PhD Math Camp: Estimation

Shangze Rudy Dai

Food and Resource Economics Department
University of Florida

Aug 12, 2025

Contents

Section 1

Introduction to Estimation

Introduction to Estimation

Example: A Farmer's Question
A local farmer asks an extension specialist:

> *"If the inflation rate is 5% this year, how much should I expect egg prices to increase?"*

You have access to historical data on:

- Annual inflation rates
- Corresponding changes in egg prices

How should you answer?

- Guess based on experience?
- Use the average change?
- Or use statistical estimation to establish a relationship?

This leads us to the need for **estimation techniques** — turning data into informed answers.

What Can Estimation Do?

Once we apply estimation techniques, we can achieve the following:

1. Estimate the Relationship

- Example: For every 1% increase in inflation, egg prices rise by **0.8% on average**.
- This gives us a quantified relationship between variables.

2. Make Predictions

- Given 5% inflation this year, we can **predict** an expected 4% increase in egg prices.
- Helps farmers and policymakers plan ahead.

3. Provide Confidence Intervals

- Instead of just a point prediction, we can say: *"There is a 95% chance the increase in egg prices will be between 2.5% and 5.5%."*
- This accounts for uncertainty and improves decision-making.

Why Estimation Matters in Economics/ Econometrics

Estimation = Turning Data into Answers

In economics, we use estimation to:

- Quantify relationships (e.g., price with demand)
- Test theories (e.g., rational expectations)
- Measure policy impacts (e.g., minimum wage on employment)
- Make predictions (e.g., inflation, growth)

Without estimation, theory stays abstract. With estimation, economics becomes actionable.

The Importance of Randomness

Why can't we just use exact formulas or fixed rules from past data?
Because real-world data — like egg prices and inflation — are influenced by many **random factors**.

Examples of randomness in our problem:

- Weather shocks affecting egg production
- Sudden changes in consumer demand
- Unexpected policy changes

As a result:

- The same inflation rate may lead to different egg price changes in different years
- We must treat variables as **random variables**, not fixed numbers

This is why estimation methods rely on *probability and statistics*:
To extract signal from noisy data, and to measure uncertainty.

From Randomness to Reliability: Large Sample Theory

Question: If data are random and noisy, how can we trust our estimates?

Answer: Thanks to **Large Sample Theory** (also called *asymptotic theory*):

- As the sample size grows, the effect of randomness **averages out**.
- Our estimates become more **accurate** and **stable**.

Key ideas:

- **Law of Large Numbers (LLN)**: The sample average converges to the true population value.
- **Central Limit Theorem (CLT)**: The distribution of the estimator becomes approximately normal as $n \to \infty$.

Implication: With enough data, we can make reliable predictions — even in a random world.

Section 2

Large Sample Theory

Convergence

What does it mean for an estimator to be "good" as the sample size increases? We need the estimator to **converge** to the true value.

Two important types of convergence:

- **Convergence in Probability:**

$$\hat{\theta}_n \xrightarrow{p} \theta$$

  Means that as $n \to \infty$, the probability that $\hat{\theta}_n$ is close to θ approaches 1. $\to$ Used in the **Law of Large Numbers**.

- **Convergence in Distribution:**

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

  Means the distribution of the scaled estimator approaches a normal distribution. $\to$ Used in the **Central Limit Theorem**.

Formal Definition: Convergence in Probability

Definition: A sequence of random variables X_n **converges in probability** to a constant μ, written as

$$X_n \xrightarrow{p} \mu,$$

if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - \mu| > \varepsilon) = 0.$$

Intuition:

- As the sample size n grows, the probability that X_n differs from μ by more than any small number ε becomes negligible.
- X_n becomes more and more "concentrated" around μ.

Example: Sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges in probability to the true mean μ, under mild conditions.

$$\bar{X}_n \xrightarrow{p} \mu$$

Example: Sample Mean Converges in Probability

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

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

Define the sample mean:

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

Claim:

$$\bar{X}_n \xrightarrow{p} \mu$$

Proof Sketch (via Chebyshev's Inequality)*:

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

Interpretation: As we collect more data, the sample mean becomes very likely to be close to the true mean.

Formal Definition: Convergence in Distribution

Definition: A sequence of random variables X_n **converges in distribution** to a random variable X, written as

$$X_n \xrightarrow{d} X,$$

if for every point x where the cumulative distribution function (CDF) $F_X(x)$ is continuous,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

where $F_{X_n}(x) = \mathbb{P}(X_n \leq x)$.

Intuition:

- The distribution of X_n gets closer and closer to that of X, in shape.
- Unlike convergence in probability, this does *not* require X_n to be close to X with high probability — only their distributions need to be close.

Simple Example: Convergence in Distribution

Example: Let $X_n \sim \textbf{Uniform}(0, 1/n)$ That is, for each n, X_n is a random variable uniformly distributed on the interval $[0, 1/n]$.

Claim:

$$X_n \xrightarrow{d} 0$$

(i.e., converges in distribution to a constant random variable equal to 0)

Why? Let $F_{X_n}(x)$ be the CDF of X_n. Then for $x > 0$:

$$F_{X_n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{1/n} = nx & \text{if } 0 \leq x \leq 1/n \\ 1 & \text{if } x > 1/n \end{cases}$$

As $n \to \infty$, for any $x > 0$,

$$F_{X_n}(x) \to 1, \quad F_{X_n}(x) \to 0 \text{ for } x < 0$$

So:

$$F_{X_n}(x) \to \mathbf{1}\{x \geq 0\} = \text{CDF of constant } 0$$

Therefore: $X_n \xrightarrow{d} 0$

Other Types of Convergence

1. Almost Sure Convergence (a.s.):

$$X_n \xrightarrow{\text{a.s.}} X$$

Means:

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Interpretation: The sequence X_n converges to X for almost every outcome (sample path). $\rightarrow$ Strongest form of convergence.

2. L^r Convergence (e.g., Mean Square Convergence):

$$X_n \xrightarrow{L^r} X \quad \text{if } \mathbb{E}[|X_n - X|^r] \to 0$$

Special case:

$$r = 2 \Rightarrow \text{Mean Square Convergence}$$

Relationship:

$$\text{a.s. } \textit{and } L^r \Rightarrow \text{in probability} \Rightarrow \text{in distribution}$$

(The reverse implications do *not* generally hold.)

Weak Law of Large Numbers (WLLN)

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

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

Define the sample mean:

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

Then, as $n \to \infty$,

$$\bar{X}_n \xrightarrow{p} \mu$$

Interpretation:

- The sample mean becomes close to the population mean with high probability as sample size increases.
- Even though each X_i is random, their average becomes predictable.

Why "weak"? Because it guarantees **convergence in probability**, not almost sure convergence. (The "Strong Law"gives a.s. convergence.)
Proof idea (via Chebyshev's inequality):

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

Example: WLLN with Fair Coin Tosses

Setup: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables representing coin tosses:

$$X_i = \begin{cases} 1 & \text{Heads} \\ 0 & \text{Tails} \end{cases} \quad \text{with } \mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = 0.5$$

Then:

$$\mathbb{E}[X_i] = 0.5, \quad \mathrm{Var}(X_i) = 0.25$$

Define the sample mean:

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

By the Weak Law of Large Numbers:

$$\bar{X}_n \xrightarrow{p} 0.5$$

Interpretation:

- As you toss more coins, the proportion of heads gets closer to 50%.
- For large n, it's very unlikely to see a proportion far from 0.5.

Central Limit Theorem (CLT)

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

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

Then the standardized sample mean:

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

converges in distribution to the standard normal:

$$Z_n \xrightarrow{d} \mathcal{N}(0,1)$$

Interpretation:

- For large n, the sampling distribution of $\bar{X}_n$ is approximately normal.
- This holds even if the original X_i are **not normally distributed**.
- Makes statistical inference possible (confidence intervals, hypothesis tests).

Key takeaway: *The average of many small, independent random effects tends to look Gaussian.*

Proof Sketch of Central Limit Theorem (CLT)*

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

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

Define the standardized sum:

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

Goal: Show that $Z_n \xrightarrow{d} \mathcal{N}(0,1)$

Sketch of Proof (via Lindeberg–Feller / Lyapunov idea):

1. Define $Y_i = \frac{X_i - \mu}{\sigma}$, so that $\mathbb{E}[Y_i] = 0$, $\mathrm{Var}(Y_i) = 1$
2. Then $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$
3. By Lyapunov's condition: If $\mathbb{E}[|Y_i|^{2+\delta}] < \infty$ for some $\delta > 0$, then:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \xrightarrow{d} \mathcal{N}(0,1)$$

4. Hence, $Z_n \xrightarrow{d} \mathcal{N}(0,1)$

Example: CLT with Rolling Dice

Setup: Let $X_1, X_2, \dots, X_n$ be i.i.d. outcomes from rolling a fair six-sided die.

$$X_i \in \{1, 2, 3, 4, 5, 6\}, \quad \mathbb{P}(X_i = k) = \frac{1}{6}$$

Then:

$$\mathbb{E}[X_i] = 3.5, \quad \mathrm{Var}(X_i) = \frac{35}{12}$$

Standardize:

$$Z_n = \frac{\bar{X}_n - 3.5}{\sqrt{\frac{35}{12n}}}$$

By CLT:

$$Z_n \xrightarrow{d} \mathcal{N}(0, 1)$$

Interpretation:

- Although individual rolls are clearly **not normal**, the sample mean becomes approximately normal when n is large.
- This allows us to compute confidence intervals and p-values using the normal distribution.

Extension: Law of Large Numbers beyond i.i.d.*

So far: We stated the Weak Law of Large Numbers (WLLN) under the assumption that $X_1, X_2, \ldots, X_n$ are i.i.d.

But the law still holds under more general conditions:

- **Non-i.i.d. (independent but not identically distributed):** If $X_1, X_2, \ldots$ are independent with

$$\sup_n \mathbb{E}[|X_n|] < \infty, \quad \text{and } \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] \to \mu,$$

  then $\bar{X}_n \xrightarrow{p} \mu$

- **Weighted Averages (not $1/n$):** Let weights $w_i \geq 0$, and $\sum w_i = 1$. Then under regularity conditions,

$$\hat{\theta}_n = \sum_{i=1}^{n} w_i X_i \xrightarrow{p} \mu$$

Takeaway: The LLN is surprisingly robust — it often holds beyond the ideal i.i.d. case!

Extension: Central Limit Theorem beyond i.i.d.*

Recall: The classic CLT requires i.i.d. random variables with finite variance.

But CLT still holds under more general conditions:

- **Independent, Non-Identical:** Let $X_1, \ldots, X_n$ be independent but not identically distributed, with

  $$\mathbb{E}[X_i] = \mu_i, \quad \mathrm{Var}(X_i) = \sigma_i^2$$

  Define:

  $$S_n = \sum_{i=1}^{n} (X_i - \mu_i), \quad V_n^2 = \sum_{i=1}^{n} \sigma_i^2$$

  If Lyapunov's condition holds (e.g. finite $(2+\delta)$-th moments), then:

  $$\frac{S_n}{\sqrt{V_n^2}} \xrightarrow{d} \mathcal{N}(0,1)$$

Key Point: CLT is powerful and flexible — but conditions must be checked carefully in non-i.i.d. settings.

Section 3

Point Estimation

Introduction to Point Estimation

Goal of Statistics: Use sample data to learn about unknown population parameters.

What is a Point Estimator? A point estimator is a rule or formula that provides a single number to estimate an unknown parameter.

Estimation in General: A point estimator is a function of the sample:

$$\hat{\theta}_n = \varphi(X_1, X_2, \ldots, X_n)$$

That is, it uses data $X_1, X_2, \ldots, X_n$ to estimate the unknown parameter θ.

Examples:

- $\bar{X}_n = \frac{1}{n} \sum X_i$: estimates population mean μ
- $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2$: estimates variance σ^2
- $\hat{\beta} = (X'X)^{-1}X'Y$: OLS estimate of regression coefficients

Example: Estimating Average Corn Yield

Scenario: An agricultural extension specialist wants to estimate the **average corn yield per acre** in a region.

Data: The specialist randomly samples $n = 10$ farms and records their yields (in bushels per acre):

$$X_1 = 150,\ X_2 = 170,\ \ldots,\ X_{10} = 165$$

Estimator: We use the sample mean as a point estimator for the population mean:

$$\hat{\mu} = \bar{X}_{10} = \frac{1}{10}\sum_{i=1}^{10} X_i$$

Interpretation:

- $\hat{\mu}$ is a single number — the estimated average yield in the population.
- This is our best guess, based on limited data.

How Good Is an Estimator? Measures of Closeness

Goal: We want our estimator $\hat{\theta}$ to be **close** to the true value θ

A Common Measure: Mean Squared Error (MSE)

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right]$$

Why squared error?

- Penalizes large deviations more heavily
- Always non-negative
- Easy to analyze mathematically

Interpretation:

- Low MSE means the estimator is *on average close* to the true value
- A trade-off often exists between bias and variance

Two Key Properties: Unbiasedness and Efficiency

1. Unbiasedness

$$\hat{\theta} \text{ is unbiased if } \mathbb{E}[\hat{\theta}] = \theta$$

Interpretation: On average, the estimator hits the true parameter value.

Example: The sample mean $\bar{X}_n$ is an unbiased estimator of the population mean μ.

Note: An estimator can still be useful even if it's slightly biased, as long as its MSE is low.

2. Efficiency
Among all unbiased estimators, an efficient estimator has the **lowest variance**.

Definition: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased, and

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2),$$

then $\hat{\theta}_1$ is more efficient.

Example: Calculate Bias, Variance, and MSE

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

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

Consider the following estimator of μ:

$$\hat{\mu}_1 = \frac{1}{4}X_1 + \frac{3}{4}X_2 \quad \text{(use only first two observations)}$$

Questions:

1. Is $\hat{\mu}_1$ unbiased for μ?
2. What is $\mathrm{Var}(\hat{\mu}_1)$?
3. What is $\mathrm{MSE}(\hat{\mu}_1)$?

Hints:

- Use linearity of expectation: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- Use independence: $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$

Maximum Likelihood Estimation (MLE)

Idea: Choose the parameter value that makes the observed data most "likely."

Setup: Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations from a distribution with density $f(x; \theta)$

Likelihood Function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

Often easier to work with the **log-likelihood:**

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(X_i; \theta)$$

MLE:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(\theta)$$

MLE Example: Coin with 3 Observations

We observe $X_1 = 1, \quad X_2 = 0, \quad X_3 = 1$ Assume: $X_i \sim \text{Bernoulli}(p)$, i.i.d.

Step 1: Likelihood Function

$$L(p) = p^{X_1}(1-p)^{1-X_1} \cdot p^{X_2}(1-p)^{1-X_2} \cdot p^{X_3}(1-p)^{1-X_3}$$
$$= p^1(1-p)^0 \cdot p^0(1-p)^1 \cdot p^1(1-p)^0 = p^2(1-p)^1$$

Step 2: Log-Likelihood

$$\ell(p) = \log L(p) = \log(p^2) + \log(1-p) = 2\log p + \log(1-p)$$

Step 3: Maximize Log-Likelihood Take derivative:

$$\frac{d\ell}{dp} = \frac{2}{p} - \frac{1}{1-p}$$

Set derivative to zero:

$$\frac{2}{p} = \frac{1}{1-p} \Rightarrow 2(1-p) = p \Rightarrow 2 - 2p = p \Rightarrow 3p = 2 \Rightarrow \hat{p}_{\text{MLE}} = \frac{2}{3}$$

Conclusion: The MLE for p based on this sample is $\frac{2}{3}$

Key Properties of MLE

1. Consistency

$$\hat{\theta}_{\text{MLE}} \xrightarrow{p} \theta$$

As sample size $n \to \infty$, the MLE converges in probability to the true parameter value.

2. Asymptotic Normality

$$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta) \xrightarrow{d} \mathcal{N}(0, I(\theta)^{-1})$$

Where $I(\theta)$ is the Fisher information. $\to$ Enables confidence intervals and hypothesis tests .

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} \log f(X;\theta)\right)^2\right]$$

3. Invariance Property

If $\hat{\theta}_{\text{MLE}}$ is the MLE of θ, then for any function $g(\cdot)$,

$$g(\hat{\theta}_{\text{MLE}}) \text{ is the MLE of } g(\theta)$$

Beyond Point Estimation: Interval Estimation

Motivation: A point estimate (like $\hat{\theta} = 2.3$) gives us a single best guess — but tells us nothing about uncertainty.

Question: How can we quantify the **uncertainty** in our estimator?

Answer: Interval Estimation

Section 4

Interval Estimation

What Is a Confidence Interval?

Definition: A $(1-\alpha)$ confidence interval for a parameter θ is a random interval $[L(X), U(X)]$ such that:

$$\mathbb{P}(L(X) \leq \theta \leq U(X)) = 1 - \alpha$$

Key Components:

- θ: unknown parameter (fixed)
- $L(X), U(X)$: functions of the sample (random)
- $1-\alpha$: confidence level (e.g., 0.95 for 95%)

Interpretation: If we repeated the experiment many times and built a confidence interval each time, about $(1-\alpha) \times 100\%$ of the intervals would contain the true value θ.

Note: The confidence is in the *procedure*, not in the specific interval from one sample.

Example: Confidence Interval for the Mean

We know $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d., with known variance σ^2 and unknown mean. And we have observed the values for these Xs.

Goal: Construct a $(1-\alpha)$ confidence interval for the population mean μ

Step 1: Standardize the sample mean

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

Step 2: Rearranging gives the confidence interval

$$\mathbb{P}\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Result:

$\boxed{\bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$ is a $(1-\alpha)$ confidence interval for μ

How Do We Construct a Confidence Interval?

1. What do we want to estimate? The unknown parameter (e.g., the population mean μ)

2. What do we already know?

- The distribution of the estimator (e.g., $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$)
- Known quantities (e.g., $\bar{X}_n, \sigma, n$)

3. What distribution can we use? Standardize the estimator to get a known distribution:

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

4. Use that distribution to form a probability statement:

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

5. Rearranging the inequality gives the confidence interval for μ:

$$\mu \in \left[\bar{X}_n - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X}_n + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Key idea: We construct the interval by "inverting" a known distribution.

Exercise: Confidence Interval for the Mean

Problem:

Suppose you observe 3 independent draws from a normal distribution:

$$X_1 = -1, \quad X_2 = 0, \quad X_3 = 1, \quad \text{with } X_i \sim \mathcal{N}(\mu, 1)$$

Tasks:

1. Find the theoretical distribution of the sample mean $\bar{X}$
2. Construct a 95% confidence interval for μ, assuming known variance

Hints:

- $\bar{X} = \frac{X_1+X_2+X_3}{3}$
- Use the fact that $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n) = \mathcal{N}(\mu, 1/3)$
- Use $z_{0.025} \approx 1.96$ for 95% confidence

Important Distribution: Chi-Square (χ^2) Distribution

Definition: If $Z_1, Z_2, \ldots, Z_k$ are i.i.d. standard normal random variables, then the sum of their squares follows a chi-square distribution with k degrees of freedom:

$$\chi_k^2 = \sum_{i=1}^{k} Z_i^2$$

Notation:

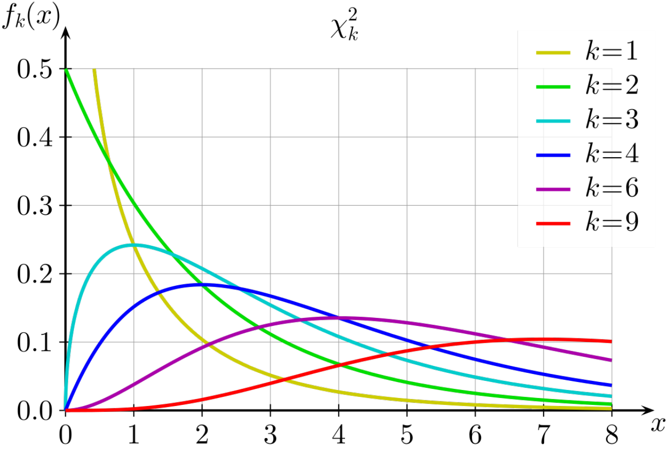
$$\chi_k^2 \sim \text{Chi-Square distribution with } k \text{ degrees of freedom}$$

Key Properties:

- Non-negative: $\chi^2 \geq 0$
- Right-skewed, especially for small k
- Mean $= k$, Variance $= 2k$

Visual intuition: As k increases, χ_k^2 becomes more symmetric and resembles a normal distribution.

Important Distribution: Chi-Square (χ^2) Distribution



Important Distribution: Student's t-Distribution

Motivation: When the population variance σ^2 is unknown, we estimate it using the sample variance. This introduces additional uncertainty — the normal distribution is no longer exact.

Definition: Let:

- $Z \sim \mathcal{N}(0,1)$ (standard normal)
- $U \sim \chi^2_k$ (chi-square with k degrees of freedom)
- Z and U are independent

Then the random variable

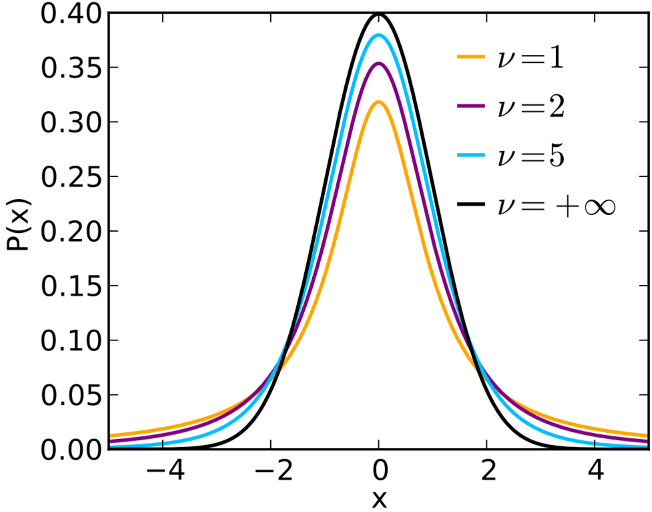
$$T = \frac{Z}{\sqrt{U/k}} \sim t_k$$

follows a Student's t-distribution with k degrees of freedom.

Key Properties:

- Symmetric and bell-shaped, like normal
- Heavier tails $\rightarrow$ more uncertainty from estimating σ^2
- As $k \to \infty$, $t_k \to \mathcal{N}(0,1)$

Important Distribution: Student's t-Distribution



Inference When Both μ and σ^2 Are Unknown

Setup: Suppose we observe:

$$X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2) \quad \text{i.i.d.}$$

Both μ and σ^2 are unknown.

Step 1: Estimate the mean and variance from data

- Sample mean: $\bar{X}_n = \frac{1}{n} \sum X_i$
- Sample variance: $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2$

Step 2: Use the t-distribution

$$T = \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

This result holds exactly because the sample comes from a normal distribution.

Step 3: Construct a confidence interval for μ

Example: 95% Confidence Interval for μ (Unknown σ^2)

Given data:

$$X_1 = 2, \quad X_2 = 4, \quad X_3 = 6, \quad n = 3$$

Example: 95% Confidence Interval for μ (Unknown σ^2)

Given data:

$$X_1 = 2, \quad X_2 = 4, \quad X_3 = 6, \quad n = 3$$

Step 1: Compute sample mean and variance

$$\bar{X} = \frac{2+4+6}{3} = 4$$

$$S^2 = \frac{1}{2}\left[(2-4)^2 + (4-4)^2 + (6-4)^2\right] = \frac{1}{2}(4+0+4) = 4 \quad \Rightarrow \quad S = 2$$

Step 2: Use $t_{n-1} = t_2$ distribution

$$\text{For 95\% CI:} \quad t_{2,0.025} \approx 4.303$$

Step 3: Construct the confidence interval

$$\bar{X} \pm t_{2,0.025} \cdot \frac{S}{\sqrt{n}} = 4 \pm 4.303 \cdot \frac{2}{\sqrt{3}} \approx 4 \pm 4.97$$

Result:

$$[-0.97,\ 8.97] \quad \text{is the 95\% confidence interval for } \mu$$

Wrap-Up and What's Next

Today's Topics: Estimation and Inference

- Why we need **estimation** in statistics
- Point estimation: sample mean, MLE, their properties (bias, variance, MSE)
- Interval estimation: confidence intervals and how to construct them
- Key distributions in inference:
  - Normal
  - Chi-square
  - t-distribution
- Full example: 95% confidence interval for unknown μ and σ^2

Tomorrow's Topics:

- **Linear Regression:** modeling relationships between variables
- **Hypothesis Testing:** making decisions under uncertainty

See you then!

References

- Moss, C. B. (2014). *Mathematical Statistics for Applied Econometrics*. CRC Press.
- Wikipedia contributors. *Various entries on statistical distributions and estimation*. Retrieved from `https://en.wikipedia.org`
- Stanford University. *Lecture 18 — Confidence Intervals*. STATS 200: Introduction to Statistical Inference, Autumn 2016. Retrieved from `https://web.stanford.edu/class/archive/stats/stats200/stats200.1172/Lecture18.pdf`