Topic 3: Estimation

Ethan P. Marzban University of California, Santa Barbara PSTAT 120B

Outline

1. Transition to Statistics

2. Estimation

3. Assessing the Performance of Estimators

Transition to Statistics

UC SANTA BARBARA
Department of Statistics and Applied Probability

Leadup

- It is finally time for us to begin our transition from **probability** to **statistics**!

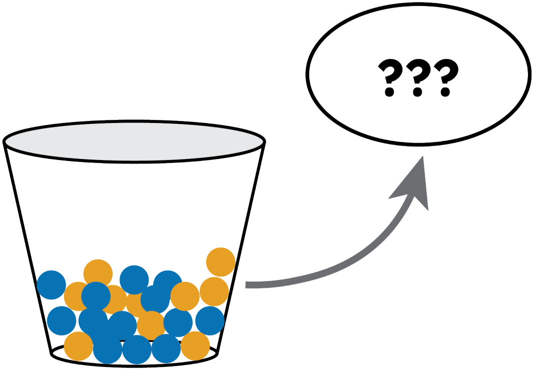
UC SANTA BARBARA
Department of Statistics and Applied Probability

Leadup

- It is finally time for us to begin our transition from **probability** to **statistics**!
- Let's start off with an analogy I posed on the first day of the quarter.

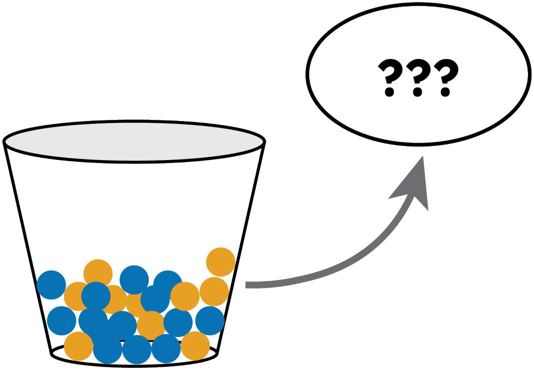
UC SANTA BARBARA
Department of Statistics and Applied Probability

Scenario 1



UC SANTA BARBARA
Department of Statistics and Applied Probability

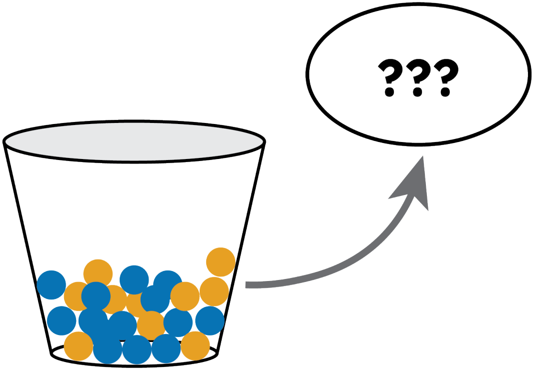
Scenario 1



- We know that a bucket contains some (known) number of blue and gold marbles. From this bucket we take a sample.

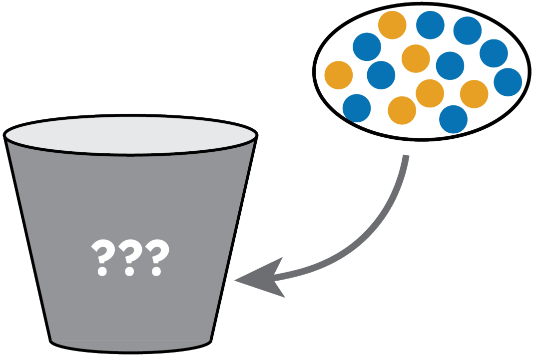
UC SANTA BARBARA
Department of Statistics and Applied Probability

Scenario 1

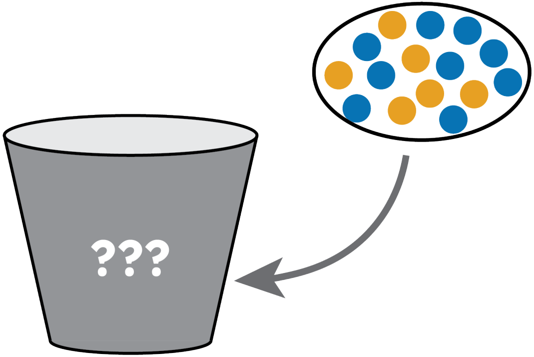


- We know that a bucket contains some (known) number of blue and gold marbles. From this bucket we take a sample.
- Given our knowledge of what's in the bucket, we want to inform what's in our hand (e.g. number of gold marbles, probability of having more than 3 blue marbles, etc.)

Scenario 2

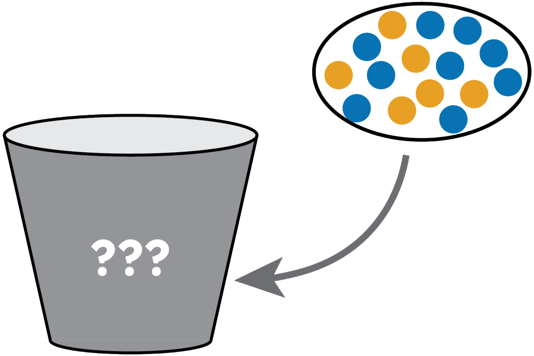


Scenario 2



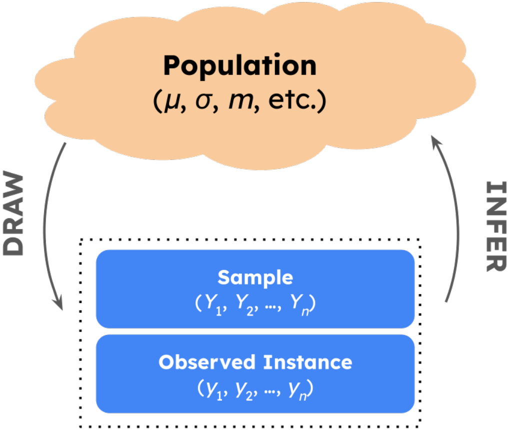
- We have a sample of blue and gold marbles (and we know how many of each are in our sample), that we know came from a bucket.

Scenario 2



- We have a sample of blue and gold marbles (and we know how many of each are in our sample), that we know came from a bucket.
- Given our knowledge of what's in our hand, we want to inform what's in the bucket (e.g. number of gold marbles, probability of having more than 3 blue marbles, etc.)

Cycle



Population
(μ, σ, m, etc.)
DRAW
INFER
Sample
(Y_1, Y_2, ..., Y_n)
Observed Instance
(y_1, y_2, ..., y_n)

Example

Goal

Determine the true average (mean) commute time (in minutes) of college students.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Goal

Determine the true average (mean) commute time (in minutes) of college students.

- First, note that it is next-to-impossible to compute this quantity exactly. Doing so would require us to survey every single college student in the US, ensure they are reporting accurate commute times, etc.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Goal

Determine the true average (mean) commute time (in minutes) of college students.

- First, note that it is next-to-impossible to compute this quantity exactly. Doing so would require us to survey every single college student in the US, ensure they are reporting accurate commute times, etc.
- A much better strategy would be to take a **random sample** of college students and try to use this sample to make inferences about the true average commute times of college students.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- In this situation, our **population** can be thought of as the set of all college students.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- In this situation, our **population** can be thought of as the set of all college students.
- The commute times of n randomly-selected college students could then be modeled by a collection $\{Y_i\}_{i=1}^{n}$ of random variables.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- In this situation, our **population** can be thought of as the set of all college students.
- The commute times of n randomly-selected college students could then be modeled by a collection $\{Y_i\}_{i=1}^{n}$ of random variables.
- But wait - why random? Aren't commute times deterministic?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- In this situation, our **population** can be thought of as the set of all college students.
- The commute times of n randomly-selected college students could then be modeled by a collection $\{Y_i\}_{i=1}^n$ of random variables.
- But wait - why random? Aren't commute times deterministic?
- Well yes... and no. The commute time *of a randomly-selected person* will clearly be random - it varies from person-to-person!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example

- In this situation, our **population** can be thought of as the set of all college students.
- The commute times of n randomly-selected college students could then be modeled by a collection $\{Y_i\}_{i=1}^{n}$ of random variables.
- But wait - why random? Aren't commute times deterministic?
- Well yes... and no. The commute time *of a randomly-selected person* will clearly be random - it varies from person-to-person!
- But, you're right in that, say, Angela's (or any specific *individual*'s) commute time is fixed. (Admittedly, it may vary from day to day, but let's ignore that subtlety for now.)

Example

- This is where we need to be very careful about our notation.

Example

- This is where we need to be very careful about our notation.
- Specifically, we can let Y_i denote the commute time of a randomly-selected person; then the collection $\vec{\boldsymbol{Y}} := \{Y_i\}_{i=1}^{n}$ is a collection of random variables. [Following the 120A convention, we use capital letters to denote random variables.]

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- This is where we need to be very careful about our notation.
- Specifically, we can let Y_i denote the commute time of a randomly-selected person; then the collection $\vec{\boldsymbol{Y}} := \{Y_i\}_{i=1}^{n}$ is a collection of random variables. [Following the 120A convention, we use capital letters to denote random variables.]
- But, once we collect a *specific* sample, we obtain a list of n numbers (commute times), which we can denote as $\vec{\boldsymbol{y}} = \{y_i\}_{i=1}^{n}$. [Note that we are using *lowercase* letters here.]

Example

- This is where we need to be very careful about our notation.
- Specifically, we can let Y_i denote the commute time of a randomly-selected person; then the collection $\vec{\boldsymbol{Y}} := \{Y_i\}_{i=1}^{n}$ is a collection of random variables. [Following the 120A convention, we use capital letters to denote random variables.]
- But, once we collect a *specific* sample, we obtain a list of n numbers (commute times), which we can denote as $\vec{\boldsymbol{y}} = \{y_i\}_{i=1}^{n}$. [Note that we are using *lowercase* letters here.]
- I often refer to $\vec{\boldsymbol{y}}$ as an "observed instance" or "realization" of the random vector $\vec{\boldsymbol{Y}}$.

Example

Goal

Determine the true average (mean) weight (in lbs) of all domestic shorthair (DSH) cats

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Goal

Determine the true average (mean) weight (in lbs) of all domestic shorthair (DSH) cats

- **Population:**

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Goal

Determine the true average (mean) weight (in lbs) of all domestic shorthair (DSH) cats

- **Population:** set of all DSH cats

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example

Goal

Determine the true average (mean) weight (in lbs) of all domestic shorthair (DSH) cats

- **Population:** set of all DSH cats
- **Random Sample:**

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Goal

Determine the true average (mean) weight (in lbs) of all domestic shorthair (DSH) cats

- **Population:** set of all DSH cats
- **Random Sample:** $\vec{Y} := \{Y_i\}_{i=1}^n$, the weights of n randomly-selected cats

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Goal

Determine the true average (mean) weight (in lbs) of all domestic shorthair (DSH) cats

- **Population:** set of all DSH cats
- **Random Sample:** $\vec{Y} := \{Y_i\}_{i=1}^{n}$, the weights of n randomly-selected cats
- **Observed Instance:**

UC SANTA BARBARA
Department of Statistics and Applied Probability

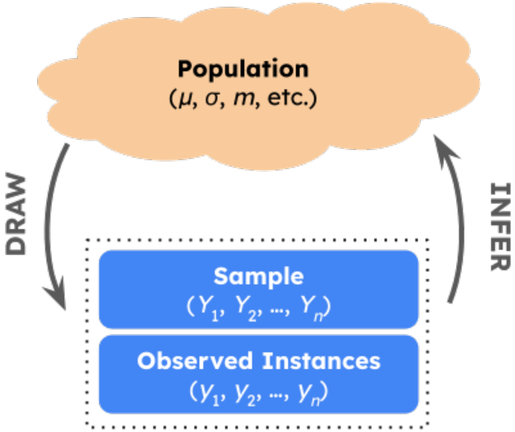
Example

Goal

Determine the true average (mean) weight (in lbs) of all domestic shorthair (DSH) cats

- **Population:** set of all DSH cats
- **Random Sample:** $\vec{\boldsymbol{Y}} := \{Y_i\}_{i=1}^n$, the weights of n randomly-selected cats
- **Observed Instance:** $\vec{\boldsymbol{y}} := \{y_i\}_{i=1}^n$, the weights of Kitty, Shiro, Bean, etc.

Back to the Cycle



Population Parameters

- Now, a **population parameter** is some quantity that governs the population.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Population Parameters

- Now, a **population parameter** is some quantity that governs the population.
- For the purposes of this class, we'll think of this as a number (e.g. the true average weight of all DSH cats; the longest time a human can hold their breath; etc.)

UC SANTA BARBARA
Department of Statistics and Applied Probability

Population Parameters

- Now, a **population parameter** is some quantity that governs the population.
- For the purposes of this class, we'll think of this as a number (e.g. the true average weight of all DSH cats; the longest time a human can hold their breath; etc.)
- We'll even go a step further and impose a *distributional* assumption on our population, after which we can interpret population parameters as, well, the parameters of the population distribution! (More on that later.)

UC SANTA BARBARA
Department of Statistics and Applied Probability

Inferences

- Now, notice that I've used the word "infer" to describe what we seek to do with our sample.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Inferences

- Now, notice that I've used the word "infer" to describe what we seek to do with our sample.
- Inference (typically) means one of two things:

UC SANTA BARBARA
Department of Statistics and Applied Probability

Inferences

- Now, notice that I've used the word "infer" to describe what we seek to do with our sample.
- Inference (typically) means one of two things:
  - **Estimation:** where we seek to use our sample to *estimate* the value of a population parameter

UC SANTA BARBARA
Department of Statistics and Applied Probability

Inferences

- Now, notice that I've used the word "infer" to describe what we seek to do with our sample.
- Inference (typically) means one of two things:
  - **Estimation:** where we seek to use our sample to *estimate* the value of a population parameter
  - **Hypothesis Testing:** where we seek to use our sample to assess the potential validity of claims surrounding a particular population parameter.

Inferences

- Now, notice that I've used the word "infer" to describe what we seek to do with our sample.
- Inference (typically) means one of two things:
  - **Estimation:** where we seek to use our sample to *estimate* the value of a population parameter
  - **Hypothesis Testing:** where we seek to use our sample to assess the potential validity of claims surrounding a particular population parameter.
- Both of these are examples (or subcases) of **inferential statistics**, where we seek to take a sample and make *inferences* about the population.

Example

- To further elucidate the distinction between estimation and hypothesis testing, let's return to our Domestic Shorthair Cat example.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- To further elucidate the distinction between estimation and hypothesis testing, let's return to our Domestic Shorthair Cat example.
- Specifically, we take our population to be the set of all Domestic Shorthair Cats, and seek to perform inference on $\mu :=$ the true average weight (in lbs) of a randomly-selected Domestic Shorthair Cat.

Example

- To further elucidate the distinction between estimation and hypothesis testing, let's return to our Domestic Shorthair Cat example.
- Specifically, we take our population to be the set of all Domestic Shorthair Cats, and seek to perform inference on $\mu :=$ the true average weight (in lbs) of a randomly-selected Domestic Shorthair Cat.
- Suppose our goal is to <u>estimate the true value of μ</u>.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- To further elucidate the distinction between estimation and hypothesis testing, let's return to our Domestic Shorthair Cat example.
- Specifically, we take our population to be the set of all Domestic Shorthair Cats, and seek to perform inference on $\mu :=$ the true average weight (in lbs) of a randomly-selected Domestic Shorthair Cat.
- Suppose our goal is to <u>estimate the true value of μ</u>.
    - To do so, we might collect a random sample of DSH cats, compute their observed sample mean weight, and use this to try and say something about the true value of μ.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- To further elucidate the distinction between estimation and hypothesis testing, let's return to our Domestic Shorthair Cat example.
- Specifically, we take our population to be the set of all Domestic Shorthair Cats, and seek to perform inference on $\mu :=$ the true average weight (in lbs) of a randomly-selected Domestic Shorthair Cat.
- Suppose our goal is to <u>estimate the true value of μ</u>.
  - To do so, we might collect a random sample of DSH cats, compute their observed sample mean weight, and use this to try and say something about the true value of μ.
  - This is an <u>estimation</u> problem, as our goal is to *estimate* the value of μ.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- Now, suppose we have the same population (set of all DSH cats) and the same population parameter (μ, the true average weight in lbs of a randomly-selected DSH cats).

Example

- Now, suppose we have the same population (set of all DSH cats) and the same population parameter (μ, the true average weight in lbs of a randomly-selected DSH cats).
- Suppose a veterinarian has told us that the average DSH cat weights 6.2 lbs.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- Now, suppose we have the same population (set of all DSH cats) and the same population parameter (μ, the true average weight in lbs of a randomly-selected DSH cats).
- Suppose a veterinarian has told us that the average DSH cat weights 6.2 lbs.
- We might then ask ourselves: if we collect data of our own (i.e. random samples of cats, and compute the observed sample average weight), will our data agree with this claim?

Example

- Now, suppose we have the same population (set of all DSH cats) and the same population parameter (μ, the true average weight in lbs of a randomly-selected DSH cats).
- Suppose a veterinarian has told us that the average DSH cat weights 6.2 lbs.
- We might then ask ourselves: if we collect data of our own (i.e. random samples of cats, and compute the observed sample average weight), will our data agree with this claim?
  - Here, we don't really care what the *exact* value of μ is; we're just trying to ascertain whether or not it is equal to 6.2.

Example

- Now, suppose we have the same population (set of all DSH cats) and the same population parameter (μ, the true average weight in lbs of a randomly-selected DSH cats).
- Suppose a veterinarian has told us that the average DSH cat weights 6.2 lbs.
- We might then ask ourselves: if we collect data of our own (i.e. random samples of cats, and compute the observed sample average weight), will our data agree with this claim?
    - Here, we don't really care what the *exact* value of μ is; we're just trying to ascertain whether or not it is equal to 6.2.
    - This is an example of hypothesis testing, as we are using our data to try and determine the validity of a given claim about μ (in this case, the claim that $\mu = 6.2$).

The Road Ahead

- We'll start off with a discussion of estimation, which will last us several lectures.

UC SANTA BARBARA
Department of Statistics and Applied Probability

The Road Ahead

- We'll start off with a discussion of estimation, which will last us several lectures.
- After that, we'll tackle Hypothesis Testing.

Estimation

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Goal

Goal

Given a population with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Goal

Goal

Given a population with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

- Again, as a concrete example, think in terms of our Cats example: we seek to use samples $\vec{\boldsymbol{Y}}$ of randomly-selected cat weights to estimate the value of μ, the true population mean.

Goal

Goal

Given a population with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

- Again, as a concrete example, think in terms of our Cats example: we seek to use samples $\vec{\boldsymbol{Y}}$ of randomly-selected cat weights to estimate the value of μ, the true population mean.
- Note that it is customary to use the letter θ to denote an *arbitrary* population parameter.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Notation

- We use μ to denote population means

Notation

- We use μ to denote population means
- We use σ^2 to denote population variances, and σ to denote population standard deviations

UC SANTA BARBARA
Department of Statistics and Applied Probability

Notation

- We use μ to denote population means
- We use σ^2 to denote population variances, and σ to denote population standard deviations
- We use p to denote population proportions (e.g. true proportion of all cats that weight above 5 lbs)

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Notation

- We use μ to denote population means
- We use σ^2 to denote population variances, and σ to denote population standard deviations
- We use p to denote population proportions (e.g. true proportion of all cats that weight above 5 lbs)

- Again, we use θ to denote an arbitrary parameter.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Assumptions

Goal

Given a population with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Assumptions

Goal

Given a population with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

- This is a pretty lofty goal as it stands!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Assumptions

Goal

Given a population with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

- This is a pretty lofty goal as it stands!
- So, to make our lives easier, we are going to make some additional assumptions.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Assumptions

Goal

Given a population with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

- This is a pretty lofty goal as it stands!
- So, to make our lives easier, we are going to make some additional assumptions.
- Specifically, we are going to assume that our random draws from the population follow a predetermined distribution.

Goal, Restated

- For example, we may say things like "assume the weight of a randomly-selected DSH cat follows a normal distribution"

UC SANTA BARBARA
Department of Statistics and Applied Probability

Goal, Restated

- For example, we may say things like "assume the weight of a randomly-selected DSH cat follows a normal distribution"
- Or, "suppose the commute time of a randomly-selected person follows an Exponential distribution with mean θ"

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Goal, Restated

- For example, we may say things like "assume the weight of a randomly-selected DSH cat follows a normal distribution"
- Or, "suppose the commute time of a randomly-selected person follows an Exponential distribution with mean θ"
- We also often impose an independence constraint; that is, we'll assume our sample $\vec{\boldsymbol{Y}} := \{Y_i\}_{i=1}^{n}$ is a collection of *independent* random variables.

Goal, Restated

- For example, we may say things like "assume the weight of a randomly-selected DSH cat follows a normal distribution"
- Or, "suppose the commute time of a randomly-selected person follows an Exponential distribution with mean θ"
- We also often impose an independence constraint; that is, we'll assume our sample $\vec{\boldsymbol{Y}} := \{Y_i\}_{i=1}^{n}$ is a collection of *independent* random variables.
  - In fact, more often than not, we'll even assume that our sample is i.i.d. (independent and identically distributed)

UC SANTA BARBARA
Department of Statistics and Applied Probability

Assumptions

Goal

Given a population, from which random variables are assumed to follow a distribution $\mathcal{F}$ with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

Assumptions

Goal

Given a population, from which random variables are assumed to follow a distribution $\mathcal{F}$ with parameter θ, we seek to take random samples $\vec{\boldsymbol{Y}} := (Y_1, \cdots, Y_n)$ from this population and use them to estimate the true value of θ.

- Here, I use $\mathcal{F}$ to denote an arbitrary distribution, with CDF $F()$ and PDF $f()$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Intuition

- Alright, let's establish some intuition.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Intuition

- Alright, let's establish some intuition.
- In fact, to ground things even more, let's go back to our cat example.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Intuition

- Alright, let's establish some intuition.
- In fact, to ground things even more, let's go back to our cat example.
- We assume we have a sample $\vec{\boldsymbol{Y}}$ of randomly-selected cat weights following some distribution $\mathcal{F}$ with some true mean μ, and we seek to estimate μ.

Intuition

- Alright, let's establish some intuition.
- In fact, to ground things even more, let's go back to our cat example.
- We assume we have a sample $\vec{\boldsymbol{Y}}$ of randomly-selected cat weights following some distribution $\mathcal{F}$ with some true mean μ, and we seek to estimate μ.
- Since we want to estimate the value of a *population* mean, doesn't it make sense to consider using the *sample* mean as a proxy?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Intuition

- Said differently, say I collect a random sample of 5 cats and find their weights (in lbs) to be $(8.2, 6.1, 10.2, 8.4, 9.2)$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Intuition

- Said differently, say I collect a random sample of 5 cats and find their weights (in lbs) to be $(8.2, 6.1, 10.2, 8.4, 9.2)$.
- If I ask you, based on just this sample along, to give me your best guess for the average weight of all cats in the world, wouldn't you just say 8.42 (i.e. the sample mean weight)?

Intuition

- Said differently, say I collect a random sample of 5 cats and find their weights (in lbs) to be $(8.2, 6.1, 10.2, 8.4, 9.2)$.
- If I ask you, based on just this sample along, to give me your best guess for the average weight of all cats in the world, wouldn't you just say 8.42 (i.e. the sample mean weight)?
- So, it seems like however we estimate μ with $\vec{\boldsymbol{Y}}$, we should somehow use $\overline{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i$, right?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Statistics

Definition (Statistic)

Given a random sample $\vec{\boldsymbol{Y}} = \{Y_i\}_{i=1}^n$, a **statistic** T is simply a function of $\vec{\boldsymbol{Y}}$:

$$T := T(\vec{\boldsymbol{Y}}) = T(Y_1, \cdots, Y_n)$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Statistics

Definition (Statistic)

Given a random sample $\vec{\boldsymbol{Y}} = \{Y_i\}_{i=1}^n$, a **statistic** T is simply a function of $\vec{\boldsymbol{Y}}$:

$$T := T(\vec{\boldsymbol{Y}}) = T(Y_1, \cdots, Y_n)$$

- Example: sample mean $\overline{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i$

Statistics

Definition (Statistic)

Given a random sample $\vec{\boldsymbol{Y}} = \{Y_i\}_{i=1}^n$, a **statistic** T is simply a function of $\vec{\boldsymbol{Y}}$:

$$T := T(\vec{\boldsymbol{Y}}) = T(Y_1, \cdots, Y_n)$$

- Example: sample mean $\overline{Y}_n := \frac{1}{n} \sum_{i=1}^n Y_i$
- Example: sample variance $\frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Statistics

Definition (Statistic)

Given a random sample $\vec{\boldsymbol{Y}} = \{Y_i\}_{i=1}^n$, a **statistic** T is simply a function of $\vec{\boldsymbol{Y}}$:

$$T := T(\vec{\boldsymbol{Y}}) = T(Y_1, \cdots, Y_n)$$

- Example: sample mean $\overline{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i$
- Example: sample variance $\frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$
- Example: sample maximum $Y_{(n)}$

Statistics

Definition (Estimator)

An **estimator** $\widehat{\theta}_n$ for a population parameter θ is a statistic (computed from a random sample taken from this population) that is being used to estimate θ.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Statistics

Definition (Estimator)

An **estimator** $\widehat{\theta}_n$ for a population parameter θ is a statistic (computed from a random sample taken from this population) that is being used to estimate θ.

- Example: we can use the sample mean as an estimator of the population mean: $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$

Statistics

Definition (Estimator)

An **estimator** $\widehat{\theta}_n$ for a population parameter θ is a statistic (computed from a random sample taken from this population) that is being used to estimate θ.

- Example: we can use the sample mean as an estimator of the population mean: $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$
- Example: we can use the sample variance as an estimator of the population variance: $\widehat{\sigma^2}_n := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Estimators

- Hey look - **functions of random variables!**

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Estimators

- Hey look - **functions of random variables!**
- Yup, that's right - we will most *certainly* be using our Transformations-related results going forward.

Estimators

- Hey look - **functions of random variables!**
- Yup, that's right - we will most *certainly* be using our Transformations-related results going forward.

Definition (Sampling Distribution)

Given a random sample $\vec{\boldsymbol{Y}} = \{Y_i\}_{i=1}^{n}$ and a statistic $T := T(\vec{\boldsymbol{Y}})$, the **sampling distribution** of T is simply the distribution of T.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Estimators

- Hey look - **functions of random variables!**
- Yup, that's right - we will most *certainly* be using our Transformations-related results going forward.

Definition (Sampling Distribution)

Given a random sample $\vec{\boldsymbol{Y}} = \{Y_i\}_{i=1}^{n}$ and a statistic $T := T(\vec{\boldsymbol{Y}})$, the **sampling distribution** of T is simply the distribution of T.

- To find sampling distributions of specific statistics, we will (perhaps unsurprisingly) rely on our techniques from Topic 02 of this course (i.e. our unit on Transformations).

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Example

Suppose the weight (in lbs) of a randomly-selected DSH cat follows a normal distribution with unknown mean μ and known standard deviation $\sigma = 1.8$ lbs. Let $\vec{\boldsymbol{Y}} := \{Y_i\}_{i=1}^n$ denote an i.i.d. random sample of DSH cats, and consider using the sample mean as an estimator for μ:

$$\widehat{\mu}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i$$

What is the sampling distribution of $\widehat{\mu}_n$?

Solution

- We know that, by assumption, $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1.8^2)$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solution

- We know that, by assumption, $Y_1, \cdots, Y_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1.8^2)$
- Hence, by (Linear Combinations of Independent Normals), we have

$$\widehat{\mu}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i \sim \mathcal{N}\left(\mu, \frac{1.8^2}{n}\right) \sim \mathcal{N}\left(\mu, \frac{3.24}{n}\right)$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Estimators vs. Estimates

- The textbook defines an estimator to be “a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.”

UC SANTA BARBARA
Department of Statistics and Applied Probability

Estimators vs. Estimates

- The textbook defines an estimator to be "a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample."
- This is essentially the same definition I proposed above - it's a *rule*, meaning it's not really a number *per se*.

Estimators vs. Estimates

- The textbook defines an estimator to be "a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample."
- This is essentially the same definition I proposed above - it's a *rule*, meaning it's not really a number *per se*.
- Rather, estimators are **random variables**! (Otherwise, why would we talk about sampling distributions of estimators?)

Estimators vs. Estimates

- The textbook defines an estimator to be "a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample."
- This is essentially the same definition I proposed above - it's a *rule*, meaning it's not really a number *per se*.
- Rather, estimators are **random variables**! (Otherwise, why would we talk about sampling distributions of estimators?)
- Of course, once we obtain an observed instance of a sample, we will be able to compute an actual *numerical* estimate for θ. This is what we call a **estimate** - an estimate is just an observed instance of an estimator.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Estimators vs. Estimates

- Again, maybe it helps to think in terms of our cat example.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Estimators vs. Estimates

- Again, maybe it helps to think in terms of our cat example.
- Say I take one random sample of cat weights $\vec{\boldsymbol{Y}}$. This represents the act of taking an *arbitrary* sample of cat weights - it's random, because different samples of cats will have different weights!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Estimators vs. Estimates

- Again, maybe it helps to think in terms of our cat example.
- Say I take one random sample of cat weights $\vec{\boldsymbol{Y}}$. This represents the act of taking an *arbitrary* sample of cat weights - it's random, because different samples of cats will have different weights!
  - This random sample is what we use to construct an *estimator*, like $\widehat{\mu}_n := \overline{Y}_n$. "I take a random sample of cats and compute the mean weight -" different samples will have different observed values of the mean!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Estimators vs. Estimates

- But, once I pin down a *specific* sample of cats (e.g. Kitty, Shiro, Bean), I can compute the mean of this *deterministic* collection of weights, which will in turn give me a *deterministic* number.

Estimators vs. Estimates

- But, once I pin down a *specific* sample of cats (e.g. Kitty, Shiro, Bean), I can compute the mean of this *deterministic* collection of weights, which will in turn give me a *deterministic* number.
  - This is what we use to construct our *estimate*.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Estimators vs. Estimates

- But, once I pin down a *specific* sample of cats (e.g. Kitty, Shiro, Bean), I can compute the mean of this *deterministic* collection of weights, which will in turn give me a *deterministic* number.
  - This is what we use to construct our *estimate*.

- So, the big moral is: estimators $\widehat{\theta}_n$ of a parameter θ are random, whereas estimates are deterministic (specifically, observed instances of $\widehat{\theta}_n$).

UC SANTA BARBARA
Department of Statistics and Applied Probability

Non-Uniqueness of Estimators

- Given a single parameter θ, we can consider constructing several *different* estimators!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Non-Uniqueness of Estimators

- Given a single parameter θ, we can consider constructing several *different* estimators!
- For example, given a sample $\vec{\boldsymbol{Y}}$ to estimate the mean μ, each of the following could be used as an estimator of μ:

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Non-Uniqueness of Estimators

- Given a single parameter θ, we can consider constructing several *different* estimators!
- For example, given a sample $\vec{\boldsymbol{Y}}$ to estimate the mean μ, each of the following could be used as an estimator of μ:
  - $\widehat{\mu}_{n,1} := \overline{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i$
- So, a natural question arises: can we pinpoint what makes one estimator *better* than another? Or, analogously, can we pinpoint what makes a given estimator *good*?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Non-Uniqueness of Estimators

- Given a single parameter θ, we can consider constructing several *different* estimators!
- For example, given a sample $\vec{\boldsymbol{Y}}$ to estimate the mean μ, each of the following could be used as an estimator of μ:
  - $\widehat{\mu}_{n,1} := \overline{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i$
  - $\widehat{\mu}_{n,2} := (Y_1 + Y_3)/2$
- So, a natural question arises: can we pinpoint what makes one estimator *better* than another? Or, analogously, can we pinpoint what makes a given estimator *good*?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Non-Uniqueness of Estimators

- Given a single parameter θ, we can consider constructing several *different* estimators!
- For example, given a sample $\vec{\boldsymbol{Y}}$ to estimate the mean μ, each of the following could be used as an estimator of μ:
  - $\widehat{\mu}_{n,1} := \overline{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i$
  - $\widehat{\mu}_{n,2} := (Y_1 + Y_3)/2$
  - $\widehat{\mu}_{n,3} := Y_5$
- So, a natural question arises: can we pinpoint what makes one estimator *better* than another? Or, analogously, can we pinpoint what makes a given estimator *good*?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Assessing the Performance of Estimators

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Intuition/Analogy

- The textbook presents a very nice analogy: we can think of parameter estimation as trying to fire a revolver at a target.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Intuition/Analogy

- The textbook presents a very nice analogy: we can think of parameter estimation as trying to fire a revolver at a target.
- The target is like the parameter - we want to "hit it" (i.e. estimate it) as closely as we can.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Intuition/Analogy

- The textbook presents a very nice analogy: we can think of parameter estimation as trying to fire a revolver at a target.
- The target is like the parameter - we want to “hit it” (i.e. estimate it) as closely as we can.
- Any particular shot can be thought of as an estimate, and our marksperson can be thought of as our estimator (we don’t know exactly where any arbitrary shot is going to land until after it is made!)

Intuition/Analogy

- Under this analogy, assessing how well a particular estimator does at estimating the parameter is akin to assessing how good of a shot our marksperson is.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Intuition/Analogy

- Under this analogy, assessing how well a particular estimator does at estimating the parameter is akin to assessing how good of a shot our marksperson is.
- Say the marksperson takes a single shot, and ends up hitting the target exactly. Can we definitively say they're a perfect marksperson?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Intuition/Analogy

- Under this analogy, assessing how well a particular estimator does at estimating the parameter is akin to assessing how good of a shot our marksperson is.
- Say the marksperson takes a single shot, and ends up hitting the target exactly. Can we definitively say they're a perfect marksperson?
- I think most of us would agree "no-" we need more data! Specifically, the marksperson could have gotten incredibly lucky and happened to hit the target by pure chance.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Intuition/Analogy

- Under this analogy, assessing how well a particular estimator does at estimating the parameter is akin to assessing how good of a shot our marksperson is.
- Say the marksperson takes a single shot, and ends up hitting the target exactly. Can we definitively say they're a perfect marksperson?
- I think most of us would agree "no-" we need more data! Specifically, the marksperson could have gotten incredibly lucky and happened to hit the target by pure chance.
- As such, this is why **sampling distributions** of estimators are so important - they are our attempt at modeling our beliefs about how well an estimator would perform after having taken many samples (a.k.a. assessing how good of a shot our makrsperson is after having observed them taking *multiple* shots).

Bias

Definition (Bias)

The **bias** of an estimator $\widehat{\theta}_n$ that is being used to estimate θ is defined to be

$$\mathrm{Bias}(\widehat{\theta}_n, \theta) = \mathbb{E}[\widehat{\theta}_n] - \theta$$

Bias

Definition (Bias)

The **bias** of an estimator $\widehat{\theta}_n$ that is being used to estimate θ is defined to be

$$\text{Bias}(\widehat{\theta}_n, \theta) = \mathbb{E}[\widehat{\theta}_n] - \theta$$

- If it is obvious what parameter $\widehat{\theta}_n$ is being used as an estimator for, then it is customary to simply write $\text{Bias}(\widehat{\theta}_n)$.

Bias

Definition (Bias)

The **bias** of an estimator $\widehat{\theta}_n$ that is being used to estimate θ is defined to be

$$\text{Bias}(\widehat{\theta}_n, \theta) = \mathbb{E}[\widehat{\theta}_n] - \theta$$

- If it is obvious what parameter $\widehat{\theta}_n$ is being used as an estimator for, then it is customary to simply write $\text{Bias}(\widehat{\theta}_n)$.
- Note that the bias of an estimator is just the signed distance between the center of its sampling distribution and the parameter being estimated.

Unbiasedness

Definition (Unbiased Estimator)

An estimator $\widehat{\theta}_n$ is said to be an **unbiased estimator** of a parameter θ if $\mathrm{Bias}(\widehat{\theta}_n, \theta) = 0$, which is equivalent to $\mathbb{E}[\widehat{\theta}_n] = \theta$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

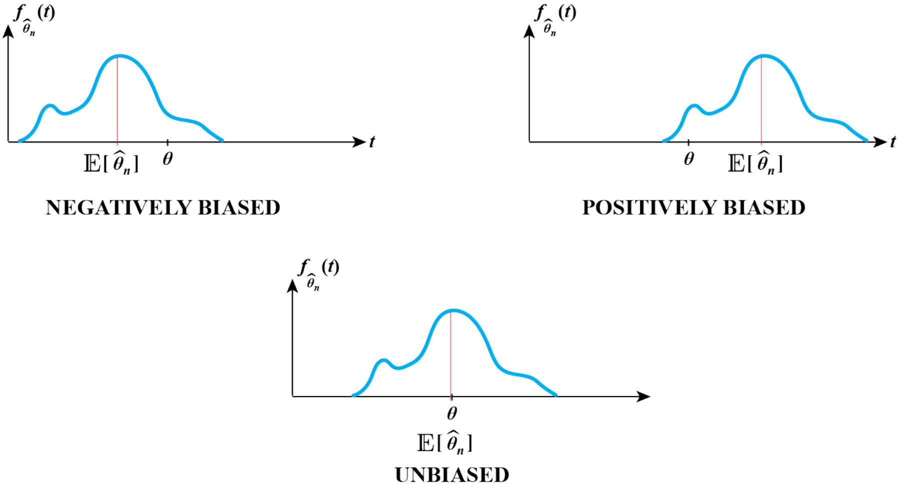
Unbiasedness

Definition (Unbiased Estimator)

An estimator $\widehat{\theta}_n$ is said to be an **unbiased estimator** of a parameter θ if $\mathrm{Bias}(\widehat{\theta}_n, \theta) = 0$, which is equivalent to $\mathbb{E}[\widehat{\theta}_n] = \theta$.

- As such, an unbiased estimator is one whose sampling distribution is centered at the true value of the parameter.

Bias



UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example

Example

Suppose $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$, and consider the following three estimators of μ:

$$\widehat{\mu}_{n,1} := Y_1; \qquad \widehat{\mu}_{n,2} := \frac{Y_1 + Y_2}{2}; \qquad \widehat{\mu}_{n,3} := \frac{Y_1 - Y_2}{2}; \qquad \widehat{\mu}_{n,4} := \overline{Y}_n$$

Which (if any) are unbiased estimators for μ? For those which are biased, compute the bias.

Solution

- We need only to compute the expectation of each estimator, and compare that to μ.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solution

- We need only to compute the expectation of each estimator, and compare that to μ.
- For example, $\mathbb{E}[\widehat{\mu}_{n,1}] = \mathbb{E}[Y_1] = \mu$,

Solution

- We need only to compute the expectation of each estimator, and compare that to μ.
- For example, $\mathbb{E}[\widehat{\mu}_{n,1}] = \mathbb{E}[Y_1] = \mu$, so $\mathrm{Bias}(\widehat{\mu}_{n,1}, \mu) := \mathbb{E}[\widehat{\mu}_{n,1}] - \mu = \mu - \mu = 0$;

Solution

- We need only to compute the expectation of each estimator, and compare that to μ.
- For example, $\mathbb{E}[\widehat{\mu}_{n,1}] = \mathbb{E}[Y_1] = \mu$, so $\mathrm{Bias}(\widehat{\mu}_{n,1}, \mu) := \mathbb{E}[\widehat{\mu}_{n,1}] - \mu = \mu - \mu = 0$; hence $\widehat{\mu}_{n,1}$ is an unbiased estimator for μ

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solution

- We need only to compute the expectation of each estimator, and compare that to μ.
- For example, $\mathbb{E}[\widehat{\mu}_{n,1}] = \mathbb{E}[Y_1] = \mu$, so $\mathrm{Bias}(\widehat{\mu}_{n,1}, \mu) := \mathbb{E}[\widehat{\mu}_{n,1}] - \mu = \mu - \mu = 0$; hence $\widehat{\mu}_{n,1}$ is an unbiased estimator for μ
- For $\widehat{\mu}_{n,2}$:

$$\mathbb{E}[\widehat{\mu}_{n,2}] = \mathbb{E}\left[\frac{Y_1 + Y_2}{2}\right] = \frac{\mathbb{E}[Y_1] + \mathbb{E}[Y_2]}{2} = \frac{\mu + \mu}{2} = \frac{2\mu}{2} = \mu$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solution

- We need only to compute the expectation of each estimator, and compare that to μ.
- For example, $\mathbb{E}[\widehat{\mu}_{n,1}] = \mathbb{E}[Y_1] = \mu$, so
  $\text{Bias}(\widehat{\mu}_{n,1}, \mu) := \mathbb{E}[\widehat{\mu}_{n,1}] - \mu = \mu - \mu = 0$; hence
  $\widehat{\mu}_{n,1}$ is an unbiased estimator for μ
- For $\widehat{\mu}_{n,2}$:

$$\mathbb{E}[\widehat{\mu}_{n,2}] = \mathbb{E}\left[\frac{Y_1 + Y_2}{2}\right] = \frac{\mathbb{E}[Y_1] + \mathbb{E}[Y_2]}{2} = \frac{\mu + \mu}{2} = \frac{2\mu}{2} = \mu$$

  so $\text{Bias}(\widehat{\mu}_{n,2}, \mu) := \mathbb{E}[\widehat{\mu}_{n,1}] - \mu = \mu - \mu = 0$ and
  $\widehat{\mu}_{n,2}$ is an unbiased estimator for μ

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solution

- For $\widehat{\mu}_{n,3}$:

$$\mathbb{E}[\widehat{\mu}_{n,3}] = \mathbb{E}\left[\frac{Y_1 - Y_2}{2}\right] = \frac{\mathbb{E}[Y_1] - \mathbb{E}[Y_2]}{2} = \frac{\mu - \mu}{2} = 0$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Solution

- For $\widehat{\mu}_{n,3}$:

$$\mathbb{E}[\widehat{\mu}_{n,3}] = \mathbb{E}\left[\frac{Y_1 - Y_2}{2}\right] = \frac{\mathbb{E}[Y_1] - \mathbb{E}[Y_2]}{2} = \frac{\mu - \mu}{2} = 0$$

so $\text{Bias}(\widehat{\mu}_{n,3}, \mu) = \mu$ meaning $\widehat{\mu}_{n,3}$ is a biased estimator for μ

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Solution

- For $\widehat{\mu}_{n,3}$:

$$\mathbb{E}[\widehat{\mu}_{n,3}] = \mathbb{E}\left[\frac{Y_1 - Y_2}{2}\right] = \frac{\mathbb{E}[Y_1] - \mathbb{E}[Y_2]}{2} = \frac{\mu - \mu}{2} = 0$$

  so $\text{Bias}(\widehat{\mu}_{n,3}, \mu) = \mu$ meaning $\widehat{\mu}_{n,3}$ is a biased estimator for μ

- Finally, for $\widehat{\mu}_{n,4}$, we can use our familiar result that the expectation of the sample mean is the population mean to conclude

$$\mathbb{E}[\widehat{\mu}_{n,4}] = \mathbb{E}\left[\overline{Y}_n\right] = \mu$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Solution

- For $\widehat{\mu}_{n,3}$:

$$\mathbb{E}[\widehat{\mu}_{n,3}] = \mathbb{E}\left[\frac{Y_1 - Y_2}{2}\right] = \frac{\mathbb{E}[Y_1] - \mathbb{E}[Y_2]}{2} = \frac{\mu - \mu}{2} = 0$$

  so $\text{Bias}(\widehat{\mu}_{n,3}, \mu) = \mu$ meaning $\widehat{\mu}_{n,3}$ is a biased estimator for μ

- Finally, for $\widehat{\mu}_{n,4}$, we can use our familiar result that the expectation of the sample mean is the population mean to conclude

$$\mathbb{E}[\widehat{\mu}_{n,4}] = \mathbb{E}\left[\overline{Y}_n\right] = \mu$$

  so $\text{Bias}(\widehat{\mu}_{n,4}, \mu) = 0$ and $\widehat{\mu}_{n,4}$ is an unbiased estimator for μ

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Asymptotically Unbiased

Definition (Asymptotically Unbiased)

An estimator $\widehat{\theta}_n$ for a parameter θ is said to be **asymptotically unbiased** if

$$\lim_{n \to \infty} \text{Bias}(\widehat{\theta}_n, \theta) = 0$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Asymptotically Unbiased

Definition (Asymptotically Unbiased)

An estimator $\widehat{\theta}_n$ for a parameter θ is said to be **asymptotically unbiased** if

$$\lim_{n \to \infty} \text{Bias}(\widehat{\theta}_n, \theta) = 0$$

- Note that all unbiased estimators are also asymptotically unbiased. The converse does not hold, though.

Asymptotically Unbiased

Definition (Asymptotically Unbiased)

An estimator $\widehat{\theta}_n$ for a parameter θ is said to be **asymptotically unbiased** if

$$\lim_{n \to \infty} \mathrm{Bias}(\widehat{\theta}_n, \theta) = 0$$

- Note that all unbiased estimators are also asymptotically unbiased. The converse does not hold, though.
- All asymptotic unbiasedness is saying is: as my sample size grows larger and larger, any discrepancies between $\mathbb{E}[\widehat{\theta}_n]$ and θ get washed out.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Chalkboard Example

Example

Let $Y_1, \cdots, Y_n \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$ for some unknown parameter $\theta > 0$. Consider the following two estimators for θ:

$$\widehat{\theta}_{n,1} = \overline{Y}_n; \qquad \widehat{\theta}_{n,2} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

Determine which (if either) is an unbiased estimator for θ. Is either of the two estimators biased but asymptotically unbiased?