Asymptotics and Uncertainty

Isaac Mehlhaff

PLSC 30500, Fall 2025

Last Time

- Repeated sampling motivates **sampling distributions**
- Sample mean: unbiased, variance σ^2/n
- Plug-in estimators: replace population values (θ) with sample analogues ($\hat{\theta}$)
- Bias-variance decomposition: $MSE(\hat{\theta}) = Var(\theta) + \text{Bias}(\theta)^2$; tradeoff illustrated via shrinkage
- Correcting bias in variance estimator via degrees of freedom $(n-1)$

Today: finite-sample distributions often intractable; large-n theory provides workable approximations

Convergence

Problem: In finite samples, exact distributions of estimators are often unknown or complex

Solution: Study behavior as $n \to \infty$ to derive approximations useful for finite (but "large") n

- i.e. examine estimator behavior at asymptotes—values approached in the limit

Convergence

Problem: In finite samples, exact distributions of estimators are often unknown or complex

Solution: Study behavior as $n \to \infty$ to derive approximations useful for finite (but "large") n

- i.e. examine estimator behavior at asymptotes—values approached in the limit

Key questions:

- Does $\hat{\theta}_n$ get close to θ as n increases? (consistency)
- What happens to the variance of $\hat{\theta}_n$ as n increases? (efficiency)
- What is the distribution of $\hat{\theta}_n$? (asymptotic normality)
- How do we quantify uncertainty in finite samples using asymptotic approximations?

Convergence in Probability

A sequence $\{X_n\}$ **converges in probability** to constant c if for every $\epsilon > 0$:

$$\lim_{n\to\infty} P(|X_n - c| > \epsilon) = 0$$

Alternative notations: $X_n \xrightarrow{p} c$, $\text{plim}_{n\to\infty} X_n = c$

Convergence in Probability

A sequence $\{X_n\}$ **converges in probability** to constant c if for every $\epsilon > 0$:

$$\lim_{n \to \infty} P(|X_n - c| > \epsilon) = 0$$

Alternative notations: $X_n \xrightarrow{p} c$, $\text{plim}_{n\to\infty} X_n = c$

Interpretation: as n grows, X_n becomes arbitrarily close to c with probability approaching 1

Example: Sample mean $\overline{X}_n \xrightarrow{p} \mu$

Convergence in Distribution

A sequence $\{X_n\}$ **converges in distribution** to random variable X if:

$$\lim_{n \to \infty} F_n(x) = F(x)$$

$\forall \; x$ where F is continuous, and where F_n is the CDF of X_n

Convergence in Distribution

A sequence $\{X_n\}$ **converges in distribution** to random variable X if:

$$\lim_{n \to \infty} F_n(x) = F(x)$$

$\forall\ x$ where F is continuous, and where F_n is the CDF of X_n

Alternative notations: $X_n \xrightarrow{d} X$, $X_n \Rightarrow X$, or $\mathcal{L}(X_n) \to \mathcal{L}(X)$

Convergence in Distribution

A sequence $\{X_n\}$ **converges in distribution** to random variable X if:

$$\lim_{n \to \infty} F_n(x) = F(x)$$

$\forall\ x$ where F is continuous, and where F_n is the CDF of X_n

Alternative notations: $X_n \xrightarrow{d} X$, $X_n \Rightarrow X$, or $\mathcal{L}(X_n) \to \mathcal{L}(X)$

Interpretation: as n grows, the shape of the probability distribution of X_n gets very similar to the shape of the probability distribution of X

Example: Standardized sample mean $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$

Convergence in distribution is weaker than convergence in probability; latter implies former, but not vice versa

Standard Errors Redux

Recall: the standard deviation of the sampling distribution of an estimator $\hat{\theta}$ is the **standard error:** $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$

Standard Errors Redux

Recall: the standard deviation of the sampling distribution of an estimator $\hat{\theta}$ is the **standard error:** $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$

Example: For sample mean $\overline{X}_n$,

$$se(\overline{X}_n) = \sqrt{Var(\overline{X}_n)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Standard Errors Redux

Recall: the standard deviation of the sampling distribution of an estimator $\hat{\theta}$ is the **standard error:** $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$

Example: For sample mean $\overline{X}_n$,

$$se(\overline{X}_n) = \sqrt{Var(\overline{X}_n)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

In practice, σ^2 unknown, so use $\widehat{se}(\overline{X}_n) = \frac{s}{\sqrt{n}}$ where $s^2 = \frac{1}{n-1} \sum (X_i - \overline{X}_n)^2$

Standard Errors Redux

Recall: the standard deviation of the sampling distribution of an estimator $\hat{\theta}$ is the **standard error:** $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$

Example: For sample mean $\overline{X}_n$,

$$se(\overline{X}_n) = \sqrt{Var(\overline{X}_n)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

In practice, σ^2 unknown, so use $\widehat{se}(\overline{X}_n) = \frac{s}{\sqrt{n}}$ where $s^2 = \frac{1}{n-1}\sum(X_i - \overline{X}_n)^2$

Since $E(\bar{X}) = E(X)$ and $V(\bar{X}) = \frac{V(X)}{n}$, what happens to $\bar{X}$ as $n \to \infty$?

Standard Errors Redux

Recall: the standard deviation of the sampling distribution of an estimator $\hat{\theta}$ is the **standard error:** $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$

Example: For sample mean $\overline{X}_n$,

$$se(\overline{X}_n) = \sqrt{Var(\overline{X}_n)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

In practice, σ^2 unknown, so use $\widehat{se}(\overline{X}_n) = \frac{s}{\sqrt{n}}$ where $s^2 = \frac{1}{n-1}\sum(X_i - \overline{X}_n)^2$

Since $E(\bar{X}) = E(X)$ and $V(\bar{X}) = \frac{V(X)}{n}$, what happens to $\bar{X}$ as $n \to \infty$? It gets arbitrarily close to $E(X)$!

Laws of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be i.i.d. RVs with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Then the **weak** law of large numbers (WLLN) states:

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

Interpretation: sample mean **approaches** population mean as $n \to \infty$ (foundation of plug-in principle)

Laws of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be i.i.d. RVs with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Then the **weak** law of large numbers (WLLN) states:

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

Interpretation: sample mean **approaches** population mean as $n \to \infty$ (foundation of plug-in principle)

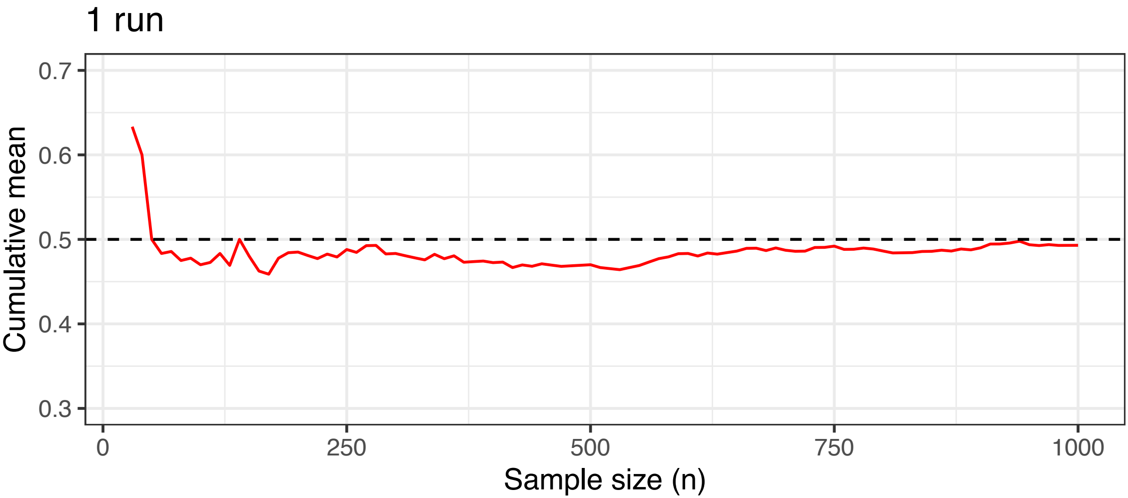
And the **strong** law of large numbers states (SLLN):

$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1$$

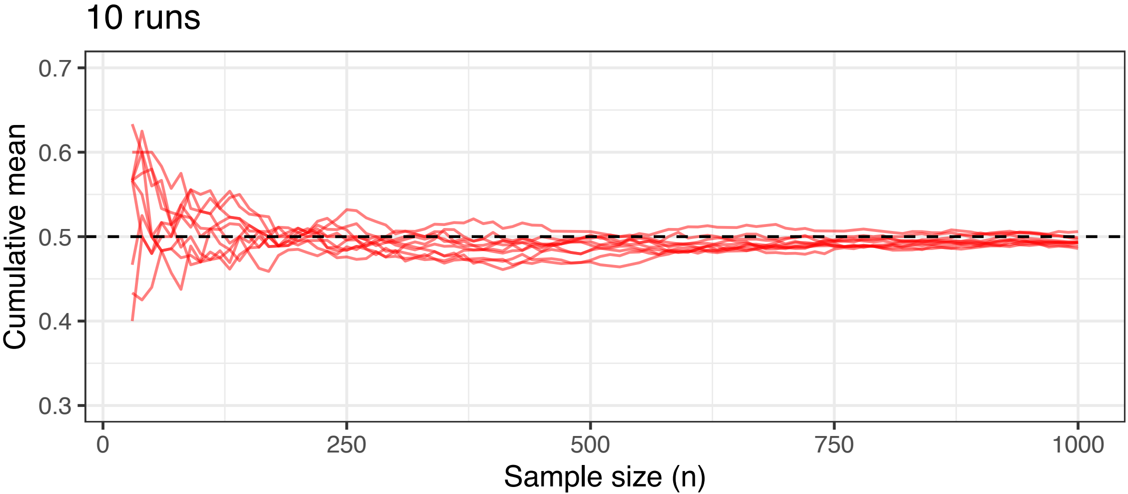
Interpretation: sample mean converges to population mean with certainty as $n \to \infty$ (stronger than WLLN)

We normally rely only on the WLLN

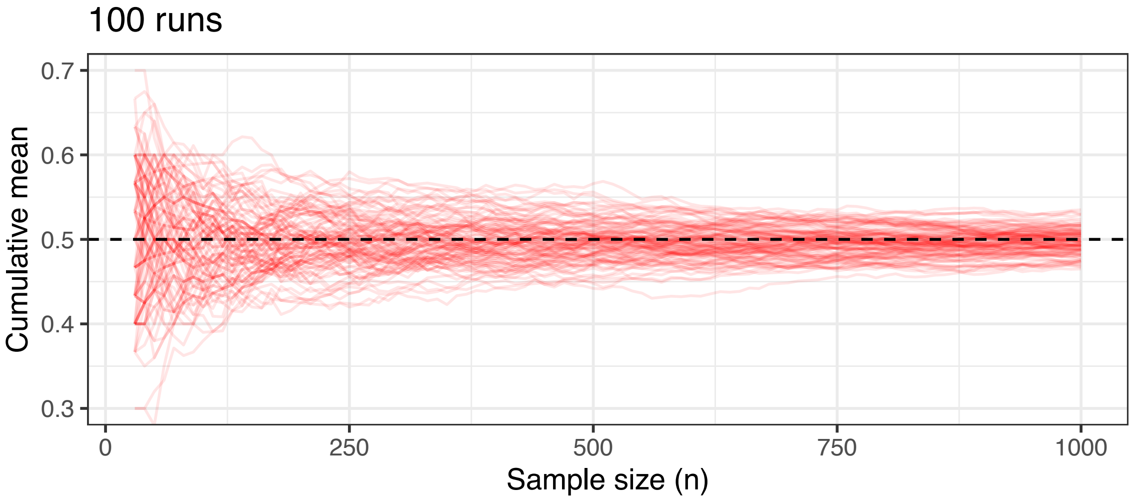
WLLN Example: Binomial $(p = 0.5)$



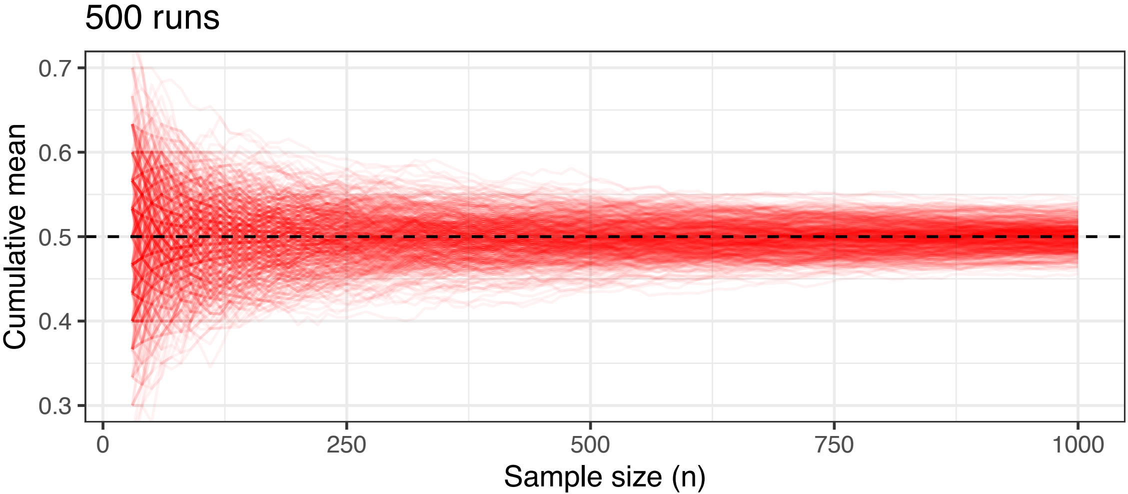
WLLN Example: Binomial $(p = 0.5)$



WLLN Example: Binomial $(p = 0.5)$



WLLN Example: Binomial $(p = 0.5)$



Misinterpretations of the WLLN

Gambler's Fallacy: "if roulette lands black many times, red is 'due' to balance out"

- i.i.d. trials have no memory; each spin is independent, has same probability of red/black
- Correct: across **infinite spins**, proportion of reds converges to p

Misinterpretations of the WLLN

Gambler's Fallacy: "if roulette lands black many times, red is 'due' to balance out"

- i.i.d. trials have no memory; each spin is independent, has same probability of red/black
- Correct: across **infinite spins**, proportion of reds converges to p

Hot Hand Fallacy: "if LeBron James makes several shots in a row, he's more likely to make the next one too"

- **Law of small numbers** bias (Tversky and Kahneman 1971)
  - People expect small samples to reflect population properties; short streaks are too extreme to be random and must indicate some hidden cause
- Correct: across **infinite shots**, proportion of makes converges to p (mean regression)

Misinterpretations of the WLLN

Gambler's Fallacy: "if roulette lands black many times, red is 'due' to balance out"

- i.i.d. trials have no memory; each spin is independent, has same probability of red/black
- Correct: across **infinite spins**, proportion of reds converges to p

Hot Hand Fallacy: "if LeBron James makes several shots in a row, he's more likely to make the next one too"

- **Law of small numbers** bias (Tversky and Kahneman 1971)
  - People expect small samples to reflect population properties; short streaks are too extreme to be random and must indicate some hidden cause
- Correct: across **infinite shots**, proportion of makes converges to p (mean regression)

WLLN describes the **distribution** of $\overline{X}_n$, not probabilities of single realizations

Consistency of Estimators

Following from WLLN, estimator $\hat{\theta}_n$ is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$

To be consistent, must show: $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ and $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

Consistency of Estimators

Following from WLLN, estimator $\hat{\theta}_n$ is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$

To be consistent, must show: $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ and $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

Four estimators of μ:

$\hat{\mu} = \overline{X}_n$:

Consistency of Estimators

Following from WLLN, estimator $\hat{\theta}_n$ is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$

To be consistent, must show: $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ and $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

Four estimators of μ:

$\hat{\mu} = \overline{X}_n$:

- $E(\overline{X}_n) = \mu$ (unbiased)
- $Var(\overline{X}_n) = \frac{\sigma^2}{n} \to 0$ (consistent)

$\hat{\mu} = X_1$ (first observation only):

Consistency of Estimators

Following from WLLN, estimator $\hat{\theta}_n$ is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$

To be consistent, must show: $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ and $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

Four estimators of μ:

$\hat{\mu} = \overline{X}_n$:

- $E(\overline{X}_n) = \mu$ (unbiased)
- $Var(\overline{X}_n) = \frac{\sigma^2}{n} \to 0$ (consistent)

$\hat{\mu} = X_1$ (first observation only):

- $E(X_1) = \mu$ (unbiased)
- $Var(X_1) = \sigma^2 \not\to 0$
  (inconsistent—constant w.r.t. n)

$\hat{\mu} = \overline{X}_n + \frac{1}{n}$:

Consistency of Estimators

Following from WLLN, estimator $\hat{\theta}_n$ is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$

To be consistent, must show: $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ and $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

Four estimators of μ:

$\hat{\mu} = \overline{X}_n$:

- $E(\overline{X}_n) = \mu$ (unbiased)
- $Var(\overline{X}_n) = \frac{\sigma^2}{n} \to 0$ (consistent)

$\hat{\mu} = X_1$ (first observation only):

- $E(X_1) = \mu$ (unbiased)
- $Var(X_1) = \sigma^2 \not\to 0$ (inconsistent—constant w.r.t. n)

$\hat{\mu} = \overline{X}_n + \frac{1}{n}$:

- $E(\overline{X}_n + \frac{1}{n}) = \mu + \frac{1}{n} \to \mu$ (asymptotically unbiased)
- $Var(\overline{X}_n + \frac{1}{n}) = \frac{\sigma^2}{n} \to 0$ (consistent)

$\hat{\mu} = \overline{X}_n + 1$:

Consistency of Estimators

Following from WLLN, estimator $\hat{\theta}_n$ is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$

To be consistent, must show: $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ and $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

Four estimators of μ:

$\hat{\mu} = \overline{X}_n$:

- $E(\overline{X}_n) = \mu$ (unbiased)
- $Var(\overline{X}_n) = \frac{\sigma^2}{n} \to 0$ (consistent)

$\hat{\mu} = X_1$ (first observation only):

- $E(X_1) = \mu$ (unbiased)
- $Var(X_1) = \sigma^2 \not\to 0$ (inconsistent—constant w.r.t. n)

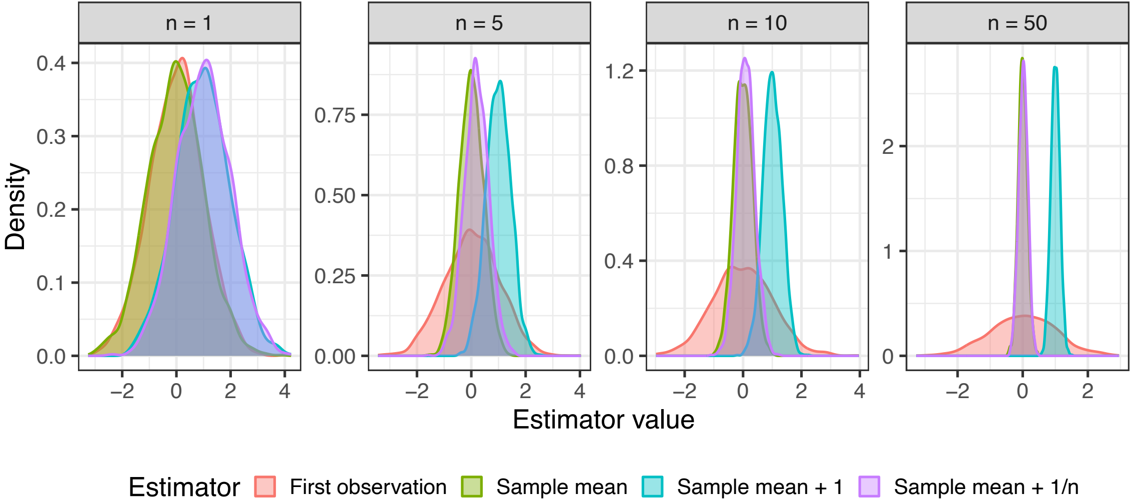
$\hat{\mu} = \overline{X}_n + \frac{1}{n}$:

- $E(\overline{X}_n + \frac{1}{n}) = \mu + \frac{1}{n} \to \mu$ (asymptotically unbiased)
- $Var(\overline{X}_n + \frac{1}{n}) = \frac{\sigma^2}{n} \to 0$ (consistent)

$\hat{\mu} = \overline{X}_n + 1$:

- $E(\overline{X}_n + 1) = \mu + 1 \not\to \mu$ (biased)
- $Var(\overline{X}_n + 1) = \frac{\sigma^2}{n} \to 0$ (still inconsistent due to bias)

Example: Bias and Consistency



Efficiency of Estimators

Among consistent estimators of θ, which is "best"?

Efficiency of Estimators

Among consistent estimators of θ, which is "best"?

$\hat{\theta}_n$ is **efficient** if it achieves the smallest variance among all consistent estimators of θ

Efficiency of Estimators

Among consistent estimators of θ, which is "best"?

$\hat{\theta}_n$ is **efficient** if it achieves the smallest variance among all consistent estimators of θ

Imagine two runners are running toward a finish line

- Consistency: **whether** each runner will eventually reach the finish
- Efficiency: which runner is **faster**

Efficiency implies we are using data optimally, inefficient estimators "waste" information

Relative Efficiency

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two consistent estimators of θ with variances σ_1^2 and σ_2^2, the **relative efficiency** of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is:

$$\mathrm{RE}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\sigma_2^2}{\sigma_1^2}$$

If RE $> 1 \rightarrow \hat{\theta}_1$ is more efficient. If RE $= 0.9, \rightarrow \hat{\theta}_1$ requires $\approx 10\%$ more data than $\hat{\theta}_2$ to achieve the same precision

Relative Efficiency

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two consistent estimators of θ with variances σ_1^2 and σ_2^2, the **relative efficiency** of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is:

$$\mathrm{RE}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\sigma_2^2}{\sigma_1^2}$$

If RE $> 1 \to \hat{\theta}_1$ is more efficient. If RE $= 0.9, \to \hat{\theta}_1$ requires $\approx 10\%$ more data than $\hat{\theta}_2$ to achieve the same precision

Example: For estimating mean μ of normal distribution:

- Sample mean: Var$(\overline{X}_n) = \sigma^2/n$
- Sample median: Var$(\widetilde{X}_n) = \pi\sigma^2/(2n)$ (for normal)
- RE(mean, median) $=$

Relative Efficiency

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two consistent estimators of θ with variances σ_1^2 and σ_2^2, the **relative efficiency** of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is:

$$\text{RE}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\sigma_2^2}{\sigma_1^2}$$

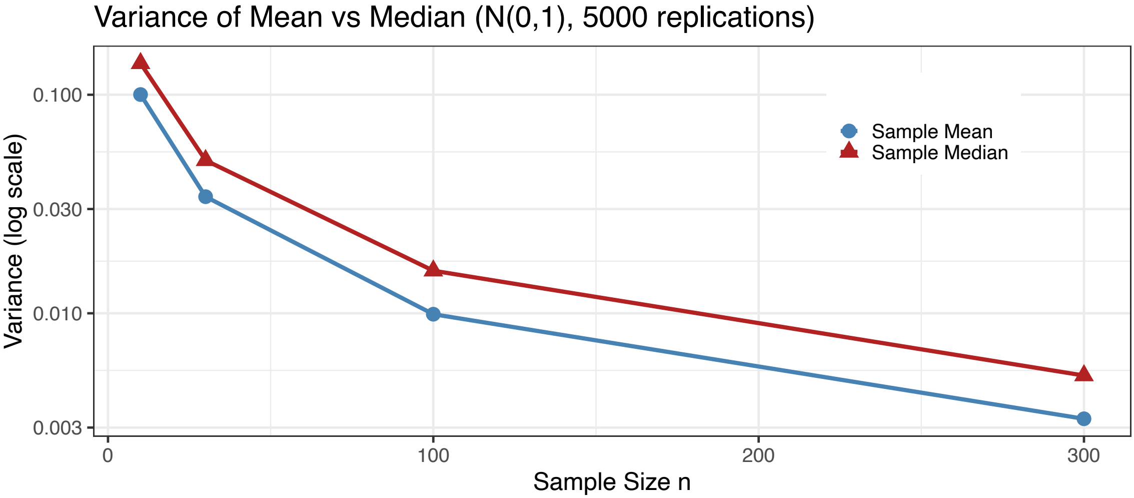
If RE $> 1 \rightarrow \hat{\theta}_1$ is more efficient. If RE $= 0.9, \rightarrow \hat{\theta}_1$ requires $\approx 10\%$ more data than $\hat{\theta}_2$ to achieve the same precision

Example: For estimating mean μ of normal distribution:

- Sample mean: Var$(\overline{X}_n) = \sigma^2/n$
- Sample median: Var$(\widetilde{X}_n) = \pi\sigma^2/(2n)$ (for normal)
- RE(mean, median) $= \frac{\pi\sigma^2/(2n)}{\sigma^2/n} = \frac{\pi}{2} \approx 1.57$

Mean is $\sim 57\%$ more efficient than median for normal data

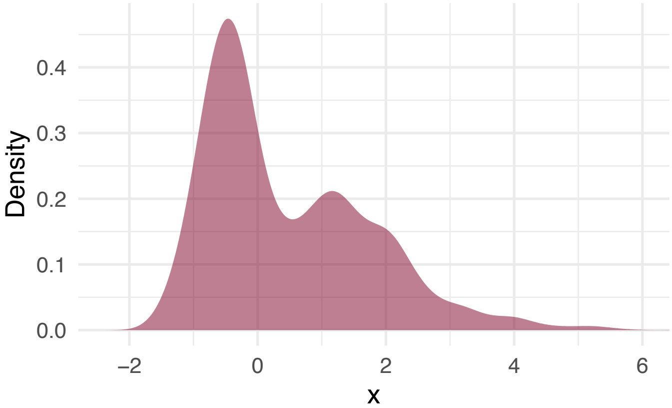
Relative Efficiency: Mean vs. Median



Central Limit Theorem

Suppose the population distribution of X looks like this

What would the sampling distribution of $\bar{X}$ look like?

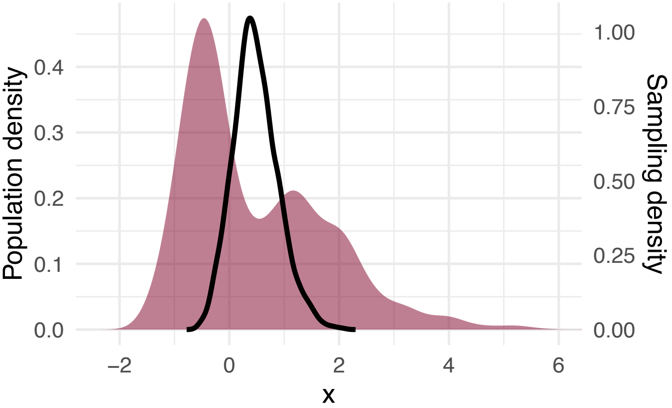


Central Limit Theorem

Suppose the population distribution of X looks like this

What would the sampling distribution of $\bar{X}$ look like?

Very different!



Central Limit Theorem

Already know that for any sample size, $E(\overline{X}_n) = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$. Central limit theorem (CLT) describes the **distribution** of $\overline{X}_n$

If X_i is i.i.d. with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, then:

$$\overline{X}_n \xrightarrow{d} \mathrm{N}(\mu, \frac{\sigma^2}{n})$$

Standardized form: $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathrm{N}(0, 1)$ ($\sqrt{n}$ is a shrinkage correction—stay tuned)

Central Limit Theorem

Already know that for any sample size, $E(\overline{X}_n) = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$. Central limit theorem (CLT) describes the **distribution** of $\overline{X}_n$

If X_i is i.i.d. with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, then:

$$\overline{X}_n \xrightarrow{d} \mathrm{N}(\mu, \frac{\sigma^2}{n})$$

Standardized form: $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathrm{N}(0, 1)$ ($\sqrt{n}$ is a shrinkage correction—stay tuned)

Regardless of distribution of X, $\overline{X}_n$ is approximately normal for large n

- Holds for all plug-in estimators: $\frac{\sqrt{n}(\hat{\theta} - \theta)}{se(\hat{\theta})} \xrightarrow{d} \mathrm{N}(0, 1)$

Central Limit Theorem

Already know that for any sample size, $E(\overline{X}_n) = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$. Central limit theorem (CLT) describes the **distribution** of $\overline{X}_n$

If X_i is i.i.d. with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, then:

$$\overline{X}_n \xrightarrow{d} \mathrm{N}(\mu, \frac{\sigma^2}{n})$$

Standardized form: $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathrm{N}(0,1)$ ($\sqrt{n}$ is a shrinkage correction—stay tuned)

Regardless of distribution of X, $\overline{X}_n$ is approximately normal for large n

- Holds for all plug-in estimators: $\frac{\sqrt{n}(\hat{\theta} - \theta)}{se(\hat{\theta})} \xrightarrow{d} \mathrm{N}(0,1)$

Why? By definition, there are more ways to get a sample mean close to $E(X)$ than far away from it. If i.i.d., $\overline{X}_n$ is therefore more likely to be near $E(X)$ than far away from it

How Large is "Large Enough" for CLT?

How large is "large enough?" Rule of thumb: $n \geq 30$, but depends on skewness and kurtosis of X

How Large is "Large Enough" for CLT?

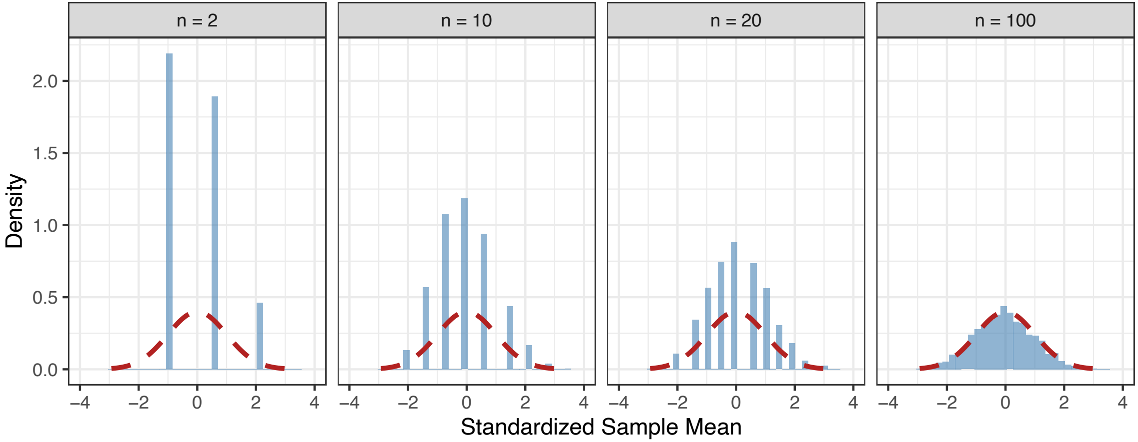
How large is "large enough?" Rule of thumb: $n \geq 30$, but depends on skewness and kurtosis of X

Factors affecting convergence speed:

- **Symmetry:** Symmetric distributions converge faster
- **Light tails:** Bounded or thin-tailed X (e.g., uniform, Bernoulli) converge quickly
- **Heavy tails:** Skewed or thick-tailed (e.g., exponential, Cauchy) require larger n

Doesn't hold when $E(X) = \infty$, $Var(X) = 0$, non-i.i.d. data, mode/median/quantile of a discrete RV

Example: CLT for Binomial Distribution



From Point Estimates to Interval Estimates

Even unbiased/consistent/efficient estimators generate estimates with error

- Sample is finite—if $n = \infty$, $\hat{\theta}_n \xrightarrow{p} \theta$ (population parameter)

Goal: figure out when sample statistic is giving useful info about population parameter

From Point Estimates to Interval Estimates

Even unbiased/consistent/efficient estimators generate estimates with error

- Sample is finite—if $n = \infty$, $\hat{\theta}_n \xrightarrow{p} \theta$ (population parameter)

Goal: figure out when sample statistic is giving useful info about population parameter

Three (main) ways to quantify/communicate uncertainty:

1. **Standard errors:** $se(\hat{\theta}_n) = \frac{\hat{\sigma}}{\sqrt{n}}$
2. **Confidence intervals:** interval centered on $\bar{X}$ that will contain $E(X)$ in e.g. 95% of samples
3. ***p*-values:** probability of observing a value at least as extreme as the one we did (coming up later)

Confidence Intervals

A $100 \cdot (1 - \alpha)\%$ **confidence interval** (CI) for θ is a random interval $[L_n, U_n]$ (function of data) such that:

$$\lim_{n \to \infty} P(L_n \leq \theta \leq U_n) = 1 - \alpha$$

for all possible values of θ, where α is the amount of error we are willing to tolerate (Type I error—risk of false positive)

- **Tradeoff:** higher confidence $\rightarrow$ fewer misses but less informative (wider) intervals

Confidence Intervals

A $100 \cdot (1-\alpha)\%$ **confidence interval** (CI) for θ is a random interval $[L_n, U_n]$ (function of data) such that:

$$\lim_{n \to \infty} P(L_n \leq \theta \leq U_n) = 1 - \alpha$$

for all possible values of θ, where α is the amount of error we are willing to tolerate (Type I error—risk of false positive)

- **Tradeoff:** higher confidence $\rightarrow$ fewer misses but less informative (wider) intervals

CI is therefore a **procedure** (pair of estimators) that contains θ in $100 \cdot (1-\alpha)\%$ of samples

θ is fixed, $[L_n, U_n]$ is random (varies across samples)

In case of $\theta = \mu$: range of values likely to include μ, given the $\bar{X}$ and $se(\bar{X})$ we observe

Confidence Intervals

Recall $Z \sim \mathrm{N}(0,1)$. Then from the CLT:

$$\frac{\sqrt{n}(\hat{\theta} - \theta)}{se(\hat{\theta})} \xrightarrow{d} \mathrm{N}(0,1)$$

Confidence Intervals

Recall $Z \sim \text{N}(0,1)$. Then from the CLT:

$$\frac{\sqrt{n}(\hat{\theta} - \theta)}{se(\hat{\theta})} \xrightarrow{d} \text{N}(0,1)$$

$$\rightarrow \lim_{n \to \infty} P\left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{se(\hat{\theta})} \leq z_{\alpha/2}\right) = 1 - \alpha$$

Confidence Intervals

Recall $Z \sim \mathrm{N}(0,1)$. Then from the CLT:

$$\frac{\sqrt{n}(\hat{\theta}-\theta)}{se(\hat{\theta})} \xrightarrow{d} \mathrm{N}(0,1)$$

$$\rightarrow \lim_{n\to\infty} P\left(-z_{\alpha/2} \leq \frac{\hat{\theta}-\theta}{se(\hat{\theta})} \leq z_{\alpha/2}\right) = 1-\alpha$$

$$\rightarrow \lim_{n\to\infty} P\left(\hat{\theta} - z_{\alpha/2}\cdot se(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\alpha/2}\cdot se(\hat{\theta})\right) = 1-\alpha$$

where $z_{\alpha/2}$ is the $\frac{1-\alpha}{2}$ quantile of $N(0,1)$

Confidence Intervals

Recall $Z \sim \mathrm{N}(0,1)$. Then from the CLT:

$$\frac{\sqrt{n}(\hat{\theta}-\theta)}{se(\hat{\theta})} \xrightarrow{d} \mathrm{N}(0,1)$$

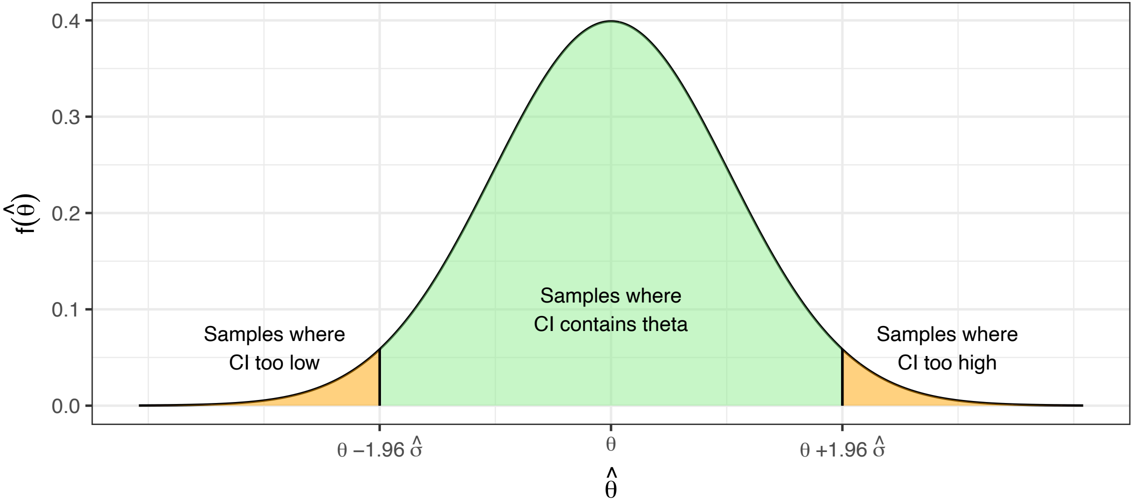
$$\rightarrow \lim_{n\rightarrow\infty} P\left(-z_{\alpha/2} \leq \frac{\hat{\theta}-\theta}{se(\hat{\theta})} \leq z_{\alpha/2}\right) = 1-\alpha$$

$$\rightarrow \lim_{n\rightarrow\infty} P\left(\hat{\theta} - z_{\alpha/2}\cdot se(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\alpha/2}\cdot se(\hat{\theta})\right) = 1-\alpha$$

where $z_{\alpha/2}$ is the $\frac{1-\alpha}{2}$ quantile of $N(0,1)$

95% CI of μ: $\left[\overline{X}_n - 1.96\frac{s}{\sqrt{n}},\ \ \overline{X}_n + 1.96\frac{s}{\sqrt{n}}\right]$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

Confidence Intervals



Calculating Quantiles

Common quantiles: 90% CI: $z_{0.05} \approx 1.645$, 95% CI: $z_{0.025} \approx 1.96$, 99% CI: $z_{0.005} \approx 2.576$

If $Z \sim \mathrm{N}(0, 1)$, then: $P(-1.96 \leq Z \leq 1.96) = \Phi(1.96) - \Phi(-1.96) = 0.975 - 0.025 = 0.95$

Calculating Quantiles

Common quantiles: 90% CI: $z_{0.05} \approx 1.645$, 95% CI: $z_{0.025} \approx 1.96$, 99% CI: $z_{0.005} \approx 2.576$

If $Z \sim \mathrm{N}(0, 1)$, then: $P(-1.96 \leq Z \leq 1.96) = \Phi(1.96) - \Phi(-1.96) = 0.975 - 0.025 = 0.95$

```
qnorm(0.95) # 95th percentile (for 90% CI)

## [1] 1.644854

qnorm(0.975) # 97.5th percentile (for 95% CI)

## [1] 1.959964

qnorm(0.995) # 99.5th percentile (for 99% CI)

## [1] 2.575829

pnorm(1.96) - pnorm(-1.96) # area between -1.96 and 1.96

## [1] 0.9500042
```

Interpreting Confidence Intervals

Suppose we estimate μ with $\bar{X} = 0$ and a 95% CI of $[-5,\ 5]$. How would you interpret this?

Interpreting Confidence Intervals

Suppose we estimate μ with $\bar{X} = 0$ and a 95% CI of $[-5,\ 5]$. How would you interpret this?

Ostensibly reasonable but **wrong** interpretations:

1. "There is a 95% probability that μ is within the interval $[-5,\ 5]$"
    - μ is fixed, so μ is either in the interval or it's not. Probability is either 0 or 1

Interpreting Confidence Intervals

Suppose we estimate μ with $\bar{X} = 0$ and a 95% CI of $[-5,\ 5]$. How would you interpret this?

Ostensibly reasonable but **wrong** interpretations:

1. "There is a 95% probability that μ is within the interval $[-5,\ 5]$"
   - μ is fixed, so μ is either in the interval or it's not. Probability is either 0 or 1
2. "We are 95% confident that $\bar{X}$ is in the interval $[-5,\ 5]$"
   - $\bar{X}$ is fixed for a given sample. We know what it is

Interpreting Confidence Intervals

Suppose we estimate μ with $\bar{X} = 0$ and a 95% CI of $[-5,\ 5]$. How would you interpret this?

Ostensibly reasonable but **wrong** interpretations:

1. "There is a 95% probability that μ is within the interval $[-5,\ 5]$"
   - μ is fixed, so μ is either in the interval or it's not. Probability is either 0 or 1
2. "We are 95% confident that $\bar{X}$ is in the interval $[-5,\ 5]$"
   - $\bar{X}$ is fixed for a given sample. We know what it is
3. "95% of the data are in the interval $[-5,\ 5]$"
   - CI is about a parameter, not the data

Interpreting Confidence Intervals

Suppose we estimate μ with $\bar{X} = 0$ and a 95% CI of $[-5,\ 5]$. How would you interpret this?

Ostensibly reasonable but **wrong** interpretations:

1. "There is a 95% probability that μ is within the interval $[-5,\ 5]$"
   - μ is fixed, so μ is either in the interval or it's not. Probability is either 0 or 1
2. "We are 95% confident that $\bar{X}$ is in the interval $[-5,\ 5]$"
   - $\bar{X}$ is fixed for a given sample. We know what it is
3. "95% of the data are in the interval $[-5,\ 5]$"
   - CI is about a parameter, not the data

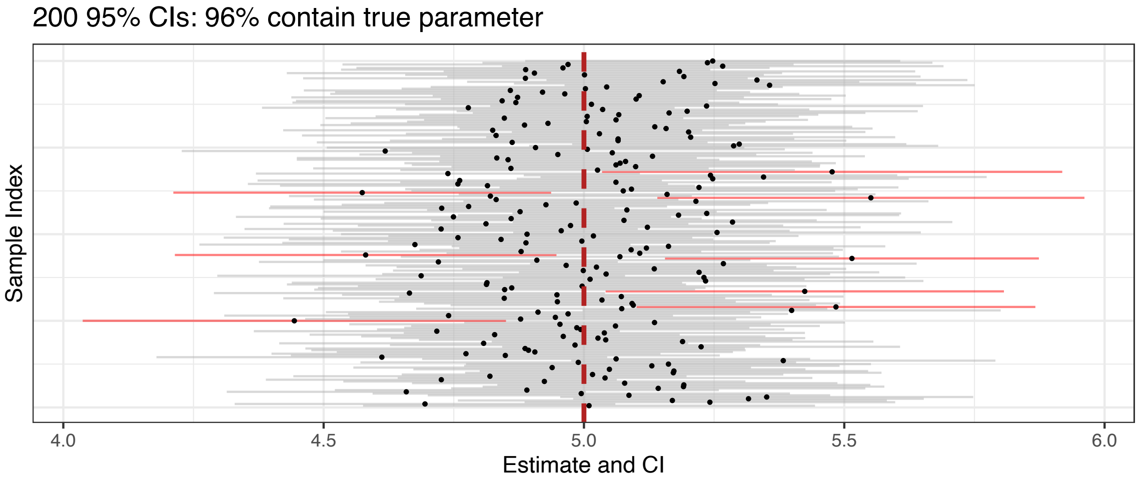
"Across repeated samples of size n from population F, 95% of the confidence intervals of the form $\bar{X} \pm z_{0.025} \cdot se(\bar{X})$ will contain μ"

- Inference depends on sampling distribution, describes probabilities over **many samples**

CI Coverage Simulation: 95% Level

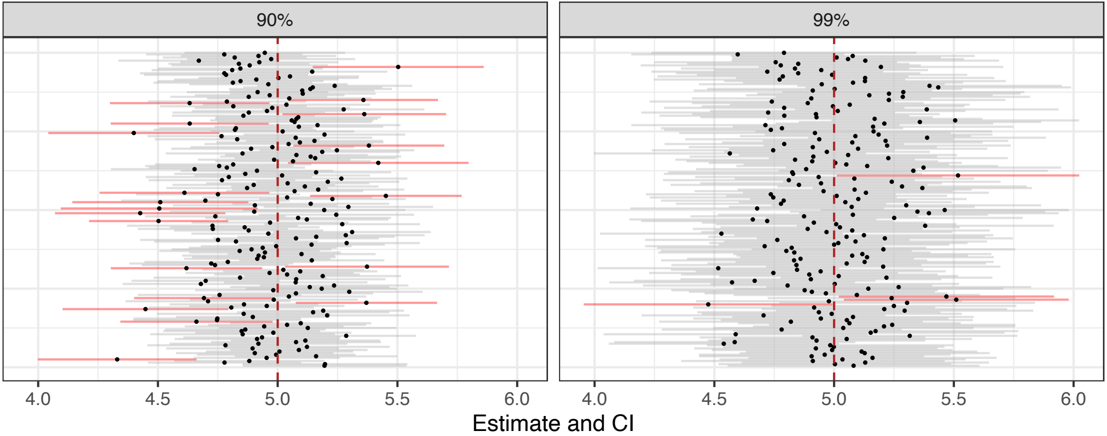
```
set.seed(77)
M <- 200  # number of CIs to visualize
n <- 100  # sample size
mu_true <- 5; sigma_true <- 2  # true parameters

ci_data <- purrr::map_dfr(1:M, function(i) {
    samp <- rnorm(n, mu_true, sigma_true)
    xbar <- mean(samp)
    se <- sd(samp) / sqrt(n)
    ci_lower <- xbar - qnorm(0.975) * se
    ci_upper <- xbar + qnorm(0.975) * se
    contains <- (mu_true >= ci_lower) & (mu_true <= ci_upper)
    tibble(i = i, xbar = xbar, ci_lower = ci_lower, ci_upper = ci_upper,
           contains = contains)
})
coverage <- mean(ci_data$contains)
```

CI Coverage Simulation: 95% Level



CI Coverage Simulation: 90% vs 99%

Coverage: 90% (89.5% empirical), 99% (98% empirical)



CI Width and Sample Size

What happens to CI width as n increases (holding α constant)?

Width of $100 \cdot (1 - \alpha)\%$ CI: $2z_{\alpha/2} se(\hat{\theta}_n) = 2z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$

CI Width and Sample Size

What happens to CI width as n increases (holding α constant)?

Width of $100 \cdot (1-\alpha)\%$ CI: $2z_{\alpha/2} se(\hat{\theta}_n) = 2z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}$

Width therefore decreases at rate $1/\sqrt{n}$

- Same reason standard errors decrease as n grows
- Also why we add $\sqrt{n}$ shrinkage correction in CLT. Otherwise, variance would go to zero (owing to consistency) and we couldn't say anything useful about the distribution

Implications:

- To halve CI width, need $4\times$ sample size
- To reduce width by factor of 10, need $100\times$ sample size
- Diminishing returns: large n gives small improvements

Estimating More Complex Variances

Problem: constructed CI for $\hat{\theta}_n$, but interested in $g(\theta)$

- Or the estimator is very complex/difficult to express analytically, non-i.i.d. sample, etc.

Estimating More Complex Variances

Problem: constructed CI for $\hat{\theta}_n$, but interested in $g(\theta)$

- Or the estimator is very complex/difficult to express analytically, non-i.i.d. sample, etc.

Solution 1: Report $[g(L_n),\ g(U_n)]$ where $[L_n,\ U_n]$ is CI for θ

- Only valid for monotone g; coverage not guaranteed otherwise

Estimating More Complex Variances

Problem: constructed CI for $\hat{\theta}_n$, but interested in $g(\theta)$

- Or the estimator is very complex/difficult to express analytically, non-i.i.d. sample, etc.

Solution 1: Report $[g(L_n),\ g(U_n)]$ where $[L_n,\ U_n]$ is CI for θ

- Only valid for monotone g; coverage not guaranteed otherwise

Solution 2: Take advantage of randomness in DGP, "bootstrap" $se(\hat{\theta}_n)$ (and thus the CI) by resampling m samples of size n **with replacement**, calculate variance of $\hat{\theta}_n^{(m)}$ across m

Estimating More Complex Variances

Problem: constructed CI for $\hat{\theta}_n$, but interested in $g(\theta)$

- Or the estimator is very complex/difficult to express analytically, non-i.i.d. sample, etc.

Solution 1: Report $[g(L_n),\ g(U_n)]$ where $[L_n,\ U_n]$ is CI for θ

- Only valid for monotone g; coverage not guaranteed otherwise

Solution 2: Take advantage of randomness in DGP, "bootstrap" $se(\hat{\theta}_n)$ (and thus the CI) by resampling m samples of size n **with replacement**, calculate variance of $\hat{\theta}_n^{(m)}$ across m

Solution 3 (later courses): By delta method, $\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, [g'(\theta)]^2\sigma^2)$. Therefore, $100 \cdot (1-\alpha)\%$ CI for $g(\theta)$: $\left[g(\hat{\theta}_n) - z_{\alpha/2}|g'(\hat{\theta}_n)|\frac{\hat{\sigma}}{\sqrt{n}}, \quad g(\hat{\theta}_n) + z_{\alpha/2}|g'(\hat{\theta}_n)|\frac{\hat{\sigma}}{\sqrt{n}}\right]$

Bootstrapping

Why does bootstrapping work? It's a **plug-in estimator**

Estimand is $Var(\hat{\theta}_n)$, the sampling variance of $\hat{\theta}_n$

Don't know the population, but can continuously regenerate $\hat{\theta}_n$ to reverse-engineer the sampling distribution

Sample resembles population as $n \to \infty$ and $Var(\hat{\theta}_n)$ converges to the estimand

Usually don't need to bootstrap common estimators like $Var(\bar{X})$, but it's a general, distribution-agnostic solution

Example: Survey Data

Calculate 95% CI for jobs/environment tradeoff variable (`env`) from 2012 CES

```
# population SE (sqrt(Var(X)/n))
pop_std_error <- sqrt(var(data$env, na.rm = TRUE) / sum(!is.na(data$env)))

# take m samples of size n with replacement, store X-bar
m <- 1000; samp_means <- rep(NA, times = m)
for(i in 1:m){
    resamp_data <- data[sample(1:nrow(data), size = nrow(data), replace = TRUE),]
    samp_means[i] <- mean(resamp_data$env, na.rm = TRUE)
}

boot_std_error <- sd(samp_means) # bootstrapped SE
```

Example: Survey Data

```
# sanity check: population and bootstrapped means
mean(data$env, na.rm = TRUE); mean(samp_means)

## [1] 3.195822
## [1] 3.196187

# empirical and bootstrapped CIs should be very close
mean(data$env, na.rm = TRUE) + qnorm(0.975)*pop_std_error*c(-1, 1)

## [1] 3.168248 3.223396

mean(samp_means) + qnorm(0.975)*boot_std_error*c(-1, 1)

## [1] 3.167954 3.224419
```