8. Sampling & Estimation

Spring 2021

Matthew Blackwell

Gov 2002 (Harvard)

Where are we? Where are we going?

- Last few weeks: probability, learning how to think about r.v.s

Where are we? Where are we going?

- Last few weeks: probability, learning how to think about r.v.s
- Now: how to estimate features of underlying distributions with data.

Where are we? Where are we going?

- Last few weeks: probability, learning how to think about r.v.s
- Now: how to estimate features of underlying distributions with data.
- How do we construct estimators? What are their properties?

1/ Point Estimation

Motivating example

- Gerber, Green, and Larimer (APSR, 2008)

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	______
9995 JENNIFER KAY SMITH		Voted	______
9997 RICHARD B JACKSON		Voted	______
9999 KATHY MARIE JACKSON		Voted	______

Motivating Example

```r
load("../assets/gerber_green_larimer.RData")
## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")
neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])
neigh.mean
```

```
## [1] 0.378
```

Motivating Example

```r
load("../assets/gerber_green_larimer.RData")
## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")
neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])
neigh.mean
```

```
## [1] 0.378
```

```r
contr.mean <- mean(social$voted[social$treatment == "Civic Duty"])
contr.mean
```

```
## [1] 0.315
```

Motivating Example

```r
load("../assets/gerber_green_larimer.RData")
## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")
neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])
neigh.mean
```

```
## [1] 0.378
```

```r
contr.mean <- mean(social$voted[social$treatment == "Civic Duty"])
contr.mean
```

```
## [1] 0.315
```

```r
neigh.mean - contr.mean
```

```
## [1] 0.0634
```

Motivating Example

```r
load("../assets/gerber_green_larimer.RData")
## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")
neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])
neigh.mean
```

```
## [1] 0.378
```

```r
contr.mean <- mean(social$voted[social$treatment == "Civic Duty"])
contr.mean
```

```
## [1] 0.315
```

```r
neigh.mean - contr.mean
```

```
## [1] 0.0634
```

- Is this a “real”? Is it big?

Why study estimators?

- **Goal 1: Inference**

Why study estimators?

- **Goal 1: Inference**
    - What is our best guess about some quantity of interest?

Why study estimators?

- **Goal 1: Inference**
  - What is our best guess about some quantity of interest?
  - What are a set of plausible values of the quantity of interest?

Why study estimators?

- **Goal 1: Inference**
    - What is our best guess about some quantity of interest?
    - What are a set of plausible values of the quantity of interest?
- **Goal 2: Compare estimators**

Why study estimators?

- **Goal 1: Inference**
    - What is our best guess about some quantity of interest?
    - What are a set of plausible values of the quantity of interest?
- **Goal 2: Compare estimators**
    - In an experiment, use simple difference in sample means $(\overline{Y} - \overline{X})$?

Why study estimators?

- **Goal 1: Inference**
  - What is our best guess about some quantity of interest?
  - What are a set of plausible values of the quantity of interest?
- **Goal 2: Compare estimators**
  - In an experiment, use simple difference in sample means ($\overline{Y} - \overline{X}$)?
  - Or the **post-stratification estimator**, where we estimate the estimate the difference among two subsets of the data (male and female, for instance) and then take the weighted average of the two ($\overline{Z}$ is the share of women):
  $$(\overline{Y}_f - \overline{X}_f)\overline{Z} + (\overline{Y}_m - \overline{X}_m)(1 - \overline{Z})$$

Why study estimators?

- **Goal 1: Inference**
  - What is our best guess about some quantity of interest?
  - What are a set of plausible values of the quantity of interest?
- **Goal 2: Compare estimators**
  - In an experiment, use simple difference in sample means ($\overline{Y} - \overline{X}$)?
  - Or the **post-stratification estimator**, where we estimate the estimate the difference among two subsets of the data (male and female, for instance) and then take the weighted average of the two ($\overline{Z}$ is the share of women):
$$(\overline{Y}_f - \overline{X}_f)\overline{Z} + (\overline{Y}_m - \overline{X}_m)(1 - \overline{Z})$$
  - Which (if either) is better? How would we know?

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
    - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
    - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
    - n is the **sample size**

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
    - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
    - n is the **sample size**
    - i.i.d. can be justified through random sampling from an inifinite population.

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
    - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
    - n is the **sample size**
    - i.i.d. can be justified through random sampling from an inifinite population.
    - F is often called the **population distribution** or just **population**

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
  - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
  - n is the **sample size**
  - i.i.d. can be justified through random sampling from an inifinite population.
  - F is often called the **population distribution** or just **population**
  - Model-based because we are assuming the probability model F

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
    - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
    - n is the **sample size**
    - i.i.d. can be justified through random sampling from an inifinite population.
    - F is often called the **population distribution** or just **population**
    - Model-based because we are assuming the probability model F
- Two metaphors:

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
    - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
    - n is the **sample size**
    - i.i.d. can be justified through random sampling from an inifinite population.
    - F is often called the **population distribution** or just **population**
    - Model-based because we are assuming the probability model F
- Two metaphors:
    - Actual/potential population of size $N >> n$ and we randomly sample n.

Samples from the population

- **Model-based inferece**: random vectors $X_1, \dots, X_n$ are i.i.d. draws from c.d.f. F
    - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
    - n is the **sample size**
    - i.i.d. can be justified through random sampling from an inifinite population.
    - F is often called the **population distribution** or just **population**
    - Model-based because we are assuming the probability model F
- Two metaphors:
    - Actual/potential population of size $N >> n$ and we randomly sample n.
    - F represents the **data generating process**, we repeat n times

Samples from the population

- **Model-based inferece**: random vectors $X_1, \ldots, X_n$ are i.i.d. draws from c.d.f. F
  - e.g.: $X_i = 1$ if citizen i votes, $X_i = 0$ otherwise.
  - n is the **sample size**
  - i.i.d. can be justified through random sampling from an inifinite population.
  - F is often called the **population distribution** or just **population**
  - Model-based because we are assuming the probability model F
- Two metaphors:
  - Actual/potential population of size $N >> n$ and we randomly sample n.
  - F represents the **data generating process**, we repeat n times
- **Statistical inference** or **learning** is using data to infer F.

Point estimation

- Goal of inference: learn about the features of the population.

Point estimation

- Goal of inference: learn about the features of the population.
- **Parameter**: θ is any function of the population distribution F

Point estimation

- Goal of inference: learn about the features of the population.
- **Parameter**: θ is any function of the population distribution F
    - Also called: quantities of interest, estimands.

Point estimation

- Goal of inference: learn about the features of the population.
- **Parameter**: θ is any function of the population distribution F
    - Also called: quantities of interest, estimands.
- Examples of parameters:

Point estimation

- Goal of inference: learn about the features of the population.
- **Parameter**: θ is any function of the population distribution F
  - Also called: quantities of interest, estimands.
- Examples of parameters:
  - $\mu = \mathbb{E}[X_i]$: the mean (turnout rate in the population).

Point estimation

- Goal of inference: learn about the features of the population.
- **Parameter**: θ is any function of the population distribution F
    - Also called: quantities of interest, estimands.
- Examples of parameters:
    - $\mu = \mathbb{E}[X_i]$: the mean (turnout rate in the population).
    - $\sigma^2 = \mathbb{V}[X_i]$: the variance.

Point estimation

- Goal of inference: learn about the features of the population.
- **Parameter**: θ is any function of the population distribution F
  - Also called: quantities of interest, estimands.
- Examples of parameters:
  - $\mu = \mathbb{E}[X_i]$: the mean (turnout rate in the population).
  - $\sigma^2 = \mathbb{V}[X_i]$: the variance.
  - $\mu_y - \mu_x = \mathbb{E}[Y_i] - \mathbb{E}[X_i]$: the difference in mean turnout between two groups.

Point estimation

- Goal of inference: learn about the features of the population.
- **Parameter**: θ is any function of the population distribution F
  - Also called: quantities of interest, estimands.
- Examples of parameters:
  - $\mu = \mathbb{E}[X_i]$: the mean (turnout rate in the population).
  - $\sigma^2 = \mathbb{V}[X_i]$: the variance.
  - $\mu_y - \mu_x = \mathbb{E}[Y_i] - \mathbb{E}[X_i]$: the difference in mean turnout between two groups.
- **Point estimation**: providing a single "best guess" about these parameters.

Estimators

- A **statistic** is any function of the sample $\{X_1, \ldots, X_n\}$.

Estimators

- A **statistic** is any function of the sample $\{X_1, \dots, X_n\}$.
    - Before we see the data, statistics are random and have distributions, etc.

Estimators

- A **statistic** is any function of the sample $\{X_1, \dots, X_n\}$.
  - Before we see the data, statistics are random and have distributions, etc.
  - After we see the data, statistic is realized and we see the specific value.

Estimators

- A **statistic** is any function of the sample $\{X_1, \dots, X_n\}$.
  - Before we see the data, statistics are random and have distributions, etc.
  - After we see the data, statistic is realized and we see the specific value.

Definition

An **estimator** $\hat{\theta}_n$ for some parameter θ, is a statistic intended as a guess about θ.

Estimators

- A **statistic** is any function of the sample $\{X_1, \dots, X_n\}$.
  - Before we see the data, statistics are random and have distributions, etc.
  - After we see the data, statistic is realized and we see the specific value.

Definition

An **estimator** $\hat{\theta}_n$ for some parameter θ, is a statistic intended as a guess about θ.

- $\hat{\theta}_n$ is a r.v. because it is a function of r.v.s.

Estimators

- A **statistic** is any function of the sample $\{X_1, \ldots, X_n\}$.
  - Before we see the data, statistics are random and have distributions, etc.
  - After we see the data, statistic is realized and we see the specific value.

Definition

An **estimator** $\hat{\theta}_n$ for some parameter θ, is a statistic intended as a guess about θ.

- $\hat{\theta}_n$ is a r.v. because it is a function of r.v.s.
  - $\rightsquigarrow$ $\hat{\theta}_n$ has a distribution.

Estimators

- A **statistic** is any function of the sample $\{X_1, \ldots, X_n\}$.
    - Before we see the data, statistics are random and have distributions, etc.
    - After we see the data, statistic is realized and we see the specific value.

Definition

An **estimator** $\hat{\theta}_n$ for some parameter θ, is a statistic intended as a guess about θ.

- $\hat{\theta}_n$ is a r.v. because it is a function of r.v.s.
    - $\rightsquigarrow \hat{\theta}_n$ has a distribution.
- An **estimate** is one particular realization of the estimator

Estimators

- A **statistic** is any function of the sample $\{X_1, \ldots, X_n\}$.
    - Before we see the data, statistics are random and have distributions, etc.
    - After we see the data, statistic is realized and we see the specific value.

Definition

An **estimator** $\hat{\theta}_n$ for some parameter θ, is a statistic intended as a guess about θ.

- $\hat{\theta}_n$ is a r.v. because it is a function of r.v.s.
    - $\rightsquigarrow$ $\hat{\theta}_n$ has a distribution.
- An **estimate** is one particular realization of the estimator
    - Why is the following statement wrong: "My estimate was the sample mean and my estimator was 0.38"?

Examples of Estimators

- For the population expectation, $\mathbb{E}[X_i]$, many possible estimators:

Examples of Estimators

- For the population expectation, $\mathbb{E}[X_i]$, many possible estimators:
  - $\hat{\theta}_n = \overline{X}_n$ the sample mean

Examples of Estimators

- For the population expectation, $\mathbb{E}[X_i]$, many possible estimators:
  - $\hat{\theta}_n = \overline{X}_n$ the sample mean
  - $\hat{\theta}_n = X_1$ just use the first observation

Examples of Estimators

- For the population expectation, $\mathbb{E}[X_i]$, many possible estimators:
    - $\widehat{\theta}_n = \overline{X}_n$ the sample mean
    - $\widehat{\theta}_n = X_1$ just use the first observation
    - $\widehat{\theta}_n = \max(X_1, \ldots, X_n)$

Examples of Estimators

- For the population expectation, $\mathbb{E}[X_i]$, many possible estimators:
  - $\widehat{\theta}_n = \overline{X}_n$ the sample mean
  - $\widehat{\theta}_n = X_1$ just use the first observation
  - $\widehat{\theta}_n = \max(X_1, \ldots, X_n)$
  - $\widehat{\theta}_n = 3$ always guess 3

The three distributions

- **Population Distribution**: the data-generating process

The three distributions

- **Population Distribution**: the data-generating process
    - Bernoulli in the case of the social pressure/voter turnout example)

The three distributions

- **Population Distribution**: the data-generating process
  - Bernoulli in the case of the social pressure/voter turnout example)
- **Empirical distribution**: $X_1, \dots, X_n$

The three distributions

- **Population Distribution**: the data-generating process
  - Bernoulli in the case of the social pressure/voter turnout example)
- **Empirical distribution**: $X_1, \ldots, X_n$
  - series of 1s and 0s in the sample

The three distributions

- **Population Distribution**: the data-generating process
    - Bernoulli in the case of the social pressure/voter turnout example)
- **Empirical distribution**: $X_1, \ldots, X_n$
    - series of 1s and 0s in the sample
- **Sampling distribution**: distribution of the estimator over repeated samples from the population distribution

The three distributions

- **Population Distribution**: the data-generating process
    - Bernoulli in the case of the social pressure/voter turnout example)
- **Empirical distribution**: $X_1, \ldots, X_n$
    - series of 1s and 0s in the sample
- **Sampling distribution**: distribution of the estimator over repeated samples from the population distribution
    - the 0.38 sample mean in the "Neighbors" group is one draw from this distribution

Sampling distribution, in pictures

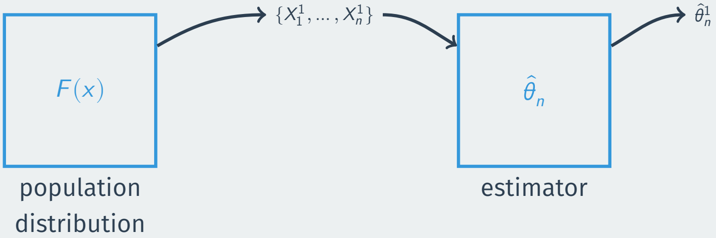
$F(x)$

population distribution

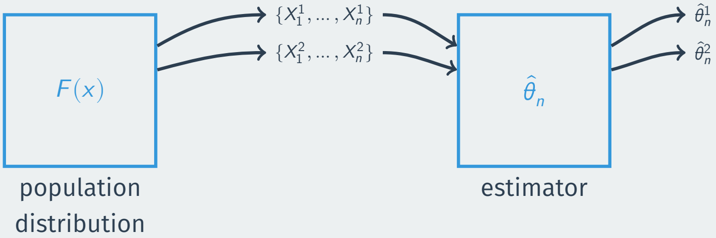
$\hat{\theta}_n$

estimator

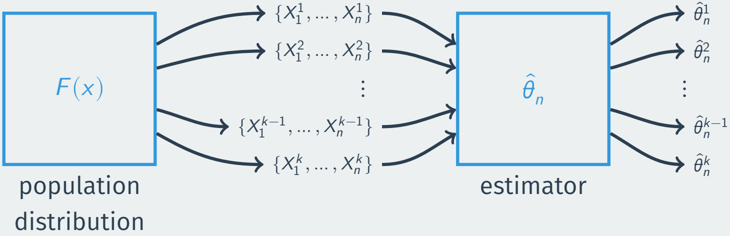
Sampling distribution, in pictures



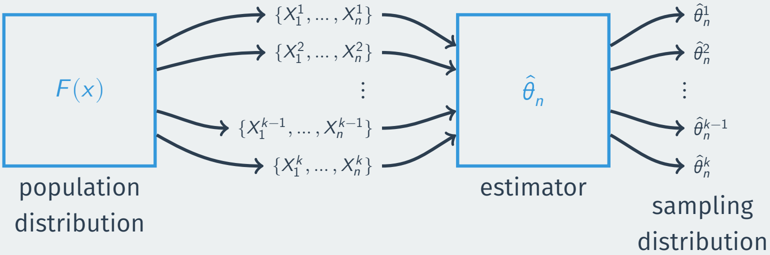
Sampling distribution, in pictures



Sampling distribution, in pictures



Sampling distribution, in pictures



Sampling distribution

```r
## now we take the mean of one sample, which is
## one draw from the  **sampling distribution**
my.samp <- rbinom(n = 10, size = 1, prob = 0.4)
mean(my.samp)
```

```
## [1] 0.3
```

Sampling distribution

```r
## now we take the mean of one sample, which is
## one draw from the  **sampling distribution**
my.samp <- rbinom(n = 10, size = 1, prob = 0.4)
mean(my.samp)
```

```
## [1] 0.3
```

```r
## let's take another draw from the population dist
my.samp.2 <- rbinom(n = 10, size = 1, prob = 0.4)
```

Sampling distribution

```r
## now we take the mean of one sample, which is
## one draw from the  **sampling distribution**
my.samp <- rbinom(n = 10, size = 1, prob = 0.4)
mean(my.samp)
```

```
## [1] 0.3
```

```r
## let's take another draw from the population dist
my.samp.2 <- rbinom(n = 10, size = 1, prob = 0.4)
```

```r
## Let's feed this sample to the sample mean estimator
## to get another estimate, which is another draw from
## the sampling distribution
mean(my.samp.2)
```

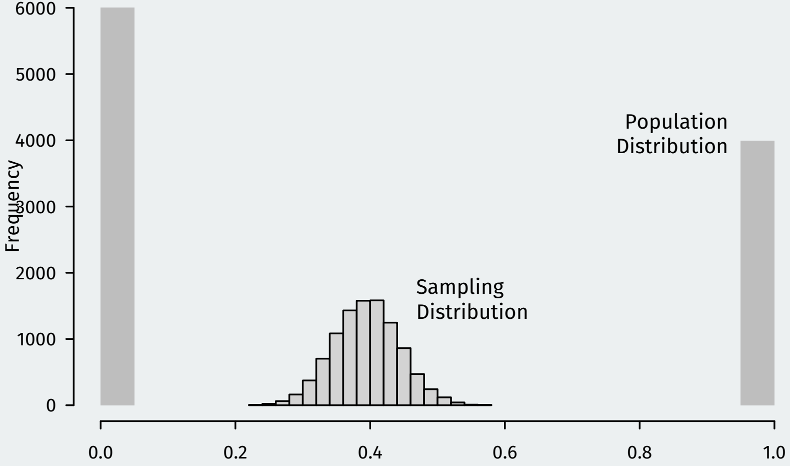
```
## [1] 0.2
```

Sampling distribution by simulation

- Let's generate 10,000 draws from the sampling distribution of the sample mean here when $n = 100$.

```r
nsims <- 10000
mean.holder <- rep(NA, times = nsims)
for (i in 1:nsims) {
    my.samp <- rbinom(n = 100, size = 1, prob = 0.4)
    mean.holder[i] <- mean(my.samp) ## sample mean
    first.holder[i] <- my.samp[1] ## first obs
}
```

Sampling distribution versus population distribution



Question The sampling distribution refers to the distribution of θ, true or false.

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.
  - Example: F is $\text{Pois}(\lambda)$.

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.
    - Example: F is $\text{Pois}(\lambda)$.
    - Construct estimator $\hat{\lambda}$ using **maximum likelihood**

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.
    - Example: F is $\text{Pois}(\lambda)$.
    - Construct estimator $\hat{\lambda}$ using **maximum likelihood**
    - Downside: inferences are **model dependent**

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.
    - Example: F is $\text{Pois}(\lambda)$.
    - Construct estimator $\hat{\lambda}$ using **maximum likelihood**
    - Downside: inferences are **model dependent**
- **Nonparametric inference**: make minimal assumptions on F.

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.
    - Example: F is $\text{Pois}(\lambda)$.
    - Construct estimator $\hat{\lambda}$ using **maximum likelihood**
    - Downside: inferences are **model dependent**
- **Nonparametric inference**: make minimal assumptions on F.
- **Plug-in/analogy principle**: replace F with the empirical distribution.

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.
    - Example: F is $\text{Pois}(\lambda)$.
    - Construct estimator $\hat{\lambda}$ using **maximum likelihood**
    - Downside: inferences are **model dependent**
- **Nonparametric inference**: make minimal assumptions on F.
- **Plug-in/analogy principle**: replace F with the empirical distribution.
    - Empirical distribution: probability $1/n$ at each observed value of X_i:

$$\widehat{F}_n(x) = \frac{\sum_{i=1}^{n} \mathbb{I}(X_i \leq x)}{n}$$

Where do estimators come from?

- **Parametric modeling:** assume $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$ and specify what family F is from.
    - Example: F is $\text{Pois}(\lambda)$.
    - Construct estimator $\hat{\lambda}$ using **maximum likelihood**
    - Downside: inferences are **model dependent**
- **Nonparametric inference**: make minimal assumptions on F.
- **Plug-in/analogy principle**: replace F with the empirical distribution.
    - Empirical distribution: probability $1/n$ at each observed value of X_i:

$$\widehat{F}_n(x) = \frac{\sum_{i=1}^n \mathbb{I}(X_i \leq x)}{n}$$

    - $\rightsquigarrow$ if $\theta = \mathbb{E}[g(X)]$ replace $\mathbb{E}$ sample means: $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n g(X_i)$

Plug-in estimators, examples

- Expectation:

$$\mu = \mathbb{E}[X_i] \rightsquigarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$$

Plug-in estimators, examples

- Expectation:

$$\mu = \mathbb{E}[X_i] \rightsquigarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$$

- Variance:

$$\sigma^2 = \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] \rightsquigarrow \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Plug-in estimators, examples

- Expectation:

$$\mu = \mathbb{E}[X_i] \rightsquigarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$$

- Variance:

$$\sigma^2 = \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] \rightsquigarrow \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

- Covariance:

$$\sigma_{xy} = \mathrm{Cov}[X_i, Y_i] = \mathbb{E}[(X_i - \mathbb{E}[X_i])(Y_i - \mathbb{E}[Y_i])] \rightsquigarrow \widehat{\sigma}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

2/ Finite-Sample Properties of Estimators

Properties of estimators

- We only get one draw from the sampling distribution, $\hat{\theta}_n$.

Properties of estimators

- We only get one draw from the sampling distribution, $\widehat{\theta}_n$.
- Want to use estimators whose distribution is "close" to the true value.

Properties of estimators

- We only get one draw from the sampling distribution, $\widehat{\theta}_n$.
- Want to use estimators whose distribution is "close" to the true value.
- There are two ways we evaluate estimators:

Properties of estimators

- We only get one draw from the sampling distribution, $\hat{\theta}_n$.
- Want to use estimators whose distribution is "close" to the true value.
- There are two ways we evaluate estimators:
    - **Finite sample**: the properties of its sampling distribution for a fixed sample size n.

Properties of estimators

- We only get one draw from the sampling distribution, $\hat{\theta}_n$.
- Want to use estimators whose distribution is "close" to the true value.
- There are two ways we evaluate estimators:
    - **Finite sample**: the properties of its sampling distribution for a fixed sample size n.
    - **Large sample**: the properties of the sampling distribution as we let $n \to \infty$.

Bias

- The **bias** of estimator $\hat{\theta}$ for parameter θ is

$$\text{bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta.$$

Bias

- The **bias** of estimator $\hat{\theta}$ for parameter θ is

$$\text{bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta.$$

- An estimator is **unbiased** if $\text{bias}[\hat{\theta}] = 0$.

Bias

- The **bias** of estimator $\hat{\theta}$ for parameter θ is

$$\mathrm{bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta.$$

- An estimator is **unbiased** if $\mathrm{bias}[\hat{\theta}] = 0$.
- Sample mean of i.i.d. $X_1, \ldots, X_n$ with $\mathbb{E}[X_i] = \mu$

$$\mathbb{E}\left[\overline{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

Bias

- The **bias** of estimator $\widehat{\theta}$ for parameter θ is

$$\text{bias}[\widehat{\theta}] = \mathbb{E}[\widehat{\theta}] - \theta.$$

- An estimator is **unbiased** if $\text{bias}[\widehat{\theta}] = 0$.
- Sample mean of i.i.d. $X_1, \ldots, X_n$ with $\mathbb{E}[X_i] = \mu$

$$\mathbb{E}\left[\overline{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

- Thus, $\overline{X}_n$ is unbiased for μ if $\mathbb{E}[|X|] < \infty$

Bias

- The **bias** of estimator $\hat{\theta}$ for parameter θ is

$$\text{bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta.$$

- An estimator is **unbiased** if $\text{bias}[\hat{\theta}] = 0$.
- Sample mean of i.i.d. $X_1, \ldots, X_n$ with $\mathbb{E}[X_i] = \mu$

$$\mathbb{E}\left[\overline{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

- Thus, $\overline{X}_n$ is unbiased for μ if $\mathbb{E}[|X|] < \infty$
    - What about a weighted average?

Bias

- The **bias** of estimator $\hat{\theta}$ for parameter θ is

$$\text{bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta.$$

- An estimator is **unbiased** if $\text{bias}[\hat{\theta}] = 0$.
- Sample mean of i.i.d. $X_1, \ldots, X_n$ with $\mathbb{E}[X_i] = \mu$

$$\mathbb{E}\left[\overline{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

- Thus, $\overline{X}_n$ is unbiased for μ if $\mathbb{E}[|X|] < \infty$
    - What about a weighted average?
- Unbiasedness is preserved under linear transformations.

Estimation variance

- **Sampling variance**: the variance of an estimator $\mathbb{V}[\widehat{\theta}]$.

Estimation variance

- **Sampling variance**: the variance of an estimator $\mathbb{V}[\hat{\theta}]$.
    - Measure of how spread the estimator it is around its mean.

Estimation variance

- **Sampling variance**: the variance of an estimator $\mathbb{V}[\hat{\theta}]$.
    - Measure of how spread the estimator it is around its mean.
- Sampling variance of the sample mean:

$$\mathbb{V}\left[\overline{X}_n\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

Estimation variance

- **Sampling variance**: the variance of an estimator $\mathbb{V}[\hat{\theta}]$.
    - Measure of how spread the estimator it is around its mean.
- Sampling variance of the sample mean:

$$\mathbb{V}\left[\overline{X}_n\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

- **Standard error**: standard deviation of the estimator $\text{se}(\hat{\theta}) = \sqrt{\mathbb{V}[\hat{\theta}]}$

Estimation variance

- **Sampling variance**: the variance of an estimator $\mathbb{V}[\hat{\theta}]$.
    - Measure of how spread the estimator it is around its mean.
- Sampling variance of the sample mean:

$$\mathbb{V}\left[\overline{X}_n\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

- **Standard error**: standard deviation of the estimator $\text{se}(\hat{\theta}) = \sqrt{\mathbb{V}[\hat{\theta}]}$
    - Like all SDs, nice that it's on the same scale.

Estimation variance

- **Sampling variance**: the variance of an estimator $\mathbb{V}[\hat{\theta}]$.
    - Measure of how spread the estimator it is around its mean.
- Sampling variance of the sample mean:

$$\mathbb{V}\left[\overline{X}_n\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

- **Standard error**: standard deviation of the estimator $\text{se}(\hat{\theta}) = \sqrt{\mathbb{V}[\hat{\theta}]}$
    - Like all SDs, nice that it's on the same scale.
- Standard error of the sample mean: $\sigma/\sqrt{n}$

Mean squared error

- **Mean squared error** or **MSE** is

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

Mean squared error

- **Mean squared error** or **MSE** is

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

- The MSE assesses the quality of an estimator.

Mean squared error

- **Mean squared error** or **MSE** is

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

- The MSE assesses the quality of an estimator.
  - How big are (squared) deviations from the true parameter?

Mean squared error

- **Mean squared error** or **MSE** is

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

- The MSE assesses the quality of an estimator.
    - How big are (squared) deviations from the true parameter?
    - Ideally, this would be as low as possible!

Mean squared error

- **Mean squared error** or **MSE** is

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

- The MSE assesses the quality of an estimator.
  - How big are (squared) deviations from the true parameter?
  - Ideally, this would be as low as possible!
- Useful decomposition result:

$$\text{MSE} = \text{bias}[\hat{\theta}_n]^2 + \mathbb{V}[\hat{\theta}_n]$$

Mean squared error

- **Mean squared error** or **MSE** is

$$\mathsf{MSE} = \mathbb{E}[(\hat{\theta}_n - \theta)^2]$$

- The MSE assesses the quality of an estimator.
  - How big are (squared) deviations from the true parameter?
  - Ideally, this would be as low as possible!
- Useful decomposition result:

$$\mathsf{MSE} = \mathsf{bias}[\hat{\theta}_n]^2 + \mathbb{V}[\hat{\theta}_n]$$

- $\rightsquigarrow$ for unbiased estimators, MSE is the sampling variance.

Mean squared error

- **Mean squared error** or **MSE** is

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

- The MSE assesses the quality of an estimator.
  - How big are (squared) deviations from the true parameter?
  - Ideally, this would be as low as possible!
- Useful decomposition result:

$$\text{MSE} = \text{bias}[\hat{\theta}_n]^2 + \mathbb{V}[\hat{\theta}_n]$$

- $\rightsquigarrow$ for unbiased estimators, MSE is the sampling variance.
- Might accept some bias for large reductions in variance for lower overall MSE.

3/ Design-based inference

Survey sampling

- Up to now: focus on **model-based inference**.

Survey sampling

- Up to now: focus on **model-based inference**.
    - $X_1, \ldots, X_n$ are iid draws from an infinite population modeled by cdf F

Survey sampling

- Up to now: focus on **model-based inference**.
    - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.

Survey sampling

- Up to now: focus on **model-based inference**.
  - $X_1, \ldots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \ldots, N$.
- Population characteristics: $x_1, x_2, \ldots, x_N$ (list of fixed numbers)

Survey sampling

- Up to now: focus on **model-based inference**.
  - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.
- Population characteristics: $x_1, x_2, \dots, x_N$ (list of fixed numbers)
  - We'll think of the population and everything about it as **fixed**

Survey sampling

- Up to now: focus on **model-based inference**.
  - $X_1, \ldots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \ldots, N$.
- Population characteristics: $x_1, x_2, \ldots, x_N$ (list of fixed numbers)
  - We'll think of the population and everything about it as **fixed**
- **Assumption**: simple random sample (eg, with replacement) of size n from this population

Survey sampling

- Up to now: focus on **model-based inference**.
  - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.
- Population characteristics: $x_1, x_2, \dots, x_N$ (list of fixed numbers)
  - We'll think of the population and everything about it as **fixed**
- **Assumption**: simple random sample (eg, with replacement) of size n from this population
  - Number of possible samples: $\binom{N}{n}$

Survey sampling

- Up to now: focus on **model-based inference**.
  - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.
- Population characteristics: $x_1, x_2, \dots, x_N$ (list of fixed numbers)
  - We'll think of the population and everything about it as **fixed**
- **Assumption**: simple random sample (eg, with replacement) of size n from this population
  - Number of possible samples: $\binom{N}{n}$
  - Sampling inclusion indicators: $I_1, I_2, \dots, I_N$

Survey sampling

- Up to now: focus on **model-based inference**.
    - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.
- Population characteristics: $x_1, x_2, \dots, x_N$ (list of fixed numbers)
    - We'll think of the population and everything about it as **fixed**
- **Assumption**: simple random sample (eg, with replacement) of size n from this population
    - Number of possible samples: $\binom{N}{n}$
    - Sampling inclusion indicators: $I_1, I_2, \dots, I_N$
    - These are random because of the random sampling (uppercase!)

Survey sampling

- Up to now: focus on **model-based inference**.
  - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.
- Population characteristics: $x_1, x_2, \dots, x_N$ (list of fixed numbers)
  - We'll think of the population and everything about it as **fixed**
- **Assumption**: simple random sample (eg, with replacement) of size n from this population
  - Number of possible samples: $\binom{N}{n}$
  - Sampling inclusion indicators: $I_1, I_2, \dots, I_N$
  - These are random because of the random sampling (uppercase!)
  - Total sample size is fixed: $\sum_{i=1}^{N} I_i = n$

Survey sampling

- Up to now: focus on **model-based inference**.
    - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.
- Population characteristics: $x_1, x_2, \dots, x_N$ (list of fixed numbers)
    - We'll think of the population and everything about it as **fixed**
- **Assumption**: simple random sample (eg, with replacement) of size n from this population
    - Number of possible samples: $\binom{N}{n}$
    - Sampling inclusion indicators: $I_1, I_2, \dots, I_N$
    - These are random because of the random sampling (uppercase!)
    - Total sample size is fixed: $\sum_{i=1}^{N} I_i = n$
    - **Inclusion probabilities**: $\pi = \mathbb{P}(I_i = 1) = n/N$

Survey sampling

- Up to now: focus on **model-based inference**.
  - $X_1, \dots, X_n$ are iid draws from an infinite population modeled by cdf F
- Alternative: a large, but **finite sample** of size N indexed $i = 1, \dots, N$.
- Population characteristics: $x_1, x_2, \dots, x_N$ (list of fixed numbers)
  - We'll think of the population and everything about it as **fixed**
- **Assumption**: simple random sample (eg, with replacement) of size n from this population
  - Number of possible samples: $\binom{N}{n}$
  - Sampling inclusion indicators: $I_1, I_2, \dots, I_N$
  - These are random because of the random sampling (uppercase!)
  - Total sample size is fixed: $\sum_{i=1}^{N} I_i = n$
  - **Inclusion probabilities**: $\pi = \mathbb{P}(I_i = 1) = n/N$
- Different **sampling designs** lead to different inclusion probabilities and difference inferences.

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
    - Fixed quantity because the population is fixed and finite.

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
  - Fixed quantity because the population is fixed and finite.
  - But we don't observe all x_i, so we cannot calculate it.

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
  - Fixed quantity because the population is fixed and finite.
  - But we don't observe all x_i, so we cannot calculate it.
- Estimator: sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{N} I_i x_i$

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
  - Fixed quantity because the population is fixed and finite.
  - But we don't observe all x_i, so we cannot calculate it.
- Estimator: sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{N} I_i x_i$
  - This estimator is random because the sample is random.

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
  - Fixed quantity because the population is fixed and finite.
  - But we don't observe all x_i, so we cannot calculate it.
- Estimator: sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{N} I_i x_i$
  - This estimator is random because the sample is random.
- **Design-based inference**: randomness comes from sampling alone and depends on sampling design.

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
  - Fixed quantity because the population is fixed and finite.
  - But we don't observe all x_i, so we cannot calculate it.
- Estimator: sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{N} I_i x_i$
  - This estimator is random because the sample is random.
- **Design-based inference**: randomness comes from sampling alone and depends on sampling design.
- Unbiasedness proof is illustrative:

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{N} I_i x_i\right] = \frac{1}{n}\sum_{i=1}^{N}\mathbb{E}[I_i]x_i = \frac{1}{n}\sum_{i=1}^{N}\frac{n}{N}x_i = \frac{1}{N}\sum_{i=1}^{N}x_i = \overline{x}$$

Estimands and estimators

- Estimand: population mean $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
    - Fixed quantity because the population is fixed and finite.
    - But we don't observe all x_i, so we cannot calculate it.
- Estimator: sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{N} I_i x_i$
    - This estimator is random because the sample is random.
- **Design-based inference**: randomness comes from sampling alone and depends on sampling design.
- Unbiasedness proof is illustrative:

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{N} I_i x_i\right] = \frac{1}{n}\sum_{i=1}^{N}\mathbb{E}[I_i]x_i = \frac{1}{n}\sum_{i=1}^{N}\frac{n}{N}x_i = \frac{1}{N}\sum_{i=1}^{N}x_i = \overline{x}$$

- Remember: unbiased across repeated samples from the sampling design.

Variance of the sample mean

- Variance of $\overline{X}_n$ across repeated samples:

$$\mathbb{V}[\overline{X}_n] = \underbrace{\left(1 - \frac{n}{N}\right)}_{\text{finite pop. correction}} \frac{s^2}{n}$$

Variance of the sample mean

- Variance of $\overline{X}_n$ across repeated samples:

$$\mathbb{V}[\overline{X}_n] = \underbrace{\left(1 - \frac{n}{N}\right)}_{\text{finite pop. correction}} \frac{s^2}{n}$$

- s^2 is the **population variance** of x_i (a fixed quantity!!):

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

Variance of the sample mean

- Variance of $\overline{X}_n$ across repeated samples:

$$\mathbb{V}[\overline{X}_n] = \underbrace{\left(1 - \frac{n}{N}\right)}_{\text{finite pop. correction}} \frac{s^2}{n}$$

- s^2 is the **population variance** of x_i (a fixed quantity!!):

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2$$

- We can still apply the plug-in principle and use the **sample variance** S^2

$$\widehat{\mathbb{V}}[\overline{X}_n] = \left(1 - \frac{n}{N}\right) \frac{S^2}{n} \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{N} I_i (x_i - \overline{X}_n)^2$$

Variance of the sample mean

- Variance of $\overline{X}_n$ across repeated samples:

$$\mathbb{V}[\overline{X}_n] = \underbrace{\left(1 - \frac{n}{N}\right)}_{\text{finite pop. correction}} \frac{s^2}{n}$$

- s^2 is the **population variance** of x_i (a fixed quantity!!):

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2$$

- We can still apply the plug-in principle and use the **sample variance** S^2

$$\widehat{\mathbb{V}}[\overline{X}_n] = \left(1 - \frac{n}{N}\right) \frac{S^2}{n} \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{N} I_i (x_i - \overline{X}_n)^2$$

- We can show that this is unbiased so that $\mathbb{E}[\widehat{\mathbb{V}}[\overline{X}_n]] = \mathbb{V}[\overline{X}_n]$

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i
  - Typically to oversample groups that are difficult to reach

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i
    - Typically to oversample groups that are difficult to reach
    - Or to ensure sufficient sample sizes for smaller minority groups

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i
    - Typically to oversample groups that are difficult to reach
    - Or to ensure sufficient sample sizes for smaller minority groups
- **Horvitz-Thompson estimator**:

$$\widetilde{X}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{I_i x_i}{\pi_i}$$

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i
    - Typically to oversample groups that are difficult to reach
    - Or to ensure sufficient sample sizes for smaller minority groups
- **Horvitz-Thompson estimator**:

$$\widetilde{X}_{HT} = \frac{1}{N}\sum_{i=1}^{N}\frac{I_i x_i}{\pi_i}$$

- The HT estimator is unbiased: $\mathbb{E}[\widetilde{X}_{HT}] = \overline{x}$

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i
  - Typically to oversample groups that are difficult to reach
  - Or to ensure sufficient sample sizes for smaller minority groups
- **Horvitz-Thompson estimator**:

$$\widetilde{X}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{I_i x_i}{\pi_i}$$

  - The HT estimator is unbiased: $\mathbb{E}[\widetilde{X}_{HT}] = \overline{x}$
  - But be very unstable and high variance if a low π_i actually gets sampled

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i
    - Typically to oversample groups that are difficult to reach
    - Or to ensure sufficient sample sizes for smaller minority groups
- **Horvitz-Thompson estimator**:

$$\widetilde{X}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{I_i x_i}{\pi_i}$$

    - The HT estimator is unbiased: $\mathbb{E}[\widetilde{X}_{HT}] = \overline{x}$
    - But be very unstable and high variance if a low π_i actually gets sampled
- Alternative: **Hajek estimator** (also known as the IPW estimator)

$$\widetilde{X}_{ipw} = \frac{\sum_{i=1}^{N} I_i x_i / \pi_i}{\sum_{i=1}^{N} I_i / \pi_i}$$

Inverse probability weighting

- More often, we have unequal sampling probabilities: $\pi_i = \mathbb{P}(I_i = 1)$ for each i
    - Typically to oversample groups that are difficult to reach
    - Or to ensure sufficient sample sizes for smaller minority groups
- **Horvitz-Thompson estimator**:

$$\widetilde{X}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{I_i x_i}{\pi_i}$$

    - The HT estimator is unbiased: $\mathbb{E}[\widetilde{X}_{HT}] = \overline{x}$
    - But be very unstable and high variance if a low π_i actually gets sampled
- Alternative: **Hajek estimator** (also known as the IPW estimator)

$$\widetilde{X}_{ipw} = \frac{\sum_{i=1}^{N} I_i x_i / \pi_i}{\sum_{i=1}^{N} I_i / \pi_i}$$

    - Normalizes by the sum of weights.