# Topic 3: Estimation

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

# Outline

1. Unbiasedness, and MSE

2. Other Assessments

# Unbiasedness, and MSE


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Recap

**Goal**

Given a population, from which random variables are assumed to follow a distribution $\mathcal{F}$ with parameter $\theta$, 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 $\theta$.

- **Estimator** $\widehat{\theta}_n$: a statistic being used to estimate $\theta$.
  - Alternatively, "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."
- **Estimate**: an observed instance of our estimator.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Recap

- For instance, last lecture we talked about trying to estimate a population mean $\mu$.
- Given a sample $Y_1, \cdots, Y_n$ from the population (which, again, has mean $\mu$), we can consider several different estimators for $\mu$:
  - $\widehat{\mu}_{n,1} := \overline{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i$
  - $\widehat{\mu}_{n,2} := (Y_1 + Y_3)/2$
  - $\widehat{\mu}_{n,3} := Y_5$
- Since there are many potential estimators we can use to estimate a parameter, we'd like to determine how to quantify how "well" an estimator is performing.

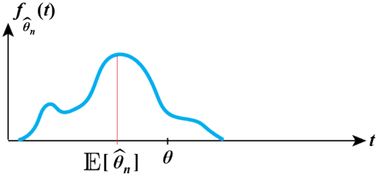

UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Recap:

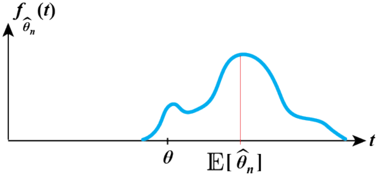
- One metric we talked about was that of **bias**, which is the signed distance between the expected value of our estimator and the true parameter value:
$$\text{Bias}(\widehat{\theta}_n\ ,\ \theta) := \mathbb{E}[\widehat{\theta}_n] - \theta$$
- An **unbiased** estimator $\widehat{\theta}_n$ of $\theta$ is one that satisfies $\mathbb{E}[\widehat{\theta}_n] = \theta$.
  - I.e., an unbiased estimator "gets it right on average."


UC SANTA BARBARA
Department of Statistics and Applied Probability

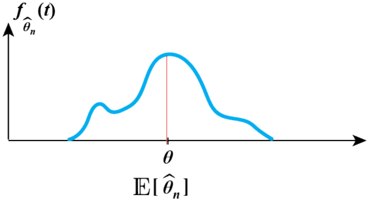
# Bias




**NEGATIVELY BIASED**




**POSITIVELY BIASED**




**UNBIASED**

# Recap

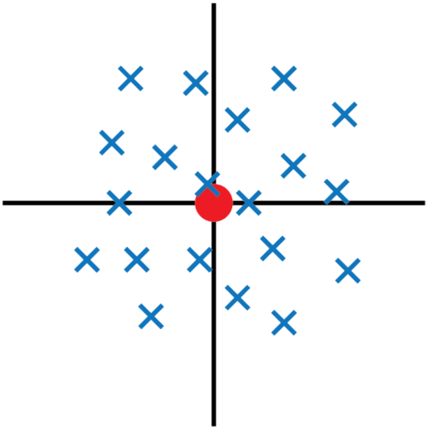
- I also introduced an analogy our textbook uses, whereby we can think of estimation as trying to hit a target with a revolver.
- The bullseye/target is the parameter we're trying to estimate; every bullet we fire is an estimate, and our shooting prowess is essentially the estimator.
- Assessing how well an estimator is performing is, then, akin to assessing how good of a shot we are!

# Recap

- An unbiased estimator is akin to a marksman who, on average, hits the target.
- More specifically, an unbiased estimator is akin to a marksperson whose average location of many shots is right on the target.


UC SANTA BARBARA
Department of Statistics and Applied Probability

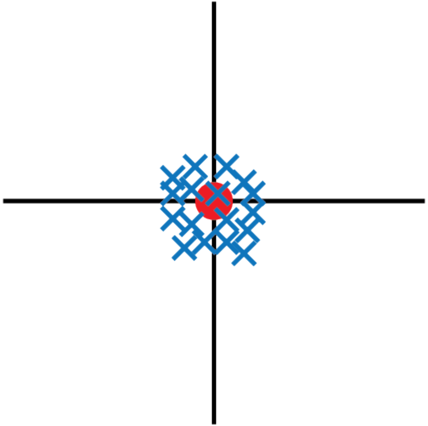
# Unbiasedness



- This marksperson is an example of an unbiased estimator - the average location of all of their shots (depicted as blue $\times$'s) is quite close to the target (indicated in red).
- But would we classify them as a "*good*" marksperson? Specifically, how would we classify their performance in comparison to...


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Unbiasedness



- This marksperson is an *also* “unbiased”.
- But doesn’t our intuition tell us that they are performing “better” than the marksperson on the previous slide?



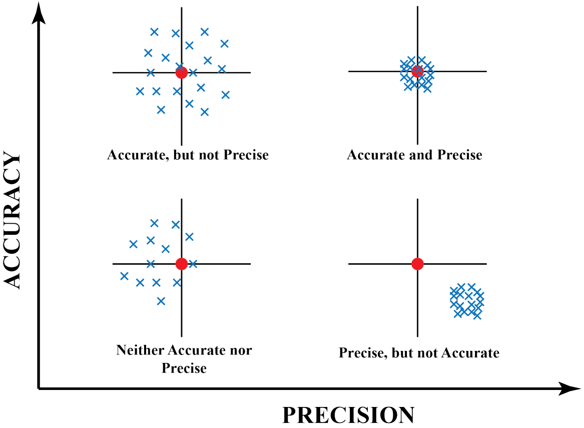
UC SANTA BARBARA
Department of Statistics and Applied Probability

# Precision vs. Accuracy

- So, this perhaps indicates to us that unbiasedness alone, though a decent critera to strive for, isn't the whole picture.
- Indeed, this relates to the distinction between two very important concepts in science (not just statistics): **precision** vs **accuracy**.
- Accuracy, more or less, corresponds to our notion of unbiasedness - it refers to "on average, how close are we to the ground truth?"
- *Precision* is the other half of the story that we're missing - it relates to "on average, how much *variability* is there from trial to trial?"


UC SANTA BARBARA
Department of Statistics and Applied Probability

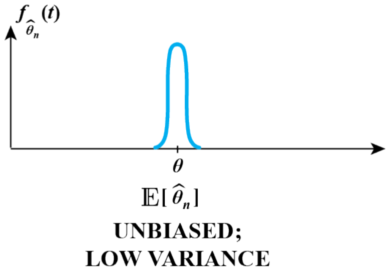
# Precision vs. Accuracy

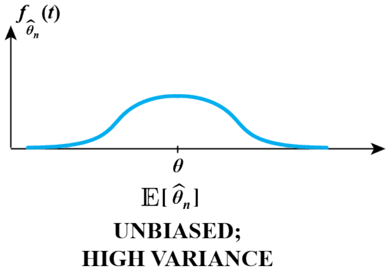


# Precision

- As was hinted at before, *precision* is linked (in the context of estimation) to the *variance* of a given estimator.
- Not only would we like our estimator to get the right value of $\theta$ on average, we'd also like to be fairly certain that on any particular draw we're close to the true value!

# Precision





UC SANTA BARBARA
Department of Statistics and Applied Probability

# Ideal Estimator

- So, based on everything we've discussed so far, it seems as though an "ideal" estimator is one that is both unbiased and also possesses a small variance.
- Thankfully, we have a metric that is able to simultaneously assess a given estimator's bias and variance - this metric is called the **mean square error** (MSE).

**Definition (MSE)**

The mean square error (MSE) of an estimator $\widehat{\theta}_n$ for a parameter $\theta$ is defined to be

$$\mathrm{MSE}(\widehat{\theta}_n \,,\, \theta) := \mathbb{E}\left[\left(\widehat{\theta}_n - \theta\right)^2\right]$$


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Bias-Variance Decomposition

**Theorem (Bias-Variance Decomposition)**

Given an estimator $\widehat{\theta}_n$ for a parameter $\theta$, we have that

$$\text{MSE}(\widehat{\theta}_n, \theta) = \left[\text{Bias}(\widehat{\theta}_n - \theta)\right]^2 + \text{Var}(\widehat{\theta}_n)$$

- We'll save the proof for later.


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Bias-Variance Decomposition

**Theorem (MSE of an Unbiased Estimator)**

Given an unbiased estimator $\widehat{\theta}_n$ for a parameter $\theta$, we have that

$$\mathrm{MSE}(\widehat{\theta}_n, \theta) = \mathrm{Var}(\widehat{\theta}_n)$$

- This follows directly from the Bias-Variance Decomposition, along with the definition of unbiasedness.

# Example

## Example

Let $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, \theta]$ for some deterministic constant $\theta > 0$. Compute the mean square error of using $\overline{Y}_n$ as an estimator for $\theta$.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Solutions

- When trying to compute the MSE of a given estimator, it's usually a good idea to start off by computing the expected value of the estimator.
- We know that the expected value of the sample mean is the population mean, which in this case is $(\theta + 0)/2 = \theta/2$ [we get this from the formula for the expectation of the Uniform distribution]. Hence,

$$\mathbb{E}[\overline{Y}_n] = \frac{\theta}{2}$$


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Solutions

- Let's now compute the bias of using $\overline{Y}_n$ as an estimator for $\theta$. By definition,
$$\text{Bias}(\overline{Y}_n, \theta) = \mathbb{E}[\overline{Y}_n] - \theta = \frac{\theta}{2} - \theta = -\frac{\theta}{2}$$
- Finally, we can compute the variance of $\overline{Y}_n$:

$$\text{Var}(\overline{Y}_n) = \frac{\text{Var}(Y_1)}{n} = \frac{\left(\frac{\theta^2}{12}\right)}{n} = \frac{\theta^2}{12n}$$


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Solutions

- So, by the Bias-Variance Decomposition,

$$\begin{aligned} \mathrm{MSE}(\overline{Y}_n, \theta) &= \left[ \mathrm{Bias}(\widehat{\theta}_n - \theta) \right]^2 + \mathrm{Var}(\widehat{\theta}_n) \\ &= \left( -\frac{\theta}{2} \right)^2 + \frac{\theta^2}{12n} = \frac{\theta^2(3n+1)}{12n} \end{aligned}$$


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Clicker Question

**Clicker Question 1**

Which of the following statements is true?

(A) An ideal estimator has a very large MSE

(B) An ideal estimator has an MSE that is very close to 0

(C) An ideal estimator has an MSE that is very negative


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Clicker Question

## Clicker Question 2

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

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

In terms of MSE, which (if either) estimator performs better?

(A) $\widehat{\mu}_{n,1}$

(B) $\widehat{\mu}_{n,2}$

(C) The two estimators perform equally well in terms of MSE

# Result

**Theorem (Sample Variance is an U.B.E. for Population Variance)**

Given an i.i.d. sample $\{Y_i\}_{i=1}^n$ from a distribution with unknown variance $\sigma^2$, then

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

is an unbiased estimator for $\sigma^2$.

# Example

## Example

Given $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ for some unknown $\sigma^2 > 0$, compute the MSE of using

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

as an estimator for $\sigma^2$.


UC SANTA BARBARA
Department of Statistics and Applied Probability

## Solutions

- Since $S_n^2$ is an unbiased estimator for $\sigma^2$ (by the previous theorem), we know (by the theorem pertaining to the MSE of an unbiased estimator) that $\mathrm{MSE}(S_n^2, \sigma^2) = \mathrm{Var}(S_n^2)$.
- By the result pertaining to the sampling distribution of $S_n^2$ (mentioned a few lectures ago), we have

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1} \sim \mathrm{Gamma}\left( \frac{n-1}{2} \, , \, 2 \right)$$


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Interstitial Result

**Theorem (Closure of Gamma Distribution under Scalar Multiplication)**

Given $Y \sim \text{Gamma}(\alpha, \beta)$ and $U := (cY)$ for some $c > 0$, we have that $U \sim \text{Gamma}(\alpha, c\beta)$.

**Proof.**
Use the MGF method. $\square$


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Solutions

$$\frac{n-1}{\sigma^2} S_n^2 \sim \text{Gamma}\left( \frac{n-1}{2} , 2 \right)$$

$$\implies S_n^2 \sim \text{Gamma}\left( \frac{n-1}{2} , 2 \cdot \frac{\sigma^2}{n-1} \right)$$

$$\implies \text{Var}(S_n^2) = \left( \frac{n-1}{2} \right) \cdot \left( 2 \cdot \frac{\sigma^2}{n-1} \right)^2$$

$$= \frac{n-1}{2} \cdot \frac{4\sigma^4}{(n-1)^2} = \boxed{\frac{2\sigma^4}{n-1}}$$

# Other Assessments



UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Leadup

- MSE is a very useful metric for measuring how well a given estimator is performing!
- Indeed, as we've seen, it even allows us to compare the performance of two estimators, by simply comparing their MSE's (remember the result of our clicker questions!)
- But, there are other properties we might seek to impose on our estimators.

# Leadup

- Recall last lecture that I introduced the notion of an *asymptotically* unbiased estimator.
  - As a review, an estimator $\widehat{\theta}_n$ for a parameter $\theta$ is said to be asymptotically unbiased if
$$\lim_{n \to \infty} \mathrm{Bias}(\widehat{\theta}_n, \theta) = 0$$
- Indeed, the field of **asymptotics** is the subfield of statistics dedicated to studying what happens as our sample size ($n$) becomes very large.
- Borrowing from asymptotics, we may seek to impose certain *large-sample* properties we would like our "good" estimators to obey.

# Disclaimer

- Disclaimer - things are about to get pretty math-y.
- I'll do my best to translate these results into words - I urge you to think through these definitions carefully on your own!


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Consistency

**Definition (Consistent Estimator)**

An estimator $\widehat{\theta}_n$ is said to be a **consistent** estimator for $\theta$ if

$$\widehat{\theta}_n \xrightarrow{p} \theta$$

That is, if either of the two equivalent conditions hold for any $\varepsilon > 0$:

$$\lim_{n \to \infty} \mathbb{P}(|\widehat{\theta}_n - \theta| \geq \varepsilon) = 0$$

$$\lim_{n \to \infty} \mathbb{P}(|\widehat{\theta}_n - \theta| < \varepsilon) = 1$$


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Interpretation

- Okay, what the heck is this saying???
- Let's parse through the first definition:

$$\lim_{n \to \infty} \mathbb{P}(|\widehat{\theta}_n - \theta| \geq \varepsilon) = 0$$

- What is the event $\{|\widehat{\theta}_n - \theta| \geq \varepsilon\}$ saying?
- Well, $|\widehat{\theta}_n - \theta|$ is essentially the distance between $\widehat{\theta}_n$ and $\theta$.
- Hence, the event $\{|\widehat{\theta}_n - \theta| \geq \varepsilon\}$ is essentially just the event "$\widehat{\theta}_n$ is very far away from $\theta$."


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Interpretation

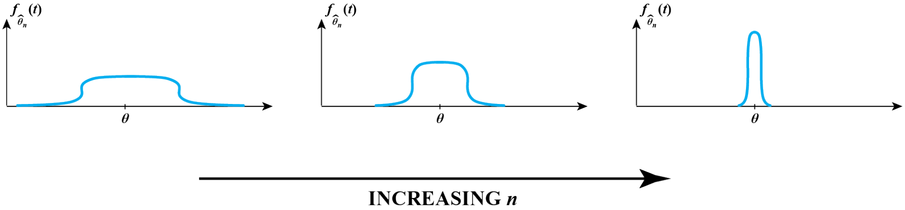
- Therefore, $\mathbb{P}(|\widehat{\theta}_n - \theta| \geq \varepsilon)$ is just the probability that $\widehat{\theta}_n$ is very far away from $\theta$.
- What the definition of consistency is saying is: this probability goes to zero as our sample size increases.
- Equivalently, $\mathbb{P}(|\widehat{\theta}_n - \theta| \geq \varepsilon)$ is just the probability that $\widehat{\theta}_n$ is very close to $\theta$.
- The definition of consistency also asserts that this probability goes to 1 as our sample size increases.

# Interpretation

- So, all in all, consistency is saying: as we keep taking samples of larger and larger size, we become more and more *certain* that $\widehat{\theta}_n$ is very close to $\theta$.
- That sounds like a pretty desirable property for an estimator to have, doesn't it?


UC SANTA BARBARA
Department of Statistics
and Applied Probability

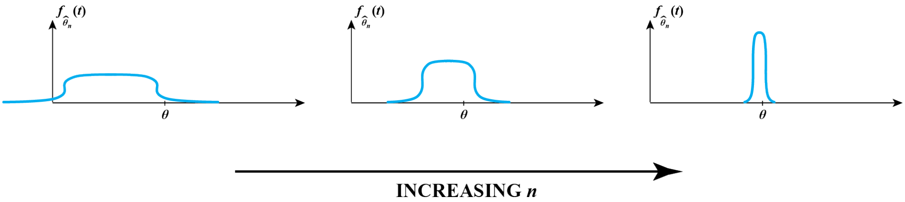
# Consistent and Unbiased




- This is a (cartoon) example of an estimator that is unbiased *and* consistent.
- There do exist consistent estimators that are biased:


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Consistent yet Biased




- You can (and will) show that $S_n$, the sample standard deviation, is a biased yet consistent estimator for $\sigma$, the population standard deviation.
- Fun fact - the background of our course logo contains an example of a biased yet consistent estimator!

# Example

## Example

Consider an i.i.d. sample $\{Y_i\}_{i=1}^{n}$ from a population with (unknown) mean $\mu$. Show that $\overline{Y}_n$ is a consistent estimator for $\mu$.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Solutions

- What we want to show is that, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) = 0$$

- First note that, by virtue of being a probability,

$$0 \leq \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon)$$

- Additionally, by Chebyshev's Inequality,

$$\mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\overline{Y}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

# Proof

- So, combining these two statements, we have

$$0 \leq \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

- Note that $[\sigma^2/(n\varepsilon^2)] \to 0$ as $n \to \infty$; additionally, $0 \to 0$ as $n \to \infty$. Hence, by the Squeeze Theorem (from Calculus),

$$\lim_{n \to \infty} \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) = 0$$

which, by definition, implies

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


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Result

**Theorem (Consistency and Unbiasedness, I)**

Consider an unbiased estimator $\widehat{\theta}_n$ for $\theta$. Then, $\widehat{\theta}_n$ is a consistent estimator for $\theta$ if $\lim_{n \to \infty} \text{Var}(\widehat{\theta}_n) = 0$.

- We'll prove this on the board together - please note that the techniques behind this proof are very important!


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Example

## Example

Consider an i.i.d. sample $\{Y_i\}_{i=1}^{n}$ from a population with mean $\mu$ and finite variance $\sigma^2 < \infty$.

(a) Show that $\overline{Y}_n$, the sample mean, is a consistent estimator for $\mu$.

(b) Show that $S_n^2$, the sample variance, is a consistent estimator for $\sigma^2$.

# Convergence in Probability

- Consistency is actually related to another important statistical notion, known as **convergence in probability**.

# Convergence in Probability

**Definition (Convergence in Probability)**

A sequence $\{X_n\}_{n \geq 0}$ of random variables is said to **converge in probability** to a constant $x$ if for every $\varepsilon > 0$ either of the equivalent conditions hold:

$$\lim_{n \to \infty} \mathbb{P}(|X_n - x| \geq \varepsilon) = 0$$

$$\lim_{n \to \infty} \mathbb{P}(|X_n - x| < \varepsilon) = 1$$

Convergence in probability is notated as

$$X_n \xrightarrow{p} x$$


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Properties

**Theorem (Properties of Convergence in Probability)**

Suppose that $X_n \xrightarrow{p} x$ and $Y_n \xrightarrow{p} y$. Then:

(I) $(X_n + Y_n) \xrightarrow{p} (x + y)$

(II) $(X_n \cdot Y_n) \xrightarrow{p} (x \cdot y)$

(III) $(X_n / Y_n) \xrightarrow{p} (x/y)$ provided that $y \neq 0$

(IV) **Continuous Mapping Theorem**: $g(X_n) \xrightarrow{p} g(x)$ for any real-valued function.

# Example

## Example

Consider an i.i.d. sample $\{Y_i\}_{i=1}^{n}$ from a population with mean $\mu$ and finite variance $\sigma^2 < 0$.

(a) Propose a consistent estimator for $\mu^2$, and show explicitly that your estimator *is* consistent.

(b) Propose a consistent estimator for $\mathbb{E}[Y_1^2]$ and show explicitly that your estimator *is* consistent.


UC SANTA BARBARA
Department of Statistics and Applied Probability