

Lecture 7: Sampling Distributions

Monte Carlo Simulation · CLT in Action · Standard Errors

Previously, on Lectures 4–6...

MLE: $\hat{\theta} = \arg \max \ell(\theta)$. Find the parameter that makes data most likely.

MAP: $\hat{\theta} = \arg \max [\ell(\theta) + \log P(\theta)]$. MLE with a prior = regularization.

Fisher info: $I(\theta) = -E[\ell''(\theta)]$. Curvature of the log-likelihood.

CR bound: $\text{Var}(\hat{\theta}) \geq 1/(nI(\theta))$. A floor on how precise we can be.

Asymptotic normality: $\hat{\theta}_{\text{MLE}} \sim N(\theta_0, 1/(nI(\theta_0)))$ for large n .

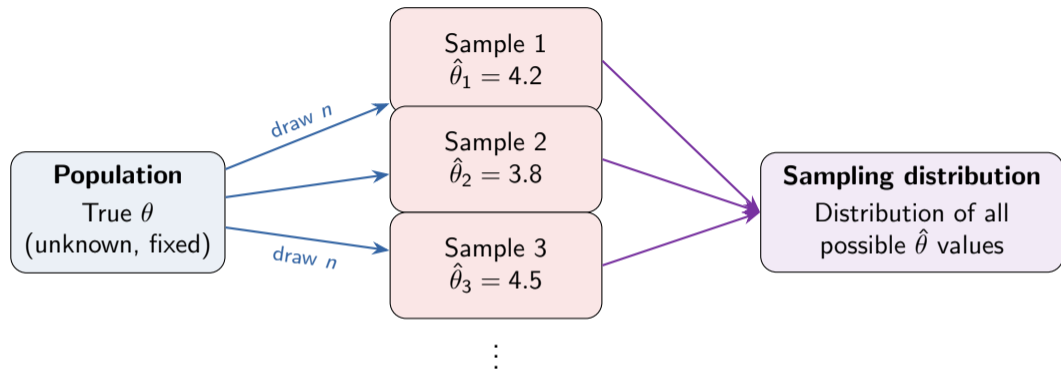
Today: You computed $\hat{\theta} = 3.2$. But how **reliable** is this number?

If you repeated the experiment, would you get 3.2 again? 3.1? 5.7?

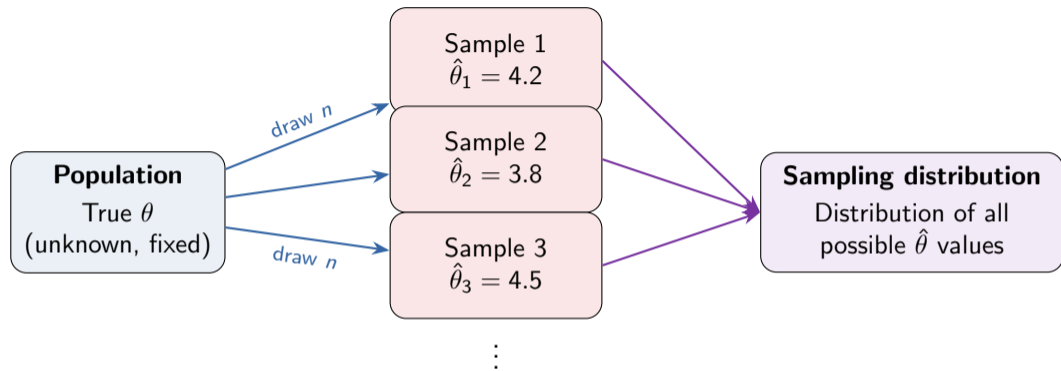
Your Estimator Is Random

Different sample \Rightarrow different $\hat{\theta}$

The Thought Experiment



The Thought Experiment



Key insight: $\hat{\theta}$ is a *function of random data*, so $\hat{\theta}$ itself is a random variable.

Its distribution is called the **sampling distribution**. Understanding it is the foundation of all inference.

But We Only Have One Sample!

In real life, you run the experiment **once**. You get **one** $\hat{\theta}$.

The problem:

We can't actually repeat the experiment 1,000 times to see the sampling distribution.

Today: solutions 1 and 2. Next lecture: solution 3 (the bootstrap).

But We Only Have One Sample!

In real life, you run the experiment **once**. You get **one** $\hat{\theta}$.

The problem:

We can't actually repeat the experiment 1,000 times to see the sampling distribution.

Three solutions:

1. **Theory** — derive it mathematically
2. **Simulation** — fake it on a computer
3. **Bootstrap** — re-sample the data itself

But We Only Have One Sample!

In real life, you run the experiment **once**. You get **one** $\hat{\theta}$.

The problem:

We can't actually repeat the experiment 1,000 times to see the sampling distribution.

Three solutions:

1. **Theory** — derive it mathematically
2. **Simulation** — fake it on a computer
3. **Bootstrap** — re-sample the data itself

Today: solutions 1 and 2. Next lecture: solution 3 (the bootstrap).

Seeing It: Monte Carlo Simulation

Let the computer repeat the experiment for you

The Monte Carlo Recipe

Setup: Pick a true population distribution F with known θ .

Why “Monte Carlo”? Named after the famous casino. We use *randomness* to learn about randomness.

The Monte Carlo Recipe

Setup: Pick a true population distribution F with known θ .

Step 1: Generate a random sample $X_1, \dots, X_n \sim F$.

Why “Monte Carlo”? Named after the famous casino. We use *randomness* to learn about randomness.

The Monte Carlo Recipe

Setup: Pick a true population distribution F with known θ .

Step 1: Generate a random sample $X_1, \dots, X_n \sim F$.

Step 2: Compute the estimator: $\hat{\theta} = T(X_1, \dots, X_n)$.

Why “Monte Carlo”? Named after the famous casino. We use *randomness* to learn about randomness.

The Monte Carlo Recipe

Setup: Pick a true population distribution F with known θ .

Step 1: Generate a random sample $X_1, \dots, X_n \sim F$.

Step 2: Compute the estimator: $\hat{\theta} = T(X_1, \dots, X_n)$.

Step 3: Repeat Steps 1–2 a total of R times (e.g., $R = 10,000$).

Why “Monte Carlo”? Named after the famous casino. We use *randomness* to learn about randomness.

The Monte Carlo Recipe

Setup: Pick a true population distribution F with known θ .

Step 1: Generate a random sample $X_1, \dots, X_n \sim F$.

Step 2: Compute the estimator: $\hat{\theta} = T(X_1, \dots, X_n)$.

Step 3: Repeat Steps 1–2 a total of R times (e.g., $R = 10,000$).

Step 4: Plot the histogram of $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_R$. This *is* the sampling distribution.

The Monte Carlo Recipe

Setup: Pick a true population distribution F with known θ .

Step 1: Generate a random sample $X_1, \dots, X_n \sim F$.

Step 2: Compute the estimator: $\hat{\theta} = T(X_1, \dots, X_n)$.

Step 3: Repeat Steps 1–2 a total of R times (e.g., $R = 10,000$).

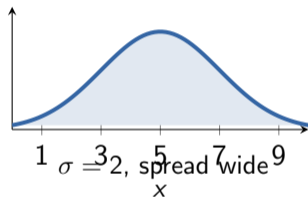
Step 4: Plot the histogram of $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_R$. This *is* the sampling distribution.

Why “Monte Carlo”? Named after the famous casino. We use *randomness* to learn about randomness.

Example: Sample Mean from $N(5, 4)$

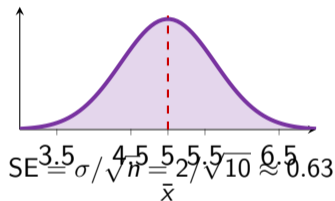
Population: $X_i \sim N(\mu = 5, \sigma^2 = 4)$, $n = 10$, estimator: $\hat{\mu} = \bar{X}$

Population $N(5, 4)$



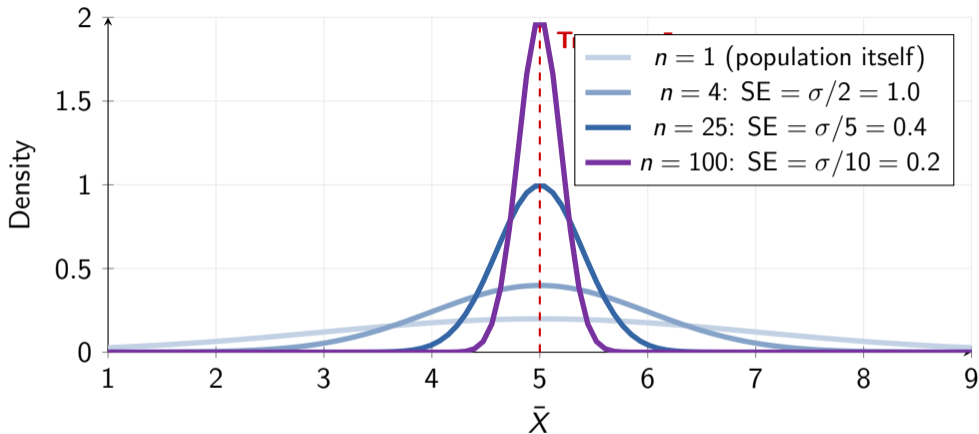
simulate $R = 10,000$

Sampling dist. of \bar{X}



The sampling distribution is **much narrower** than the population — averaging reduces the spread by \sqrt{n} .

What Happens as n Grows?

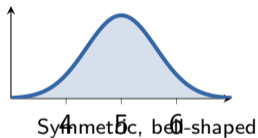


More data \Rightarrow the sampling distribution **concentrates** around the true value. This is **consistency**.

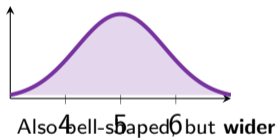
It Works for Any Estimator

Not just \bar{X} ! **Every** estimator has a sampling distribution.

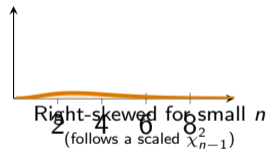
Sample Mean \bar{X}



Sample Median



Sample Variance S^2



Comparing Estimators via Their Sampling Distributions

Key observation

Observation: Different estimators have **different** sampling distributions.

The mean is narrower than the median (for Normal data) — it's a *more efficient* estimator.

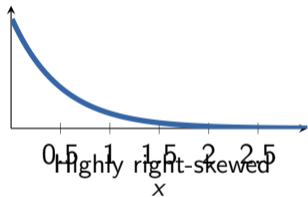
Recall from Lecture 4: the Cramér–Rao bound tells us the *best possible* variance.

Coming up: We'll see that the **Central Limit Theorem** explains *why* sampling distributions tend to look Normal — and **Fisher information** determines how narrow they are.

Example: MLE for $\text{Exp}(\lambda)$

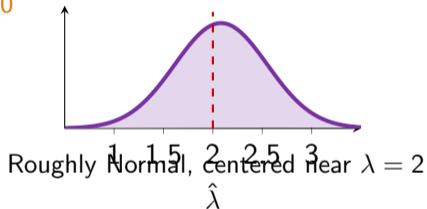
Population: $X_i \sim \text{Exp}(\lambda = 2)$, MLE: $\hat{\lambda} = 1/\bar{X}$

Population $\text{Exp}(2)$



$R = 10,000, n = 20$

Sampling dist. of $\hat{\lambda}_{\text{MLE}}$



The population is **exponential** (skewed), but the MLE is **approximately Normal**. Why? **CLT!**

The Central Limit Theorem

Why everything looks Normal

CLT: The Most Important Theorem in Statistics

Central Limit Theorem: If X_1, \dots, X_n are i.i.d.
with mean μ and variance $\sigma^2 < \infty$, then

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

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

CLT: The Most Important Theorem in Statistics

Central Limit Theorem: If X_1, \dots, X_n are i.i.d. with mean μ and variance $\sigma^2 < \infty$, then

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

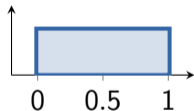
Equivalently: $\bar{X} \overset{\sim}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ for large n .

The magic: It doesn't matter what the population looks like — uniform, exponential, Bernoulli, bimodal, anything. As long as the variance is finite, \bar{X} becomes Normal.

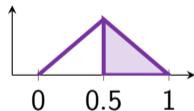
CLT in Action: Uniform Population

$X_i \sim \text{Uniform}(0, 1)$ — completely flat, yet \bar{X} becomes Normal

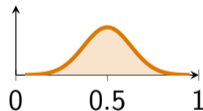
$n = 1$



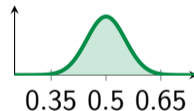
$n = 2$



$n = 5$



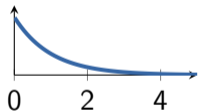
$n = 30$



CLT in Action: Exponential Population

$X_i \sim \text{Exp}(1)$ — strongly right-skewed, yet \bar{X} becomes Normal (needs larger n)

$n = 1$



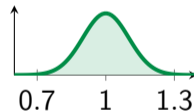
$n = 5$



$n = 30$



$n = 100$



How Large Must n Be?

Symmetric
(Uniform, Normal)

$n \geq 5$ is often
enough

Mildly skewed
(Exponential, Poisson)

$n \geq 25-30$
usually works

Heavily skewed
(Pareto, lognormal)

May need $n \geq 100$
or more

How Large Must n Be?

Symmetric
(Uniform, Normal)

$n \geq 5$ is often
enough

Mildly skewed
(Exponential, Poisson)

$n \geq 25-30$
usually works

Heavily skewed
(Pareto, lognormal)

May need $n \geq 100$
or more

The common “ $n \geq 30$ ” rule is a useful heuristic, not a law.
For symmetric data, even $n = 5$ is fine. For heavy tails, even $n = 100$ may
not suffice.
When in doubt — **simulate!** That's what Monte Carlo is for.

CLT: What It Does **Not** Say

The CLT is powerful, but often over-applied. Three common misconceptions:

✗ “With large n , the **data** becomes Normal.”

✓ The data stays whatever it is. Only the *sample mean* (and sums) become Normal.

CLT: What It Does **Not** Say

The CLT is powerful, but often over-applied. Three common misconceptions:

✗ “With large n , the **data** becomes Normal.”

✓ The data stays whatever it is. Only the *sample mean* (and sums) become Normal.

✗ “CLT works for any distribution.”

✓ It requires **finite variance**. The **Cauchy** distribution has no mean or variance — averaging Cauchy data does *not* converge to Normal. Try it in the practical!

CLT: What It Does **Not** Say

The CLT is powerful, but often over-applied. Three common misconceptions:

✗ “With large n , the **data** becomes Normal.”

✓ The data stays whatever it is. Only the *sample mean* (and sums) become Normal.

✗ “CLT works for any distribution.”

✓ It requires **finite variance**. The **Cauchy** distribution has no mean or variance — averaging Cauchy data does *not* converge to Normal. Try it in the practical!

✗ “CLT works for any sample, even if observations are dependent.”

✓ The basic CLT requires **independence** (or at least weak dependence). Time series, spatial data, and clustered samples need special CLT versions.

CLT for MLEs: Even More Powerful

The CLT doesn't just apply to \bar{X} . For **any** MLE (under regularity conditions):

$$\hat{\theta}_{\text{MLE}} \quad \dot{\sim} \quad N\left(\theta_0, \frac{1}{n \cdot I(\theta_0)}\right)$$

The MLE is approximately Normal, centered at the truth, with variance determined by the **Fisher information**.

CLT for MLEs: Even More Powerful

The CLT doesn't just apply to \bar{X} . For **any** MLE (under regularity conditions):

$$\hat{\theta}_{\text{MLE}} \sim N\left(\theta_0, \frac{1}{n \cdot I(\theta_0)}\right)$$

The MLE is approximately Normal, centered at the truth, with variance determined by the **Fisher information**.

Bernoulli(p)

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Poisson(λ)

$$\hat{\lambda} \sim N\left(\lambda, \frac{\lambda}{n}\right)$$

Exp(λ)

$$\hat{\lambda} \sim N\left(\lambda, \frac{\lambda^2}{n}\right)$$

Each follows from $I(\theta)$: Bernoulli $I(p) = 1/(p(1-p))$, Poisson $I(\lambda) = 1/\lambda$, etc.

Standard Error

The most important number after the estimate itself

SD \neq SE: The Most Confused Pair in Statistics

Standard Deviation (SD)

Spread of the **data**

$$\sigma = \sqrt{E[(X - \mu)^2]}$$

“How variable are individual observations?”

Does not shrink with n .
More data doesn't make individuals less variable.

Standard Error (SE)

Spread of the **estimator**

$$SE(\bar{X}) = \sigma / \sqrt{n}$$

“How variable is $\hat{\theta}$ across repeated experiments?”

Shrinks with n .
More data \Rightarrow more precise estimate.

SD \neq SE: The Most Confused Pair in Statistics

Standard Deviation (SD)

Spread of the **data**

$$\sigma = \sqrt{E[(X - \mu)^2]}$$

“How variable are individual observations?”

Does not shrink with n .
More data doesn't make individuals less variable.

Standard Error (SE)

Spread of the **estimator**

$$SE(\bar{X}) = \sigma / \sqrt{n}$$

“How variable is $\hat{\theta}$ across repeated experiments?”

Shrinks with n .
More data \Rightarrow more precise estimate.

Example: Heights of adults: SD \approx 7 cm (people vary).
Average of 100 heights: SE = $7 / \sqrt{100} = 0.7$ cm (the average is precise).

Standard Error of the Sample Mean

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

The **standard error** is the standard deviation of the sampling distribution.

It measures how much \bar{X} would vary if we repeated the experiment.

Standard Error of the Sample Mean

$$\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

The **standard error** is the standard deviation of the sampling distribution.

It measures how much \bar{X} would vary if we repeated the experiment.

Derivation:

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

Standard Error of the Sample Mean

$$\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

The **standard error** is the standard deviation of the sampling distribution. It measures how much \bar{X} would vary if we repeated the experiment.

Derivation:

$$\begin{aligned}\text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \\ \Rightarrow \text{SE}(\bar{X}) &= \sqrt{\text{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}\end{aligned}$$

Uses i.i.d. assumption: observations are independent, so variances **add**.

SE in Practice: A Polling Example

Scenario: A poll of $n = 1,000$ voters finds $\hat{p} = 0.52$ support for a candidate.

SE in Practice: A Polling Example

Scenario: A poll of $n = 1,000$ voters finds $\hat{p} = 0.52$ support for a candidate.

Step 1: Compute the estimated SE.

$$\widehat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0.52 \times 0.48}{1000}} \approx 0.016$$

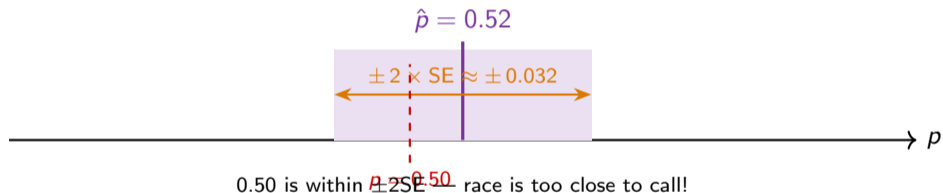
SE in Practice: A Polling Example

Scenario: A poll of $n = 1,000$ voters finds $\hat{p} = 0.52$ support for a candidate.

Step 1: Compute the estimated SE.

$$\widehat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0.52 \times 0.48}{1000}} \approx 0.016$$

Step 2: Interpret. The estimate is $\hat{p} = 0.52 \pm 0.016$.



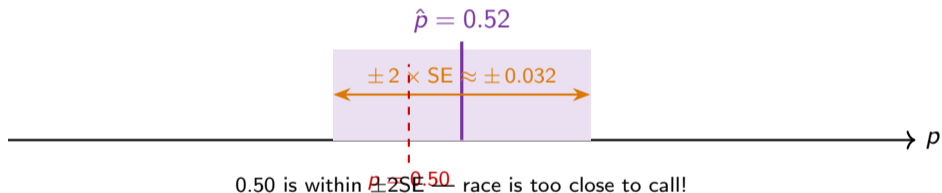
SE in Practice: A Polling Example

Scenario: A poll of $n = 1,000$ voters finds $\hat{p} = 0.52$ support for a candidate.

Step 1: Compute the estimated SE.

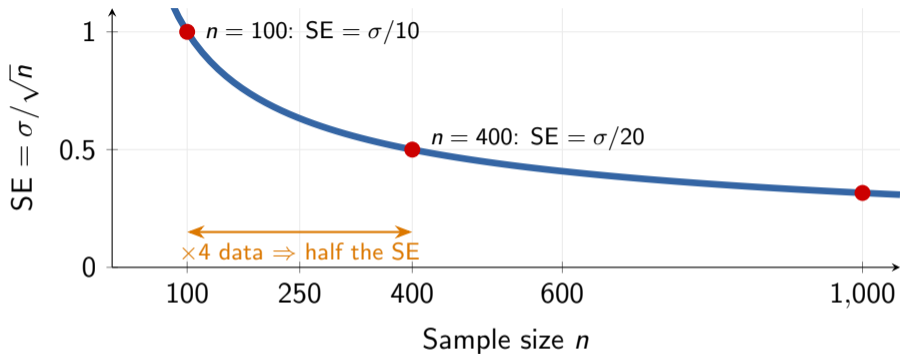
$$\widehat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0.52 \times 0.48}{1000}} \approx 0.016$$

Step 2: Interpret. The estimate is $\hat{p} = 0.52 \pm 0.016$.

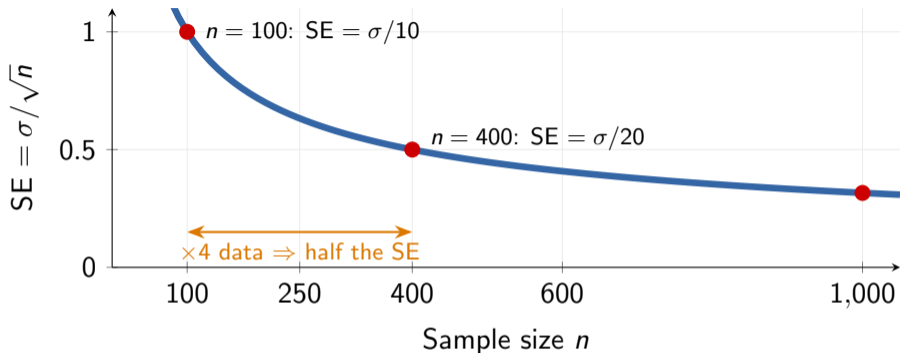


Rule of thumb: always report $\hat{\theta} \pm SE$. An estimate without a standard error is like a measurement without units — it tells you nothing about reliability.

The \sqrt{n} Law: Precision Is Expensive



The \sqrt{n} Law: Precision Is Expensive



To **halve** the SE, you must **quadruple** the sample size.
Going from $n = 100$ to $n = 10,000$ only reduces SE by $10\times$.
This is why sample size planning matters!

SE for Any MLE: The General Formula

From Lecture 4 (Cramér–Rao) and Lecture 5 (MLE asymptotics):

$$\text{SE}(\hat{\theta}_{\text{MLE}}) \approx \frac{1}{\sqrt{n \cdot I(\theta)}}$$

The Fisher information $I(\theta)$ determines the precision of the MLE.

More informative data (larger I) \Rightarrow smaller SE.

SE for Any MLE: The General Formula

From Lecture 4 (Cramér–Rao) and Lecture 5 (MLE asymptotics):

$$\text{SE}(\hat{\theta}_{\text{MLE}}) \approx \frac{1}{\sqrt{n \cdot I(\theta)}}$$

The Fisher information $I(\theta)$ determines the precision of the MLE.

More informative data (larger I) \Rightarrow smaller SE.

Problem: This formula contains the *unknown* θ !

SE for Any MLE: The General Formula

From Lecture 4 (Cramér–Rao) and Lecture 5 (MLE asymptotics):

$$\text{SE}(\hat{\theta}_{\text{MLE}}) \approx \frac{1}{\sqrt{n \cdot I(\theta)}}$$

The Fisher information $I(\theta)$ determines the precision of the MLE.

More informative data (larger I) \Rightarrow smaller SE.

Problem: This formula contains the *unknown* θ !

Solution: Plug-in. Replace θ with $\hat{\theta}$:

$$\widehat{\text{SE}}(\hat{\theta}) = \frac{1}{\sqrt{n \cdot I(\hat{\theta})}}$$

Estimated standard error — this
is what we report in practice.

SE for Any MLE: The General Formula

From Lecture 4 (Cramér–Rao) and Lecture 5 (MLE asymptotics):

$$\text{SE}(\hat{\theta}_{\text{MLE}}) \approx \frac{1}{\sqrt{n \cdot I(\theta)}}$$

The Fisher information $I(\theta)$ determines the precision of the MLE.

More informative data (larger I) \Rightarrow smaller SE.

Problem: This formula contains the *unknown* θ !

Solution: Plug-in. Replace θ with $\hat{\theta}$:

$$\widehat{\text{SE}}(\hat{\theta}) = \frac{1}{\sqrt{n \cdot I(\hat{\theta})}}$$

Estimated standard error — this is what we report in practice.

Known vs Estimated SE

If σ is known:

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Rare in practice.

Use the z-distribution (Normal).

But what *is* the *t*-distribution? And why do we need it?

Known vs Estimated SE

If σ is known:

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Rare in practice.

Use the z-distribution (Normal).

If σ is unknown (typical):

$$\widehat{SE}(\bar{X}) = \frac{S}{\sqrt{n}}$$

Replace σ with $S =$

$$\sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2}.$$

Use the t -distribution.

Known vs Estimated SE

If σ is known:

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Rare in practice.

Use the z-distribution (Normal).

If σ is unknown (typical):

$$\widehat{SE}(\bar{X}) = \frac{S}{\sqrt{n}}$$

Replace σ with $S =$

$$\sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2}.$$

Use the t -distribution.

But what *is* the t -distribution? And why do we need it?

The t -Distribution: What It Is and Why It Exists

The problem: When σ is known, $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ exactly (for Normal data).

But when we **replace** σ **with** S , the ratio is *no longer* Normal — it has **heavier tails**:

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

The t -Distribution: What It Is and Why It Exists

The problem: When σ is known, $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ exactly (for Normal data).

But when we **replace** σ **with** S , the ratio is *no longer* Normal — it has **heavier tails**:

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

Why heavier tails? Sometimes S underestimates σ , making T too large. Sometimes it overestimates, making T too small. This **extra randomness from** S inflates the tails.

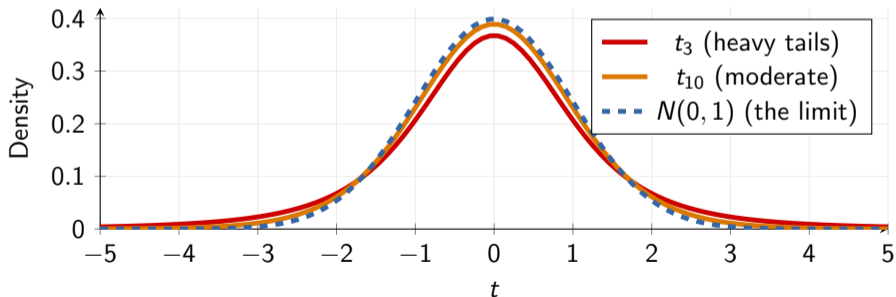
The t -Distribution: What It Is and Why It Exists

The problem: When σ is known, $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ exactly (for Normal data).

But when we **replace** σ **with** S , the ratio is *no longer* Normal — it has **heavier tails**:

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

Why heavier tails? Sometimes S underestimates σ , making T too large. Sometimes it overestimates, making T too small. This **extra randomness from** S inflates the tails.



Fewer observations (n small) $\Rightarrow S$ is noisy $\Rightarrow t_{n-1}$ has fat tails.

More observations (n large) $\Rightarrow S \approx \sigma \Rightarrow t_{n-1} \rightarrow N(0, 1)$. By $n \geq 30$, nearly identical. 27 / 33

The t -Distribution: Key Facts

Shape: Symmetric, bell-shaped, like a Normal but with **heavier tails**.

Parameter: Degrees of freedom $\nu = n - 1$. Smaller $\nu \Rightarrow$ heavier tails.

Mean: 0 (for $\nu > 1$). **Variance:** $\nu/(\nu - 2)$ (for $\nu > 2$), slightly > 1 .

Convergence: As $\nu \rightarrow \infty$, $t_\nu \rightarrow N(0, 1)$. For $\nu \geq 30$, the difference is tiny.

When to use: Normal data, unknown σ . This is the standard setting in practice.

Why “degrees of freedom” = $n - 1$?

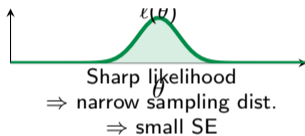
We estimate σ using $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$. The deviations from \bar{X} must sum to 0, so only $n - 1$ of them are “free.” This lost degree of freedom is exactly what makes T not Normal.

Connecting the Dots

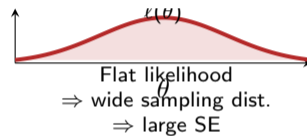
From data to uncertainty

Fisher Information \leftrightarrow Sampling Distribution

High Fisher info

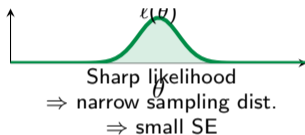


Low Fisher info

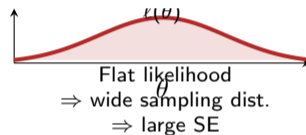


Fisher Information \leftrightarrow Sampling Distribution

High Fisher info



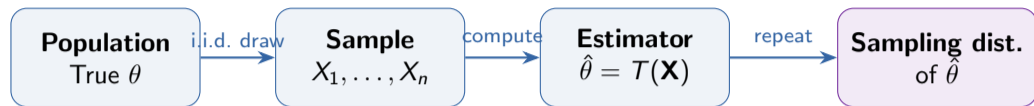
Low Fisher info



The chain: More curvature ($I(\theta) \uparrow$) \Rightarrow Smaller variance ($1/(nI) \downarrow$) \Rightarrow Smaller SE \Rightarrow More precise $\hat{\theta}$

This is why Fisher information is called “information” — it literally measures how much the data tells us about θ .

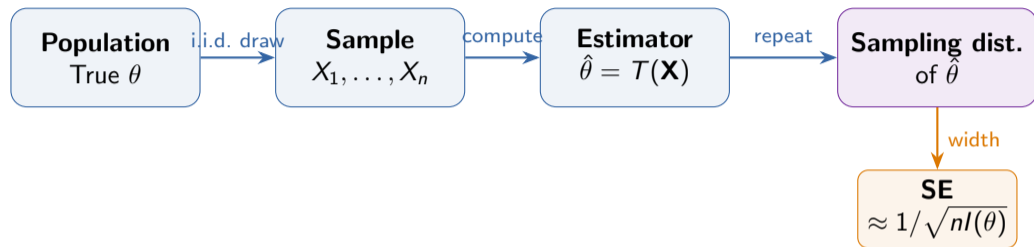
The Full Picture



Two paths to quantify uncertainty: **analytical** (formulas, today) and **computational** (bootstrap, next lecture).

Preview: Since $\hat{\theta} \sim N(\theta, SE^2)$, about **68%** of estimates fall within ± 1 SE, and **95%** within ± 1.96 SE of the truth. That's where confidence intervals come from — next lecture!

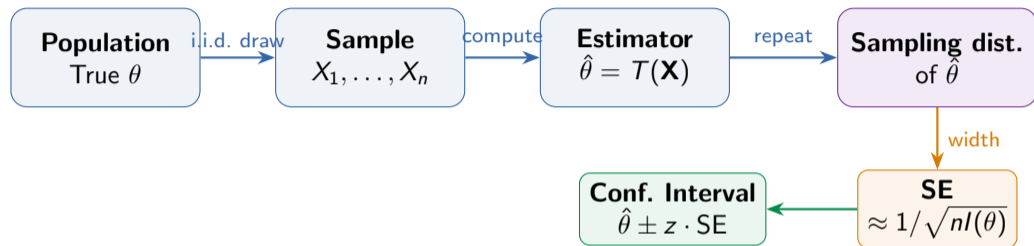
The Full Picture



Two paths to quantify uncertainty: **analytical** (formulas, today) and **computational** (bootstrap, next lecture).

Preview: Since $\hat{\theta} \sim N(\theta, SE^2)$, about **68%** of estimates fall within ± 1 SE, and **95%** within ± 1.96 SE of the truth. That's where confidence intervals come from — next lecture!

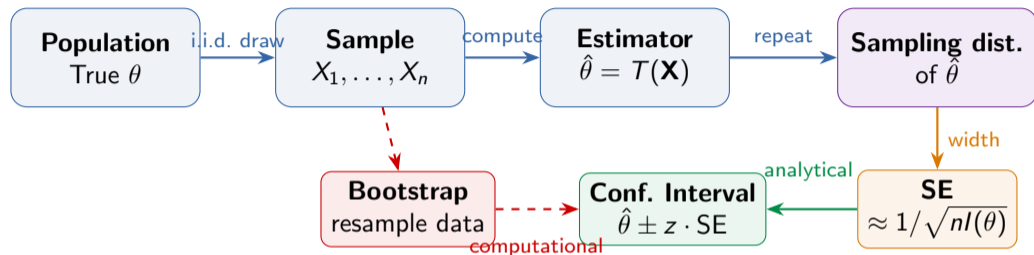
The Full Picture



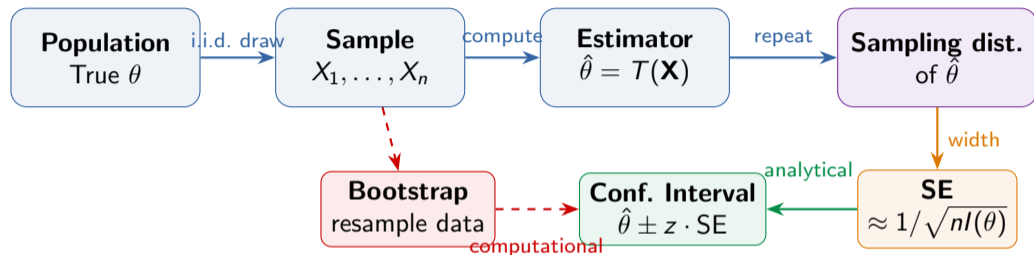
Two paths to quantify uncertainty: **analytical** (formulas, today) and **computational** (bootstrap, next lecture).

Preview: Since $\hat{\theta} \sim N(\theta, \text{SE}^2)$, about **68%** of estimates fall within ± 1 SE, and **95%** within ± 1.96 SE of the truth. That's where confidence intervals come from — next lecture!

The Full Picture



The Full Picture



Two paths to quantify uncertainty: **analytical** (formulas, today) and **computational** (bootstrap, next lecture).

Preview: Since $\hat{\theta} \sim N(\theta, \text{SE}^2)$, about **68%** of estimates fall within ± 1 SE, and **95%** within ± 1.96 SE of the truth. That's where confidence intervals come from — next lecture!

Recommended Visualizations & Resources

Interactive: Seeing Theory — Basic Probability (Brown University)

seeing-theory.brown.edu/basic-probability — drag sliders to change n and watch the sampling distribution form in real time.

Interactive: CLT Visualization (Online Stat Book)

onlinestatbook.com/stat_sim/sampling_dist — choose any population shape and see the sampling distribution for different n .

Video: 3Blue1Brown — But what is the Central Limit Theorem?

Beautiful visual explanation of *why* CLT works, using convolutions. Watch after the lecture.

Video: StatQuest — Standard Error

Clear explanation of the difference between SD and SE, with worked examples.

Reading: Wasserman, “All of Statistics” — Ch. 3 & 5

Convergence, CLT, and the delta method. Concise, mathematically clean, excellent reference.

Questions?

Next: Lecture 8 — Confidence intervals and the bootstrap