbb{E}(X)$. (c) Find $\text{Var}(X)$.

Practice: Expected Value (Continuous)

Problem: Let X have pdf $f(x) = 4x^3$ for $0 < x < 1$, zero elsewhere.

(a) Verify valid pdf. (b) Find $\mathbb{E}(X)$. (c) Find $\text{Var}(X)$.

Solution:

$$(a) \quad \int_0^1 4x^3 dx = [x^4]_0^1 = 1 \checkmark$$

$$(b) \quad \mathbb{E}(X) = \int_0^1 x \cdot 4x^3 dx = 4 \int_0^1 x^4 dx = 4 \cdot \frac{1}{5} = \boxed{0.8}$$

$$(c) \quad \mathbb{E}(X^2) = \int_0^1 x^2 \cdot 4x^3 dx = 4 \cdot \frac{1}{6} = \frac{2}{3}$$

$$\text{Var}(X) = \frac{2}{3} - (0.8)^2 = 0.667 - 0.64 = \boxed{0.027}$$

Covariance and Correlation

Definition

Covariance: $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Correlation: $\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad -1 \leq \rho \leq 1$

- $\text{Cov} > 0$: X and Y move in the **same direction**
- $\text{Cov} < 0$: they move in **opposite directions**
- $\text{Cov} = 0$: no *linear* association (nonlinear still possible!)
- Independent $\Rightarrow \text{Cov} = 0$, but **not** vice versa

Variance of a Linear Combination

Two variables:

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$$

General (n variables): If $Y = \sum_{i=1}^n a_i X_i$, then:

$$\text{Var}(Y) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j)$$

If $X_1, \dots, X_n$ independent: $\text{Var}(\sum a_i X_i) = \sum a_i^2 \text{Var}(X_i)$

Portfolio diversification: If two assets have **negative** covariance, combining them *reduces* total risk below the sum of individual risks. This is exactly why investors diversify!

Joint Distributions

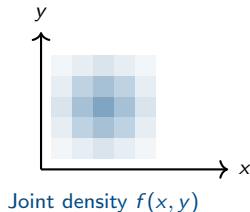
Definition

The **joint distribution** describes the probability of all (x, y) combinations.

Discrete: $f(x, y) = \mathbb{P}(X=x, Y=y)$

Continuous:

$$\mathbb{P}((X, Y) \in A) = \iint_A f(x, y) dx dy$$



Marginal Distributions

From the joint, recover the distribution of **one** variable by “integrating out” the other:

Discrete

$$f_X(x) = \sum_{\text{all } y} f(x, y)$$

$$f_Y(y) = \sum_{\text{all } x} f(x, y)$$

Continuous

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

“Marginal” comes from writing row/column totals **in the margins** of the joint probability table. Sum across a row $\rightarrow$ marginal of X ; sum down a column $\rightarrow$ marginal of Y .

From Joint to Marginal: A Visual Summary

Joint table

$f(x_1, y_1)$	$f(x_1, y_2)$	$f(x_1, y_3)$	$\longrightarrow f_X(x_1)$
$f(x_2, y_1)$	$f(x_2, y_2)$	$f(x_2, y_3)$	$\longrightarrow f_X(x_2) \leftarrow \text{row sums}$
$f(x_3, y_1)$	$f(x_3, y_2)$	$f(x_3, y_3)$	$\longrightarrow f_X(x_3)$

$\downarrow \quad \downarrow \quad \downarrow$

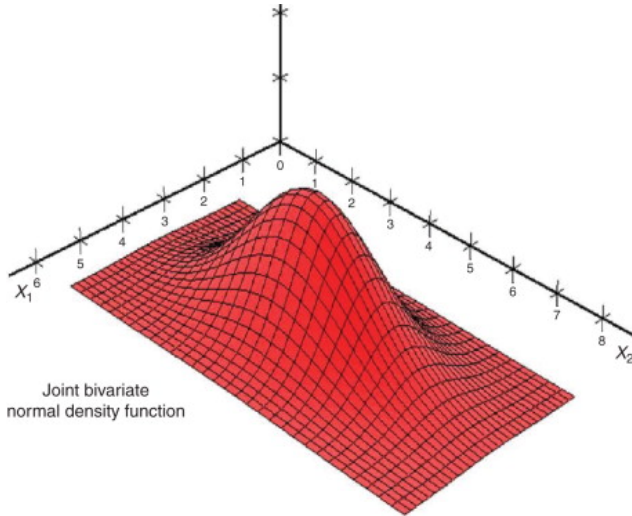
$f_Y(y_1) f_Y(y_2) f_Y(y_3)$

$\uparrow \text{column sums}$

Row sums $\rightarrow$ marginal distribution of X

Column sums $\rightarrow$ marginal distribution of Y

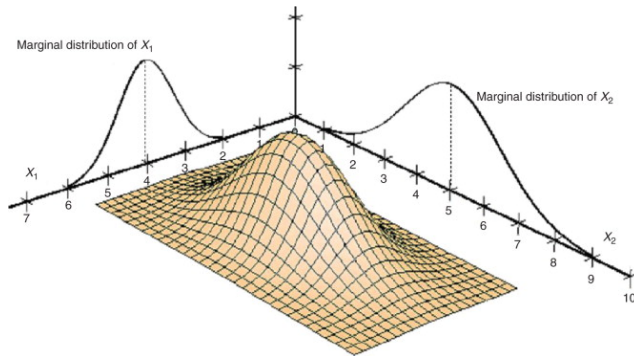
Visualizing a Joint Distribution



The surface shows the **joint bivariate normal density** $f(x_1, x_2)$. The height at any point (x_1, x_2) represents how likely that combination is.

The joint density captures the **complete** relationship between two variables.

From Joint to Marginal Distributions



The curves along each axis are the **marginal distributions**:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

To get the marginal, “collapse” the 3D surface onto one axis by integrating out the other variable. Each marginal tells you about *one* variable, ignoring the other.

Example: Joint & Marginal (Committee Problem)

Problem: From 3 Republicans, 2 Democrats, 1 Independent, a committee of 2 is randomly selected. Let $X = \#$ Republicans, $Y = \#$ Democrats. Total: $\binom{6}{2} = 15$.

Example: Joint & Marginal (Committee Problem)

Problem: From 3 Republicans, 2 Democrats, 1 Independent, a committee of 2 is randomly selected. Let $X = \#$ Republicans, $Y = \#$ Democrats. Total: $\binom{6}{2} = 15$.

	$Y=0$	$Y=1$	$Y=2$	$f_X(x)$
$X=0$	0	2/15	1/15	3/15
$X=1$	3/15	6/15	0	9/15
$X=2$	3/15	0	0	3/15
$f_Y(y)$	6/15	8/15	1/15	1

Row sums $\rightarrow$ marginal of X

Column sums $\rightarrow$ marginal of Y

Conditional Distributions

Definition

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad f_Y(y) > 0$$

- Conditional distributions answer: “**Given what I know** about Y , what can I say about X ?”
- If $X \perp Y$: $f_{X|Y}(x|y) = f_X(x)$ — knowing Y tells you *nothing* new
- This is the foundation of **regression**: $\mathbb{E}[Y|X]$ is the conditional mean

Conditional Expectation & the Law of Iterated Expectations

Definition

$$\mathbb{E}[X \mid Y = y] = \sum_x x \mathbb{P}(X = x \mid Y = y) \quad \text{or} \quad \int x f_{X|Y}(x|y) dx$$

Theorem (Law of Iterated Expectations (LIE))

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$$

Compute an overall average by first computing **group averages**, then taking a **weighted average** of those. Used constantly in labor economics (e.g., wage gaps by education, gender).

Example: Law of Iterated Expectations

Setup: A class has 10 students. We observe each student's GPA and gender:

Student	1	2	3	4	5	6	7	8	9	10
Gender	M	M	M	M	F	F	F	F	F	F
GPA	2.5	2.7	3.0	3.0	2.8	3.0	3.2	3.3	3.5	3.4

Example: Law of Iterated Expectations

Setup: A class has 10 students. We observe each student's GPA and gender:

Student	1	2	3	4	5	6	7	8	9	10
Gender	M	M	M	M	F	F	F	F	F	F
GPA	2.5	2.7	3.0	3.0	2.8	3.0	3.2	3.3	3.5	3.4

Step 1 — Compute group averages (conditional expectations):

$$\mathbb{E}[\text{GPA} \mid \text{Male}] = \frac{2.5 + 2.7 + 3.0 + 3.0}{4} = \frac{11.2}{4} = 2.80$$

$$\mathbb{E}[\text{GPA} \mid \text{Female}] = \frac{2.8 + 3.0 + 3.2 + 3.3 + 3.5 + 3.4}{6} = \frac{19.2}{6} = 3.20$$

Example: Law of Iterated Expectations (cont.)

Step 2 — Compute the overall average directly:

$$\mathbb{E}[\text{GPA}] = \frac{2.5 + 2.7 + 3.0 + 3.0 + 2.8 + 3.0 + 3.2 + 3.3 + 3.5 + 3.4}{10} = \frac{30.4}{10} = 3.04$$

Example: Law of Iterated Expectations (cont.)

Step 2 — Compute the overall average directly:

$$\mathbb{E}[\text{GPA}] = \frac{2.5 + 2.7 + 3.0 + 3.0 + 2.8 + 3.0 + 3.2 + 3.3 + 3.5 + 3.4}{10} = \frac{30.4}{10} = 3.04$$

Step 3 — Verify using the Law of Iterated Expectations:

$$\begin{aligned}\mathbb{E}[\text{GPA}] &= \mathbb{E}[\text{GPA} \mid \text{Male}] \cdot \mathbb{P}(\text{Male}) + \mathbb{E}[\text{GPA} \mid \text{Female}] \cdot \mathbb{P}(\text{Female}) \\ &= 2.80 \times \frac{4}{10} + 3.20 \times \frac{6}{10} \\ &= 1.12 + 1.92 = \boxed{3.04} \quad \checkmark\end{aligned}$$

Both methods give the **same answer**! The LIE says: compute **group averages** first, then take a **weighted average** by group shares. This decomposition is used constantly in labor economics (e.g., gender wage gaps, returns to education by sector).

What Is a Random Sample?

Definition

If $X_1, X_2, \dots, X_n$ are **independent and identically distributed (i.i.d.)**, they constitute a **random sample** of size n from the common distribution.

Independence: knowing one observation gives **no info** about others

Identical distribution: every observation comes from the **same population**

When Does i.i.d. Fail?

- Time series (autocorrelation)
- Non-random sampling (selection bias)
- Population changes during collection

Statistics: Functions of the Sample

Definition

A **statistic** is any function $T = T(X_1, \dots, X_n)$ that **does not involve unknown parameters**.

Most common statistics:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

A statistic **summarizes** the sample. The big question: *how good is this summary?* Is $\bar{X}$ close to μ ? Can we quantify uncertainty? This is what the rest of this course answers!

Key Properties of the Sample Mean

Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$.

Properties of $\bar{X}$

- 1 $\mathbb{E}(\bar{X}) = \mu$ (unbiased — on average, it hits the target)
- 2 $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ (more data $\Rightarrow$ less uncertainty)

Proof of (2):

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} \sum \text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Quadrupling the sample size **halves** the standard error. Diminishing returns: going from $n = 10$ to $n = 40$ helps a lot; from $n = 1000$ to $n = 4000$, less so.

Convergence in Probability

Definition

X_n **converges in probability** to X if for every $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0 \quad \text{Written: } X_n \xrightarrow{P} X$$

As you collect **more data**, the random quantity X_n gets **closer and closer** to the target. Not perfectly every time, but the probability of being far away shrinks to zero.

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

Let X be a random variable with $\mathbb{E}(X) = \mu$ and $\text{Var}(X) = \sigma^2 < \infty$. Then for any $\varepsilon > 0$:

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Equivalently, for any $k > 0$: $\mathbb{P}(|X - \mu| \geq k\sigma) \leq 1/k^2$.

Why is this useful?

- Works for **any** distribution — no normality required!
- Gives a **guaranteed upper bound** on the probability of being far from the mean
- Main tool for proving **convergence in probability**

If the variance is small, X is **unlikely to be far from its mean**. We will use this to prove the Law of Large Numbers next.

The Weak Law of Large Numbers (WLLN)

Theorem (Weak Law of Large Numbers)

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

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

The Weak Law of Large Numbers (WLLN)

Theorem (Weak Law of Large Numbers)

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

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

Proof: We know $\mathbb{E}(\bar{X}_n) = \mu$ and $\text{Var}(\bar{X}_n) = \sigma^2/n$. Applying Chebyshev:

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0 \quad \text{as } n \rightarrow \infty \quad \square$$

With enough data, $\bar{X}_n$ will be **as close to μ as you want**. It works because $\text{Var}(\bar{X}_n) = \sigma^2/n \rightarrow 0$. This is the theoretical foundation of surveys, polling, and empirical economics.

Example 1: Sample Mean Converges to μ

The WLLN applies to **any** i.i.d. sequence with finite variance:

- Survey n households about income: $\bar{X}_n \xrightarrow{P} \mu_{\text{income}}$

In every case, Chebyshev guarantees convergence:

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$$

The key mechanism: The variance of $\bar{X}_n$ shrinks at rate $1/n$. No matter how large σ^2 is, we can make $\text{Var}(\bar{X}_n)$ as small as we want by taking n large enough. This is why **collecting more data always helps** (assuming the data is i.i.d.).

Example 2: Sample Proportion Converges to p

Let $X_1, X_2, \dots$ be i.i.d. Bernoulli(p) (e.g., each voter supports a policy or not).

Define $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$ (sample proportion).

Claim: $\hat{p}_n \xrightarrow{P} p$.

Example 2: Sample Proportion Converges to p

Let $X_1, X_2, \dots$ be i.i.d. Bernoulli(p) (e.g., each voter supports a policy or not). Define $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$ (sample proportion).

Claim: $\hat{p}_n \xrightarrow{P} p$.

Proof: Since $\text{Var}(\hat{p}_n) = p(1 - p)/n$, Chebyshev gives:

$$\mathbb{P}(|\hat{p}_n - p| \geq \varepsilon) \leq \frac{p(1 - p)}{n\varepsilon^2} \longrightarrow 0 \quad \text{as } n \rightarrow \infty \quad \square$$

This is a special case of the WLLN (Bernoulli is i.i.d. with $\mu = p$ and $\sigma^2 = p(1 - p)$). It explains why **opinion polls** become more accurate with larger samples: the sample proportion “locks on” to the true proportion.

The Continuous Mapping Theorem (CMT)

Theorem (Continuous Mapping Theorem)

If $X_n \xrightarrow{P} c$ (a constant) and g is a continuous function, then:

$$g(X_n) \xrightarrow{P} g(c)$$

Similarly, if $X_n \xrightarrow{d} X$ and g is continuous, then $g(X_n) \xrightarrow{d} g(X)$.

Continuous functions **preserve convergence**. Since $\bar{X}_n \xrightarrow{P} \mu$ by the WLLN:

- $\bar{X}_n^2 \xrightarrow{P} \mu^2, \quad e^{\bar{X}_n} \xrightarrow{P} e^\mu, \quad \log(\bar{X}_n) \xrightarrow{P} \log(\mu) \text{ (if } \mu > 0)$

Once you prove convergence of one quantity, you get convergence of many related quantities **for free**.

Convergence in Distribution

Definition

X_n **converges in distribution** to X if for every continuity point x of F_X :

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad \text{Written: } X_n \xrightarrow{d} X$$

Convergence in distribution is **weaker** than convergence in probability. It says the *shape* of the CDF approaches a limit, but the random variables themselves need not get close to each other. The CLT (coming next) is a statement about convergence in distribution.

Relationships Between Modes of Convergence

The Key Implications

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$$

Convergence in probability is **stronger** — it implies convergence in distribution, but **not** vice versa.

Important special case: If the limit is a **constant** c , then:

$$X_n \xrightarrow{d} c \iff X_n \xrightarrow{P} c$$

The two notions **coincide** when the limit is a constant.

Summary of our tools:

- **WLLN:** $\bar{X}_n \xrightarrow{P} \mu$ (conv. in probability to a constant)
- **CLT:** $(\bar{X}_n - \mu)/(\sigma/\sqrt{n}) \xrightarrow{d} N(0, 1)$ (conv. in distribution)
- **CMT:** continuous functions preserve both types

The Central Limit Theorem (CLT)

Theorem (Central Limit Theorem)

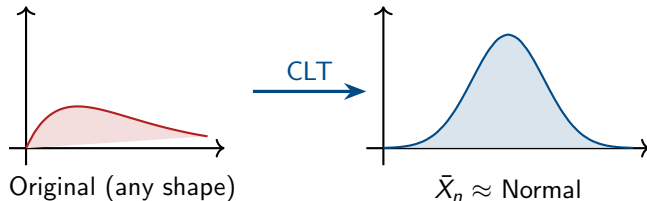
Let $X_1, X_2, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$. Then:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1) \quad \text{as } n \rightarrow \infty$$

Equivalently: $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ for large n .

The most important theorem in statistics. No matter what the original distribution looks like (skewed, bimodal, discrete), the sample mean becomes approximately normal for large n . This is why the normal distribution appears everywhere!

Why Is the CLT So Powerful?



What makes it powerful:

- Works for **any** distribution (as long as $\sigma^2 < \infty$)
- Tells us the **exact distribution** to use for inference
- Makes confidence intervals and hypothesis tests possible

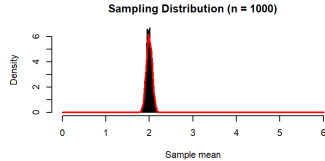
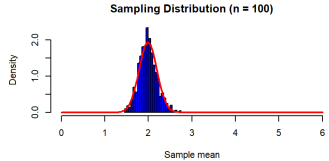
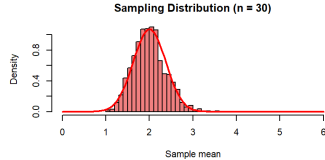
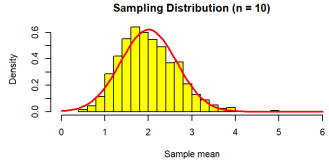
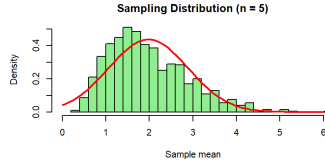
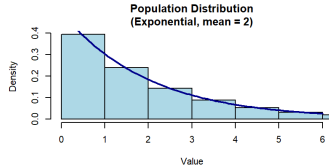
CLT: How Large Is “Large Enough”?

Rule of thumb: $n \geq 30$, but it depends on the shape:

Distribution	n needed
Symmetric, light tails	$n \geq 10$
Moderate skewness	$n \geq 30$
Heavy skewness	$n \geq 50+$
Already normal	$n = 1$ (exact!)

The more “non-normal” the original distribution, the larger n needs to be. For symmetric distributions, the CLT kicks in very fast!

CLT in Action: Simulation



Example: Applying the CLT

Problem: A factory produces bolts with mean length $\mu = 10$ cm and $\sigma = 0.5$ cm. A sample of $n = 64$ bolts is taken. What is $\mathbb{P}(\bar{X} > 10.1)$?

Example: Applying the CLT

Problem: A factory produces bolts with mean length $\mu = 10$ cm and $\sigma = 0.5$ cm. A sample of $n = 64$ bolts is taken. What is $\mathbb{P}(\bar{X} > 10.1)$?

Solution: By CLT, $\bar{X} \sim N\left(10, \frac{0.5^2}{64}\right) = N(10, 0.003906)$.

$$Z = \frac{\bar{X} - 10}{0.5/\sqrt{64}} = \frac{10.1 - 10}{0.0625} = 1.6$$

$$\mathbb{P}(\bar{X} > 10.1) = \mathbb{P}(Z > 1.6) = 1 - \Phi(1.6) = 1 - 0.9452 = \boxed{0.0548}$$

About 5.5% chance the sample mean exceeds 10.1 cm. If we actually observe $\bar{x} = 10.1$, should we question whether $\mu = 10$? This is the logic behind **hypothesis testing** (Weeks 9–11)!

Example: CLT for Sums

Problem: An insurance company has 10,000 policies. Each policy has an expected annual claim of \$500 with $\sigma = \$800$. What is the probability total claims exceed \$5.2 million?

Example: CLT for Sums

Problem: An insurance company has 10,000 policies. Each policy has an expected annual claim of \$500 with $\sigma = \$800$. What is the probability total claims exceed \$5.2 million?

Solution: Let $T = \sum_{i=1}^{10000} X_i$.

$$\mathbb{E}(T) = 10000 \times 500 = \$5,000,000$$

$$\sigma_T = 800\sqrt{10000} = \$80,000$$

By CLT: $T \sim N(5,000,000, (80,000)^2)$.

$$\mathbb{P}(T > 5,200,000) = \mathbb{P}\left(Z > \frac{200,000}{80,000}\right) = \mathbb{P}(Z > 2.5) = \boxed{0.0062}$$

Less than 1% risk. This is how insurance companies set **premiums**: use the CLT to quantify the probability of extreme total losses.

What's Coming Next

Week 1 (Today): Overview & CLT ✓

Week 2: Sampling Distributions (χ^2 , t , F , Order Statistics)

Weeks 3–6: Estimation Theory (Bias, MSE, MLE, ...)

Weeks 7–8: Confidence Intervals

Weeks 9–11: Hypothesis Testing

Weeks 12–14: Regression & OLS

Key Takeaways from Today

- 1 **Random samples** (i.i.d.) are the foundation of statistical inference
- 2 **Statistics** ($\bar{X}$, S^2) are random variables with their own distributions
- 3 **Chebyshev's inequality** bounds the probability of being far from the mean
- 4 The **Law of Large Numbers**: $\bar{X}_n \xrightarrow{P} \mu$
- 5 The **Central Limit Theorem**: $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$
- 6 The **CMT**: continuous functions preserve convergence

Reading Assignment

- Wackerly, Mendenhall & Scheaffer: Chapters 7.1–7.2
- Hogg, Tanis & Zimmerman: Chapter 5.6–5.8