Estimation: Unbiasedness and Consistency

Zeyang (Arthur) Yu

Princeton University

November 6, 2024

Outline

Outline

Estimation: Motivation

Recall: one canonical problems in mathematical statistics

- Sample: $\{X_i\}_{i=1}^n$ distributed according to P (population *dist.*)
- Assume $\{X_i\}_{i=1}^n$ is an (aka, *iid*) sample
- "Learn" some "features" of P (e.g., a *param.* $\theta(P)$) from the data
- $\theta(P)$ can be either a descriptive or a causal *param.*
- Provides a "best guess" $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ for $\theta(P)$
- $\hat{\theta}_n$ is an estimator for $\theta(P)$
- $\hat{\theta}_n$ is a function maps from data $\{X_i\}_{i=1}^n$ to a number

Natural question: what is a "best guess"

- Naturally, to think about the following criteria
- Does the estimator center around the true parameter $\theta(P)$?
- With a larger sample size, is the estimator close to $\theta(P)$?
- Does the estimator have the least variability (i.e., more precise)?

Estimation: Motivation (Conti.)

"Bad guess" for $\theta(P)$: setup

- P: population *dist.* for annual income for U.S. residents
- Just like the counterfactual, we do not know P
- Independent and identically distributed (aka, *iid*) sample: $\{X_i\}_{i=1}^n$
- *iid*: $P(X_1, \ldots, X_n) = P(X_1) \times \ldots \times P(X_n) = P^n$
- Suppose the sample size $n = 5000$
- "Learn" some "features" of P (e.g., a *param.* $\theta(P)$) from the data
- Suppose the *param.* of interest is $E(X_i)$

  $E(X_i)$ is the average income for all U.S. residents

  This parameter is a descriptive *param.*
- We want to give a "guess" $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ for $\theta(P)$
- Given the 5000 sample, we want to guess a number for $\theta(P)$
- Now, let us consider the following bad guesses

Estimation: Motivation (Conti.)

"Bad guess" for $\theta(P)$: example 1

- $\hat{\theta}_n = \mathbb{R}$

- This guess tells me: the average income can be any real number
- Doesn't sound like a good guess given how uninformative it is

"Bad guess" for $\theta(P)$: example 2 and 3

- $\hat{\theta}_n = \max\left\{\{X_i\}_{i=1}^{5000}\right\}$

- Use the richest person in the sample as a guess for $E(X_i)$
- Doesn't sound like a good guess, since we might over-estimate
  What if we sampled Bill Gates in this sample?

- $\hat{\theta}_n = \min\left\{\{X_i\}_{i=1}^{5000}\right\}$

- Use the poorest person in the same as a guess for $E(X_i)$
- Doesn't sound like a good guess, since we might under-estimate

Outline

Unbiasedness

Unbiasedness: definition

An estimator $\hat{\theta}_n$ is said to be an unbiased estimator for $\theta(P)$ if its expectation equals to $\theta(P)$:

$$E\left(\hat{\theta}_n\right) = \theta(P).$$

Unbiasedness: example

Suppose that $\{X_i\}_{i=1}^n$ is an *iid* sample from P, then, the sample average, $\frac{\sum_{i=1}^n X_i}{n}$, is an unbiased estimator for $E(X_i)$.

Unbiasedness: proof for the example

$$E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{n \times E(X_i)}{n} = E(X_i).$$

Unbiasedness (Conti.)

Biasedness: definition

An estimator $\hat{\theta}_n$ is said to be a biased estimator for $\theta(P)$ if its expectation does not equal to $\theta(P)$:

$$E\left(\hat{\theta}_n\right) \neq \theta(P).$$

Biasedness: examples

- Suppose that $\{X_i\}_{i=1}^n$ is an *iid* sample from P, then:

- $E(X_i) \leq E\left(\max\left\{\{X_i\}_{i=1}^{5000}\right\}\right)$

  Therefore, this max estimator has an upward bias ($\hat{\theta}_n > E(X_i)$)

- $\frac{\sum_{i=1}^n X_i}{n} - \frac{1}{n} = E(X_i) - \frac{1}{n} < E(X_i)$

  Therefore, this estimator has a downward bias ($\hat{\theta}_n < E(X_i)$)

- $\frac{1}{\frac{\sum_{i=1}^n X_i}{n}}$ is not an unbiased estimator for $\frac{1}{E(X_i)}$ because $E\left(\frac{1}{X_i}\right) \neq \frac{1}{E(X_i)}$

Outline

Consistency

Consistency: definition

An estimator $\hat{\theta}_n$ is said to be a consistent estimator for $\theta(P)$ if it converges in probability $\theta(P)$:

$$\hat{\theta}_n \xrightarrow{P} \theta(P).$$

As sample size n increases, $\hat{\theta}_n$ becomes more likely to be close $\theta(P)$.

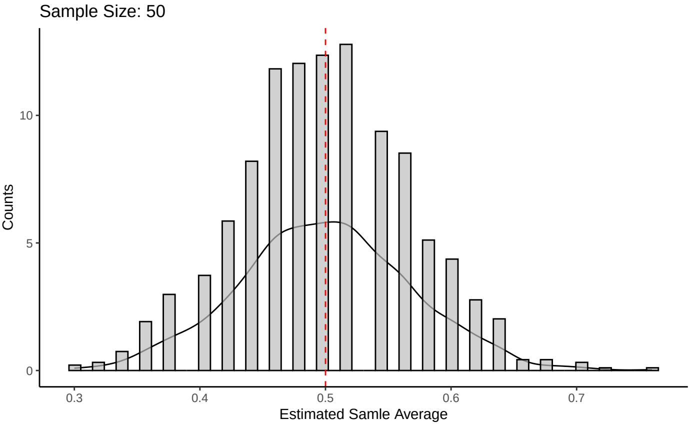
Consistency: example

- Suppose that $\{X_i\}_{i=1}^n$ is an *iid* sample from P, then:

- $\frac{\sum_{i=1}^n X_i}{n}$, is a consistent estimator for $E(X_i)$, by WLLN
- $\frac{\sum_{i=1}^n X_i}{n} - \frac{1}{n}$ and $\frac{\sum_{i=1}^n X_i}{n} + \frac{1}{n}$, are consistent estimators for $E(X_i)$
- $\frac{1}{\frac{\sum_{i=1}^n X_i}{n}}$ is a consistent estimator for $\frac{1}{E(X_i)}$

Outline

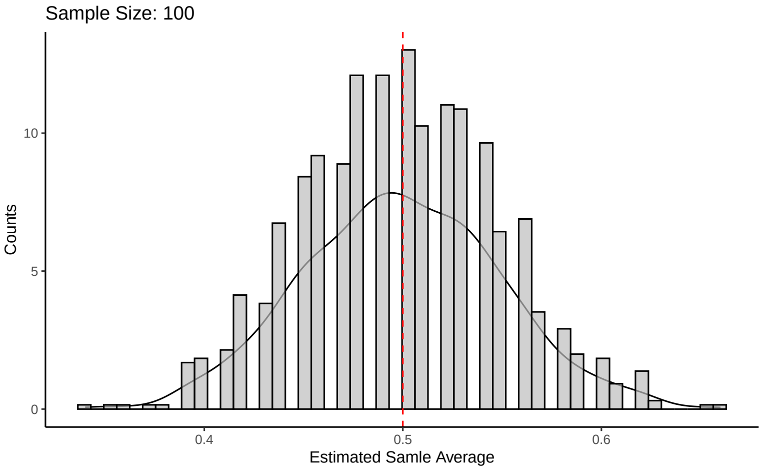
Unbiased Estimator: Simulation



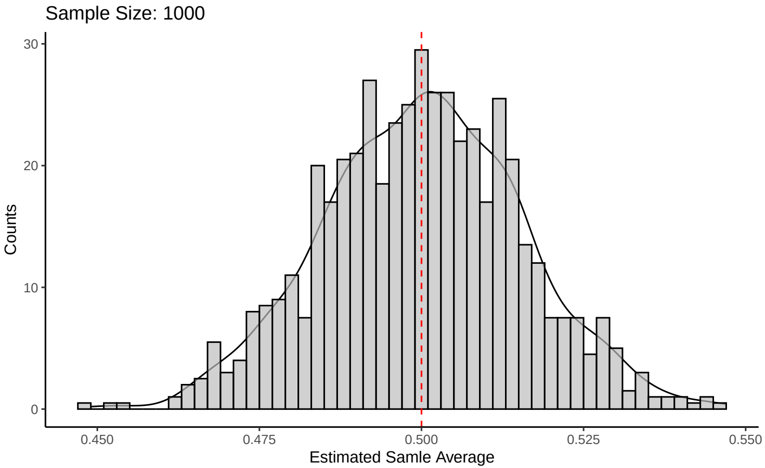
Simulation Setup

- $\{X_i\}_{i=1}^{50}$ is *iid* draw from Ber(0.5)
- $\theta(P) = E(X_i)$, $\hat{\theta}_n = \frac{\sum_{i=1}^{n} X_i}{n}$

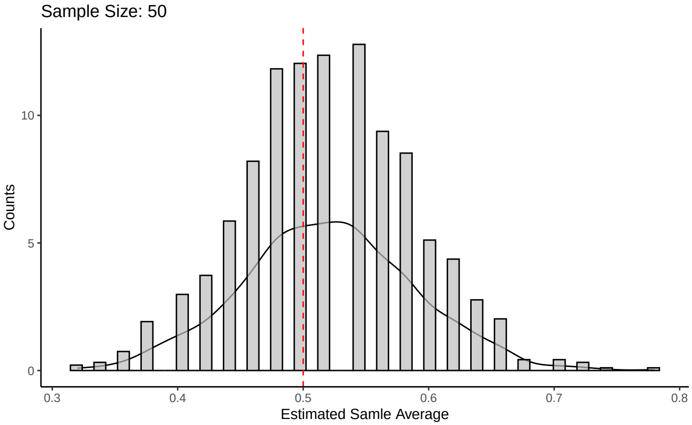
Unbiased Estimator: Simulation (Conti.)



Unbiased Estimator: Simulation (Conti.)



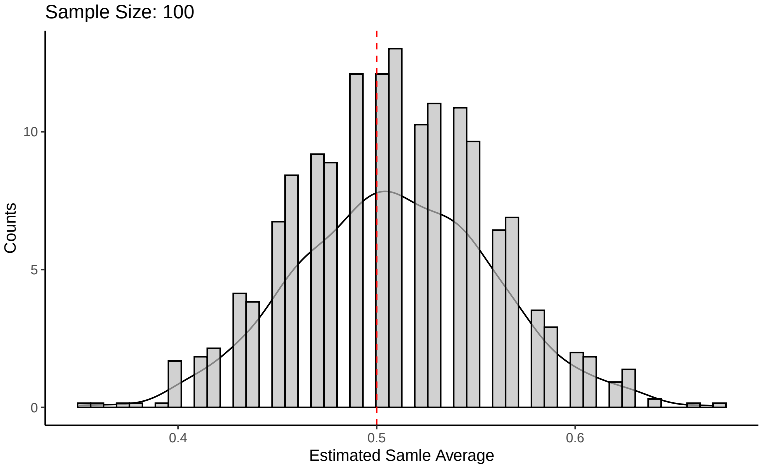
Biased but Consistent Estimator: Simulation



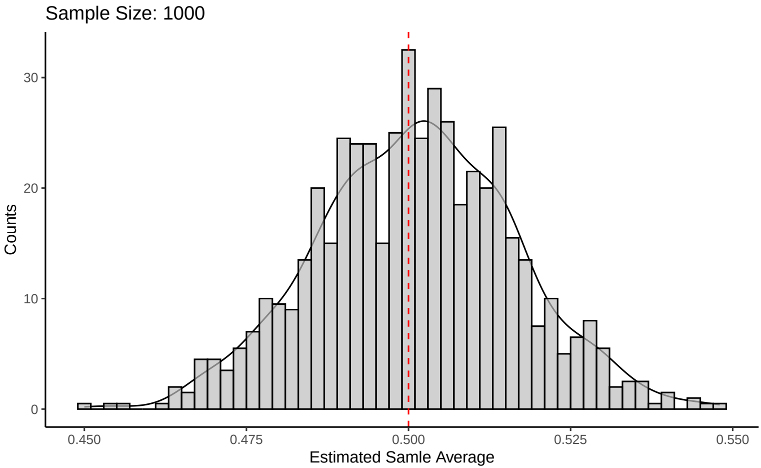
Simulation Setup

- $\{X_i\}_{i=1}^{50}$ is *iid* draw from Ber(0.5)
- $\theta(P) = E(X_i)$, $\hat{\theta}_n = \frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{n}$

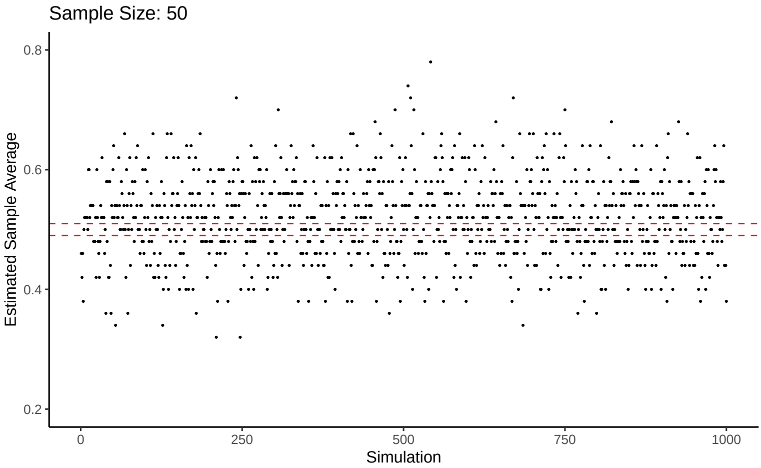
Biased but Consistent Estimator: Simulation (Conti.)



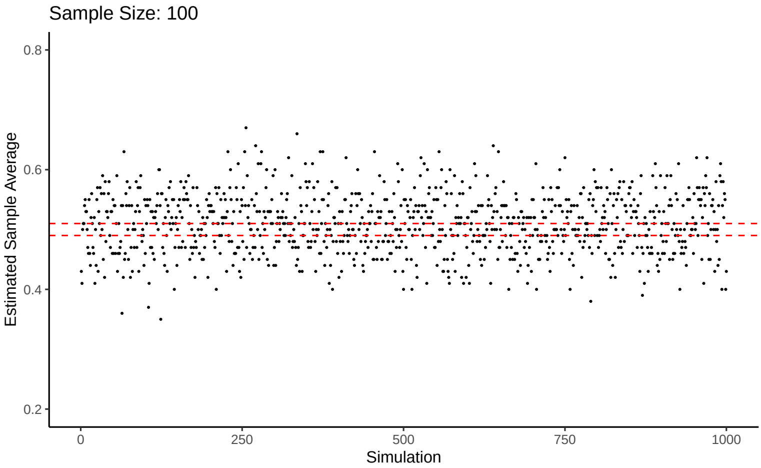
Biased but Consistent Estimator: Simulation (Conti.)



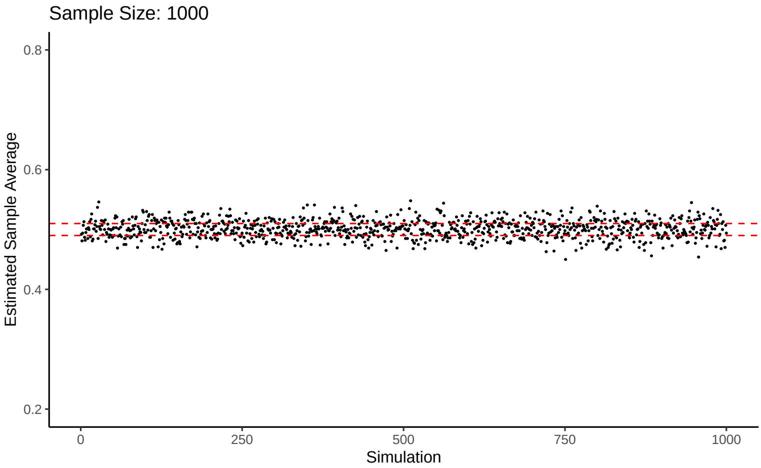
Biased but Consistent Estimator: Simulation (Conti.)



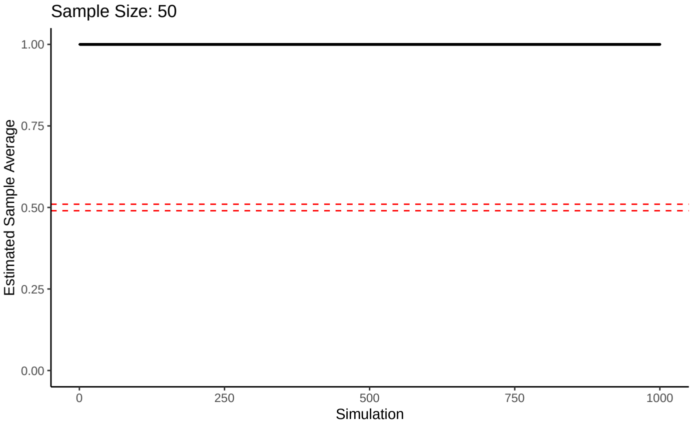
Biased but Consistent Estimator: Simulation (Conti.)



Biased but Consistent Estimator: Simulation (Conti.)



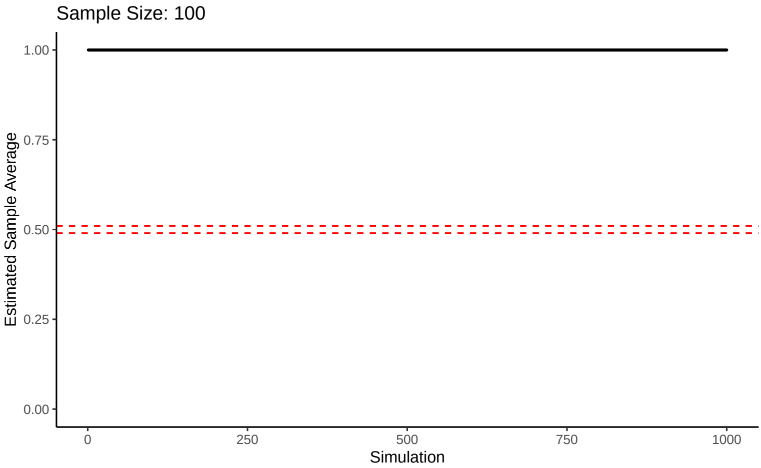
Biased and Inconsistent: Max Estimator



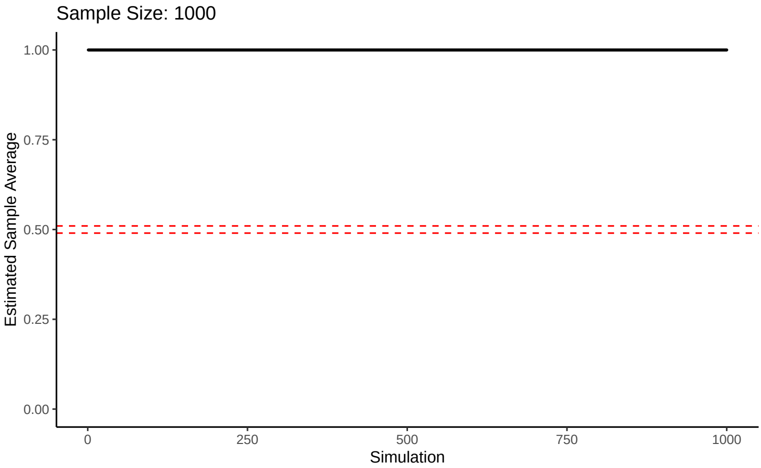
Simulation Setup

- $\{X_i\}_{i=1}^{50}$ is *iid* draw from Ber(0.5)
- $\theta(P) = E(X_i)$, $\hat{\theta}_n = \max\left\{\{X_i\}_{i=1}^{n}\right\}$

Biased and Inconsistent: Max Estimator (Conti.)



Biased and Inconsistent: Max Estimator (Conti.)



Learning Goals: Unbiased and Consistent

Students will be able to:

- Understand the definition of unbiasedness
  - Know that sample mean is an unbiased estimator
- Understand the definition of consistency
  - Know that sample mean is a consistent estimator