9. Asymptotics

Fall 2023

Matthew Blackwell

Gov 2002 (Harvard)

Where are we? Where are we going?

- Last time: introducing estimators, looking at finite-sample properties.

Where are we? Where are we going?

- Last time: introducing estimators, looking at finite-sample properties.
- Now: can we say more as sample size grows?

Political canvassing study

POLITICAL SCIENCE

Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman[1*] and Joshua Kalla[2]

Existing research depicts intergroup prejudices as deeply ingrained, requiring intense intervention to lastingly reduce. Here, we show that a single approximately 10-minute conversation encouraging actively taking the perspective of others can markedly reduce prejudice for at least 3 months. We illustrate this potential with a door-to-door canvassing intervention in South Florida targeting antitransgender prejudice. Despite declines in homophobia, transphobia remains pervasive. For the intervention, 56 canvassers went door to door encouraging active perspective-taking with 501 voters at voters' doorsteps. A randomized trial found that these conversations substantially reduced transphobia, with decreases greater than Americans' average decrease in homophobia from 1998 to 2012. These effects persisted for 3 months, and both transgender and nontransgender canvassers were effective. The intervention also increased support for a nondiscrimination law, even after exposing voters to counterarguments.

- Can canvassers change minds about topics like transgender rights?

Credit: Fabrice Florian via Flickr

Political canvassing study

POLITICAL SCIENCE

Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman[1*] and Joshua Kalla[2]

Existing research depicts intergroup prejudices as deeply ingrained, requiring intense intervention to lastingly reduce. Here, we show that a single approximately 10-minute conversation encouraging actively taking the perspective of others can markedly reduce prejudice for at least 3 months. We illustrate this potential with a door-to-door canvassing intervention in South Florida targeting antitransgender prejudice. Despite declines in homophobia, transphobia remains pervasive. For the intervention, 56 canvassers went door to door encouraging active perspective-taking with 501 voters at voters' doorsteps. A randomized trial found that these conversations substantially reduced transphobia, with decreases greater than Americans' average decrease in homophobia from 1998 to 2012. These effects persisted for 3 months, and both transgender and nontransgender canvassers were effective. The intervention also increased support for a nondiscrimination law, even after exposing voters to counterarguments.

- Can canvassers change minds about topics like transgender rights?
- Experimental setting:

Credit: Fabrice Florian via Flickr

Political canvassing study

POLITICAL SCIENCE

Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman[1*] and Joshua Kalla[2]

Existing research depicts intergroup prejudices as deeply ingrained, requiring intense intervention to lastingly reduce. Here, we show that a single approximately 10-minute conversation encouraging actively taking the perspective of others can markedly reduce prejudice for at least 3 months. We illustrate this potential with a door-to-door canvassing intervention in South Florida targeting antitransgender prejudice. Despite declines in homophobia, transphobia remains pervasive. For the intervention, 56 canvassers went door to door encouraging active perspective-taking with 501 voters at voters' doorsteps. A randomized trial found that these conversations substantially reduced transphobia, with decreases greater than Americans' average decrease in homophobia from 1998 to 2012. These effects persisted for 3 months, and both transgender and nontransgender canvassers were effective. The intervention also increased support for a nondiscrimination law, even after exposing voters to counterarguments.

- Can canvassers change minds about topics like transgender rights?
- Experimental setting:
    - Randomly assign canvassers to have a conversation about transgender right or a conversation about recycling.

Credit: Fabrice Florian via Flickr

Political canvassing study

POLITICAL SCIENCE

Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman[1*] and Joshua Kalla[2]

Existing research depicts intergroup prejudices as deeply ingrained, requiring intense intervention to lastingly reduce. Here, we show that a single approximately 10-minute conversation encouraging actively taking the perspective of others can markedly reduce prejudice for at least 3 months. We illustrate this potential with a door-to-door canvassing intervention in South Florida targeting antitransgender prejudice. Despite declines in homophobia, transphobia remains pervasive. For the intervention, 56 canvassers went door to door encouraging active perspective-taking with 501 voters at voters' doorsteps. A randomized trial found that these conversations substantially reduced transphobia, with decreases greater than Americans' average decrease in homophobia from 1998 to 2012. These effects persisted for 3 months, and both transgender and nontransgender canvassers were effective. The intervention also increased support for a nondiscrimination law, even after exposing voters to counterarguments.

- Can canvassers change minds about topics like transgender rights?
- Experimental setting:
    - Randomly assign canvassers to have a conversation about transgender right or a conversation about recycling.
    - Trans rights conversations focused on "perspective taking"

Credit: Fabrice Florian via Flickr

Political canvassing study

POLITICAL SCIENCE

Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman[1*] and Joshua Kalla[2]

Existing research depicts intergroup prejudices as deeply ingrained, requiring intense intervention to lastingly reduce. Here, we show that a single approximately 10-minute conversation encouraging actively taking the perspective of others can markedly reduce prejudice for at least 3 months. We illustrate this potential with a door-to-door canvassing intervention in South Florida targeting antitransgender prejudice. Despite declines in homophobia, transphobia remains pervasive. For the intervention, 56 canvassers went door to door encouraging active perspective-taking with 501 voters at voters' doorsteps. A randomized trial found that these conversations substantially reduced transphobia, with decreases greater than Americans' average decrease in homophobia from 1998 to 2012. These effects persisted for 3 months, and both transgender and nontransgender canvassers were effective. The intervention also increased support for a nondiscrimination law, even after exposing voters to counterarguments.

- Can canvassers change minds about topics like transgender rights?
- Experimental setting:
  - Randomly assign canvassers to have a conversation about transgender right or a conversation about recycling.
  - Trans rights conversations focused on "perspective taking"
- Outcome of interest: support for trans rights policies.

Credit: Fabrice Florian via Flickr

Translating into math

- Outcome: $Y_i \in \{1 \text{ (least supportive)}, 2, 3, 4, 5 \text{ (most supportive)}\}$

Translating into math

- Outcome: $Y_i \in \{1 \text{ (least supportive)}, 2, 3, 4, 5 \text{ (most supportive)}\}$
- Treatment: $D_i \in \{0 \text{ (recycling script)}, 1 \text{ (trans rights script)}\}$

Translating into math

- Outcome: $Y_i \in \{1 \text{ (least supportive)}, 2, 3, 4, 5 \text{ (most supportive)}\}$
- Treatment: $D_i \in \{0 \text{ (recycling script)}, 1 \text{ (trans rights script)}\}$
- Goal is to learn **something** about the joint distribution of (Y_i, D_i).

Translating into math

- Outcome: $Y_i \in \{1 \text{ (least supportive)}, 2, 3, 4, 5 \text{ (most supportive)}\}$
- Treatment: $D_i \in \{0 \text{ (recycling script)}, 1 \text{ (trans rights script)}\}$
- Goal is to learn **something** about the joint distribution of (Y_i, D_i).
- Typical estimand would be the difference in conditional expectations:

$$\tau = \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$$

Translating into math

- Outcome: $Y_i \in \{1 \text{ (least supportive)}, 2, 3, 4, 5 \text{ (most supportive)}\}$
- Treatment: $D_i \in \{0 \text{ (recycling script)}, 1 \text{ (trans rights script)}\}$
- Goal is to learn **something** about the joint distribution of (Y_i, D_i).
- Typical estimand would be the difference in conditional expectations:

$$\tau = \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$$

- Typical plug in estimator would be the difference in sample means:

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i(1 - D_i)}{\sum_{i=1}^n (1 - D_i)}$$

Translating into math

- Outcome: $Y_i \in \{1 \text{ (least supportive)}, 2, 3, 4, 5 \text{ (most supportive)}\}$
- Treatment: $D_i \in \{0 \text{ (recycling script)}, 1 \text{ (trans rights script)}\}$
- Goal is to learn **something** about the joint distribution of (Y_i, D_i).
- Typical estimand would be the difference in conditional expectations:

$$\tau = \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$$

- Typical plug in estimator would be the difference in sample means:

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i (1 - D_i)}{\sum_{i=1}^n (1 - D_i)}$$

- Today: what happens to the distribution of $\widehat{\tau}_n$ as n grows?

1/ Asymptotics

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:

Current knowledge

- For i.i.d. r.v.s, $X_1, \dots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\overline{X}_n$ is **unbiased**, $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\overline{X}_n$ is **unbiased**, $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$
  - Sampling variance is $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathbb{V}[X_i]$

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
    - $\overline{X}_n$ is **unbiased**, $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$
    - Sampling variance is $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathbb{V}[X_i]$
    - None of these rely on a **specific distribution** for X_i!

Current knowledge

- For i.i.d. r.v.s, $X_1, \dots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
    - $\overline{X}_n$ is **unbiased**, $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$
    - Sampling variance is $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathbb{V}[X_i]$
    - None of these rely on a **specific distribution** for X_i!
- Assuming $X_i \sim \mathcal{N}(\mu, \sigma^2)$, we know the exact distribution of $\overline{X}_n$.

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
    - $\overline{X}_n$ is **unbiased**, $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$
    - Sampling variance is $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathbb{V}[X_i]$
    - None of these rely on a **specific distribution** for X_i!
- Assuming $X_i \sim \mathcal{N}(\mu, \sigma^2)$, we know the exact distribution of $\overline{X}_n$.
    - What if the data isn't normal? What is the sampling distribution of $\overline{X}_n$?

Current knowledge

- For i.i.d. r.v.s, $X_1, \dots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
    - $\overline{X}_n$ is **unbiased**, $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$
    - Sampling variance is $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathbb{V}[X_i]$
    - None of these rely on a **specific distribution** for X_i!
- Assuming $X_i \sim \mathcal{N}(\mu, \sigma^2)$, we know the exact distribution of $\overline{X}_n$.
    - What if the data isn't normal? What is the sampling distribution of $\overline{X}_n$?
- **Asymptotics**: approximate the sampling distribution of $\overline{X}_n$ as n gets big.

Sequence of sample means

- What can we say about the sample mean n gets large?

Sequence of sample means

- What can we say about the sample mean n gets large?
- Need to think about sequences of sample means with increasing n:

Sequence of sample means

- What can we say about the sample mean n gets large?
- Need to think about sequences of sample means with increasing n:

$$
\begin{aligned}
\overline{X}_1 &= X_1 \\
\overline{X}_2 &= (1/2) \cdot (X_1 + X_2) \\
\overline{X}_3 &= (1/3) \cdot (X_1 + X_2 + X_3) \\
\overline{X}_4 &= (1/4) \cdot (X_1 + X_2 + X_3 + X_4) \\
\overline{X}_5 &= (1/5) \cdot (X_1 + X_2 + X_3 + X_4 + X_5) \\
&\vdots \\
\overline{X}_n &= (1/n) \cdot (X_1 + X_2 + X_3 + X_4 + X_5 + \cdots + X_n)
\end{aligned}
$$

Sequence of sample means

- What can we say about the sample mean n gets large?
- Need to think about sequences of sample means with increasing n:

$$\begin{aligned}
\overline{X}_1 &= X_1 \\
\overline{X}_2 &= (1/2) \cdot (X_1 + X_2) \\
\overline{X}_3 &= (1/3) \cdot (X_1 + X_2 + X_3) \\
\overline{X}_4 &= (1/4) \cdot (X_1 + X_2 + X_3 + X_4) \\
\overline{X}_5 &= (1/5) \cdot (X_1 + X_2 + X_3 + X_4 + X_5) \\
&\vdots \\
\overline{X}_n &= (1/n) \cdot (X_1 + X_2 + X_3 + X_4 + X_5 + \cdots + X_n)
\end{aligned}$$

- Note: this is a sequence of random variables!

Asymptotics and Limits

- Asymptotic analysis is about making **approximations** to finite sample properties.

Asymptotics and Limits

- Asymptotic analysis is about making **approximations** to finite sample properties.
- Useful to know some properties of deterministic sequences:

Asymptotics and Limits

- Asymptotic analysis is about making **approximations** to finite sample properties.
- Useful to know some properties of deterministic sequences:

Definition

A sequence $\{a_n : n = 1, 2, ...\}$ has the **limit** a written $a_n \to a$ as $n \to \infty$ if for all $\delta > 0$ there is some $n_\delta < \infty$ such that for all $n \geq n_\delta$, $|a_n - a| \leq \delta$.

Asymptotics and Limits

- Asymptotic analysis is about making **approximations** to finite sample properties.
- Useful to know some properties of deterministic sequences:

Definition

A sequence $\{a_n : n = 1, 2, ...\}$ has the **limit** a written $a_n \rightarrow a$ as $n \rightarrow \infty$ if for all $\delta > 0$ there is some $n_\delta < \infty$ such that for all $n \geq n_\delta$, $|a_n - a| \leq \delta$.

- a_n gets closer and closer to a as n gets larger (a_n **converges** to a)

Asymptotics and Limits

- Asymptotic analysis is about making **approximations** to finite sample properties.
- Useful to know some properties of deterministic sequences:

Definition

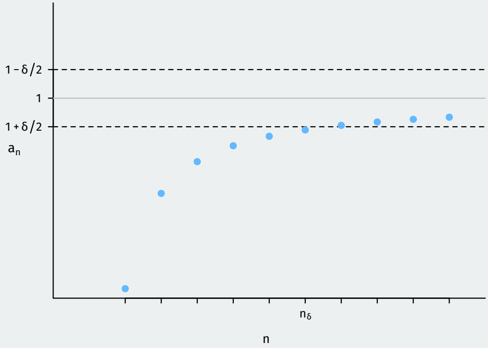
A sequence $\{a_n : n = 1, 2, ...\}$ has the **limit** a written $a_n \to a$ as $n \to \infty$ if for all $\delta > 0$ there is some $n_\delta < \infty$ such that for all $n \geq n_\delta$, $|a_n - a| \leq \delta$.

- a_n gets closer and closer to a as n gets larger (a_n **converges** to a)
- $\{a_n : n = 1, 2, ...\}$ is **bounded** if there is $b < \infty$ such that $|a_n| < b$ for all n.

Limit example: (n-1)/n

Definition

A sequence $\{a_n : n = 1, 2, \ldots\}$ has the **limit** a written $a_n \to a$ as $n \to \infty$ if for all $\delta > 0$ there is some $n_\delta < \infty$ such that for all $n \geq n_\delta$, $|a_n - a| \leq \delta$.



Convergence in Probability

Definition

A sequence of random variables, $\{Z_n : n = 1, 2, ...\}$, is said to **converge in probability** to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

as $n \to \infty$. We write this $Z_n \overset{p}{\to} b$.

Convergence in Probability

Definition

A sequence of random variables, $\{Z_n : n = 1, 2, ...\}$, is said to **converge in probability** to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

as $n \to \infty$. We write this $Z_n \overset{p}{\to} b$.

- Basically: probability that Z_n lies outside any (teeny, tiny) interval around b approaches 0 as $n \to \infty$

Convergence in Probability

Definition

A sequence of random variables, $\{Z_n : n = 1, 2, ...\}$, is said to **converge in probability** to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

as $n \to \infty$. We write this $Z_n \overset{p}{\to} b$.

- Basically: probability that Z_n lies outside any (teeny, tiny) interval around b approaches 0 as $n \to \infty$
- Economists writes $\text{plim}(Z_n) = b$ if $Z_n \overset{p}{\to} b$.

Convergence in Probability

Definition

A sequence of random variables, $\{Z_n : n = 1, 2, ...\}$, is said to **converge in probability** to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

as $n \to \infty$. We write this $Z_n \xrightarrow{p} b$.

- Basically: probability that Z_n lies outside any (teeny, tiny) interval around b approaches 0 as $n \to \infty$
- Economists writes $\text{plim}(Z_n) = b$ if $Z_n \xrightarrow{p} b$.
- An estimator is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$.

Convergence in Probability

Definition

A sequence of random variables, $\{Z_n : n = 1, 2, ...\}$, is said to **converge in probability** to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

as $n \to \infty$. We write this $Z_n \overset{p}{\to} b$.

- Basically: probability that Z_n lies outside any (teeny, tiny) interval around b approaches 0 as $n \to \infty$
- Economists writes $\text{plim}(Z_n) = b$ if $Z_n \overset{p}{\to} b$.
- An estimator is **consistent** if $\hat{\theta}_n \overset{p}{\to} \theta$.
    - Distribution of $\hat{\theta}_n$ collapses on θ as $n \to \infty$.

Convergence in Probability

Definition

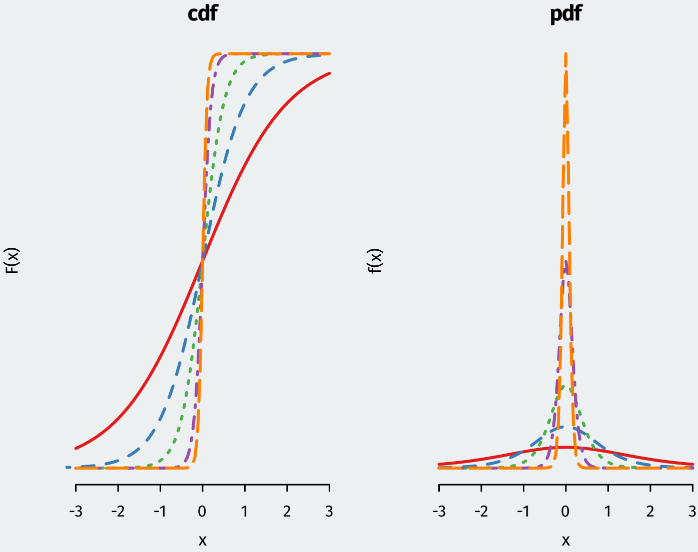
A sequence of random variables, $\{Z_n : n = 1, 2, ...\}$, is said to **converge in probability** to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

as $n \to \infty$. We write this $Z_n \overset{p}{\to} b$.

- Basically: probability that Z_n lies outside any (teeny, tiny) interval around b approaches 0 as $n \to \infty$
- Economists writes $\text{plim}(Z_n) = b$ if $Z_n \overset{p}{\to} b$.
- An estimator is **consistent** if $\hat{\theta}_n \overset{p}{\to} \theta$.
    - Distribution of $\hat{\theta}_n$ collapses on θ as $n \to \infty$.
    - Inconsistent estimator are bad bad bad: more data gives worse answers!

Convergence in probability visually



Law of large numbers

Weak Law of Large Numbers

Let $X_1, \ldots, X_n$ be a an i.i.d. draws from a distribution with mean $\mathbb{E}[|X_i|] < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\overline{X}_n \xrightarrow{p} \mathbb{E}[X_i]$.

Law of large numbers

Weak Law of Large Numbers

Let $X_1, \dots, X_n$ be a an i.i.d. draws from a distribution with mean $\mathbb{E}[|X_i|] < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\overline{X}_n \xrightarrow{p} \mathbb{E}[X_i]$.

- Note: we don't assume finite variance, only finite expectation.

Law of large numbers

Weak Law of Large Numbers

Let $X_1, \ldots, X_n$ be a an i.i.d. draws from a distribution with mean $\mathbb{E}[|X_i|] < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\overline{X}_n \xrightarrow{p} \mathbb{E}[X_i]$.

- Note: we don't assume finite variance, only finite expectation.
- Intuition: The probability of $\overline{X}_n$ being "far away" from μ goes to 0 as n gets big.

Law of large numbers

Weak Law of Large Numbers

Let $X_1, \ldots, X_n$ be a an i.i.d. draws from a distribution with mean $\mathbb{E}[|X_i|] < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\overline{X}_n \xrightarrow{p} \mathbb{E}[X_i]$.

- Note: we don't assume finite variance, only finite expectation.
- Intuition: The probability of $\overline{X}_n$ being "far away" from μ goes to 0 as n gets big.
- Implies general consistency of **plug-in estimators**

Law of large numbers

Weak Law of Large Numbers

Let $X_1, \ldots, X_n$ be a an i.i.d. draws from a distribution with mean $\mathbb{E}[|X_i|] < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\overline{X}_n \overset{p}{\to} \mathbb{E}[X_i]$.

- Note: we don't assume finite variance, only finite expectation.
- Intuition: The probability of $\overline{X}_n$ being "far away" from μ goes to 0 as n gets big.
- Implies general consistency of **plug-in estimators**
  - If $\mathbb{E}[|g(X_i)|] < \infty$, then $\frac{1}{n}\sum_{i=1}^{n} g(X_i) \overset{p}{\to} \mathbb{E}[g(X_i)]$

LLN by simulation in R

- Draw different sample sizes from Exponential distribution with rate 0.5

LLN by simulation in R

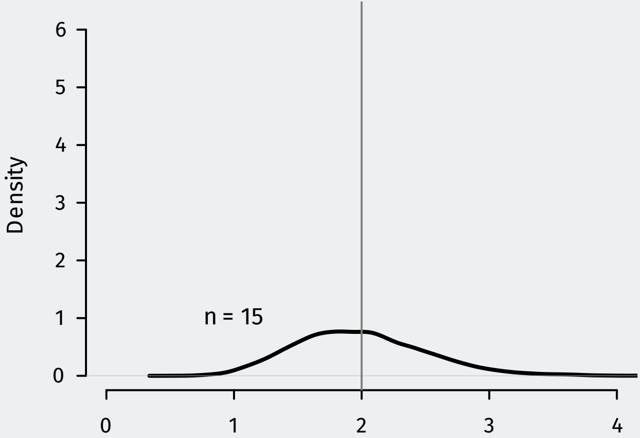
- Draw different sample sizes from Exponential distribution with rate 0.5
- $\rightsquigarrow \mathbb{E}[X_i] = 2$

LLN by simulation in R

- Draw different sample sizes from Exponential distribution with rate 0.5
- $\rightsquigarrow \mathbb{E}[X_i] = 2$

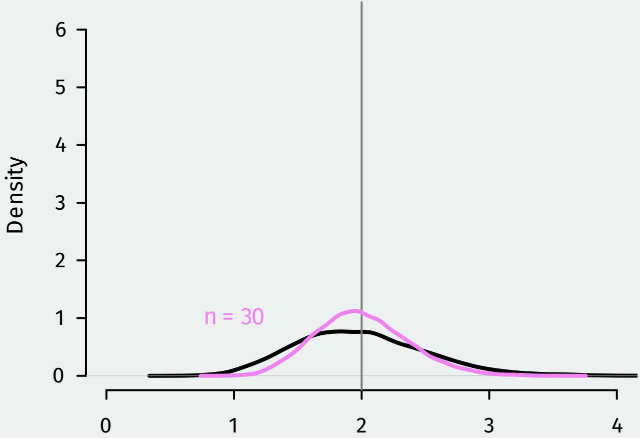
```
nsims <- 10000
holder <- matrix(NA, nrow = nsims, ncol = 6)
for (i in 1:nsims) {
  s5 <- rexp(n = 5, rate = 0.5)
  s15 <- rexp(n = 15, rate = 0.5)
  s30 <- rexp(n = 30, rate = 0.5)
  s100 <- rexp(n = 100, rate = 0.5)
  s1000 <- rexp(n = 1000, rate = 0.5)
  s10000 <- rexp(n = 10000, rate = 0.5)

  holder[i,1] <- mean(s5)
  holder[i,2] <- mean(s15)
  holder[i,3] <- mean(s30)
  holder[i,4] <- mean(s100)
  holder[i,5] <- mean(s1000)
  holder[i,6] <- mean(s10000)
}
```

LLN in action



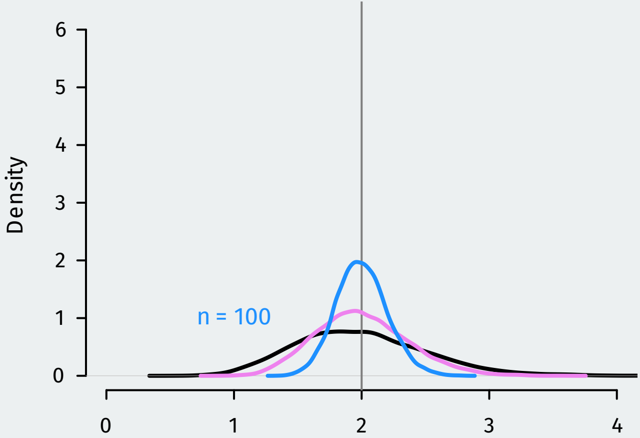
- Distribution of $\overline{X}_{15}$

LLN in action



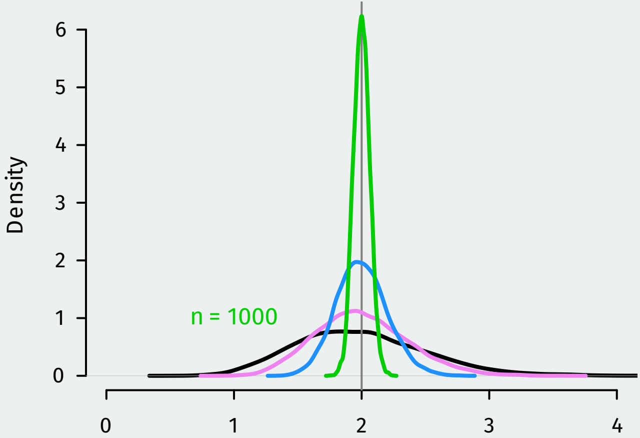
- Distribution of $\overline{X}_{30}$

LLN in action



- Distribution of $\overline{X}_{100}$

LLN in action



- Distribution of $\overline{X}_{1000}$

Chebyshev Inequality

- How can we show convergence in probability? Can verify if we know specific distribution of $\hat{\theta}$.

Chebyshev Inequality

- How can we show convergence in probability? Can verify if we know specific distribution of $\hat{\theta}$.
- But can we say anything for arbitrary distributions?

Chebyshev Inequality

- How can we show convergence in probability? Can verify if we know specific distribution of $\hat{\theta}$.
- But can we say anything for arbitrary distributions?

Chebyshev Inequality

Suppose that X is r.v. for which $\mathbb{V}[X] < \infty$. Then, for every real number $\delta > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta) \leq \frac{\mathbb{V}[X]}{\delta^2}.$$

Chebyshev Inequality

- How can we show convergence in probability? Can verify if we know specific distribution of $\hat{\theta}$.
- But can we say anything for arbitrary distributions?

Chebyshev Inequality

Suppose that X is r.v. for which $\mathbb{V}[X] < \infty$. Then, for every real number $\delta > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta) \leq \frac{\mathbb{V}[X]}{\delta^2}.$$

- Variance places limits on how far an observation can be from its mean.

Proof of Chebyshev

- Let $Z = X - \mathbb{E}[X]$ with density $f_Z(x)$. Probability is just integral over the region:

$$\mathbb{P}\left(|Z| \geq \delta\right) = \int_{|x| \geq \delta} f_Z(x) dx$$

Proof of Chebyshev

- Let $Z = X - \mathbb{E}[X]$ with density $f_Z(x)$. Probability is just integral over the region:

$$\mathbb{P}\left(|Z| \geq \delta\right) = \int_{|x|\geq\delta} f_Z(x)dx$$

- Note that where $|x| \geq \delta$, we have $1 \leq x^2/\delta^2$, so

$$\mathbb{P}\left(|Z| \geq \delta\right) \leq \int_{|x|\geq\delta} \frac{x^2}{\delta^2} f_Z(x)dx \leq \int_{-\infty}^{\infty} \frac{x^2}{\delta^2} f_Z(x)dx = \frac{\mathbb{E}[Z^2]}{\delta^2} = \frac{\mathbb{V}[X]}{\delta^2}$$

Proof of Chebyshev

- Let $Z = X - \mathbb{E}[X]$ with density $f_Z(x)$. Probability is just integral over the region:

$$\mathbb{P}\left(|Z| \geq \delta\right) = \int_{|x| \geq \delta} f_Z(x)dx$$

- Note that where $|x| \geq \delta$, we have $1 \leq x^2/\delta^2$, so

$$\mathbb{P}\left(|Z| \geq \delta\right) \leq \int_{|x| \geq \delta} \frac{x^2}{\delta^2} f_Z(x)dx \leq \int_{-\infty}^{\infty} \frac{x^2}{\delta^2} f_Z(x)dx = \frac{\mathbb{E}[Z^2]}{\delta^2} = \frac{\mathbb{V}[X]}{\delta^2}$$

- Under finite variance, applying this to $|\overline{X}_n - \mu|$ proves the LLN.

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then
   - $X_n + Z_n \xrightarrow{p} a + b$

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then
   - $X_n + Z_n \xrightarrow{p} a + b$
   - $X_n Z_n \xrightarrow{p} ab$

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then
   - $X_n + Z_n \xrightarrow{p} a + b$
   - $X_n Z_n \xrightarrow{p} ab$
   - $X_n / Z_n \xrightarrow{p} a/b$ if $b > 0$

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then
    - $X_n + Z_n \xrightarrow{p} a + b$
    - $X_n Z_n \xrightarrow{p} ab$
    - $X_n / Z_n \xrightarrow{p} a/b$ if $b > 0$

- Thus, by LLN and CMT:

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \overset{p}{\to} c$, then $g(X_n) \overset{p}{\to} g(c)$ for any continuous function g.
2. if $X_n \overset{p}{\to} a$ and $Z_n \overset{p}{\to} b$, then
   - $X_n + Z_n \overset{p}{\to} a + b$
   - $X_n Z_n \overset{p}{\to} ab$
   - $X_n / Z_n \overset{p}{\to} a/b$ if $b > 0$

- Thus, by LLN and CMT:
  - $\left(\overline{X}_n\right)^2 \overset{p}{\to} \mu^2$

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then
    - $X_n + Z_n \xrightarrow{p} a + b$
    - $X_n Z_n \xrightarrow{p} ab$
    - $X_n / Z_n \xrightarrow{p} a/b$ if $b > 0$

- Thus, by LLN and CMT:
    - $\left(\overline{X}_n\right)^2 \xrightarrow{p} \mu^2$
    - $\log(\overline{X}_n) \xrightarrow{p} \log(\mu)$

Difference in means example

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i (1 - D_i)}{\sum_{i=1}^n (1 - D_i)}$$

- What about our difference in means estimator for the transphobia example?

Difference in means example

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i (1 - D_i)}{\sum_{i=1}^n (1 - D_i)}$$

- What about our difference in means estimator for the transphobia example?
- Let's take the sample mean for the treated units:

$$\frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i D_i}{\frac{1}{n}\sum_{i=1}^n D_i} \overset{p}{\to} \frac{\mathbb{E}[Y_i D_i]}{\mathbb{E}[D_i]} = \mathbb{E}[Y_i \mid D_i = 1]$$

Difference in means example

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i(1 - D_i)}{\sum_{i=1}^n (1 - D_i)}$$

- What about our difference in means estimator for the transphobia example?
- Let's take the sample mean for the treated units:

$$\frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i D_i}{\frac{1}{n}\sum_{i=1}^n D_i} \overset{p}{\to} \frac{\mathbb{E}[Y_i D_i]}{\mathbb{E}[D_i]} = \mathbb{E}[Y_i \mid D_i = 1]$$

  - Last step uses iterated expectations and the fundamental bridge.

Difference in means example

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i(1 - D_i)}{\sum_{i=1}^n (1 - D_i)}$$

- What about our difference in means estimator for the transphobia example?
- Let's take the sample mean for the treated units:

$$\frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i D_i}{\frac{1}{n}\sum_{i=1}^n D_i} \overset{p}{\to} \frac{\mathbb{E}[Y_i D_i]}{\mathbb{E}[D_i]} = \mathbb{E}[Y_i \mid D_i = 1]$$

  - Last step uses iterated expectations and the fundamental bridge.
- Same idea for the other sample mean implies,

$$\widehat{\tau}_n \overset{p}{\to} \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \tau$$

Difference in means example

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i(1-D_i)}{\sum_{i=1}^n (1-D_i)}$$

- What about our difference in means estimator for the transphobia example?
- Let's take the sample mean for the treated units:

$$\frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i D_i}{\frac{1}{n}\sum_{i=1}^n D_i} \xrightarrow{p} \frac{\mathbb{E}[Y_i D_i]}{\mathbb{E}[D_i]} = \mathbb{E}[Y_i \mid D_i = 1]$$

  - Last step uses iterated expectations and the fundamental bridge.
- Same idea for the other sample mean implies,

$$\widehat{\tau}_n \xrightarrow{p} \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \tau$$

- Interpretation: Under iid sampling, adding more units gets us closer and closer to the truth.

Unbiased versus consistent

- By Chebyshev, unbiased estimators are consistent if $\mathbb{V}[\hat{\theta}_n] \to 0$.

Unbiased versus consistent

- By Chebyshev, unbiased estimators are consistent if $\mathbb{V}[\widehat{\theta}_n] \to 0$.
- **Unbiased, not consistent**: "first observation" estimator, $\widehat{\theta}^f_n = X_1$.

Unbiased versus consistent

- By Chebyshev, unbiased estimators are consistent if $\mathbb{V}[\hat{\theta}_n] \to 0$.
- **Unbiased, not consistent**: "first observation" estimator, $\hat{\theta}_n^f = X_1$.
    - Unbiased because $\mathbb{E}[\hat{\theta}_n^f] = \mathbb{E}[X_1] = \mu$

Unbiased versus consistent

- By Chebyshev, unbiased estimators are consistent if $\mathbb{V}[\hat{\theta}_n] \to 0$.
- **Unbiased, not consistent**: "first observation" estimator, $\hat{\theta}^f_n = X_1$.
  - Unbiased because $\mathbb{E}[\hat{\theta}^f_n] = \mathbb{E}[X_1] = \mu$
  - Not consistent: $\hat{\theta}^f_n$ is constant in n so its distribution never collapses.

Unbiased versus consistent

- By Chebyshev, unbiased estimators are consistent if $\mathbb{V}[\hat{\theta}_n] \to 0$.
- **Unbiased, not consistent**: "first observation" estimator, $\hat{\theta}_n^f = X_1$.
    - Unbiased because $\mathbb{E}[\hat{\theta}_n^f] = \mathbb{E}[X_1] = \mu$
    - Not consistent: $\hat{\theta}_n^f$ is constant in n so its distribution never collapses.
    - Said differently: the variance of $\hat{\theta}_n^f$ never shrinks.

Unbiased versus consistent

- By Chebyshev, unbiased estimators are consistent if $\mathbb{V}[\hat{\theta}_n] \to 0$.
- **Unbiased, not consistent**: "first observation" estimator, $\hat{\theta}_n^f = X_1$.
  - Unbiased because $\mathbb{E}[\hat{\theta}_n^f] = \mathbb{E}[X_1] = \mu$
  - Not consistent: $\hat{\theta}_n^f$ is constant in n so its distribution never collapses.
  - Said differently: the variance of $\hat{\theta}_n^f$ never shrinks.
- **Consistent, but biased**: sample mean with n replaced by $n - 1$:

$$\frac{1}{n-1}\sum_{i=1}^{n} X_i = \frac{n}{n-1}\overline{X}_n \xrightarrow{p} 1 \times \mu$$

Unbiased versus consistent

- By Chebyshev, unbiased estimators are consistent if $\mathbb{V}[\widehat{\theta}_n] \to 0$.
- **Unbiased, not consistent**: "first observation" estimator, $\widehat{\theta}^f_n = X_1$.
    - Unbiased because $\mathbb{E}[\widehat{\theta}^f_n] = \mathbb{E}[X_1] = \mu$
    - Not consistent: $\widehat{\theta}^f_n$ is constant in n so its distribution never collapses.
    - Said differently: the variance of $\widehat{\theta}^f_n$ never shrinks.
- **Consistent, but biased**: sample mean with n replaced by $n-1$:

$$\frac{1}{n-1}\sum_{i=1}^{n} X_i = \frac{n}{n-1}\overline{X}_n \overset{p}{\to} 1 \times \mu$$

    - Consistent because $n/(n-1) \to 1$ as $n \to \infty$.

Multivariate LLN

- Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$ be a random vectors of length k.

Multivariate LLN

- Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$ be a random vectors of length k.
- Random (iid) sample of n of these k vectors, $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Multivariate LLN

- Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$ be a random vectors of length k.
- Random (iid) sample of n of these k vectors, $\mathbf{X}_1, \ldots, \mathbf{X}_n$.
- Vector sample mean:

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i = \begin{pmatrix} \overline{X}_{n,1} \\ \overline{X}_{n,2} \\ \vdots \\ \overline{X}_{n,k} \end{pmatrix}$$

Multivariate LLN

- Let $\mathbf{X}_i = (X_{i1}, \dots, X_{ik})$ be a random vectors of length k.
- Random (iid) sample of n of these k vectors, $\mathbf{X}_1, \dots, \mathbf{X}_n$.
- Vector sample mean:

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = \begin{pmatrix} \overline{X}_{n,1} \\ \overline{X}_{n,2} \\ \vdots \\ \overline{X}_{n,k} \end{pmatrix}$$

- **Vector WLLN**: if $\mathbb{E}[\|\mathbf{X}\|] < \infty$, then as $n \to \infty$, $\overline{\mathbf{X}}_n \xrightarrow{p} \mathbb{E}[\mathbf{X}]$.

Multivariate LLN

- Let $\mathbf{X}_i = (X_{i1}, \dots, X_{ik})$ be a random vectors of length k.
- Random (iid) sample of n of these k vectors, $\mathbf{X}_1, \dots, \mathbf{X}_n$.
- Vector sample mean:

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = \begin{pmatrix} \overline{X}_{n,1} \\ \overline{X}_{n,2} \\ \vdots \\ \overline{X}_{n,k} \end{pmatrix}$$

- **Vector WLLN**: if $\mathbb{E}[\|\mathbf{X}\|] < \infty$, then as $n \to \infty$, $\overline{\mathbf{X}}_n \xrightarrow{p} \mathbb{E}[\mathbf{X}]$.
    - Converge in probability of a vector is just convergence of each element.

Multivariate LLN

- Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$ be a random vectors of length k.
- Random (iid) sample of n of these k vectors, $\mathbf{X}_1, \ldots, \mathbf{X}_n$.
- Vector sample mean:

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = \begin{pmatrix} \overline{X}_{n,1} \\ \overline{X}_{n,2} \\ \vdots \\ \overline{X}_{n,k} \end{pmatrix}$$

- **Vector WLLN**: if $\mathbb{E}[\|\mathbf{X}\|] < \infty$, then as $n \to \infty$, $\overline{\mathbf{X}}_n \overset{p}{\to} \mathbb{E}[\mathbf{X}]$.
    - Converge in probability of a vector is just convergence of each element.
    - $\mathbb{E}[\|\mathbf{X}\|] < \infty$ is equivalent to $\mathbb{E}[|X_{ij}|] < \infty$ for each $j = 1, \ldots, k$

2/ Central Limit Theorem

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big

Current knowledge

- For i.i.d. r.v.s, $X_1, \dots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.
  - Still no distributional assumptions about X_i!

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.
  - Still no distributional assumptions about X_i!
- Can we say more?

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.
  - Still no distributional assumptions about X_i!
- Can we say more?
  - Can we approximate $\Pr(a < \overline{X}_n < b)$?

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.
  - Still no distributional assumptions about X_i!
- Can we say more?
  - Can we approximate $\Pr(a < \overline{X}_n < b)$?
  - What family of distributions (Binomial, Uniform, Gamma, etc)?

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
    - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
    - $\overline{X}_n$ converges to μ as n gets big
    - Chebyshev provides some bounds on probabilities.
    - Still no distributional assumptions about X_i!
- Can we say more?
    - Can we approximate $\Pr(a < \overline{X}_n < b)$?
    - What family of distributions (Binomial, Uniform, Gamma, etc)?
- Again, need to analyze when n is large.

Convergence in Distribution

Definition

Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $F_n(u)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. $F_W(u)$ if

$$\lim_{n\to\infty} F_n(u) = F_W(u),$$

which we write as $Z_n \xrightarrow{d} W$.

Convergence in Distribution

Definition

Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $F_n(u)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. $F_W(u)$ if

$$\lim_{n \to \infty} F_n(u) = F_W(u),$$

which we write as $Z_n \xrightarrow{d} W$.

- Basically: when n is big, the distribution of Z_n is very similar to the distribution of W

Convergence in Distribution

Definition

Let $Z_1, Z_2, ...,$ be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $F_n(u)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. $F_W(u)$ if

$$\lim_{n\to\infty} F_n(u) = F_W(u),$$

which we write as $Z_n \xrightarrow{d} W$.

- Basically: when n is big, the distribution of Z_n is very similar to the distribution of W
  - Also known as the **asymptotic distribution** or **large-sample distribution**

Convergence in Distribution

Definition

Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $F_n(u)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. $F_W(u)$ if

$$\lim_{n \to \infty} F_n(u) = F_W(u),$$

which we write as $Z_n \xrightarrow{d} W$.

- Basically: when n is big, the distribution of Z_n is very similar to the distribution of W
    - Also known as the **asymptotic distribution** or **large-sample distribution**
- We use c.d.f.s here to avoid messy details with discrete vs continuous.

Convergence in Distribution

Definition

Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $F_n(u)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. $F_W(u)$ if

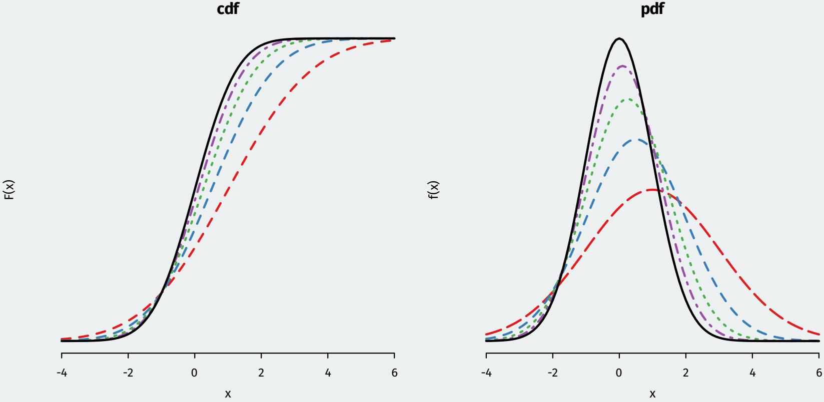
$$\lim_{n \to \infty} F_n(u) = F_W(u),$$

which we write as $Z_n \xrightarrow{d} W$.

- Basically: when n is big, the distribution of Z_n is very similar to the distribution of W
    - Also known as the **asymptotic distribution** or **large-sample distribution**
- We use c.d.f.s here to avoid messy details with discrete vs continuous.
- If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$

Convergence in distribution visualization

$$Z_n \sim N(1/n, 1 + 1/n) \xrightarrow{d} N(0,1)$$



Central Limit Theorem

Central Limit Theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Central Limit Theorem

Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?

Central Limit Theorem

Central Limit Theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?
    - The LLN implied that $\overline{X}_n \xrightarrow{p} \mu$ so $\overline{X}_n \xrightarrow{d} \mu$, which isn't very helpful!

Central Limit Theorem

Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?
  - The LLN implied that $\overline{X}_n \xrightarrow{p} \mu$ so $\overline{X}_n \xrightarrow{d} \mu$, which isn't very helpful!
  - $\sqrt{n}\left(\overline{X}_n - \mu\right)$ is more "stable" since its variance doesn't depend on n

Central Limit Theorem

Central Limit Theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \overset{d}{\to} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?
  - The LLN implied that $\overline{X}_n \overset{p}{\to} \mu$ so $\overline{X}_n \overset{d}{\to} \mu$, which isn't very helpful!
  - $\sqrt{n}\left(\overline{X}_n - \mu\right)$ is more "stable" since its variance doesn't depend on n
- But we can use the result to get an approximation: $\overline{X}_n \overset{a}{\sim} N(\mu, \sigma^2/n)$,

Central Limit Theorem

Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \overset{d}{\to} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?
  - The LLN implied that $\overline{X}_n \overset{p}{\to} \mu$ so $\overline{X}_n \overset{d}{\to} \mu$, which isn't very helpful!
  - $\sqrt{n}\left(\overline{X}_n - \mu\right)$ is more "stable" since its variance doesn't depend on n
- But we can use the result to get an approximation: $\overline{X}_n \overset{a}{\sim} N(\mu, \sigma^2/n)$,
  - $\overset{a}{\sim}$ is "approximately distributed as".

Central Limit Theorem

Central Limit Theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?
    - The LLN implied that $\overline{X}_n \xrightarrow{p} \mu$ so $\overline{X}_n \xrightarrow{d} \mu$, which isn't very helpful!
    - $\sqrt{n}\left(\overline{X}_n - \mu\right)$ is more "stable" since its variance doesn't depend on n
- But we can use the result to get an approximation: $\overline{X}_n \overset{a}{\sim} N(\mu, \sigma^2/n)$,
    - $\overset{a}{\sim}$ is "approximately distributed as".
- No assumptions about the distribution of X_i except finite variance.

Central Limit Theorem

Central Limit Theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

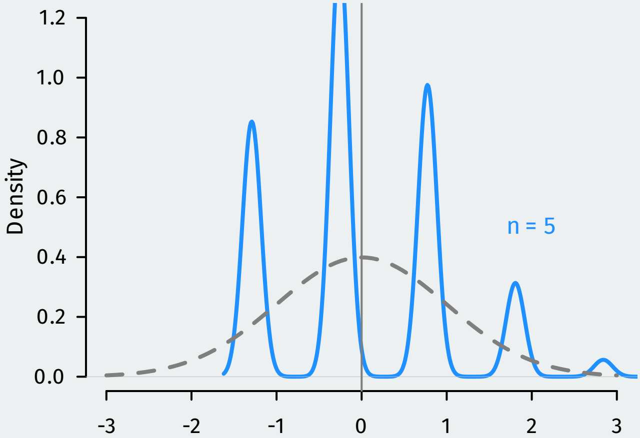
$$\sqrt{n}\left(\overline{X}_n - \mu\right) \overset{d}{\to} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?
  - The LLN implied that $\overline{X}_n \overset{p}{\to} \mu$ so $\overline{X}_n \overset{d}{\to} \mu$, which isn't very helpful!
  - $\sqrt{n}\left(\overline{X}_n - \mu\right)$ is more "stable" since its variance doesn't depend on n
- But we can use the result to get an approximation: $\overline{X}_n \overset{a}{\sim} N(\mu, \sigma^2/n)$,
  - $\overset{a}{\sim}$ is "approximately distributed as".
- No assumptions about the distribution of X_i except finite variance.
- $\rightsquigarrow$ approximations to probability statements about $\overline{X}_n$ when n is big!

CLT by simulation in R

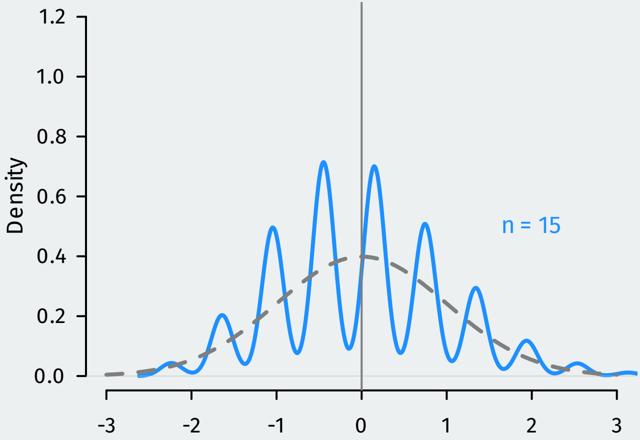
```r
set.seed(02138)
nsims <- 10000
holder2 <- matrix(NA, nrow = nsims, ncol = 6)
for (i in 1:nsims) {
  s5 <- rbinom(n = 5, size = 1, prob = 0.25)
  s15 <- rbinom(n = 15, size = 1, prob = 0.25)
  s30 <- rbinom(n = 30, size = 1, prob = 0.25)
  s100 <- rbinom(n = 100, size = 1, prob = 0.25)
  s1000 <- rbinom(n = 1000, size = 1, prob = 0.25)
  s10000 <- rbinom(n = 10000, size = 1, prob = 0.25)

  holder2[i,1] <- mean(s5)
  holder2[i,2] <- mean(s15)
  holder2[i,3] <- mean(s30)
  holder2[i,4] <- mean(s100)
  holder2[i,5] <- mean(s1000)
  holder2[i,6] <- mean(s10000)
}
```

CLT in action



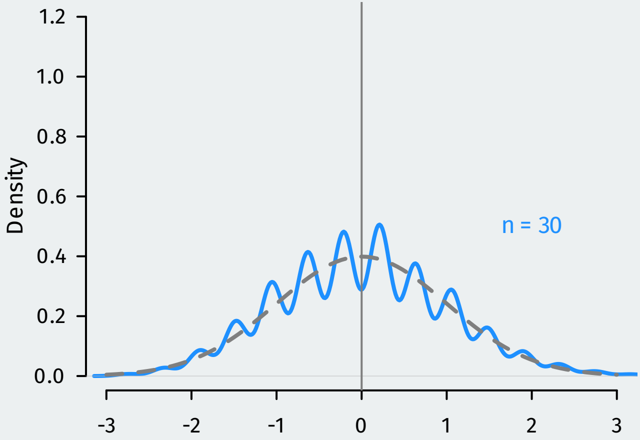
- Distribution of $\frac{\overline{X}_5 - \mu}{\sigma/\sqrt{5}}$

CLT in action



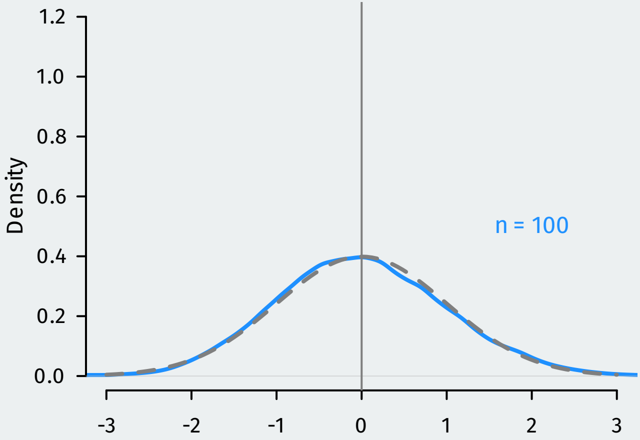
- Distribution of $\frac{\overline{X}_{15} - \mu}{\sigma / \sqrt{15}}$

CLT in action



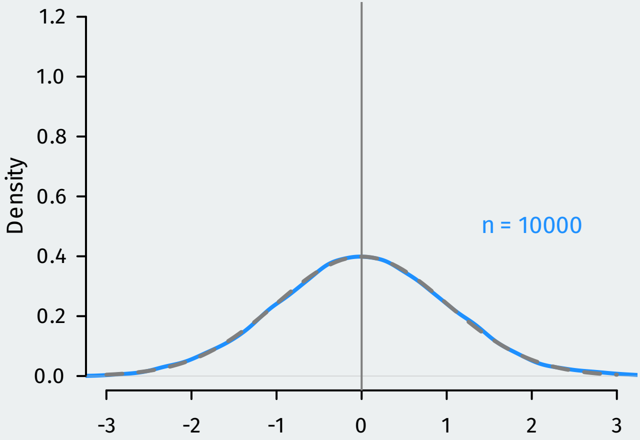
- Distribution of $\frac{\overline{X}_{30}-\mu}{\sigma/\sqrt{30}}$

CLT in action



- Distribution of $\frac{\overline{X}_{100} - \mu}{\sigma / \sqrt{100}}$

CLT in action



- Distribution of $\frac{\overline{X}_{10000}-\mu}{\sigma/\sqrt{10000}}$

CLT for plug-in estimators

- Setting: $X_1, \dots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$

CLT for plug-in estimators

- Setting: $X_1, \ldots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
  - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.

CLT for plug-in estimators

- Setting: $X_1, \dots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
  - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$

CLT for plug-in estimators

- Setting: $X_1, \ldots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
  - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n g(X_i)$
- By the CLT, if $\mathbb{E}[g(X_i)^2] < \infty$ then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

CLT for plug-in estimators

- Setting: $X_1, \ldots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
  - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$
- By the CLT, if $\mathbb{E}[g(X_i)^2] < \infty$ then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

- Any estimator that has this property is called **asymptotically normal**

CLT for plug-in estimators

- Setting: $X_1, \ldots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
    - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$
- By the CLT, if $\mathbb{E}[g(X_i)^2] < \infty$ then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \overset{d}{\to} \mathcal{N}(0, V_\theta)$$

- Any estimator that has this property is called **asymptotically normal**
- V_θ is the variance of this centered/scaled version of the estimator.

CLT for plug-in estimators

- Setting: $X_1, \dots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
  - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n g(X_i)$
- By the CLT, if $\mathbb{E}[g(X_i)^2] < \infty$ then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \overset{d}{\to} \mathcal{N}(0, V_\theta)$$

- Any estimator that has this property is called **asymptotically normal**
- V_θ is the variance of this centered/scaled version of the estimator.
  - The approximate variance of the estimator itself will be $\mathbb{V}[\hat{\theta}_n] \overset{a}{=} V_\theta / n$

CLT for plug-in estimators

- Setting: $X_1, \ldots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
    - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n g(X_i)$
- By the CLT, if $\mathbb{E}[g(X_i)^2] < \infty$ then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

- Any estimator that has this property is called **asymptotically normal**
- V_θ is the variance of this centered/scaled version of the estimator.
    - The approximate variance of the estimator itself will be $\mathbb{V}[\hat{\theta}_n] \overset{a}{=} V_\theta/n$
    - The approximate **standard error** will be $\text{se}[\hat{\theta}_n] = \sqrt{V_\theta/n}$

Why is asymptotic normality important?

- An estimator $\widehat{\theta}_n$ for θ is **asymptotically normal** when

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

Why is asymptotic normality important?

- An estimator $\widehat{\theta}_n$ for θ is **asymptotically normal** when

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

- Allows us to approximate the probability of $\widehat{\theta}_n$ being far away from θ in large samples.

Why is asymptotic normality important?

- An estimator $\widehat{\theta}_n$ for θ is **asymptotically normal** when

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

- Allows us to approximate the probability of $\widehat{\theta}_n$ being far away from θ in large samples.
    - **Warning:** you do not know if you sample is big enough for this to be a good approximation.

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \quad \implies \quad g(Z_n) \xrightarrow{d} g(Z).$$

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \qquad \Longrightarrow \qquad g(Z_n) \xrightarrow{d} g(Z).$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \stackrel{d}{\to} Z \qquad \Longrightarrow \qquad g(Z_n) \stackrel{d}{\to} g(Z).$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number, c

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \qquad \Longrightarrow \qquad g(Z_n) \xrightarrow{d} g(Z).$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number, c
- Slutsky's Theorem gives the following result:

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \quad \implies \quad g(Z_n) \xrightarrow{d} g(Z).$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number, c
- Slutsky's Theorem gives the following result:

    1. $X_n Y_n$ converges in distribution to cX

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \quad \implies \quad g(Z_n) \xrightarrow{d} g(Z).$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number, c
- Slutsky's Theorem gives the following result:

    1. $X_n Y_n$ converges in distribution to cX
    2. $X_n + Y_n$ converges in distribution to $X + c$

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \quad \implies \quad g(Z_n) \xrightarrow{d} g(Z).$$

- Let $X_1, X_2, \dots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \dots$ converge in probability to some number, c
- Slutsky's Theorem gives the following result:
    1. $X_n Y_n$ converges in distribution to cX
    2. $X_n + Y_n$ converges in distribution to $X + c$
    3. X_n / Y_n converges in distribution to X/c if $c \neq 0$

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \qquad \Longrightarrow \qquad g(Z_n) \xrightarrow{d} g(Z).$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number, c
- Slutsky's Theorem gives the following result:

  1. $X_n Y_n$ converges in distribution to cX
  2. $X_n + Y_n$ converges in distribution to $X + c$
  3. X_n / Y_n converges in distribution to X/c if $c \neq 0$

- Extremely useful when trying to figure out what the large-sample distribution of an estimator is.

Variance estimation with plug-in estimators

- Plug-in CLT:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta), \qquad V_\theta = \mathbb{E}[(g(X_i) - \theta)^2]$$

Variance estimation with plug-in estimators

- Plug-in CLT:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta), \qquad V_\theta = \mathbb{E}[(g(X_i) - \theta)^2]$$

- But we don't know V_θ?! Estimate it!

$$\widehat{V_\theta} = \frac{1}{n}\sum_{i=1}^{n}\left(g(X_i) - \hat{\theta}_n\right)^2$$

Variance estimation with plug-in estimators

- Plug-in CLT:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta), \qquad V_\theta = \mathbb{E}[(g(X_i) - \theta)^2]$$

- But we don't know V_θ?! Estimate it!

$$\widehat{V}_\theta = \frac{1}{n}\sum_{i=1}^{n}\left(g(X_i) - \hat{\theta}_n\right)^2$$

- We can show that $\widehat{V}_\theta \xrightarrow{p} V_\theta$ and so by Slutsky:

$$\frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\widehat{V}_\theta}} \xrightarrow{d} \frac{\mathcal{N}(0, V_\theta)}{\sqrt{V_\theta}} \sim \mathcal{N}(0, 1)$$

Multivariate CLT

- Convergence in distribution is the same vector $\mathbf{Z}_n$: convergence of c.d.f.s

Multivariate CLT

- Convergence in distribution is the same vector $\mathbf{Z}_n$: convergence of c.d.f.s
- Allow us to generalize the CLT to random vectors:

Multivariate CLT

- Convergence in distribution is the same vector $\mathbf{Z}_n$: convergence of c.d.f.s
- Allow us to generalize the CLT to random vectors:

Multivariate Central Limit Theorem

If $\mathbf{X}_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$, then as $n \to \infty$,

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_i]$ and $\boldsymbol{\Sigma} = \mathbb{V}[\mathbf{X}_i] = \mathbb{E}\left[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'\right]$.

Multivariate CLT

- Convergence in distribution is the same vector $\mathbf{Z}_n$: convergence of c.d.f.s
- Allow us to generalize the CLT to random vectors:

Multivariate Central Limit Theorem

If $\mathbf{X}_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$, then as $n \to \infty$,

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_i]$ and $\boldsymbol{\Sigma} = \mathbb{V}[\mathbf{X}_i] = \mathbb{E}\left[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'\right]$.

- $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$ is equivalent to $\mathbb{E}[X_{i,j}^2] < \infty$ for all $j = 1, \ldots, k$.

Multivariate CLT

- Convergence in distribution is the same vector $\mathbf{Z}_n$: convergence of c.d.f.s
- Allow us to generalize the CLT to random vectors:

Multivariate Central Limit Theorem

If $\mathbf{X}_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$, then as $n \to \infty$,

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_i]$ and $\boldsymbol{\Sigma} = \mathbb{V}[\mathbf{X}_i] = \mathbb{E}\left[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'\right]$.

- $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$ is equivalent to $\mathbb{E}[X_{i,j}^2] < \infty$ for all $j = 1, \ldots, k$.
  - Basically: multivariate CLT holds if each r.v. in the vector has finite variance.

Multivariate CLT

- Convergence in distribution is the same vector $\mathbf{Z}_n$: convergence of c.d.f.s
- Allow us to generalize the CLT to random vectors:

Multivariate Central Limit Theorem

If $\mathbf{X}_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$, then as $n \to \infty$,

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_i]$ and $\boldsymbol{\Sigma} = \mathbb{V}[\mathbf{X}_i] = \mathbb{E}\left[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'\right]$.

- $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$ is equivalent to $\mathbb{E}[X_{i,j}^2] < \infty$ for all $j = 1, \dots, k$.
  - Basically: multivariate CLT holds if each r.v. in the vector has finite variance.
- Very common for when we're estimating multiple parameters $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}_n$

3/ Confidence intervals

Interval estimation - what and why?

- $\hat{\theta}_n$ is our best guess about θ

Interval estimation - what and why?

- $\widehat{\theta}_n$ is our best guess about θ
- But $\mathbb{P}(\widehat{\theta}_n = \theta) = 0$!

Interval estimation - what and why?

- $\hat{\theta}_n$ is our best guess about θ
- But $\mathbb{P}(\hat{\theta}_n = \theta) = 0$!
- Alternative: produce a range of plausible values instead of one number.

Interval estimation - what and why?

- $\hat{\theta}_n$ is our best guess about θ
- But $\mathbb{P}(\hat{\theta}_n = \theta) = 0$!
- Alternative: produce a range of plausible values instead of one number.
  - Hopefully will increase the chance that we've captured the truth.

Interval estimation - what and why?

- $\hat{\theta}_n$ is our best guess about θ
- But $\mathbb{P}(\hat{\theta}_n = \theta) = 0$!
- Alternative: produce a range of plausible values instead of one number.
    - Hopefully will increase the chance that we've captured the truth.
- We can use the distribution of estimators (CLT!!) to derive these intervals.

What is a confidence interval?

Definition

A $1-\alpha$ **confidence interval** for a population parameter θ is a pair of statistics $L = L(X_1, \dots, X_n)$ and $U = U(X_1, \dots, X_n)$ such that $L < U$ and such that

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

What is a confidence interval?

Definition

A $1-\alpha$ **confidence interval** for a population parameter θ is a pair of statistics $L = L(X_1, \dots, X_n)$ and $U = U(X_1, \dots, X_n)$ such that $L < U$ and such that

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

- Random interval (L, U) will contain the truth $1-\alpha$ of the time.

What is a confidence interval?

Definition

A $1-\alpha$ **confidence interval** for a population parameter θ is a pair of statistics $L = L(X_1, \dots, X_n)$ and $U = U(X_1, \dots, X_n)$ such that $L < U$ and such that

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

- Random interval (L, U) will contain the truth $1-\alpha$ of the time.
  - $\mathbb{P}(L \leq \theta \leq U)$ is the **coverage probability** of the CI

What is a confidence interval?

Definition

A $1-\alpha$ **confidence interval** for a population parameter θ is a pair of statistics $L = L(X_1, \ldots, X_n)$ and $U = U(X_1, \ldots, X_n)$ such that $L < U$ and such that

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

- Random interval (L, U) will contain the truth $1-\alpha$ of the time.
    - $\mathbb{P}(L \leq \theta \leq U)$ is the **coverage probability** of the CI
- Extremely useful way to represent our uncertainty about our estimate.

What is a confidence interval?

Definition

A $1-\alpha$ **confidence interval** for a population parameter θ is a pair of statistics $L = L(X_1, \dots, X_n)$ and $U = U(X_1, \dots, X_n)$ such that $L < U$ and such that

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

- Random interval (L, U) will contain the truth $1-\alpha$ of the time.
  - $\mathbb{P}(L \leq \theta \leq U)$ is the **coverage probability** of the CI
- Extremely useful way to represent our uncertainty about our estimate.
  - Shows a range of plausible values given the data.

What is a confidence interval?

Definition

A $1-\alpha$ **confidence interval** for a population parameter θ is a pair of statistics $L = L(X_1, \dots, X_n)$ and $U = U(X_1, \dots, X_n)$ such that $L < U$ and such that

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

- Random interval (L, U) will contain the truth $1 - \alpha$ of the time.
    - $\mathbb{P}(L \leq \theta \leq U)$ is the **coverage probability** of the CI
- Extremely useful way to represent our uncertainty about our estimate.
    - Shows a range of plausible values given the data.
- A sequence of CIs, $[L_n, U_n]$ are **asymptotically valid** if the coverage probability converges to correct level:

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

Asymptotic confidence intervals

- A sequence of CIs, $[L_n, U_n]$ are **asymptotically valid** if the coverage probability converges to correct level:

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

Asymptotic confidence intervals

- A sequence of CIs, $[L_n, U_n]$ are **asymptotically valid** if the coverage probability converges to correct level:

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

- We can derive such CIs when our estimators are asymptotically normal:

$$\frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \overset{d}{\to} \mathcal{N}(0, 1)$$

Asymptotic confidence intervals

- A sequence of CIs, $[L_n, U_n]$ are **asymptotically valid** if the coverage probability converges to correct level:

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

- We can derive such CIs when our estimators are asymptotically normal:

$$\frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \overset{d}{\to} \mathcal{N}(0, 1)$$

- Then as $n \to \infty$

$$\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta})} \leq 1.96\right) \to 0.95$$

Deriving the 95% CI

$$\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq 1.96\right) \to 0.95$$

Deriving the 95% CI

$$\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq 1.96\right) \to 0.95$$

$$\mathbb{P}\left(-1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \hat{\theta}_n - \theta \leq 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) \to 0.95$$

Deriving the 95% CI

$$\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq 1.96\right) \to 0.95$$

$$\mathbb{P}\left(-1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \hat{\theta}_n - \theta \leq 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) \to 0.95$$

$$\mathbb{P}\left(-\hat{\theta}_n - 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq -\theta \leq -\hat{\theta}_n + 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) \to 0.95$$

Deriving the 95% CI

$$
\begin{aligned}
\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq 1.96\right) &\to 0.95 \\
\mathbb{P}\left(-1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \hat{\theta}_n - \theta \leq 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) &\to 0.95 \\
\mathbb{P}\left(-\hat{\theta}_n - 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq -\theta \leq -\hat{\theta}_n + 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) &\to 0.95 \\
\mathbb{P}\left(\hat{\theta}_n - 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \theta \leq \hat{\theta}_n + 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) &\to 0.95
\end{aligned}
$$

Deriving the 95% CI

$$\begin{aligned}
\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq 1.96\right) &\to 0.95 \\
\mathbb{P}\left(-1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \hat{\theta}_n - \theta \leq 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) &\to 0.95 \\
\mathbb{P}\left(-\hat{\theta}_n - 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq -\theta \leq -\hat{\theta}_n + 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) &\to 0.95 \\
\mathbb{P}\left(\hat{\theta}_n - 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \theta \leq \hat{\theta}_n + 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) &\to 0.95
\end{aligned}$$

- Lower bound: $\hat{\theta}_n - 1.96 \cdot \text{se}(\hat{\theta}_n)$

Deriving the 95% CI

$$\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq 1.96\right) \to 0.95$$

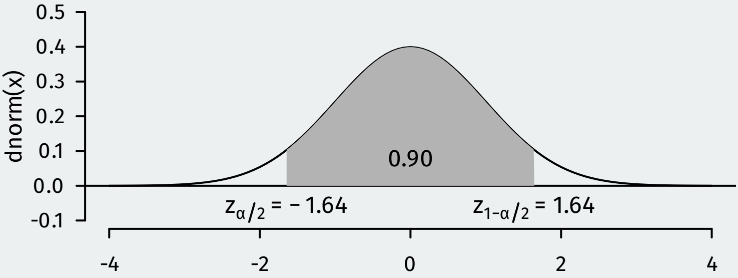
$$\mathbb{P}\left(-1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \hat{\theta}_n - \theta \leq 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) \to 0.95$$

$$\mathbb{P}\left(-\hat{\theta}_n - 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq -\theta \leq -\hat{\theta}_n + 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) \to 0.95$$

$$\mathbb{P}\left(\hat{\theta}_n - 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n) \leq \theta \leq \hat{\theta}_n + 1.96 \cdot \widehat{\text{se}}(\hat{\theta}_n)\right) \to 0.95$$

- Lower bound: $\hat{\theta}_n - 1.96 \cdot \text{se}(\hat{\theta}_n)$
- Upper bound: $\hat{\theta}_n + 1.96 \cdot \text{se}(\hat{\theta}_n)$

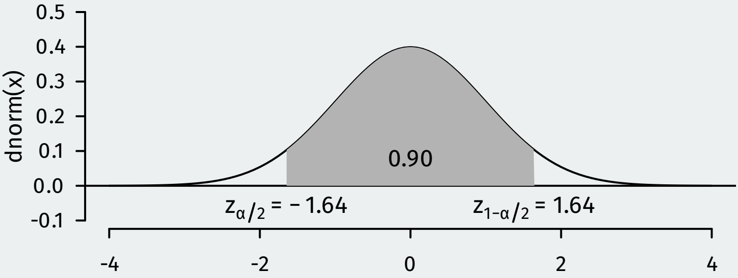
Finding the critical values



$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq z_{1-\alpha/2}\right) \to 1-\alpha \quad \Longrightarrow \quad (1-\alpha) \text{ CI: } \hat{\theta}_n \pm z_{1-\alpha/2} \cdot \widehat{\text{se}}(\hat{\theta}_n)$$

- How do we figure out what $z_{1-\alpha/2}$ will be?

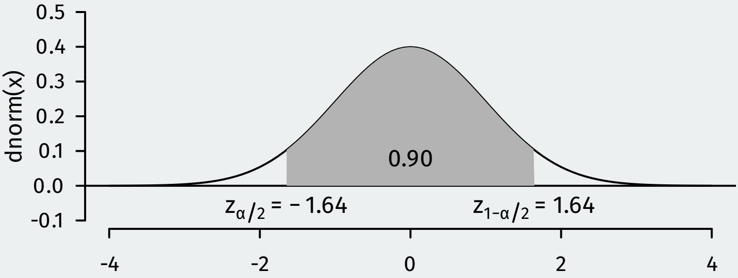
Finding the critical values



$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq z_{1-\alpha/2}\right) \to 1-\alpha \quad \Longrightarrow \quad (1-\alpha) \text{ CI: } \hat{\theta}_n \pm z_{1-\alpha/2} \cdot \widehat{\text{se}}(\hat{\theta}_n)$$

- How do we figure out what $z_{1-\alpha/2}$ will be?
- Intuitively, we want the z values that puts $\alpha/2$ in each of the tails.

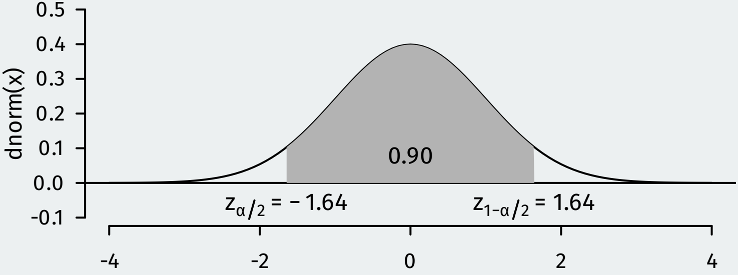
Finding the critical values



$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq z_{1-\alpha/2}\right) \to 1-\alpha \quad \Longrightarrow \quad (1-\alpha) \text{ CI: } \hat{\theta}_n \pm z_{1-\alpha/2} \cdot \widehat{\text{se}}(\hat{\theta}_n)$$

- How do we figure out what $z_{1-\alpha/2}$ will be?
- Intuitively, we want the z values that puts $\alpha/2$ in each of the tails.
  - Because normal is symmetric, we have $z_{\alpha/2} = -z_{1-\alpha/2}$

Finding the critical values



$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{\widehat{\text{se}}(\hat{\theta}_n)} \leq z_{1-\alpha/2}\right) \to 1-\alpha \quad \Longrightarrow \quad (1-\alpha) \text{ CI: } \hat{\theta}_n \pm z_{1-\alpha/2} \cdot \widehat{\text{se}}(\hat{\theta}_n)$$

- How do we figure out what $z_{1-\alpha/2}$ will be?
- Intuitively, we want the z values that puts $\alpha/2$ in each of the tails.
    - Because normal is symmetric, we have $z_{\alpha/2} = -z_{1-\alpha/2}$
    - Use the quantile function: $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ (`qnorm` in R)

CI for social pressure effect

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

CI for social pressure effect

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

```r
neigh_var <- var(social$voted[social$treatment == "Neighbors"])
neigh_n <- 38201
civic_var <- var(social$voted[social$treatment == "Civic Duty"])
civic_n <- 38218

se_diff <- sqrt(neigh_var/neigh_n + civic_var/civic_n)

## c(lower, upper)
c((0.378 - 0.315) - 1.96 * se_diff, (0.378 - 0.315) + 1.96 * se_diff)
```

```
## [1] 0.0563 0.0697
```

Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:

Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:
  - "I calculated a 95% confidence interval of [0.05,0.13], which means that there is a 95% chance that the true difference in means in is that interval."

Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:
    - "I calculated a 95% confidence interval of [0.05,0.13], which means that there is a 95% chance that the true difference in means in is that interval."
    - This is WRONG.

Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:
    - "I calculated a 95% confidence interval of [0.05,0.13], which means that there is a 95% chance that the true difference in means in is that interval."
    - This is WRONG.
- The true value of the population mean, μ, is **fixed**.

Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:
    - "I calculated a 95% confidence interval of [0.05,0.13], which means that there is a 95% chance that the true difference in means in is that interval."
    - This is WRONG.
- The true value of the population mean, μ, is **fixed**.
    - It is either in the interval or it isn't—there's no room for probability at all.

Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:
  - "I calculated a 95% confidence interval of [0.05,0.13], which means that there is a 95% chance that the true difference in means in is that interval."
  - This is WRONG.
- The true value of the population mean, μ, is **fixed**.
  - It is either in the interval or it isn't—there's no room for probability at all.
- The randomness is in the interval: $\overline{X}_n \pm 1.96 S_n / \sqrt{n}$.

Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:
  - "I calculated a 95% confidence interval of [0.05,0.13], which means that there is a 95% chance that the true difference in means in is that interval."
  - This is WRONG.
- The true value of the population mean, μ, is **fixed**.
  - It is either in the interval or it isn't—there's no room for probability at all.
- The randomness is in the interval: $\overline{X}_n \pm 1.96 S_n / \sqrt{n}$.
- Correct interpretation: **across 95% of random samples, the constructed confidence interval will contain the true value.**

Confidence interval simulation

- Draw samples of size 500 (pretty big) from $\mathcal{N}(1, 10)$

Confidence interval simulation

- Draw samples of size 500 (pretty big) from $\mathcal{N}(1, 10)$
- Calculate confidence intervals for the sample mean:

$$\overline{X}_n \pm 1.96 \times \widehat{\text{se}}[\overline{X}_n] \rightsquigarrow \overline{X}_n \pm 1.96 \times S_n/\sqrt{n}$$

Confidence interval simulation

- Draw samples of size 500 (pretty big) from $\mathcal{N}(1, 10)$
- Calculate confidence intervals for the sample mean:

$$\overline{X}_n \pm 1.96 \times \widehat{\text{se}}[\overline{X}_n] \rightsquigarrow \overline{X}_n \pm 1.96 \times S_n/\sqrt{n}$$

```
sims<- 10000
cover <- rep(0, times = sims)
low.bound <- up.bound <- rep(NA, times = sims)
for(i in 1:sims){
    draws <- rnorm(500, mean = 1, sd = sqrt(10))
    low.bound[i] <- mean(draws) - sd(draws) / sqrt(500) * 1.96
    up.bound[i] <- mean(draws) + sd(draws) / sqrt(500) * 1.96
    if (low.bound[i] < 1 & up.bound[i] > 1) {
        cover[i] <- 1
    }
}
mean(cover)
```

```
## [1] 0.95
```

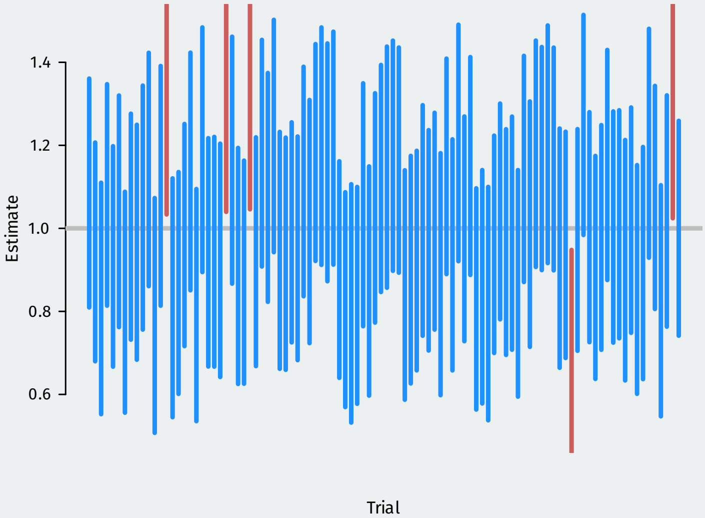
Plotting the CIs

Plotting the CIs

Plotting the CIs

Plotting the CIs

Plotting the CIs



Question

- **Question** What happens to the size of the confidence interval when we increase our confidence, from say 95% to 99%? Do confidence intervals get wider or shorter?

Question

- **Question** What happens to the size of the confidence interval when we increase our confidence, from say 95% to 99%? Do confidence intervals get wider or shorter?
- **Answer** Wider!

Question

- **Question** What happens to the size of the confidence interval when we increase our confidence, from say 95% to 99%? Do confidence intervals get wider or shorter?
- **Answer** Wider!
- Decreases α

Question

- **Question** What happens to the size of the confidence interval when we increase our confidence, from say 95% to 99%? Do confidence intervals get wider or shorter?
- **Answer** Wider!
- Decreases $\alpha \rightsquigarrow$ increases $1 - \alpha/2$

Question

- **Question** What happens to the size of the confidence interval when we increase our confidence, from say 95% to 99%? Do confidence intervals get wider or shorter?
- **Answer** Wider!
- Decreases $\alpha \rightsquigarrow$ increases $1 - \alpha/2 \rightsquigarrow$ increases $z_{\alpha/2}$

4/ Delta method

Delta method

Delta method

If $\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\widehat{\theta}_n) - h(\theta)\right) \xrightarrow{d} \mathcal{N}(0, (h'(\theta))^2 V_\theta).$$

- Why $h()$ continuously differentiable?

Delta method

Delta method

If $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\hat{\theta}_n) - h(\theta)\right) \xrightarrow{d} \mathcal{N}(0, (h'(\theta))^2 V_\theta).$$

- Why $h()$ continuously differentiable?
    - Near θ we can approximate $h()$ with a line where h' is the slope.

Delta method

Delta method

If $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\hat{\theta}_n) - h(\theta)\right) \xrightarrow{d} \mathcal{N}(0, (h'(\theta))^2 V_\theta).$$

- Why $h()$ continuously differentiable?
  - Near θ we can approximate $h()$ with a line where h' is the slope.
  - So $h(\hat{\theta}_n) - h(\theta) \approx h'(\theta)\left(\hat{\theta}_n - \theta\right)$

Delta method

Delta method

If $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\hat{\theta}_n) - h(\theta)\right) \xrightarrow{d} \mathcal{N}(0, (h'(\theta))^2 V_\theta).$$

- Why $h()$ continuously differentiable?
    - Near θ we can approximate $h()$ with a line where h' is the slope.
    - So $h(\hat{\theta}_n) - h(\theta) \approx h'(\theta)\left(\hat{\theta}_n - \theta\right)$
- Examples:

Delta method

Delta method

If $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\hat{\theta}_n) - h(\theta)\right) \xrightarrow{d} \mathcal{N}(0, (h'(\theta))^2 V_\theta).$$

- Why $h()$ continuously differentiable?
  - Near θ we can approximate $h()$ with a line where h' is the slope.
  - So $h(\hat{\theta}_n) - h(\theta) \approx h'(\theta)\left(\hat{\theta}_n - \theta\right)$
- Examples:
  - $\sqrt{n}(\overline{X}_n^2 - \mu^2) \xrightarrow{d} \mathcal{N}(0, (2\mu)^2\sigma^2)$

Delta method

Delta method

If $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \overset{d}{\to} \mathcal{N}(0, V_\theta)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\hat{\theta}_n) - h(\theta)\right) \overset{d}{\to} \mathcal{N}(0, (h'(\theta))^2 V_\theta).$$

- Why $h()$ continuously differentiable?
  - Near θ we can approximate $h()$ with a line where h' is the slope.
  - So $h(\hat{\theta}_n) - h(\theta) \approx h'(\theta)\left(\hat{\theta}_n - \theta\right)$
- Examples:
  - $\sqrt{n}(\overline{X}_n^2 - \mu^2) \overset{d}{\to} \mathcal{N}(0, (2\mu)^2\sigma^2)$
  - $\sqrt{n}(\log(\overline{X}_n) - \log(\mu)) \overset{d}{\to} \mathcal{N}(0, \sigma^2/\mu^2)$

Multivariate Delta Method

- What if we want to know the asymptotic distribution of a function of $\hat{\boldsymbol{\theta}}_n$?

Multivariate Delta Method

- What if we want to know the asymptotic distribution of a function of $\hat{\boldsymbol{\theta}}_n$?
- Let $\mathbf{h}(\boldsymbol{\theta})$ map from $\mathbb{R}^k \to \mathbb{R}^m$ and be continuously differentiable.

Multivariate Delta Method

- What if we want to know the asymptotic distribution of a function of $\hat{\boldsymbol{\theta}}_n$?
- Let $\mathbf{h}(\boldsymbol{\theta})$ map from $\mathbb{R}^k \to \mathbb{R}^m$ and be continuously differentiable.
  - Ex: $\mathbf{h}(\theta_1, \theta_2, \theta_3) = (\theta_2/\theta_1, \theta_3/\theta_1)$, from $\mathbb{R}^3 \to \mathbb{R}^2$

Multivariate Delta Method

- What if we want to know the asymptotic distribution of a function of $\hat{\boldsymbol{\theta}}_n$?
- Let $\mathbf{h}(\boldsymbol{\theta})$ map from $\mathbb{R}^k \to \mathbb{R}^m$ and be continuously differentiable.
  - Ex: $\mathbf{h}(\theta_1, \theta_2, \theta_3) = (\theta_2/\theta_1, \theta_3/\theta_1)$, from $\mathbb{R}^3 \to \mathbb{R}^2$
  - Like univariate case, we need the derivatives arranged in $m \times k$ Jacobian matrix:

$$\mathbf{H}(\boldsymbol{\theta}) = \boldsymbol{\nabla}_{\boldsymbol{\theta}}\mathbf{h}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial h_1}{\partial \theta_1} & \frac{\partial h_1}{\partial \theta_2} & \cdots & \frac{\partial h_1}{\partial \theta_k} \\ \frac{\partial h_2}{\partial \theta_1} & \frac{\partial h_2}{\partial \theta_2} & \cdots & \frac{\partial h_2}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial \theta_1} & \frac{\partial h_m}{\partial \theta_2} & \cdots & \frac{\partial h_m}{\partial \theta_k} \end{pmatrix}$$

Multivariate Delta Method

- What if we want to know the asymptotic distribution of a function of $\hat{\boldsymbol{\theta}}_n$?
- Let $\mathbf{h}(\boldsymbol{\theta})$ map from $\mathbb{R}^k \to \mathbb{R}^m$ and be continuously differentiable.
  - Ex: $\mathbf{h}(\theta_1, \theta_2, \theta_3) = (\theta_2/\theta_1, \theta_3/\theta_1)$, from $\mathbb{R}^3 \to \mathbb{R}^2$
  - Like univariate case, we need the derivatives arranged in $m \times k$ Jacobian matrix:

$$\mathbf{H}(\boldsymbol{\theta}) = \boldsymbol{\nabla}_{\boldsymbol{\theta}}\mathbf{h}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial h_1}{\partial \theta_1} & \frac{\partial h_1}{\partial \theta_2} & \cdots & \frac{\partial h_1}{\partial \theta_k} \\ \frac{\partial h_2}{\partial \theta_1} & \frac{\partial h_2}{\partial \theta_2} & \cdots & \frac{\partial h_2}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial \theta_1} & \frac{\partial h_m}{\partial \theta_2} & \cdots & \frac{\partial h_m}{\partial \theta_k} \end{pmatrix}$$

- Multivariate delta method: if $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma})$, then

$$\sqrt{n}\left(\mathbf{h}(\hat{\boldsymbol{\theta}}_n) - \mathbf{h}(\boldsymbol{\theta})\right) \xrightarrow{d} \mathcal{N}(0, \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\Sigma}\mathbf{H}(\boldsymbol{\theta})')$$

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
    - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
    - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
    - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^\lambda)$

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
    - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
    - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^{\lambda})$
    - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
    - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
    - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^{\lambda})$
    - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")
    - If $n^{-\lambda} a_n$ is bounded, we write $a_n = O(n^{\lambda})$

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
    - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
    - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^\lambda)$
    - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")
    - If $n^{-\lambda} a_n$ is bounded, we write $a_n = O(n^\lambda)$
- Stochastic order notation for random sequence, Z_n

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
    - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
    - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^\lambda)$
    - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")
    - If $n^{-\lambda} a_n$ is bounded, we write $a_n = O(n^\lambda)$
- Stochastic order notation for random sequence, Z_n
    - If $Z_n \overset{p}{\to} 0$, we write $Z_n = o_p(1)$ ("little-oh-p-one").

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
    - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
    - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^{\lambda})$
    - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")
    - If $n^{-\lambda} a_n$ is bounded, we write $a_n = O(n^{\lambda})$
- Stochastic order notation for random sequence, Z_n
    - If $Z_n \overset{p}{\to} 0$, we write $Z_n = o_p(1)$ ("little-oh-p-one").
    - For any consistent estimator, we have $\hat{\theta}_n = \theta + o_p(1)$

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
  - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
  - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^{\lambda})$
  - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")
  - If $n^{-\lambda} a_n$ is bounded, we write $a_n = O(n^{\lambda})$
- Stochastic order notation for random sequence, Z_n
  - If $Z_n \overset{p}{\to} 0$, we write $Z_n = o_p(1)$ ("little-oh-p-one").
  - For any consistent estimator, we have $\hat{\theta}_n = \theta + o_p(1)$
  - If $a_n^{-1} Z_n \overset{p}{\to} 0$, we write $Z_n = o_p(a_n)$

Bounded in probability

Definition

A random sequence Z_n is **bounded in probability**, written $Z_n = O_p(1)$ ("big-oh-p-one") for all $\delta > 0$ there exists a M_δ and n_δ, such that for $n \geq n_\delta$,

$$\mathbb{P}(|Z_n| > M_\delta) < \delta$$

- $Z_n = o_p(1)$ implies $Z_n = O_p(1)$ but not the reverse.

Bounded in probability

Definition

A random sequence Z_n is **bounded in probability**, written $Z_n = O_p(1)$ ("big-oh-p-one") for all $\delta > 0$ there exists a M_δ and n_δ, such that for $n \geq n_\delta$,

$$\mathbb{P}(|Z_n| > M_\delta) < \delta$$

- $Z_n = o_p(1)$ implies $Z_n = O_p(1)$ but not the reverse.
- If Z_n converges in distribution, it is $O_p(1)$, so if the CLT applies we have:

$$\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$$

Bounded in probability

Definition

A random sequence Z_n is **bounded in probability**, written $Z_n = O_p(1)$ ("big-oh-p-one") for all $\delta > 0$ there exists a M_δ and n_δ, such that for $n \geq n_\delta$,

$$\mathbb{P}(|Z_n| > M_\delta) < \delta$$

- $Z_n = o_p(1)$ implies $Z_n = O_p(1)$ but not the reverse.
- If Z_n converges in distribution, it is $O_p(1)$, so if the CLT applies we have:

$$\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$$

  - If $a_n^{-1} Z_n = O_p(1)$, we write $Z_n = O_p(a_n)$, so we have: $\hat{\theta}_n = \theta + O_p(n^{-1/2})$.